IJoC Translations

The Impact of Translation Technologies
on the Process and Product of Translation

STEPHEN DOHERTY[1]
The University of New South Wales, Australia

Technological advances have led to unprecedented changes in translation as a means of interlingual communication. This article discusses the impact of two major technological developments of contemporary translation: computer-assisted translation tools and machine translation. These technologies have increased productivity and quality in translation, supported international communication, and demonstrated the growing need for innovative technological solutions to the age-old problem of the language barrier. However, these tools also represent significant challenges and uncertainties for the translation profession and the industry. In highlighting the need for increased awareness and technological competencies, I propose that these challenges can be overcome and translation technologies will become even more integral in interlingual communication.

Keywords: translation technology, machine translation, international communication, globalization, localization

Background

As a constant in the development of humanity, translation has always played a crucial role in interlingual communication by allowing for the sharing of knowledge and culture between different languages. This diffusion of information can be found as far back as the ancient world through to the industrial age and into the global village of today, where technological advances opaque our perception of translation and the ascendancy of English as the lingua franca can easily lead us to believe that everything we know, and indeed everything worth knowing, somehow exists in one language. Much of the wealth of knowledge and richness of experience that is constructed and documented in our societies is, however, confined within language silos, to which access is restricted for most of us, even with our favorite Internet search engines.

Stephen Doherty: s.doherty@unsw.edu.au

Cronin (2013) argues that any form of global interaction cannot occur without interlingual activities and thus globalization denotes translation, yet many of us are simply unable or unwilling to overcome the associated language barrier and must therefore rely on translation provided by others to access information beyond our own individual linguistic reach. Traditionally, the translator (and interpreter) has played this role and provided a professional service in acting as an interlingual and intercultural communicator so that we can access the information we seek, if indeed we knew it existed there in the first place. Due to its very nature, we typically do not recognize translation even when it is right before our very eyes (e.g., Kenny, 1996). With the explosion of digital content and the maturing participatory online culture of Web 2.0 technologies (O'Reilly, 2005), traditional human translation simply cannot keep up the pace with the translation needs of today (and tomorrow).

In profiling the traits of Internet users versus online content, the most recently available data reveal that the number of English-speaking users, at 800 million (28%), is followed by Chinese-speakers, at 649 million (23%) and then drops off to 222 million (8%) for Spanish—all in a total user base of 2.8 billion—see Figure 1 (Internet World Stats, 2015; W3Techs, 2015). However, in terms of the content available to these users, English leads at 56%, with an immediate plunge to Russian and German (both 6%), Japanese and Spanish (5%), and Chinese now at 3%. This substantial disjoint between users and available content is largely explained by the dominance of English content and the unique development of Internet connectivity in the uptake of Web 2.0 technologies in different countries as well as investment in technological infrastructures such as broadband and mobile Internet.
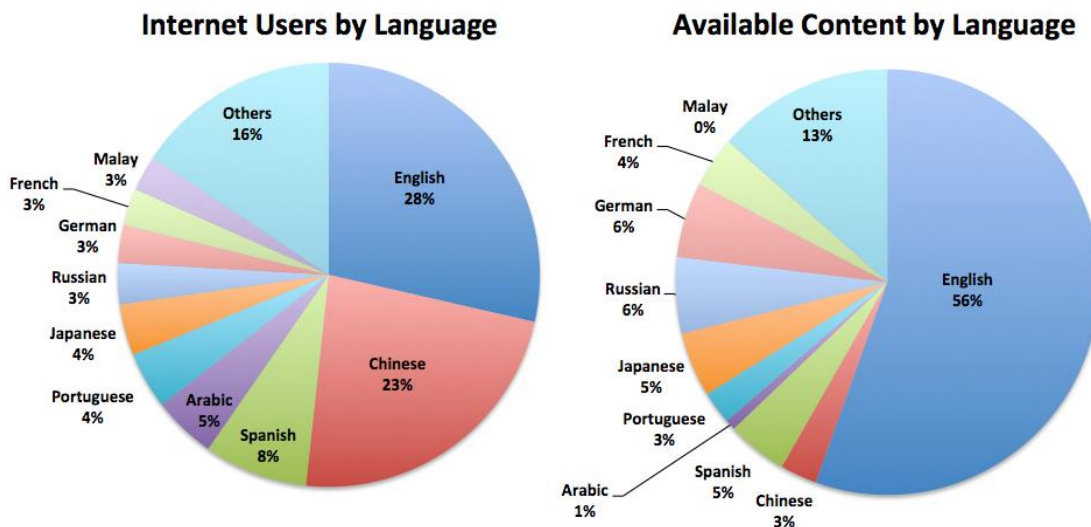
Figure 1. Internet users and available content by language (based on data from Internet World Stats, 2015; W3Techs, 2015).

Growth rates add another dimension. Although growth in the number of English-speaking users has continued steadily at a rate of about 468% from 2000 to 2013, it is overshadowed greatly by other global languages. Chinese and Spanish grew by 1,910% and 1,123%, respectively, with other languages showing considerable growth in the same period—for example, Arabic at 5,296% and Russian at 2,721% (Internet World Stats, 2015; W3Techs, 2015). This trend is mirrored in the composition of the translation industry during the same period. It has traditionally seen Europe (48.75%) and North America (35.77%) as the largest and most developed markets in the current global market, while the emerging markets of Asia (11.38%), Africa (0.29%) Latin America (1.80%), and Oceania (2.0%) have only recently begun to develop and are yet to show their full potential (DePalma, Hegde, Pielmeier, & Stewart, 2013) as is also argued here. A visual snapshot of this truly global activity is shown in Figure 2.
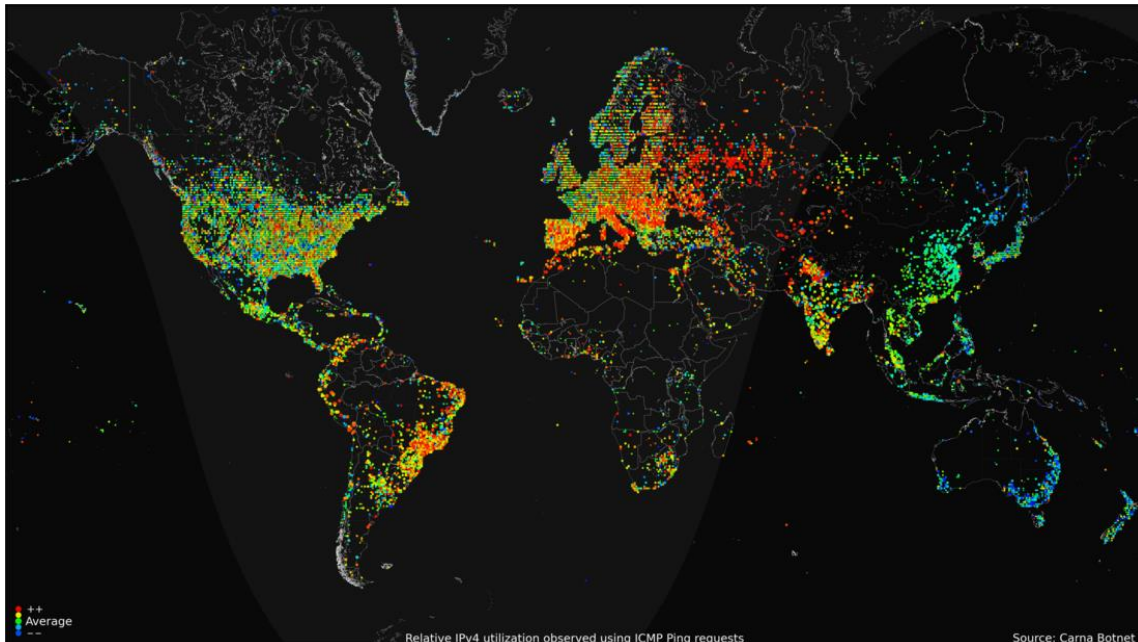
Figure 2. A snapshot of Internet activity during the European daylight hours. Retrieved from http://internetcensus2012.bitbucket.org/paper.html.

Similarly, analysts from the translation industry report that only a tiny amount of digital content, less than 0.1%, is currently being translated (DePalma et al., 2013). Indeed, the language services market as a whole has shown consistent year-on-year growth in recent years despite the global financial crisis, from US$23.50 billion in 2009 to US$34.78 in 2013—an annual growth rate of 5.13%. Translation prices per word, however, have continued to decrease by up to 50% since 2008, a diminution that analysts attribute to budgetary pressures and increased acceptance of translation technologies (DePalma et al., 2013, pp. 8–9). With Internet users now in the billions and growth far from tapering off, translation

technologies have been looked upon to provide solutions to this explosion of content that traditional human translation processes simply cannot manage. These technologies have developed vis-à-vis other information and communications technologies over recent decades and have even enabled such developments in return by providing a means of wider and more efficient interlingual communications that had hitherto been impossible (e.g., global simultaneous distribution of digital content such as computer software into tens of languages), all while transforming the very nature and practice of translation.

This article adopts an interactionist approach to demonstrate how technological developments in translation, driven by the two major technological innovations of computer-assisted translation tools and machine translation, have fundamentally changed how we communicate today. These developments and their concomitant positive and negative consequences are situated within the context of a fast-changing industry and the body of accompanying interdisciplinary translation research that focuses on process, product, and society. Thus, I critically review how translators now translate (process); what is being translated (product); and how the role of the translator has diversified to include various professional specializations and technical competencies as well as everyday users (society).

I contend that the ongoing technological evolution in translation has yielded unprecedented gains in terms of increased translator productivity and consistency, greater global language coverage, and greater support for improving international communication and distribution. However, there also exist significant knock-on effects that these technologies have on the practice and perception of translation itself, including the perceived and actual value of translation; the awareness and uptake of translation technologies; and the status and visibility of the profession.

Computer-Assisted Translation Tools

Recognizing the need to translate their products in order to be successful on international markets, software companies of the 1990s, and several other technology-related industries, sought a way to increase productivity in translation and maintain consistency of their linguistic data across a growing number of languages and countries (Esselink, 2000). As a result of this need and other factors such as the increased availability and affordability of computing power and the Internet, computer-assisted translation (CAT) tools provided the first major technological shift in the present-day translation industry with their commercial debut in the 1990s.

The core of CAT tools is a translation memory (TM), a software program that stores a translator's translated text alongside its original source text, so that these pairs can later be reused in full or in part when the translator is tasked with translating texts of a similar linguistic composition. For example, having previously translated the following sentence from English into French:

Click on the "Next" button to go to the next step.
Cliquez sur le bouton «Suivant» pour passer à l'étape suivante.

And then being presented with a new English sentence that contains:

Click on the "Back" button to go to the previous step.

The TM would show the translator the stored translation from the first sentence and highlight the lexical matches, much like the find-and-replace function in contemporary word processors, but with two languages in tandem. The matching words (as illustrated below in underlined text) in this new English sentence propose, using the stored elements, its translation into French:

Click on the "Back" button to go to the previous step.
Cliquez sur le bouton «Suivant» pour passer à l'étape suivante.

With these suggested matches, the translator can assess their quality and contextual appropriateness and use them in full or in part by editing (e.g., additions, deletions, substitutions). Here, for instance, the translation for *Back* and *previous* would need to be entered manually and the translator would substitute the verb for another to match the new context. This new English-French sentence pair would then be saved for later reuse. Over time, TMs can contain thousands, millions, and even billions of translations such as these, thereby increasing the likelihood of reuse once texts remain linguistically similar (see Figure 3).
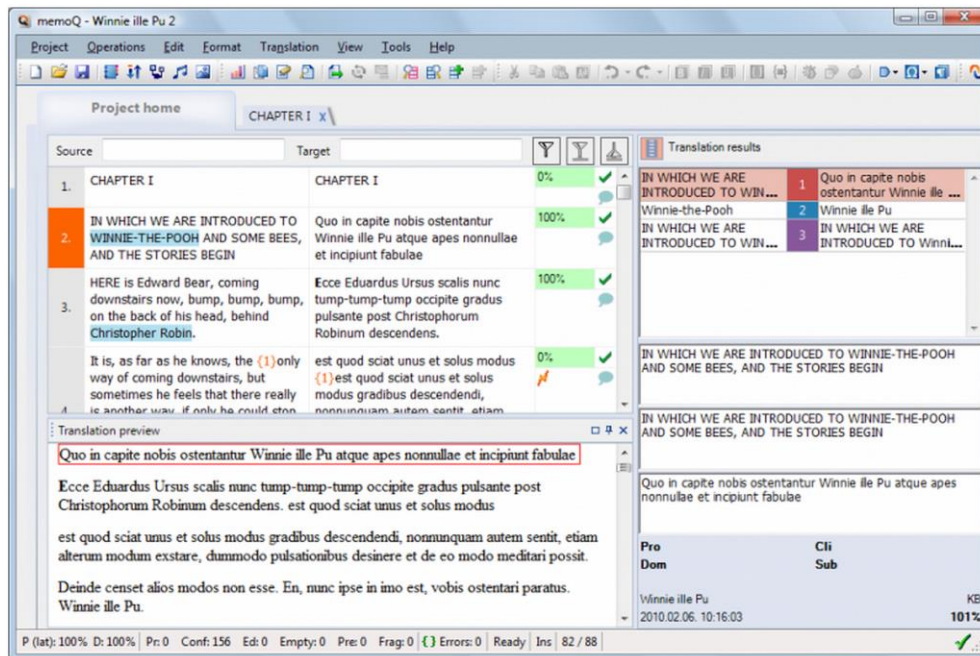
Figure 3. An example of the interface from a popular translation memory: MemoQ from Kilgray. Retrieved from http://clone.kilgray.com/products/memoq/screenshots.

In addition to this core feature, TMs are typically packaged with or integrated into additional software that allows translators to manage specialized terminology in a format similar to bilingual glossaries (e.g., medical terms, company-specific branding); search for keywords within the TM's database of stored translations; and share these linguistic data with others using project management features common to contemporary IT software.

While translation studies as a discipline and area of research has undergone many paradigm shifts (Snell-Hornby, 2006), it has been slow to adopt such translation technologies within its mainstream, resulting in a somewhat segregated subdiscipline (O'Hagan, 2013) that many scholars and industry stakeholders see as a discipline in its own right (Alcina, 2008) as it possesses many unique attributes and shares numerous fundamental commonalities with disciplines of computational linguistics and computer science, which lie far beyond traditional translation studies.

In the translation industry, too, everyday practical and commercial needs mean that theoretical models and approaches to translation are typically sidelined or ignored in favor of the more tangible and immediate gains offered by translation technology solutions. The proliferation of CAT tools in the industry and in academia quickly led to the creation of large collections of linguistic information (called *corpora*, the plural form of *corpus*) in many language pairs and across many genres. Indeed, the English-French sentences above could begin to form a small corpus to which we can add newly paired sentences as we continue to translate. With the development of CAT tools, translators could, for the first time, easily create their own collections of stored translations for later reuse in their work, for sharing with their colleagues, and for both commercial and academic research purposes. The uptake of TMs by the majority of translators has been consistently reported over the last two decades (Christensen & Schjoldager, 2010; Reinke, 2013) with saturation for many translators who work in large organizations and specialized areas.

Machine Translation

In its own parallel, machine translation (MT) had started to develop in the 1930s in the form of mechanical multilingual dictionaries. However, it was not until the 1950s that MT enjoyed a more public showcase as a limited, controlled but arguably automated translation process (see Hutchins, 2010). This was widely reported on by the media in the postwar period—a time when MT was informed principally by the disciplines of cryptography and statistics. Owing to the ever increasing availability of computing power, linguistic data, and the growing need for automation, tangible successes of MT began to emerge in the 1980s and 1990s, mostly using rule-based approaches, whereby sets of linguistic rules were written manually by linguists and translators for each language pair (see Arnold, Balkan, Meijer, Humphreys, & Sadler, 1996). Fueled by availability of the human translation data contained in the TMs that became widespread in the late 1990s, MT research experienced a further paradigm shift from prescriptive, top-down, rule-based approaches to descriptive, bottom-up, data-driven approaches chiefly in the form of statistical MT—a paradigm shift that has led to the second major technological shift in contemporary translation.

With this growing body of professional human translations in TMs becoming available in the 1990s and 2000s for an increasing number of languages, directions, genres, and text types, statistical MT

made substantial inroads into translation technology research and development. This was quickly followed by the more recent widespread adoption of MT by the translation industry and indeed by the general public in the form of freely available online systems such as Google Translate[2] and Microsoft Bing,[3] where both companies were already well-known IT providers with considerable resources to invest in the research, development, and application of MT on a global scale.

Fundamentally, these statistical MT approaches use complex statistical algorithms to analyze large amounts of data to generate a monolingual language model for each of the two given languages, and a translation model for the translation of words and phrases from one of these languages into the other. A decoder then uses these models to extrapolate the probability of a given word or phrase being translated from one language into the other, where the most probable word or phrase co-occurrences are chosen as the best translation (for a detailed description, see Hearne & Way, 2011; Kenny & Doherty, 2014).

These approaches allow for new languages to be covered without the need for handcrafted linguistic rules once parallel data for the languages are available. The downside, however, is that these systems are limited by their relative ignorance of linguistic information and their dependence on their own training data. Thus, any new terms and formulations will be difficult to translate correctly, if they absent from the systems' data.

As MT systems are typically built directly from human translations, they truly blur the borders between translation from a human and a machine. Today's systems typically contain millions of human-translated sentences from which they learn the patterns of probability, while specialized and freely available online systems can contain even more data from thousands of translators collated over many years. These systems are continually improving in terms of their quality and efficiency as their infrastructures become more refined and more high-quality translation data become available. Current and future issues lie in the quality of the TM data that the MT systems learn from, and in the trade-off between the amount of data used and in the time taken to process it. More data increases quality to a point, but takes longer to produce translations: from seconds and minutes to hours and even days depending on the computational resources available.

As part of the increasingly technology-embedded workflows of translation in the 2000s, MT has been added to the toolbox of many translators alongside TMs and other CAT tools—for some, by choice; for others, by force and necessity. An important caveat in MT is that, much like TMs before it, it typically works best with simplistic and repetitive linguistic features and within the same genres, domains, and texts types. However, manual and semiautomatic methods of domain and genre classification have begun to demonstrate improvements (e.g., Petukhova et al., 2012; Sharoff, 2007). Such texts are more easily processed by MT systems, which of course do not possess human reasoning or contextual knowledge of a text, its components, or its meaning(s). Indeed, several notable success stories of MT have come from organizations that carefully implement MT as part of a larger workflow of content creation in adherence to

---

[2] www.translate.google.com

[3] www.bing.com/translator

strict authoring guides, linguistic preprocessing, domain-specific glossaries, and the use of translators to assess and augment MT output to a high-quality, publishable standard (e.g., Roturier, 2009). Such workflows represent a shift to more automation in not only the translation technologies used to process linguistic data, but also in the overall translation project management systems required to coordinate large numbers of translators, on- and off-site, multitudinous projects, and languages.

Changing the Process of Translation

While CAT tools and, more recently, MT have been largely accepted by practitioners and researchers for their associated productivity and consistency gains, many translators are still adapting to the changes that these technologies are making to the translation industry and indeed to the process of translation itself. As most translators are freelancers or work for small-scale language service providers of between two and five employees (DePalma et al., 2013), learning how to effectively use these technologies poses a considerable challenge to most. Undeterred by calls for increased technological competencies dating back to the 1990s (O'Hagan, 2013) and the recent appearance of MT as part of the formal translation curriculum (Doherty & Kenny, 2014; Kenny & Doherty, 2014), translation technology competencies remain an underdeveloped skill set in translator education despite extensive industry surveys highlighting their absolute necessity and tremendous value (Gaspari, Almaghout, & Doherty, 2015).

As detailed above, the basic premise of TMs and MT is quite simple. Their integration into the translation process, however, has resulted in considerable alterations to how translators have traditionally worked with text. Perversely, TMs have been shown to result in a "sentence salad" (Bédard, 2000) due to the over-recycling of sentences and parts thereof that may not suit the context and cohesion of the given text to be translated but are reused by translators nevertheless. Further, focusing on text that only appears at the sentence level places great difficulties on providing an accurate and fluent translation that adheres to the cohesive and contextual norms of the target language, where, for instance, common linguistic devices of cohesion such as anaphora and cataphora typically function at the paragraph and document level. Indeed, translators may even opt for deliberate lexical repetition to decrease the variance in their expression and return more TM matches—a tactic known as "peep-hole translation" that poses a great threat to translation quality (Heyn, 1998) and consistency (Moorkens, Doherty, Kenny, & O'Brien, 2014).

Despite these dangers, Bowker (2005) points to a position of "blind faith" in TMs that has been adopted by translators who assume that the previously used human translation in TM data is of high quality and, as a result, are much less scrupulous in evaluating it than if they were translating from scratch. This is compounded by the reduced remuneration for using TMs, where the rule of thumb has been that if a certain percentage of the sentence to be translated is already provided by the TM, then the translator has that much less work to do. Despite contrary empirical evidence to this widely held belief (e.g., O'Brien, 2006), remuneration for translation using CAT tools has been decreasing consistently. Moreover, in most developed markets, clients typically insist that TMs be used and may provide their own proprietary TM data for the translator. To share these linguistic data, TMs have the function to share access both locally and internationally via local networks, servers, and cloud-based applications. There

also exist numerous collections of TM data for commercial and noncommercial use (e.g., the European Commission's TM,[4] and the Translation Automation User Society[5]). While shared TMs have great potential for leveraging existing translation data, thus increasing productivity, issues of ethics (e.g., Kenny, 2011), preservation of consistent quality (Moorkens et al., 2014), and secure storage all become inevitable points of concern that I wish to further emphasize.

With the availability of large bilingual corpora as provided by TMs and used in MT applications, other aspects of translation have come under study in addressing the calls for corpus-based and empirical research (Bowker, 2002; Holmes, 2000) of translation that emerged in the 1990s "to uncover the nature of the translated text as a mediated communicative event" (Baker, 1993, p. 243). Corpus-based approaches have since been used as an evaluative framework for translation quality assessment (Bowker, 2001) and translator training (Bowker, 2003), and can remove subjectivity and ambiguity in that they provide authentic texts that can be used by translators (and evaluators) to justify and verify choices in the translation process and in assessing the severity and impact of translation errors.
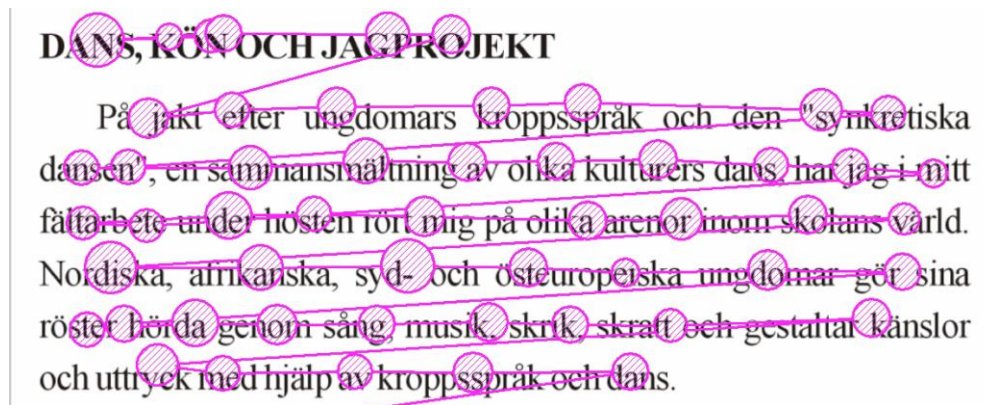
Access to this bilingual data also allows for the study of the universal features of translation as well as language- and direction-specific features of the translation process (e.g., Bowker, 2003; Olohan & Baker, 2000). Such research has uncovered insightful and useful patterns, such as lexical simplification in translation (e.g., Laviosa, 2002), explicitation (e.g., Klaudy, 1998), increased use of standard forms of language and the inescapable influence of the linguistic structure of the source text on translation choices (e.g., Toury, 1995). Similarly, these data also paved the way for comparative multidimensional evaluation of translation quality, including readability and comprehension (e.g., Doherty, 2012) and diagnostic evaluation (e.g., Gaspari et al., 2014) as well as measures of usability (e.g., Doherty & O'Brien, 2014) and cognitive effort (e.g., Doherty, O'Brien, & Carl, 2010).

Following a similar trajectory toward empiricism, translation process studies have emerged to focus on the translator and the process of translation rather than on the end product—see an example in Figure 4. These studies have been gradually mapping the cognitive and psycholinguistic elements of the translation process to uncover more about how translators work, how they use TMs and MT, and how teaching can be refined. This stream of research has incorporated qualitative, quantitative, and mixed-method designs that marry the subjective experience of this complex cognitive processing with more objective observations, all while trying to preserve the ecological validity of a real translation process. Although further development is needed in terms of methodological refinements drawn from other more mature empirical disciplines (see Doherty, in press), this body of research has nevertheless demonstrated unique advantages over psycholinguistics and cognitive sciences, which typically focus on experiments with lower ecological validity and smaller units of text that are of limited use in the real-world contexts of translation.

---

[4] http://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory

[5] https://www.taus.net/

DANS, KÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den "synkretiska dansen", en sammansmältning av olika kulturers dans, har jag i mitt fältarbete under hösten rört mig på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar gör sina röster hörda genom sång, musik, skrik, skratt och gestaltar känslor och uttryck med hjälp av kroppsspråk och dans.

Figure 4. An example of a typical reading pattern recorded by an
eye tracker. Retrieved from http://en.wikipedia.org/wiki/Eye_tracking.

Translation process studies have incorporated keystroke logging (e.g., Jakobsen & Schou, 1999; Van Maes & Leijten, 2006), eye tracking (e.g., Doherty et al., 2010; Dragsted, 2010; Jensen, 2008), brain imaging (e.g., Grabner, Brunner, Leeb, Neupera, & Pfurtscheller, 2007) and continue to present researchers with opportunities to further explore the cognitive aspects of translation (e.g., Göpferich, Jakobsen, & Mees, 2008; Shreve & Angelone, 2010). From this body of relatively recent scholarship, tangible results can already be found in the form of insights into translation subprocesses (e.g., Göpferich et al., 2008; Mossop, 2001), differences between professionals and amateurs (e.g., Dragsted, 2010), and translators' interactions with CAT tools (e.g., O'Brien, 2008). These examples are but a few of those that have yielded considerable contributions to the evidence-based teaching and practice of translation.

The Changing Product of Translation

Translation has traditionally come in the form of literary, religious, political, and technical texts. These well-defined genres have expanded to include commercial content (e.g., marketing, product descriptions, patents, support documentation, and business communications) as well as a wider range of technical genres such as scientific research, medical and pharmaceutical documentation, and patient information. Although these areas have traditionally enjoyed continuous growth, since the 1990s, an unprecedented need has arisen to translate digital content such as websites, computer software, technical documentation, video games, and subtitles. With such a wide variety of content, there is also a particular focus on the requirements of specific audiences in geographic and linguistic locales, often referred to as localization.

Often seen as an extension of traditional translation processes, localization can be characterized in terms of the three interconnected features of the product to be localized: "linguistically as translating a product to suit the target users, technically as

adjusting technology specifications to suit the local market, and culturally as following the norms and conventions of the target community" (Chan, 2013, p. 347).

The text types and formats of localized content differ considerably from traditional texts in that the former contains domain-specific neologistic terminology and language conventions, computer code, and unique file formats and structures that are also often specific to languages and regions. Thus, translators working with such content require specialized training to effectively deal with these extralinguistic features, identify translatable elements (Pym, 2010), and navigate complex software functionality and usability requirements—for example, spacing constraints on websites and text-embedded images.

Furthermore, unlike traditional texts, digital content tends to be more perishable in nature owing to the need to update information on- and off-line in a regular and continuous fashion. Cronin (2013) notes a move from "content being rolled out in a static, sequential manner" to translated content being "integrated into a dynamic system of ubiquitous delivery" (p. 498). These "living texts" (O'Hagan, 2007) mix linguistic and sociocultural information with technical content that needs to be carefully localized to specific market regions with unique requirements, functionality, and expectations, especially for software and video games (Chandler, 2005).

In line with the growth in the amount and diversity of content to be translated, globalization and expanding international markets have resulted in more languages requiring translation. In the early 2000s, the most common language combinations were from English into French, Italian, German, Spanish, Brazilian Portuguese, and Japanese (Chan, 2013). However, since then, sustained growth on a global scale, especially in Asia, has seen translation into tens of languages and hundreds of regions. A case in point is Apple, which currently localizes into about 40 languages across 150 countries with text input methods for 50 languages and their variants[6]—a model that is being viewed as the leading approach to technology-enabled simultaneous global distribution.

Recent industry data show the localization industry alone growing at an average rate of 30% each year, resulting in the proliferation of localization-specific courses at universities and professional bodies and within large companies and organizations (Chan, 2013). Much of this burgeoning digital content is audiovisual translation, principally concerning subtitling, accessibility (e.g., Gambier, 2013), and reception (e.g., Sasamoto & Doherty, 2015). Audiovisual translation, too, has seen the sometimes seamless, sometimes haphazard integration of TM and/or MT into existing proprietary and open-source audiovisual translation software (see Figure 5). Applications range from the standard usage of TMs for subtitling to using full MT (e.g., Armstrong et al., 2007; Müller & Volk, 2013). Significant quality issues include a substantial and lingering limitation to widespread application due to the vast variation in genres and user needs, especially when some users of machine-translated subtitles may be more vulnerable to errors—for example, viewers with hearing impairments.

---

[6] http://developer.apple.com/internationalization/
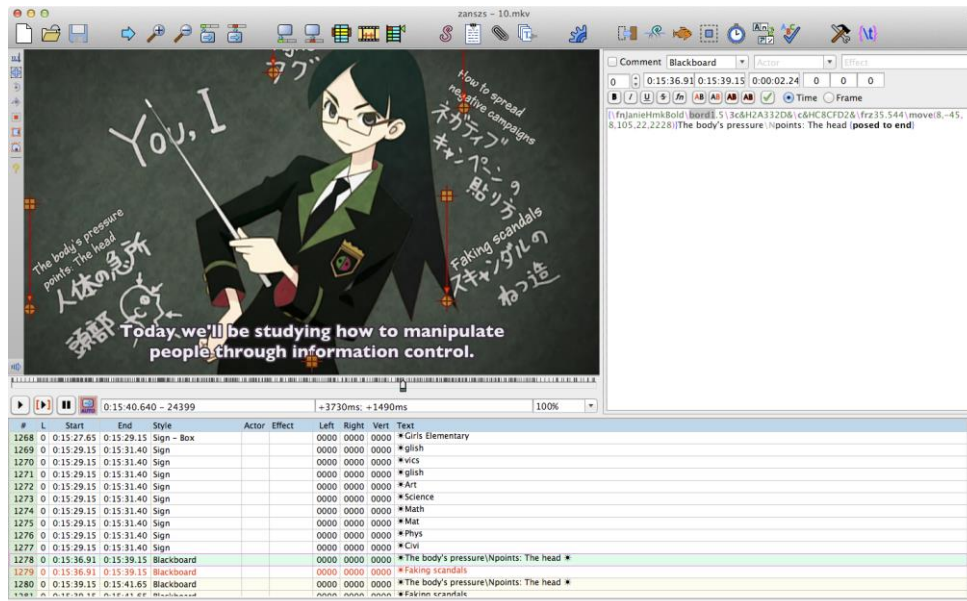
Figure 5. An example of open-source audiovisual translation software widely used in online communities: Aegisub. Retrieved from http://www.aegisub.org.

Translation Technologies and Quality

Despite the widespread and diverse adoption of MT in research and practice, most machine-translated content still requires some form of human intervention to edit the MT output to the desired level of quality and/or to verify its quality before publication, dissemination, product release, legal compliance, and so on. This question of quality, to which I now turn, has been extensively researched in the academic literature on translation and, more recently, within the translation industry given the application of the question of quality to translations produced by machines.

Throughout the long-standing debate on *what is a good (or bad) translation*, I propose that a dichotomy between accuracy and fluency is apparent across translation theory, translation technology, and in the translation industry in one guise or another, where *accuracy* typically denotes the extent to which the meaning of the source text is rendered in its translation, and *fluency* denotes the naturalness of the translated text in terms of the norms of that language. The primary goal of assessing translation quality is ensuring that a specified level of quality is reached, maintained, and delivered as part of the translation product. The debate on translation quality (e.g., House, 1997; Nord, 1991; Reiss, 2000) was far from being resolved prior to the advent of TM and MT, and, unsurprisingly, the widespread adoption of such translation technologies has only added fuel to a renewed debate on translation quality assessment, pricing for MT in the industry, and risks to everyday users.

In terms of quality assessment of MT, the industry departs from traditional academic debate due in part to a vast divergence between research and practice on this topic and also to the need for resource-

efficient means of quality assessment. Although much human evaluation of MT is carried out under the adequacy and fluency paradigm (e.g., Koehn, 2010), it remains resource-intensive and has led to the development of automatic evaluation metrics—algorithms that assess MT quality based on its comparison to a human translation by counting the number of matching words or the number of edits required to enable the MT output to match that of a human translator. Although such means of assessing quality is far from the sophistication of human judgment, it provides a *quick and dirty* solution that is especially valuable in research and development.

Automatic evaluation metrics have since become more commonplace in industry applications (e.g., in cloud-based MT systems such as KantanMT[7]), yet awareness of what they can and cannot measure remains a critical issue that cannot be understated. The absence and unintentional misuse of quality assessment in MT often occurs, and users consequently make uninformed decisions leading to incorrect judgments as to how suitable the machine-translated content is for dissemination. A simple Web search yields an endless list of examples of "bad" MT by everyday users, who are largely unable to assess its quality due to the language barrier and absence of reliable indicators, and may therefore have to blindly trust in its quality. While examples in restaurants and on billboards may be humorous (see Figure 6), MT has also gained a foothold in commercial and public-service translation, where it is increasingly being used in schools, hospitals, and public services in some countries in a desperate attempt to make content available in more languages, where, once again, human translation remains costly and slow (e.g., Randhawa, Ferreyra, Ahmed, Ezzat, & Pottie, 2013; Turner, Bergman, Brownstein, Cole, & Kirchoff, 2014).

Figure 6. An everyday example of bad machine translation: The source text reads "restaurant" in Chinese. Retrieved from http://www.flickr.com/photos/chijs/2647542165.

---

[7] http://www.kantanmt.com/features.php

Although substantial improvements in the quality of commercial MT systems are clearly evident, even the best contemporary MT systems frequently produce errors that require some degree of human intervention. This method of fixing MT output, known as *post-editing*, has become significant in translation research and throughout the industry on a global scale (DePalma, 2013). Much like the push to use TMs experienced in the 1990s and 2000s, translation buyers, hesitant to fully rely on MT, are implementing post-editing incrementally in the face of budget constraints, increased time pressure for project turnaround, and a trend toward the increased casualization of the translation profession.

Rates for post-editing tend to be even lower than translation with CAT tools, often by as much as 60% depending on the market and location (DePalma, 2013), yet the range of its applications is quite diverse. It is often the case that different levels of post-editing (light and heavy) are required to reach a designated level of quality: "gisting" (e.g., for comprehension of the main points of a text); medium quality for internal communications, knowledge, and information sharing (e.g., corporate communications across multiple sites, sharing drafts); and high-quality publishable content for direct public consumption. In addition to the various levels of post-editing, translators must master a new skill set of language-specific linguistic and technical techniques that may not be readily available to traverse the learning curve associated with post-editing MT output.

Society: Professional and Everyday Translators

Evident from the previous examples of the changes translation technologies have brought to *what* is being translated and *how* it is being translated, technologies also have changed the *who* of translation in that such technologies have opened up access and interest to translation, especially with regard to user-generated content, social media, and audiovisual translation. Indeed, one of the most substantial technological developments of the past decade has been the shift from desktop computing to distributed and ubiquitous computing (Dennis & Urry, 2007), a trend that has enabled the flourishing of Web 2.0 technologies, also known as the "user-generated web" (van Dijck, 2009).

The rise of this user-participatory culture (Jenkins, 2006) and the complex relationship between cognitive surplus (Shirky, 2010) and online social capital (Shah, Kwak & Holbert, 2001), added to the availability of translation technologies within the open-source community, has led to everyday users with varying degrees of foreign language proficiency functioning as amateur and volunteer translators: translating online content, working on large online projects, and even evaluating the quality of translations for their area of interest (e.g., social media, video games, animation). This phenomenon has had considerable impact in research and industry circles alike, leading to the widespread recognition within the translation community of specialized terms such as "user-generated translation" (O'Hagan, 2009), online "community translation" (O'Hagan, 2011) and "open translation" (DePalma & Kelly, 2008). Undoubtedly, such practices pose an additional threat to professional translators who have expressed widespread concerns about the quality (e.g., O'Hagan, 2013), and ethics of this digital ontogenesis (e.g., Drugan, 2011).

In addition to this willing and able online workforce of amateur translators, Web 2.0 technologies have opened the door to more users to access the Internet and actively create and share their own

content, which, in turn, is likely to need translation to reach a wider global audience—for example, blogging, social media, and technical support fora (e.g., Mitchell, O'Brien, & Roturier, 2014). It is for such user-generated content that users with proficiency in foreign languages become volunteer and amateur translators of their own and other users' content (see O'Hagan, 2009). Some incarnations of this so-called crowd-sourced translation have come in the form of nonprofit ventures such as the Wikipedia movement,[8] Translators without Borders,[9] and the Rosetta Foundation,[10] while others are entirely commercial operations where crowd sourcing is used as part of the marketing and/or distribution campaign for the brand, product, or service. Facebook, for example, adopted a crowd-sourcing model to allow its users to translate content from English, in which they had various degrees of proficiency, into their own native language communities (Kelly, Ray, & DePalma, 2011).

Outputs from crowd-sourced translation come in many of the same forms as traditional forms of translation, from traditional text documents, to websites, technical support documentation, instruction manuals, and audiovisual translation. Fan subtitling of popular TV programs and movies, known as "fansubs" (O'Hagan, 2009), has become a mainstream alternative to existing subtitles that fans claim can be lackluster due to the translation being carried out by professional translators who are not fans themselves. Actual and perceived censorship in official translations and subtitles are also bypassed by the sheer popularity of fan-created alternatives that are freely available on the Internet and created by amateur, volunteer translators using open-source translation technologies and techniques freely and often loosely adopted from translation studies literature—for example, presentation and timing of subtitles. Freely available (but not actually free) online technologies such as Google Translator Toolkit[11] even provide TM and MT functionalities in addition to integrated instant messaging, shared calendars, and cloud-based storage solutions, offering a comprehensive, "professional" suite of tools that can be used by amateur translators for a plethora of crowd-sourcing endeavors.

Finally, moving beyond the use of translation technologies by professional and amateur translators, everyday users also have found MT systems becoming household names—for example, Google Translate and Microsoft Bing, with Google boasting a growing user base of more than 200 million each day (Shankland, 2013). The usage scenarios of everyday users range from personal tasks such as searching for information on travel, shopping, technical support, and language learning to commercial product and market research, communicating with customers and suppliers, and opening up new markets. Various professions, including teachers and health care professionals, use freely available online MT so that they can communicate with their clients who do not speak their language—a trend especially pronounced with large-scale migration and displacement. Once again, issues of quality, legality, responsibility, and remuneration all come into play.

Although the need for translation in such cases is clear, the use of freely available online MT systems is a cause for grave concern, especially in sensitive intercultural scenarios where professional

---

[8] http://en.wikipedia.org/wiki/Wikipedia:Translate_us

[9] http://translatorswithoutborders.org/

[10] http://www.therosettafoundation.org/

[11] https://translate.google.com/toolkit/

translation (and interpreting) services are a necessity. However, given budgetary constraints and the reactive nature of providing for new and emerging languages to new geographic locations, it can take time for the provision of professional services to come into place, if they are provided at all. In such cases, many everyday users can, and do, choose MT for professional and personal use and remain unaware of the strong potential for poor quality resulting in misunderstanding, miscommunication, and liability.

Conclusion: The Obfuscation of Human and Machine Translation

In exploring the impact of translation technologies on international communication from an interactionist perspective, the effects on the translation process, its products, and its place in society are all remarkably palpable. Technological developments in the early 1990s led to the widespread uptake of CAT tools, chiefly TMs, which have created an increase in productivity and consistency in translation but a decrease in remuneration, control, and risks to overall quality. TMs then paved the way for state-of-the-art MT systems that use human translations to emulate the results of the translation process and deliver output in speeds and volumes that will never be achieved by human translators alone. MT, however, is not without its own risks to quality, misrepresentation, and misuse, and it presents another force that translators must contend with as the fixing of machine-translated output becomes the bread and butter of many professional translators.

Moreover, as the sophistication of MT improves, its reliance on human translation data is becoming more difficult to identify as the lines between human and machine are continually blurred and professional translators become more reliant and embedded into the translation process that they had hitherto controlled. This is compounded by the explosion of amateur, volunteer translators making use of such tools to diffuse the rapidly growing amount of digital content created on a daily basis in many languages, in many countries, and for many purposes. In the wake of TMs and MT software, the need for technological competencies for professional translators to remain on top, if not ahead, of change has never been more evident than it is now. With informed and effective use of TMs and MT, many of the known issues and shortcomings of these technologies can be overcome, especially in terms of translation quality, to somewhat mitigate the downward trend in pricing for translation services in line with tighter budgets and deadlines. Further empirical evidence of the effects that these tools have on productivity, consistency, and quality will add value to negotiations of fair and appropriate pricing and evidence-based best practices within the industry and academe—an agenda that is in need of much more collaborative attention.

However, these new technologies have, in turn, allowed for the creation of novel content types and newly created professional translation-related roles in the course of their own development—for example, localization, post-editing, project management, and quality assessment—and they allow (machine) translation to reach languages that were hitherto neglected due to perceived insufficient commercial viability and demand. This is a provision that many users are content with, even if the MT output is not of the best quality, because it is simply better than nothing at all.

By extension, then, the technological developments in the form of TMs and MT have had, and continue to have, considerable widespread repercussions for translators and nontranslators alike across

everyday personal and professional scenarios, where the visibility of the human translator has been opaqued by a growing selection of relatively easy-to-use and online MT systems that do not readily show users where their translations have come from and how good the quality is. To the everyday user, MT has become a household name under the guise of Google Translate and, to a lesser extent, Microsoft Bing. Such users are becoming increasingly accustomed to being able to access "free" translation services at the touch of a button as the presence of MT becomes much more commonplace and translation ergo becomes less valued and visible.

In looking ahead, what remains unclear is the particular roles that translators and everyday users of translation will play in an increasingly technology-dependent globalized society. As translation technologies intersect and sometimes subsume the translation process entirely, an important factor in moving toward the effective use of these technologies and in preparing for future changes is a critical and informed approach in understanding what such tools can and cannot do and how users should use them to achieve the desired result. It is here that I insist upon the emergent need for the fundamental awareness of and accessible education for translation technologies, their strengths and weaknesses, and their impact on international and intercultural communications for all stakeholders, including translators, buyers and sellers of translation services, and, most of all, the everyday user who is the most unaware and vulnerable.

References

Alcina, A. (2008). Translation technologies: Scope, tools and resources. *Target, 20*(1), 79–102.

Armstrong, S., Way, A., Caffrey, C., Flanagan, M., Kenny, D., & O'Hagan, M. (2007). Leading by example: Automatic translation of subtitles via EBMT. *Perspectives: Studies in Translatology, 14*(3), 163–184.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1996). *Machine translation: An introductory guide*. Oxford, UK: Blackwell.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). Amsterdam, The Netherlands: John Benjamins.

Bédard, C. (2000). Mémoire de traduction cherche traducteur de phrases [Translation memory seeks sentence translator]. *Traduire, 186*, 41–49.

Bowker, L. (2001). Towards a methodology for a corpus-based approach to translation evaluation. *Meta, 46*(2), 345–364.

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

Bowker, L. (2003). Corpus-based applications for translator training: Exploring the possibilities. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 169–184). Amsterdam, The Netherlands: Rodopi.

Bowker, L. (2005). Productivity vs quality? A pilot study on the impact of translation memory systems. *Localisation Focus, 4*(1), 13–20.

Chan, S. (2013). Approaching localization. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 347–362). London, UK: Routledge.

Chandler, H. (2005). *The game localisation handbook.* Hingham, MA: Charles River Media.

Christensen, T., & Schjoldager, A. (2010). Translation-memory (TM) research: What do we know and how do we know it. *Hermes—Journal of Language and Communication Studies*, *44*, 89–101.

Cronin, M. (2013). Translation and globalization. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 491–502). London, UK: Routledge.

Dennis, K., & Urry, J. (2007). *The digital nexus of post-automobility*. Lancaster, UK: Lancaster University.

DePalma, D. (2013). *How to add post-edited MT to your service offerings.* Cambridge, MA: Common Sense Advisory.

DePalma, D., Hegde, V., Pielmeier, H., & Stewart, R. (2013). *The languages services market.* Cambridge, MA: Common Sense Advisory.

DePalma, D., & Kelly, N. (2008). *Translation of, for, and by the people.* Lowell, MA: Common Sense Advisory.

Doherty, S. (2012). *Investigating the effects of controlled language on the reading and comprehension of machine translated texts* (Unpublished doctoral dissertation). Dublin City University, Dublin, Ireland.

Doherty, S. (in press). Improving standards in translation quality assessment with psychometric principles. In D. Kenny (Ed.), *Human issues in translation technology: The IATIS yearbook*. London, UK: Routledge.

Doherty, S., & Kenny, D. (2014). The design and evaluation of a statistical machine translation syllabus for translation students. *The Interpreter and Translator Trainer*, *8*(2), 295–315.

Doherty, S., & O'Brien, S. (2014). Assessing the usability of raw machine translated output: A user-centered study using eye tracking. *International Journal of Human-Computer Interaction*, *30*(1), 40–51.

Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine Translation*, *24*(1), 1–13.

Dragsted, B. (2010). Coordination of reading and writing processes in translation. In G. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 41–61). Amsterdam, The Netherlands John Benjamins.

Drugan, J. (2011). Translation ethics wikified: How far do professional codes of ethics and practice apply to non-professionally produced translation? *Linguistica Antverpiensia, 10*, 111–126.

Esselink, B. (2000). *A practical guide to localization.* Amsterdam, The Netherlands: John Benjamins.

Gambier, Y. (2013). The position of audiovisual translation studies. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 45–59). London, UK: Routledge.

Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives: Studies in Translatology, 23*(3), 333–358.

Gaspari, F., Toral, A., Lommel, A., Doherty, S., van Genabith, J., & Way, A. (2014, May 26). *Relating translation quality barriers to source-text properties.* Paper presented at the International Conference on Language Resources and Evaluation, Reykjavik, Iceland.

Göpferich, S., Jakobsen, A., & Mees, I. (Eds.). (2008). *Looking at eyes: Eye-tracking studies of reading and translation processing*. Copenhagen, Denmark: Samfundslitteratur.

Grabner, R., Brunner, C., Leeb, R., Neuper, C., & Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Research Bulletin, 72*(1), 57–65.

Hearne, M., & Way, A. (2011). Statistical machine translation: A guide for linguists and translators. *Language and Linguistics Compass*, *5*, 205–226.

Heyn, M. (1998). Translation memory: Insights and prospects. In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in diversity* (pp. 123–136). Manchester, UK: St. Jerome.

Holmes, J. (2000). The name and the nature of translation studies. In L. Venuti (Ed.), *The translation studies reader* (pp. 172–185). London, UK: Routledge.

House, J. (1997). *Translation quality assessment.* Tübingen, Germany: Gunter Narr Verlag.

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

Internet World Stats. (2015). *Usage and population statistics: World users by language.* Retrieved from: http://www.internetworldstats.com/stats7.htm

Jakobsen, A., & Schou, L. (1999). Translog documentation. In G. Hansen (Ed.), *Copenhagen studies in language, 24* (pp. 151–186). Copenhagen, Denmark: Samfundslitteratur.

Jenkins, H. (2006). *Convergence culture: Where old and new media collide.* New York, NY: New York University Press.

Jensen, C. (2008). Assessing eye-tracking accuracy in translation studies. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Copenhagen studies in language, 36* (pp. 157–174). Copenhagen, Denmark: Samfundslitteratur.

Kelly, N., Ray, R., & DePalma, D. (2011). From crawling to sprinting: Community translation goes mainstream. *Linguistica Antverpiensia, 10*, 75–94.

Kenny, D. (1996). It looks for all the world as if Günter Grass writes in English. *Translation Ireland*, *10*(3), 12–13.

Kenny, D. (2011, June 4). *The ethics of machine translation.* Paper presented at the New Zealand Society of Translators and Interpreters annual conference, Auckland, New Zealand.

Kenny, D., & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, *8*(2), 276–294.

Klaudy, K. (1998). Explicitation. In M. Baker (Ed.), *Encyclopedia of translation studies* (pp. 80–85). London, UK: Routledge.

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

Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Amsterdam, The Netherlands: Rodopi.

Mitchell, L., O'Brien, S., & Roturier, J. (2014). Quality evaluation in community post-editing. *Machine Translation, 28*(3–4), 237–262.

Moorkens, J., Doherty, S., Kenny, D., & O'Brien, S. (2014). A virtuous circle: Laundering translation memory data using statistical machine translation. *Perspectives: Studies in Translatology*, *22*(3), 291–303.

Mossop, B. (2001). *Revising and editing for translators.* Manchester, UK: St. Jerome.

Müller, M., & Volk, M. (2013). Statistical machine translation of subtitles: From OpenSubtitles to TED. In I. Gurevych, C. Biemann, & T. Zesch (Eds.), *Language processing and knowledge in the Web* (pp. 132–138). Berlin, Germany: Springer.

Nord, C. (1991). Scopos, loyalty and translational conventions. *Target, 3*(1), 91–109.

O'Brien, S. (2006). Eye-tracking and translation memory matching. *Perspectives: Studies in Translatology*, *14*(3), 185–205.

O'Brien, S. (2008). Processing fuzzy matches in translation memory tools—an eye-tracking analysis. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Copenhagen studies in language, 36* (pp. 79–102). Copenhagen, Denmark: Samfundslitteratur.

O'Hagan, M. (2007). Video games as a new domain for translation research: From translating text to translating experience. *Revista Tradumàtica, 5*, 1–7.

O'Hagan, M. (2009). Evolution of user-generated translation: Fansubs, translation hacking and crowdsourcing. *Journal of Internationalization and Localization, 1*(1), 94–121.

O'Hagan, M. (2011). Community translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond*. Linguistica Antverpiensia, 10*, 11–23.

O'Hagan, M. (2013). The impact of new technologies on translation studies: A technological turn? In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 503–518). London, UK: Routledge.

Olohan, M., & Baker, M. (2000). Reporting *that* in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141–158.

O'Reilly, T. (2005, September 30). *What is Web 2.0: Design patterns and business models for the next generation of software.* Retrieved from http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html

Petukhova, V., Agerri, R., Fishel, M., Georgakopoulou, Y., Penkale, S., del Pozo, A., Maučec, M., Volk, M., & Way, A. (2012, May 23). *SUMAT: Data collection and parallel corpus compilation for machine translation of subtitles.* Paper presented at the International Conference on Language Resources and Evaluation, Istanbul, Turkey.

Pym, A. (2010). *Exploring translation theories.* London, UK: Routledge.

Randhawa, G., Ferreyra, M., Ahmed, R., Ezzat, O., & Pottie, K. (2013). Using machine translation in clinical practice. *Canadian Family Physician*, *59*(4), 382–383.

Reinke, U. (2013). State of the art in translation memory technology. *Translation: Computation, Corpora, Cognition*, *3*(1), 27–48.

Reiss, K. (2000). *Translation criticism: The potentials and limitations.* Manchester, UK: St. Jerome.

Roturier, J. (2009). Deploying novel MT technology to raise the bar for quality: A review of key advantages and challenges. In L. Gerber, P. Isabelle, R. Kuhn, N. Bemish, M. Dillinger, & M. Goulet (Eds.), *Proceedings of the twelfth Machine Translation Summit* (pp. 1–8). Ottawa, Canada: Association for Machine Translation in the Americas.

Sasamoto, R., & Doherty, S. (2015). Towards the optimal use of impact captions on TV programmes. In M. O'Hagan & Q. Zhang (Eds.), *Conflict and communication: A changing Asia in a globalising world* (pp. 210–247). Bremen, Germany: EHV Academicpress.

Shah, D., Kwak, N., & Holbert, R. (2001). Connecting and disconnecting with civic life: Patterns of Internet use and the production of social capital. *Political Communication*, *18*(2), 141–162.

Shankland, S. (2013, May 18). *Google Translate now serves 200 million people daily*. Retrieved from http://www.cnet.com/news/google-translate-now-serves-200-million-people-daily/

Sharoff, S. (2007). Classifying Web corpora into domain and genre using automatic feature identification. *Cahier du Cental, 5*, 1–10.

Shirky, C. (2010). *Cognitive surplus: Creativity and generosity in a connected age.* London, UK: Allen Lane.

Shreve, G., & Angelone, E. (Eds.). (2010). *Translation and cognition*. Amsterdam, The Netherlands: John Benjamins.

Snell-Hornby, M. (2006). *The turns of translation studies.* Amsterdam, The Netherlands: John Benjamins.

Toury, G. (1995). *Descriptive translation studies and beyond*. Amsterdam, The Netherlands: John Benjamins.

Turner, A., Bergman, M., Brownstein, M., Cole, K., & Kirchhoff, K. (2014). A comparison of human and machine translation of health promotion materials for public health practice: Time, costs, and quality. *Journal of Public Health Management and Practice*, *20*(5), 523–529.

van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture and Society*, *31*, 41–58.

Van Maes, L., & Leijten, M. (2006). Logging writing processes with Inputlog. In L. Van Maes, M. Leijten, & D. Neuwirth (Eds.), *Writing and digital media* (pp. 158–165). Oxford, UK: Elsevier.

W3Techs. (2015). *Usage of content languages for websites*. Retrieved from
        http://w3techs.com/technologies/overview/content_language/all