Introduction

Accessible Health Translation Technology Matters

This book was conceived in 2018 when Pierrette and I were working together on a funded project that evaluated the performance of machine translation for migrant and minority languages. The project was co-funded by the University of Geneva and the University of Sydney as part of a strategic research partnership between the two universities. When visiting Geneva, I was fortunate to meet with the research team Pierrette leads, a group of talented multilingual researchers from around the world, passionate about improving access to translation technology for vulnerable communities: culturally and linguistically diverse refugees, migrants, and people with disabilities. Mark, meanwhile, is well known to both of us – as a long-time acquaintance of Pierrette and as a close research collaborator of mine in recent years. A seasoned developer and senior researcher of multilingual medical speech translation systems, he offers a unique perspective on the design and evaluation of health and medical translation systems that can serve multicultural communities. Together, we began to plan a volume focused upon uses of linguistic technology in healthcare, with special interest in service to widely diverse and underserved communities.

Multilingualism and multiculturalism have always been proud traditions in Europe, as these values compose the core of democracy, social equality, and sustainability. Thriving multilingualism has provided a fertile and favorable social environment for translation research and innovation in our changing world. Increasing populations of migrant minorities, refugees, and displaced people in Europe are broadening the definition of, and our traditional understanding of, this multilingual and multicultural continent. In response, translation research is developing to reflect and interact with this changing social environment. Pierrette's own agenda in speech translation research, for

https://doi.org/10.1017/9781108938976.001 Published online by Cambridge University Press

instance, exemplifies European emphasis on enhancing the social benefits, inclusiveness, and humanitarian concerns of innovation in this area of technology. Facilitation of effective, accessible health communication with disadvantaged populations through use of enhanced translation technologies is seen as a practical, innovative, and cost-effective strategy for tackling health inequality in today's multilingual and multicultural Europe.

In modern Australia, too – my own research home – multiculturalism is widely recognized as a key to prosperity and social cohesion. In 2020, 29.8 percent of Australia's population was born overseas, while more than 200 languages are spoken in Australian homes. Accordingly, the chief function of translation and interpreting in Australia is provision of essential language assistance to people with limited English proficiency to improve their access to healthcare, medical treatments, and other essential social services. However, only professional translators – interpreters who have obtained their qualifications through the National Accreditation Authority for Translators and Interpreters (NAATI) process – may provide translation services. While in Europe systematic and conventionalized use of machine translation tools and curated multilingual databases are current, authorities in Australia only adopt translation technology in medical and health settings conservatively, despite increasing demands during the pandemic crisis for more timely, direct, and effective engagement with diverse communities. (To be sure, careful adoption is warranted – a point to which we will return at length.)

And of course, Mark's home base, the United States, is more than ever a melting pot. In a favorite San Francisco restaurant, bearing the Greek name Acropolis (a word that Mark's Korean-Japanese wife still struggles to pronounce), Russian food is prepared by Chinese staff. No healthcare facilities can avoid the consequent communication issues, though the specific linguistic and cultural demands and coping strategies (whether via professional interpreters in person or via video, intercession by family members or bilingual staff members, or available translation technology) vary considerably from state to state and region to region. In California, for instance, healthcare assistance in the patient's dominant language is mandated by law (though, too often, still absent in practice). Translation requirements at a major San Francisco facility are sketched in Chapter 2.

To linguistic and cultural challenges may be added physical and mental health challenges. And, in fact, the number of people with disabilities is increasing worldwide. Currently, according to the World Health Organization, over one billion or about 15 percent of the world's population is estimated to live with some impairment in mental, sensory, or mobility functions. This finding highlights another way in which speech and translation

technology can serve: not only to narrow the gap between languages and cultures, but also to lower barriers to effective communications associated with hearing impairment, speech disorders, and related impairments.

A special case within the disabled population is the Deaf community, which can be served through sign language translation, which utilizes a repertoire of gestures, a sign language lexicon, to interpret verbal messages. Technology integrating such translation can significantly improve access to healthcare services among people with hearing loss or impairment.

A final group deserving special consideration with respect to health communication is the population with limited literacy or educational attainment. And indeed, in multicultural countries, lack of English proficiency is often associated with limited health literacy.

For all these diverse and often underserved multilingual and multicultural groups, studies of translation, whether by humans or machines, are clearly vital. And yet, until now, quality assessments have in our judgment focused too narrowly on the accuracy of linguistic details, thus overlooking the actual linguistic comprehensibility, cultural accessibility, and relevance of translations for multicultural people with limited literacy and health literacy.

Hence this volume. Here, for example, we illustrate the design of medical speech translation systems equipped with tools for increasing speech recognition and translation reliability and customization per use case; and we discuss inclusion of simplified visual aids such as health-themed pictograms for use in emergency departments to facilitate understanding of medical terms by patients with limited literacy or local-language proficiency. Overall, we stress *accessible* speech and translation technology.

Accessible translation technology can provide much-needed help in addressing and reducing healthcare inequalities due to language discordances, disabilities, and limited educational level and health literacy – inequalities often associated with entrenched socioeconomic disadvantages. However, rather than define such accessibility and inclusivity narrowly, we prefer an open, principled approach. The key principles highlighted and illustrated throughout this book are these: (1) detectability of errors to boost user confidence by health professionals; (2) adaptability or customizability for health and medical domains; (3) inclusivity of translation modalities (written, speech, sign language) to serve people with disabilities; and (4) equality of accessibility standards for localized multilingual websites of health contents. To summarize these key principles for promotion of accessible and reliable translation technology, we use the acronym I-D-E-A.

I-D-E-A: Principles of Accessible Health Translation Technology

Detectability of Machine Translation Errors to Boost User Confidence

No form of modern technology is error free, and machine translation systems are no exceptions, whether based on earlier rule-based models or the latest neural networks. Chapter 1 explains the principal types of speech and text translation systems, comparing their theoretical underpinnings and assessing their relative strengths and limitations. Chapter 2 goes on to examine their applications in the healthcare context, with emphasis on speech translation.

General-purpose translation systems are constantly improving in measurable translation accuracy, as compared to professional human translations. In specialized healthcare domains, however, lack of confidence persists: even when measurably accurate, the systems fall short in *reliability*, as this is partly a psychological matter. One reason is the difficulty of understanding these technologies, and thus the natural hesitation to trust them – an issue directly addressed in Chapter 1, where the technologies are explained. We share the view that adoption of speech and translation technology for healthcare must proceed conservatively and with *informed* caution.

Detection of errors, we believe, should be a vital element of that caution and confidence building. Mechanisms are needed to increasingly enable healthcare professionals to predict potential translation errors and to detect actual ones.

But what errors should be predicted and detected? In current machine translation performance evaluation for health and medical applications, consistent definitions and supporting systems for differentiation and classification of wide-ranging translation errors are still lacking.

For translation of general material, error assessment often draws upon experienced translators' human judgment and consensus concerning the acceptability of inherent lexical, semantic, syntactic, and pragmatic variations between the source text and the target translation. The assessment's level of granularity depends upon the purpose of the exercise. Higher granularity entails increases in breadth and scoring of details; so this approach is most suitable for interpreters at advanced skill levels, and for interpreter training. Further, for specialized text or speech translation in healthcare, assessment must also align with the evaluation of miscommunication risks and their severity levels, as judged by health professionals. Sensitivity to miscommunication on health issues entails a subtle, discreet, and risk-based approach to translation error detection. Such judgment calls can be challenging, even for skilled translators

https://doi.org/10.1017/9781108938976.001 Published online by Cambridge University Press

and interpreters without proper medical training or extensive practical experience: indeed, detection of *clinically relevant* errors in machine translation output can be handled effectively and efficiently *only* by highly experienced translators specializing in health and medicine, or by professionals with proven bilingual proficiency. These must demonstrate in-depth understanding of health concepts, health beliefs, and traditions in both source and target languages.

Our brief is that, if machine translation is to gain wide applicability in healthcare, professionals at all levels of expertise and experience need transparent systems providing instant expert feedback concerning the quality of translation output and risk levels of input. Reliable and consistent mechanisms and tools for error prediction and detection are required that can leverage the cross-lingual/cross-cultural sensitivity and keen awareness of highly experienced health translators regarding clinically significant miscommunication risks.

Chapter 2 will illustrate several tools for error *detection*, especially during translated conversations.

Prediction of clinically significant errors can boost user confidence by reducing the uncertainty and risks associated with machine translation tools. Prediction can be particularly useful for users without sufficient resources to make informed, safe decisions – that is, without linguistic proficiency, adequate health, medical knowledge, or some combination. Thus Chapters 3 and 4 illustrate a novel approach to estimating the likelihood of translation errors that could mislead users, or to estimating the danger of including expressions with negative connotations in the target language when translating sensitive content regarding mental health status. This approach to managing the risk of machine translation combines the strengths of humans and machines – the insights and sensitivity of experienced health translators on one hand, and the consistency and high prediction accuracy of machine learning algorithms on the other. We hope that its use can significantly increase users' confidence in machine translation systems.

Adaptability or Customizability

Chapter 5 introduces BabelDr, an automatic speech translation system for medical emergencies which has been piloted and evaluated with diverse end users, especially multicultural migrants and refugees admitted to University of Geneva Hospitals in Switzerland, with the goal of enabling interactions with French-speaking physicians. Chapters 1 and 2 detail the design of such phrase-based medical speech translation tools, which have significantly enhanced

reliability and adaptability for specific health and medical domains when compared to general-purpose machine translation systems.

Since their translations are prepared in advance, phrase-based translation systems may be presented in various innovative media and formats. For example, BabelDr integrates graphical images of health messages, an innovation inspired by earlier studies on the effect of pictographs on user satisfaction in healthcare communication. Feedback from multicultural patients has verified the comprehensibility of pictographs and their potential for augmenting traditional text- or speech-based medical translation tools. For example, Chapter 5 discusses the adaptation of visual aids in BabelDr, selected from open sources and initially designed for Hispanic users, for use by multicultural patients speaking Albanian, Arabic, Dari, Farsi, and Tigrinya.

Because BabelDr's visual interface simplifies speech-based interpreting, it can ease logistic issues, for example, by facilitating recruitment of adequately trained health interpreters for minority languages. At the same time, visual aids can improve attention, recall, satisfaction, and adherence among patients, thus helping to reduce health communication barriers due to patients' limited health literacy or educational levels.

Chapter 5 also shows that pictograph usability depends on various factors – not only on the comprehensibility of the pictographs per se, but also on the manner of presentation (e.g., on the order and number of images).

Per Chapters 1 and 2, development of automated translation systems tends to favor languages for which copious data is available. The resulting machine translation tools thus often perform much better on these than on low-resource languages like those studied by Pierrette and her students. However, because phrase-based speech translation systems like BabelDr can be adapted to new health situations or demographics with relative ease, adaptation for low-resource languages likewise becomes more practical – a breakthrough in automatic health translation research. What's more, the benefits extend even further: technology-enhanced communications can be made accessible to a range of disadvantaged minorities; and breakthroughs based upon work with minorities can significantly benefit high-resource languages as well.

Thus the BabelDr project, having arisen from efforts to exploit pretranslated textual or spoken translations for their reliability, has led to recognition of the usefulness in those translations of images and visual aids, since these can be universally understandable by people of many languages, cultures, educational levels, health statuses, or disabilities. Pictogram-based approaches, whether replacing or augmenting text- or speech-based methods, hold great promise for technology-enhanced healthcare communication, just as universal emojis have become indispensable in everyday textual communications.

Meanwhile, research toward increased customization of text- and speech-based translation can continue in parallel. Chapter 2 illustrates a related facility in the Converser for Healthcare speech translation prototype – the Translation Shortcuts tool. Shortcuts are prepared translations for specific use cases like pharmacy consultations and in-hospital nursing, and thus are comparable to the pretranslated phrases of BabelDr. They differ, however, in that they're designed to operate in seamless cooperation with unrestricted automatic speech translation, rather than to replace it: If an input closely enough matches a source language utterance with a prepared translation, that translation will be used; and if not, the input will be subjected to full machine translation. Shortcuts, in other words, function as translation memory – as repositories for translations known to be correct and available for reuse – but can also be quickly added on demand and efficiently browsed and searched. Thus they, too, like strictly phrase-based translation systems, answer this volume's call for adaptability and customizability, showing that these criteria can be applied to many translation approaches.

Inclusivity to Serve People of Disability

Chapter 6 introduces a sign language version of the BableDr system expressly designed for people with hearing loss or impairment.

Globally, hearing impairment is on the rise. One in six adults is affected in the UK or Australia, and one in eight people in the United States. To improve communication between impaired patients and their healthcare providers, automatic translation tools can be crucial; and sign language is widely recognized as an important translation genre.

Each sign language has a unique vocabulary repertoire and an associated set of expressive bodily gestures and facial movements, while variations may be found across communities and cultures. Accordingly, Chapter 6 discusses the creation of a Swiss French Sign Language (LSF-CH) version of BabelDr, for which sign language interpreting by virtual interpreters (avatars) was derived from videos recorded by human sign interpreters. The research team then evaluated the reception of the avatar version among Deaf people on the BabelDr platform. It was determined that, at the present stage of avatar development, the subtle nuances of sign language interpreting were better conveyed by the facial expressions and bodily movements of humans than by the avatars' comparable actions. Nevertheless, the authors conclude that the use of virtual characters does interest the target audience and does appear promising in the medical

context. They call for further research on the complex perception of sign language interpreting among Deaf people.

Equality of Accessibility Standards for Localized Multilingual Websites

Chapter 7 discusses the prevalent lack of high accessibility standards in the translation and localization of online health information websites. The level of accessibility is found to vary considerably across translated versions of original English websites. In particular, the authors studied localized Spanish versions of many English websites developed by health authorities and health-promotion organizations, finding that accessibility in the target language was often insufficient – for instance, in website titles when changing English to Spanish. The relevant suboptimal elements of major health websites following translation and localization were not recognized by existing systems for assessing accessibility but were identified only through the authors' labor-extensive manual assessment.

Unfortunately, low standards of accessibility in localized healthcare websites directly affect the target audiences – Hispanic populations worldwide – in many ways.

First, target language errors can hamper the automatic recognition of website texts by screen readers – programs that read onscreen text out loud – for users with visual loss or impairment, or by people with limited literacy and educational level who require audio versions of website information.

Second, target language errors can hinder the accessibility of website title pages, thus impeding users' searches. Numerous error types were identified: titles in Spanish (1) were too long, exceeding the recommended 64-character limit; (2) failed to identify the subject of the web page; (3) made no sense when read out of context; (4) included unnecessary repetitions; (5) included abbreviations without the expanded form, thus causing potential confusion and reading difficulties; and (6) included URL addresses.

These findings call for increased awareness of accessibility issues in website translation and localization. Generic accessibility guidelines and protocols do exist, but mechanisms are needed to effectively evaluate their proper implementation.

The COVID-19 pandemic exacerbated existing social and health inequalities around the world. Migrants, minorities, refugees, and disabled populations have been disproportionately affected by the ongoing spread of the disease. Timely, effective, and responsive communication with these communities, in

this and other health emergences, has become increasingly important to all of us.

Translation technology can help. This book aims to offer updates and insights into its design, practical applications, and evaluation, whether for written, spoken, or signed varieties. Our hope is to help enable enhanced and more accessible healthcare communication with populations of diverse languages, cultures, and physical or mental abilities.

Meng Ji, Pierrette Bouillon, and Mark Seligman
September 2022

https://doi.org/10.1017/9781108938976.001 Published online by Cambridge University Press