Author accepted manuscript. Please cite as:

Kenny, Dorothy (2019) 'Machine Translation', chapter 26 in Piers Rawling and Philip Wilson (eds) *The Routledge Handbook of Translation and Philosophy*, London and New York: Routledge, 428-445.

26

Machine translation

Dorothy Kenny

1. Introduction

Machine translation (MT) is the automatic translation of text from one human language into another. From ambitious if unpromising beginnings in the aftermath of the Second World War, MT had become ubiquitous and massively used by the first decade of the new millennium. Often lauded for its democratizing effects (Boitet et al. 2010; Goltz 2017) and its contribution to the maintenance of (online) linguistic diversity (Cronin 2013: 59), MT can be equally seen as underwriting myths of universal meaning and linguistic transparency, and symptomatic of a purely instrumentalist vision of language (see, for example, Raley 2003; Cronin 2013). Somewhat paradoxically, it is also often construed as supporting the continuing cultural hegemony of English (Raley 2003; Poibeau 2017: 168).[1] A key technology for wealthy global corporations (Poibeau 2017: 6), MT is at the same time implicated in the declining fortunes of many freelance human translators (Moorkens 2017). It is no wonder then that MT has, at times, been controversial in translation studies. It

[1] This view stems both from Warren Weaver's (1949) strategic description of foreign texts as texts that were 'really written in English' and from the fact that Google Translate systematically used English as a pivot language in its statistical machine translation systems.

remains, however, an area of great intellectual interest, not least because the way MT developers approach their task can both reflect and help construct understandings of language, meaning and translation in the world at large (Kenny 2012a), but also because it is primarily through the interface with MT that translation studies can engage with some of the most pressing questions of our time, questions linked to the resurgence of interest in artificial intelligence, and to the future of human labour.

In this chapter we first present some basic definitions of MT, and sketch its historical development from the 1940s to the present day. We then outline the main approaches to MT, before addressing some of the critical issues mentioned above. The chapter ends with some brief recommendations for practice recently espoused in the literature and with a very tentative look to the future.

In the definition of MT with which we opened this chapter, 'text' is understood as written text. Although automatic speech translation has become widely available, it normally draws on complementary technologies that first convert a speech signal into a written text in the same language (in a speech recognition phase), and then use conventional MT techniques to convert that written text into written text in a second, target, language. The translated written output may be converted into target-language speech again using speech synthesis technology, but the interlingual conversion phase—Jakobson's (1959) *translation proper*—even in so-called 'speech-to-speech translation' remains one in which operations are carried out solely on written language.[2] Recent experiments have, however, seen spoken language, in the form of

[2] The term 'spoken-language translation' is used when spoken language is automatically transcribed using speech recognition technology, and the resulting transcript is then automatically translated into a written text in the target language (see, for example, Cettolo et al. 2016). It differs from speech-to-speech translation in that there is no final speech synthesis phase.

Spanish audio files, automatically translated into English written text without prior transcription, using techniques based on artificial neural networks (Reynolds 2017). Similar techniques have also been used to integrate information from images when generating translations of written texts (see, for example, Elliott et al. 2017). In such ‘multimodal’ MT however, the image is an informational input to translation, but there is little mention thus far of automatically translating the image itself, in the same way as a human translator might replace an image in a movie or a videogame, for example, with a more appropriate image in the target culture.[3]

To complicate matters further, contemporary MT is frequently embedded in complex workflows in which it sits cheek by jowl with other technologies normally considered as aids to human translation; at the same time MT has come to rely very heavily on human translation for reasons we will address below. It has thus become commonplace to say that the lines between human translation—especially ‘computer-aided’ human translation—and MT are blurring (see, for example, Kenny 2012a; Doherty 2016; Moorkens 2017). Cronin (2003: 112) goes so far as to describe human translators as ‘*translational cyborgs* who can no longer be conceived of independently of the technologies with which they interact’ (emphasis in the original). But it is worthwhile maintaining the distinction between human and MT for the moment at least, for a number of reasons. For one, only by maintaining the distinction can we do justice to the historical development of MT. From the point of view of contemporary human translators, differentiation is also vital: it is often the most recent provenance of a translation, and especially whether it comes from a machine or a human, that

[3] A well-known example of such a replacement was the substitution of green bell peppers for broccoli in the Japanese version of the Pixar movie *Inside Out*, a replacement motivated by the fact that while American children find broccoli revolting, green bell peppers are the most despised food among Japanese children (Battersby 2015).

marks it as trustworthy or not (Karamanis et al. 2011). What is more, much contemporary translation does not use MT at all, and it would be highly inaccurate to conflate human and MT in such cases. Finally, for pedagogical purposes, the benefit of ontologies that make clear distinctions between concepts is obvious, even if those concepts are liable to change over time. In what follows then, we distinguish between 'human translation', a process in which a human being is the primary agent responsible for the conversion of a source-language text into a target-language text, even if that human has recourse to multiple technological tools and resources to assist in the process, and 'machine translation', a process in which the interlingual conversion of text is carried out by a machine, even if the proper functioning of that machine relies on the labour of human beings before or after run-time. The machines in question are, of course, digital computers that run translation programs, but the term 'computer translation' (Weaver 1949) has never gained currency.

2. Historical perspectives

John Hutchins, the preeminent historian of MT, acknowledges seventeenth-century ideas about 'universal' or 'philosophical' languages as important conceptual precursors to modern interlinguas used in some MT efforts, but notes that such early proposals 'should not be considered in any way as constituting embryonic automatic translation systems' (Hutchins 2004: 11). Rather, the early history of MT begins in the 1930s, with the filing of two patents in 1933—one by Georges Artsrouni in France, the other by Petr Trojanskij in Russia—for electromechanical devices that could be used as multilingual 'translation' dictionaries. Of the two, Trojanskij's ideas were more developed. He anticipated a three-stage architecture in which a monolingual human operator would parse a source text using a universal scheme that could capture all possible grammatical functions of words. The operator would then locate source-text words, one by one, in the part of the machine that acted as a translation dictionary, and add the relevant grammatical code for the current use of that word. The machine would output the 'equivalent' word in the target language, along with the grammatical code, information that would, in a third step, be used by a monolingual target-language speaker, to create a morphologically correct target text. Although Trojanskij anticipated a third human in the chain, namely a bilingual editor who would attend to "meaning" and "literary finishing", he later dropped the requirement that this person be bilingual (Hutchins 2004). Not only does Trojanskij's work thus stand as a forerunner to the three-stage MT architectures that would be developed later in the twentieth century, it also anticipates a debate on the role of bilingual humans in MT workflows, a debate that continues to this day.

Trojanskij's proposed translating machine differed, however, from the MT systems that were to come to prevalence later in the century, on one very important count: he envisaged a special-purpose translating machine; but it was general-purpose digital computers, unknown to Trojanskij at the time of his death in 1950, that were to dominate the field. Indeed, MT, or 'mechanical translation' as it was often known in its early days, has been described as probably the first non-numeric application of electronic computers (Hutchins 2000:1).

The main impetus for the subsequent early development of MT came in the form of a famous Memorandum circulated by Warren Weaver of the Rockefeller Foundation in 1949 (Weaver 1949). In it he bemoans the 'multiplicity of language' that 'impedes cultural interchange between the peoples of the earth, and is a serious deterrent to international understanding' and proposes the use of electronic computers to overcome the ensuing ‘world-wide translation problem’. He sets out four possible avenues that research into computer translation could pursue, in particular to overcome the shortcomings of automatic word-for-word translation (which was becoming possible thanks to the advent of electronic dictionaries) and the challenges posed by ambiguity in natural languages. He thus proposes the use of the immediate context of a given word—effectively a short window of *n* words to the left or right of the word in question—to disambiguate that word. He also proposes that MT take advantage of contemporaneous work on the logical basis of language, as well as cryptographic methods developed during the Second World War, Claude Shannon’s Information Theory and ‘the real but as yet undiscovered universal language’. Although received by some with scepticism, including Norbert Wiener, founder of cybernetics, with whom Weaver had communicated in 1947, Weaver’s ideas are credited with having sparked the first wave of MT research in the early 1950s (Hutchins 2000: 2). His Memorandum is also notable for the role it continues to play today as ‘focal point for skepticism’ (Mitchell 2010: 169), especially in the critical humanities, about MT and many of its attendant assumptions. Weaver’s contribution has been criticised (see, especially, Raley 2003; Mitchell 2010) on many grounds, including its naïve universalism, its dichotomizing approach to language (seen either as logical and functional or alogical and literary), its perceived anglocentricity and its misguided conflation of decryption with translation.

Shortly after the circulation of Weaver’s memorandum, research programmes in MT were launched in the United States and elsewhere. A key event in the early history of MT came in the guise of the first demonstration of a functioning MT system at Georgetown University, on 7 January 1954. Under the guidance of Léon Dostert, and in cooperation with IBM, a Russian-to-English translation system was demonstrated to the waiting press. While many of the reports of the day acknowledged shortcomings of the system, the demo was generally met with great enthusiasm, if not awe, with reporters extolling the virtues of the new ‘electronic brains’ and agreeing that MT systems ‘capable of translating almost everything’ would become available within five years (Hutchins 2006). The Georgetown demo is thus generally viewed as having triggered the first hype cycle in MT history, and instigating what was to become something of an in-joke in translation circles, namely that MT would be ready ‘in five years’ (Wiggins 2017).

The years following the Georgetown demo saw MT research groups spring up in the USA, the Soviet Union, the UK, and elsewhere. Their activities are described in detail in Hutchins (2000). Much of the American research was funded by the CIA and the military and concentrated on Russian-to-English translation, reflecting Cold War anxieties, while research in the Soviet Union appears to have been based on a wider range of languages (Hutchins 2000: 7-8). Different teams had different priorities: some pursued the immediate goals of building and formalizing large bilingual dictionaries, while others were more interested in investigating ‘the internal semantics of the human mind’ (ibid: 2), but all four of Weaver’s avenues of research were explored in some shape or form.

Despite the considerable practical challenges faced by MT pioneers, the years 1954 to 1959 have been characterised by Hutchins (ibid: 6) as ones of ‘innovation and enthusiasm’. By 1960, however, disillusionment was beginning to set in and some research teams had already begun either to disband or to move towards other activities (Hutchins 2000: 9) .The expected benefits of integrating syntactic analysis into MT did not materialise (ibid.) and while early protagonists such as Weaver (1949) had acknowledged the problem of ‘multiple meaning’ in translation, but had been optimistic about solving the problem for restricted domains and through the use of micro-context, it now seemed that the machine’s inability to deal with macro-context would be its downfall.

In 1960, Yehoshua Bar-Hillel argued that MT had reached an impasse, as some semantic ambiguities could not be resolved by reference to a micro-context of *n* words to the left or right of the ambiguous word, or even to neighbouring sentences or entire texts (Bar-Hillel 1960). Rather, what was needed was access to encyclopaedic knowledge, which computers did not have (and Bar-Hillel was sceptical that they could ever acquire all the encyclopaedic knowledge that a typical human has). Bar-Hillel concluded his survey of contemporary MT research on a pessimistic but pragmatic note:

> Fully automatic, high quality translation is not a reasonable goal, not even for scientific texts…Reasonable goals are then either fully automatic, low quality translation or partly automatic, high quality translation. (Bar-Hillel 1960)

The ‘high quality’ in ‘partly automatic, high quality translation’ was to be achieved by the use of human post-editors, who would fix faulty MT output.

The main body blow to MT research came however, in the form of the Automatic Language Processing Advisory Committee (ALPAC) Report in 1966. The Committee had been set up by government sponsors of MT in the United States against a backdrop of mounting bills and disappointing results in MT research. The Committee famously concluded that MT was ‘slower, less accurate and twice as expensive as human translation’ (Hutchins and Somers 1992: 7) and that there was ‘no immediate or predictable prospect of useful machine translation' (ALPAC 1966: 32). The ALPAC Report was widely criticized for being narrowly focused and short sighted (Hutchins 1986), but the damage was done. Funding for MT research in the USA and elsewhere largely dried up. The ALPAC Report remains of interest today however, providing as it does an early example of an MT evaluation campaign, and indicating how extrinsic factors such as the then abundance of human translators and relatively low cost of their services affect the overall assessment of MT. Finally, if Weaver had been concerned in the post-War period about the need for translation to ensure 'cultural interchange between the peoples of the earth', priorities had now shifted to ensuring timely delivery of those translations, and *only* those translations, desired by certain American scientists, and for whom more and more texts were becoming available in English anyway. The Report warns in particular of the possibility of an 'excess of translation', noting that 'Translation of material for which there is no definite prospective reader is not only wasteful, but it clogs the channels of translation and information flow' (ALPAC 1966: 13).

The post-ALPAC period saw reduced levels of activity in MT in the United States, but growing interest in the field in Canada, where official bilingualism had been enshrined in law in 1969 by the Official Languages Act (Canada), and in the then European Communities (now the European Union), where multilingualism was foundational, and translation demands would continue to grow with successive expansions. Translation was a political imperative in such cases, and a matter of citizens' rights, rather than primarily a means of monitoring a rival power's scientific knowledge, as had been the case in much of the government-sponsored MT in the United States before ALPAC.

One area in which MT found a firm foothold in Canada was in the translation of weather forecasts: the Météo system stands as one of the most successful MT implementations of the twentieth century, from the point of view of its longevity—a version of the English-to-French Météo system was in service from 1977 to 2002 (Poibeau 2017: 87)—but also from the point of view of: its productivity, translating up to 45,000 words a day in early the 2000s (Langlais et al. 2005: 84); the quality of its output, which reached accuracy levels of 90% (Hutchins and Somers 1992: 208); and the ideal nature of the use case. Weather bulletins require rapid translation, but have short shelf lives. They are also tedious for humans to translate. Crucially, they are linguistically simple, and draw on a naturally restricted range of vocabulary and grammatical structures. These latter properties are what allow the language of weather bulletins to be described as a 'sublanguage'. Other sublanguages proved less amenable to translation by machines however (Hutchins and Somers 1992: 219).

The European Union for its part, began its long history of engagement with MT in 1976, when the then Commission of the European Communities (CEC) first started using the Systran MT system for English-French translation (other language pairs were added later). Systran's Russian-English system had been deployed by the United States Air Force from 1970; it was used by NASA during the Apollo-Soyuz space mission, and would also be introduced at Euratom in 1976. Its use at the European Commission continued well into the current millennium, but was eventually discontinued in 2010. By this time the European Commission had introduced its own fully in-house MT system known as mt@ec (ECMT 2013), which was in turn superseded by another in-house system, eTranslation, in November 2017 (European Commission 2017). Systran, meanwhile, continues to provide MT services to other major clients, using technology that has changed considerably since it was first developed in the 1960s.

The 1980s in general are notable for a proliferation of commercial systems, in North America and Japan in particular, some of which were designed for the emerging personal computer market, and were described as ‘crude’ and requiring ‘hefty post-editing’ by contemporary reviewers (Bédard 1989). Mainframe systems also continued to be developed for multilingual organisations or to be customised for specific clients, as in the case of Systran installations at Aérospatiale and NATO (Hutchins and Somers 1992: 9).

In most cases, unless machine translations were wanted just for ‘gisting’ or information assimilation purposes, post-editors were used to improve the machine’s output. In the early days post-editors simply identified and corrected faulty machine translations by hand, sometimes working with pen and paper printouts; by the 1980s word processing programs were being used for this purpose (Sereda 1982). In some cases, users attempted to pre-empt problems in automatic translation by pre-editing source texts, and some MT implementations relied on both pre-editing and post-editing (ibid.). Pre-editing sometimes involved implementing ‘controlled languages’ or rules that placed constraints on sentence length, and the vocabulary and structures used in the source text (see Kuhn 2014). In yet other cases, attempts were made to improve MT as it happened. The ALPS Translation Support System was a good example of an early ‘interactive’ system, in which users were asked to resolve ambiguities in source texts during the MT process itself (Bateman 1983).

Addressing the state of the art in MT at the beginning of the 1990s, Harold Somers reflected on post-ALPAC, or ‘second generation’, MT systems, and observed that, despite the more principled approach they took, ‘in the 20 years since ALPAC, the second generation architecture had led to only slightly better results than the architecture it replaced’ (Somers 1993: 232).

An alternative to the overwhelmingly linguistics-oriented systems of the day had first been presented by a team of researchers from IBM in 1988 (Brown et al. 1988a, 1988b) and was subsequently developed in Brown et al. (1990; 1993). The IBM approach was based on statistics and ideas from information theory, following a lead first proposed by Weaver in his 1949 Memorandum. Using the bilingual proceedings of the Canadian Parliament, the IBM researchers showed how probabilistic models of translation could be learned directly from such ‘bitext’, without the use of linguistic knowledge. The approach was initially met with incredulity (see Way 2009), but by the early 2000s statistical MT (SMT) had gathered momentum. It was helped by the growing availability of electronic bitext, but also by increasing computer power and data storage, and the collaborative efforts of developers who made available open-source SMT toolkits from 1998, thus enabling the rapid development of SMT systems by research groups the world over (Way 2009; Koehn 2010a: 18). As has always been the case in MT research, geopolitical and economic factors were also at play. In the aftermath of the attacks on the World Trade Center in September 2001 for example, US defence agencies stepped up funding for MT research for Arabic. The 2000s also saw increased American interest in automatic translation from Mandarin. At the same time, successive expansions of the European Union (and the recognition of Irish as an official language in 2007) meant that translation needs continued to grow: by 2004 the number of official languages in the Union had grown to 20, by 2014 it had reached 24 (DGT 2016). The wider technological environment of the 1990s and 2000s also provided space for MT research to flourish. The rapid growth of the internet, which boasted one billion users by 2005, two billion users by 2010, and three billion by 2014 (internetlivestats 2018), was accompanied by increasing online linguistic diversity. Not only did the increasing number of multilingual web pages contribute to the growing stock of suitable bilingual data from which new statistics-based systems could learn how to translate, but the world wide web itself provided a platform from which MT could be easily delivered to millions of casual users. As early as 1997, Systran was being used to translate web pages retrieved through the AltaVista search engine (Joscelyne 1998). Google Translate was launched in 2006; by 2016 it claimed to have over 500 million users, translating an average of 1 billion words per day across 103 languages (Turovsky 2016). The means by which users could interact with the system had also diversified: integration of complementary technologies meant that speech input was now possible as was the translation of words recognised in images, for some languages at least. Third party software developers could also integrate Google Translate functions into their applications, using the Google Translate API.[4] Google Translate had also built a community of users who would provide feedback to improve the system's output, in line with the trend towards a kind of participatory amateur translation culture much analysed in translation studies (see, for example, O'Hagan 2011). Other major technology companies, for example Microsoft, also developed MT systems in the 2000s, with similar multi-modal input capacity, for in-house, online consumer or third party commercial use (Microsoft 2018).

At the same time as free, online MT was beginning to become a mass-consumer product in the mid to late 1990s however, MT remained of only marginal interest to

[4] Application Programming Interface.

most professional translators.[5] One reason for this was undoubtedly that the output from many of the then available MT systems did not reach quality thresholds required to make it worthwhile post-editing. Another is that the market for computer aids for translators had already been largely occupied by 'translation memory', a technology that allowed efficient reuse of human translations (see Kenny 2011). First used in the software industry, translation memory tools proliferated in the late 1990s and the 2000s (Chan 2017). But while translation memory was conceived as a way of improving, among other things, the productivity of human translators, it also eventually supported efforts to increase automation in the translation industry: on the one hand, translation memory tools enabled bitext to be created in great quantities and in a format that could be easily used in SMT; on the other hand, they provided an editing environment in which MT outputs could be presented to human translators for editing alongside human translations retrieved from memory. In this way, human translation fed into MT, and MT fed into human translation.

SMT found widespread application and was generally considered state of the art for more than a decade from the mid-2000s. It dominated the research agenda and beat all other contenders in the annual MT evaluation campaigns organised from 2004 onwards (see Bojar et al. 2016; Bentivogli et al. 2016c). Performance improvements in SMT plateaued in the mid-2010s however, and there was thus a strong incentive for research teams to explore other avenues. The breakthrough came in 2015/2016 when the leading SMT systems were outperformed by a new, albeit related approach, called neural MT, or NMT (Bentivogli et al. 2016b; Koehn 2017: 6). By the end of 2016,

[5] Poibeau (2017: 222) offers an interesting example of the marginality of MT to the translation efforts of the European Commission, for example. Despite the Commission's longstanding engagement with MT, in 2013, less than 5% of translation at the Commission's Directorate General for Translation benefited from automation.

NMT was being deployed by major technology companies to provide translation across a variety of platforms (see, for example, Google 2016; Microsoft 2016), and other translation users and specialist MT providers had begun shifting to the new technology, assisted in their development efforts by the availability of free, open-source NMT toolkits (see Forcada 2017: 302; Koehn 2017).

3. Main methods

Rule-based machine translation

As has already become apparent, MT has thus far relied on two main approaches: the first approach manipulates linguistic knowledge in the form of handcrafted grammatical and lexical rules. These systems, commonly known as 'rule-based machine translation' or RBMT systems, were state-of-the-art until the turn of the millennium, but rule-based systems continue to be developed and supported, especially for languages for which training data is less readily available. A well-known example of such a system is Apertium (Forcada et al. 2011).

RBMT systems are usually divided into three types: direct, transfer and interlingua (see Hutchins and Somers 1992). Direct systems, as exemplified by pre-ALPAC 'first generation' systems, used dictionary look-up to provide word-for-word translations, and then perhaps some local reordering rules to fix a limited number of obvious errors.

The direct approach was clearly limited, and the above-mentioned 'second generation' systems were based on the idea that a more satisfactory translation could be generated if an initial syntactic, and perhaps also semantic, analysis of source-language sentences was carried out. This analysis would result in an intermediate representation of the source sentence that would be more amenable to translation, as it would go some way towards neutralising the differences between the source and target languages. Rather than attempting to translate the surface forms of a source-language sentence, these systems would convert the more neutral 'intermediate representation' of that sentence to a similarly construed intermediate representation of a sentence in the target language, in the phase known as transfer. From the target language representation, a surface target language sentence could then be generated. This tripartite architecture involving analysis, transfer and generation, thus mirrored the workflow originally proposed by Trojanskij in the 1930s.

In interlingual systems, an analysis of the source-language sentence would be carried out in much the same way as in a transfer system, with one important exception: ideally it would culminate in a representation of the content of the source-language sentence that no longer bore traces of the source language, and from which a sentence in any target language could be generated. In other words, a true interlingua would be natural-language independent. It would involve some kind of universal notation, capable of expressing any meaning expressible in any source or target language covered by the system. But despite the efforts of a number of research teams to merely approximate an interlingua in the 1970s and 1980s (see for example, descriptions of the Rosetta system and Carnegie Mellon University's Knowledge Based MT in Hutchins and Somers (1992), and Alan Melby's (1995: 43-9) discussion of attempts to develop universal sememes), the approach has never been deployed on a large scale (Poibeau 2017: 32), with researchers ultimately finding it impractical and even theoretically misguided (Melby ibid.).[6]

Data-driven machine translation

The second broad approach to MT is data-driven, sometimes also called ‘corpus-based’. As suggested above, the basic tenet of data-driven MT is that translation knowledge can be learned directly from parallel corpora (or ‘bitexts’), that is, collections of source texts aligned with their human translations. Data-driven MT, like RBMT, can be divided into three sub-types: example-based MT, statistical MT and neural MT.

Simplifying greatly (for a more detailed overview, see Somers 1999), example-based MT involved the automatic identification and extraction from bitexts of equivalent, usually linguistically motivated, chunks in the source and target texts. Such equivalent chunks would then be stored in a database. When a new text was to be translated, it would first have to be segmented into sentences and then chunks. If the training corpus offered no existing translation of the source sentence, then the system could resort to looking for translations of chunks of the new source sentence in the database. If adequate matches were found, their target-language equivalents were taken and ‘recombined’ to make a new target-language sentence. The approach thus had much in common with translation memory, and EBMT functionality was integrated into at least one translation memory tool, Déjà Vu, but EBMT was largely eclipsed in MT research by SMT, which dominated the field by the mid-2000s.

[6] Tentative claims about the ‘discovery’ of interlingual representations in neural machine translation have recently been made by researchers at Google (Johnson et al. 2016), but have been met with some scepticism.

SMT, for its part, is conceived as a problem in which the system has to decide on the most probable translation for a given source sentence, based on a probabilistic model (or a combination of several models) of translation that it has learned from a parallel corpus, and on a probabilistic model of the target language, learned from a large monolingual corpus of texts in that language.[7] The 'learning' of these models is done in a phase known as 'training'. In a second phase, called 'tuning', system developers work out the optimal weight that should be assigned to each model to get the best outcome. When the system is called upon to translate a new sentence (in a third phase called 'decoding'), it generates many thousands of hypothetical translations for the input string, and calculates which one is most probable, given the particular source sentence, the models it has learned, and the weights assigned to them. SMT systems thus have a tri-partite architecture and involve much tuning to find the optimal weights for different models.

The original translation models proposed in the early work at IBM were based on unigrams, that is single words, but techniques were subsequently applied that could learn translations not just for single words, but for *n*-grams, that is, strings of one, two, three or *n* words that appear contiguously in the training data used. (For example 'contiguously in' is a bigram in the previous sentence; while 'contiguously in the' is a trigram.) These *n*-grams became known as 'phrases' in SMT circles, hence the description of most SMT systems as 'phrase-based SMT'.[8] Translation 'knowledge'

[7] As Hearne and Way (2011: 206) point out, two separate formulae are available to compute the most probable translation. The first, the noisy-channel model, is a straightforward application of Bayes' theorem. The second, the log-linear model, can express the same computation as the noisy-channel model, but is more flexible.

[8] The term 'phrase' here does not correspond with 'phrase' as used in much linguistic theory, and some authors, e.g. Poibeau (2017), avoid it altogether, preferring the term 'segment-based machine translation'.

in phrase-based SMT was thus recorded in so-called ‘phrase tables’ which would specify a string in the source language and a string in the target language and assign a numerical probability that the latter was a translation of the former.

The relatively limited amount of co-text used to build models, the fact that the *n*-grams are translated largely independently of each other, and that they do not necessarily correspond to any kind of structural unit in linguistic theory, mean that SMT systems can have difficulty handling discontinuous dependencies like that between ‘send’ and ‘back’ in the sentence ‘Send your certificate of motor insurance back’.[9] They are also known to perform poorly for agglutinative and highly inflected languages, as they have no principled way of handling grammatical agreement. Other problems include word drop, where a system simply fails to translate a given source word, and inconsistency, where the same source-language word is translated two different ways, sometimes in the same sentence.

Neural MT systems, like SMT systems, learn to translate from parallel corpora, but do so using very different computational methods. They use artificial neural networks, in which thousands of individual units or artificial ‘neurons’, analogous to neurons in the human brain, are linked to other such neurons, and the activation state of each neuron depends on the stimuli received from other neurons, and the strength or ‘weight’ of the connections between them. As Forcada (2017: 293) points out, the activations of individual neurons do not make much sense by themselves. It is, rather, the activation states of large sets of connected neurons that can be understood as distributed

[9] This is one of the longer ‘separations’ attested in the representative sample of British English that is the British National Corpus (Gardner and Davies 2007: 345).

representations of individual words and the contexts in which they appear, where context can refer both to source-language and target-language ‘co-text’. Training a neural network for MT basically means learning the weights that will result in those distributed representations that can best ensure that the network, when called upon to translate,[10] outputs translations that are as close as possible to the ‘gold standard’ human translations found in the training data. Representations are not built in one go, but rather in successive ‘layers’, where layers are fixed-sized lists (vectors) of numerical quantities. The external layers, which correspond to inputs to and outputs from the network, are open to the human analyst’s scrutiny, but intermediary layers remain ‘hidden’, a point to which we return again below.[11]

NMT thus shares with SMT the fact that systems learn from training data, and learning weights is similar to learning translation probabilities.[12] NMT systems have a simpler, ‘monolithic’, architecture however (see Forcada 2017: 301) and they process full sentences rather than *n*-grams. They are known to handle morphology, lexical selection and word order phenomena (including discontinuous dependencies) better than SMT (see Bentivogli et al. 2016b; Castilho et al. 2017), but they take much longer and much more computing power to train, and usually require dedicated hardware in the form of graphical processing units (Forcada 2017). NMT output can include some deceptively fluent but inaccurate passages however, and because they sometimes translate sub-word strings (in cases where they encounter previously unseen words), NMT systems can also produce non-words in the target language.

---

[10] As in statistical machine translation, the translation phase in NMT is called ‘decoding’.

[11] In the broad field of machine learning (see, e.g., Domingos 2017), the term ‘deep learning’ has come to stand for the use of neural networks with multiple hidden layers.

[12] Indeed Koehn (2017: 5) sees neural MT as a type of SMT.

4. Critical issues

Meaning

As Hutchins (2000: 2) puts it, the 'outstanding problem' of the early days of MT was meaning, 'or more precisely the differences between how different languages expressed the same objects, ideas and concepts'.[13] Rule-based MT systems are, implicitly or explicitly, inclined to see meaning as objective and residing in more-or-less discrete concepts, labelled by expressions that can in turn be combined according to the principle of compositionality and the dictates of syntax. They are thus aligned to symbolism, which has the advantage of analytical transparency, but makes RBMT prone to grounding problems[14] and problems caused by pervasive linguistic ambiguity.[15] And although there is no requirement in artificial intelligence that machines process language in the same way as humans do, RBMT has also been criticised for its cognitive implausibility (Nagao 1984).

Statistical MT, on the other hand, does not trouble itself too much with questions of meaning.[16] Its proponents tend to take it as given that the source and target texts in the

[13] This preoccupation with expressing the 'same meaning' in translation, which is frequently repeated even in contemporary MT literature (for example, Koehn 2010a: 43), was not one that was not necessarily shared with translation theorists working outside machine translation circles, and for many of whom meanings were often language specific or not amenable to capture in static formalisms anyway (see, for example, discussions in Kenny 2012b and Malmkjær 2011).

[14] As Harnad (1990: 335) puts it: 'How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?'.

[15] Ambiguity in MT arises when a word can be assigned more than one interpretation, or a string of words can be assigned more than one syntactic structure, according to the dictionaries and grammars in use. Practically every discussion of 'why translation is difficult for computers' contains a lengthy section on ambiguity. See, for example, Hutchins and Somers 1992; Arnold 2003; and Poibeau 2017.

[16] The pioneers of SMT were, in fact, careful to point out that their approach 'eschews the use of an intermediate mechanism (language) that would encode the "meaning" of the source text' (Brown et al. 1988a: 1).

parallel corpora on which MT engines are trained 'mean the same thing' – the problem of ensuring same meaning having been solved upstream by the human beings who carried out the translations in the first place. And while the phrase translation model that an SMT system learns should contain many useful pairings of source and target language *n*-grams, having no access to meaning, it might also contain many nonsensical pairings. As for the idea that SMT might somehow be cognitively plausible, Hearne and Way (2011: 206) are dismissive, opining that systems that generate target sentences 'by translating words and phrases from the source sentence in a random order using a model containing many non-sensical translations…are not intended to be either linguistically or cognitively plausible'.

But while the pioneers of SMT made few claims about meaning, others working in the broader field of statistical natural language processing see parallels between techniques that attempt to learn about words by observing their distribution in corpora–which is what even *n*-gram based language models can claim to do–and the view of meaning as use promoted by Ludwig Wittgenstein (1968) and J.R. Firth (1957) in particular. (See, for example, Manning and Schütze 1999: 17.) As the state of the art in MT shifts towards neural approaches, allusions to both Wittgenstein and Firth appear to be growing more frequent (Koehn 2017), with commentators like Poibeau (2017: 176 ff.) stressing the contribution that statistical analysis can make to our understanding of meaning.

As the description of neural MT given above might suggest, NMT is consistent with theories in which meaning is seen as associative, relational and distributed. If the meaning of a linguistic form is 'the total network of relations' it enters into, as

Catford (1965: 35) puts it,[17] then an artificial neural network appears to be a good way to represent those monolingual syntagmatic and paradigmatic relations, as well as interlingual relations that can be learned from a parallel corpus. It is even possible to represent relations between linguistic forms and extralinguistic entities, for example images, where features of such images serve to ground automatic translation.[18] Such considerations, along with the fact that NMT relies on computational techniques inspired even if only very loosely (Forcada 2017: 292; Henderson 2010: 379d) by human neurology, have, no doubt, encouraged much of the contemporary hype of NMT, in which 'artificial brains' are said to achieve near human performance in translation. Many MT researchers are eager to counteract such hype however (see, for example, Koehn 2016; Moorkens 2017: 471).

Opacity

Improved performance of MT has come at the price of increased opacity, with systems moving from the relative transparency of rule-based approaches, which explicitly manipulate knowledge, to the somewhat diminished transparency of statistical MT, which while initially difficult for non-statistically oriented researchers to understand (see Way 2009) was 'still comprehensible in its inner workings' (Bentivogli et al. 2016b: 1), to the total opacity of neural MT (ibid.). That the workings of the 'hidden' layers in NMT currently defy analysis is a point that has been made by several researchers, with Koehn and Knowles (2017: 1) observing that the answer to why the training data in a given NMT system lead the system to produce particular outputs is 'buried in large matrices of real-numbered values'.

---

[17] Catford's (1965) work is, of course, inspired by both J.R. Firth and M.A.K. Halliday.

[18] Elliot et al. (2015) give the following example: 'in the German sentence "Ein Rad steht neben dem Haus", "Rad" could refer to either "bicycle" or "wheel", but with visual context the intended meaning can be more easily translated into English'.

Increased opacity, an issue in many areas of machine learning (Domingos 2017), is a particular cause for concern for humans required to work with contemporary MT systems as it can limit their ability to intervene in translation workflows, thus undermining agendas of translator empowerment (Kenny and Doherty 2014). It may also increase the risk of inaccurate translation, or translation that is based on hidden biases. While relative opacity (from the point of view of human translators) remained largely due to 'technical illiteracy' (Burrell 2016) in the case of SMT, and could thus be tackled through translator education (Kenny and Doherty ibid.), the opacity of many contemporary machine learning techniques is rather a property that arises from the nature of the algorithms themselves , and the scale required to apply them (Burrell 2016), and so is much more difficult to solve. The problem affects MT developers as much as anyone else however, and there is thus a strong incentive to develop better analytics for NMT (Koehn and Knowles 2017).

The relationship between human translation and machine translation

Recent advances in data-driven MT have also prompted pessimistic prognoses for the human translation profession, principally from translation industry outsiders. Brynjolfsson and McAfee (2011), for example, proffer translation as an area in which computers have 'raced ahead' of humans. The incremental improvements observed with neural MT have only served to intensify speculation in the press that the days of the human translator are numbered.[19] But such prognoses assume that MT substitutes for human translation, when it is more often the case that human translation

[19] Such prognoses are, incidentally, not supported by other work in labour economics. See Frey and Osborne (2017) and The US Bureau of Labor Statistics (2016), which views translation and interpreting as 'bright future' professions.

complements MT. For one, human translators provide the data on which MT systems are trained; and they are also typically used to post-edit MT output. Although there are use cases in which ‘raw’ MT output is sufficient, such cases remain rare (Moorkens 2017), and industry sources claim that human post-editing of MT is one of the fastest growing sectors of the translation market (Common Sense Advisory 2016). Concerns about deskilling and boredom among humans called upon to post-edit MT output have been raised however, and remuneration can be poor (see Moorkens and O’Brien 2017; Moorkens 2017).

There is also a long history in certain quarters of MT research of attempting to exclude human translators from MT workflows, where ‘human translators’ are understood as professionals, one of whose essential attributes is that they are at least bilingual. This tendency goes back as far as Trojanskij (see above), and resurfaces in attempts to recast the role of the post-editor as that of a ‘monolingual translator’ or to enable crowdsourced ‘monolingual translation’ (see, for example, Koehn 2010b; Chang et al. 2011), even if some of these efforts are motivated by a desire to achieve noble translation goals despite the unavailability of qualified bilingual humans.

In some of the more pessimistic treatments, bilingual human translators thus face extinction, boredom or irrelevance. A number of broad responses to these projections can be observed in the literature however: on the one hand, human translators are advised not to try to compete with machines, but to focus their efforts on the more creative, better-paid segments of the market (Moorkens 2017); on the other hand, researchers have begun investigating how the technological environments in which human translators work can better serve their needs. Here, adaptive MT (Bentivogli et al. 2016a; Farajian et al. 2017), that is, MT that learns from post-editors' corrections, so that they are not forced to make the same edits over and over again, is seen as one of the principal ways in which translator/post-editor experience of MT can be enhanced.[20] This latter approach is favoured by commentators who seek to put human translators back at the centre of technologized professional workflows, in configurations that enable what is now being called 'augmented translation' (Lommel 2017).

5. Conclusion

MT has had a chequered history. At the time of writing, it has moved into a new phase, with neural approaches offering promising results, but still beset by the problems associated with most deep learning.[21] MT also remains a technology that cannot explain or take responsibility for its decisions in the way a human translator might be expected to, and it is still prone to errors that might only be spotted by an informed bilingual human. For these reasons, bilingual humans will continue to be important arbiters in professional workflows that use MT. And although MT is available for the world's most economically important languages, and languages that are of interest to certain national intelligence services, the vast majority of the world's written and spoken languages are not served by any machine translation system. In

[20] Other approaches rely on a version of 'interactive' MT, in which the machine attempts to predict the words that the human translator is currently typing; quality estimation, in which the machine attempts to decide whether an MT output is good enough to propose to a post-editor; and better integration of MT and translation memory (Bentivogli et al. 2016a).

[21] Pontin (2018) lists as the downsides of deep learning the fact that it is 'greedy' (systems require huge quantities of training data), 'brittle' (systems cannot easily cope with scenarios not encountered in their training data), 'opaque' (systems are difficult to debug), and 'shallow' (systems possess no innate knowledge or common sense).

such cases, human translation or non-translation remain the only options, but in the coming years, we can expect data collection initiatives to be accelerated for languages not currently catered for by contemporary data-driven MT but that become of sufficient economic, political or strategic importance to sponsors of MT research.

Related topics

current trends in philosophy and translation; translation theory and philosophy; meaning; cognitive approaches to translation

Further reading

Poibeau, T. 2017. *Machine Translation*. Cambridge, MA/London: The MIT Press. (This book provides a very accessible, non-technical overview of machine translation, covering history, major approaches (linguistic rule-based, statistical and neural), evaluation and commercial systems.)

Forcada, M. 2017. Making Sense of Neural Translation. *Translation Spaces*. 6(2): 291-309. (In this journal article, Mikel Forcada, one of the pioneers of neural machine translation, explains NMT in non-mathematical terms to a translation studies audience, focusing on basic methods, typical outputs and the implications for translators and post-editors.)

Hearne, M. and A. Way. 2011. Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass* 5/5 (2011): 205–26. (In this journal article, the authors explain SMT to a non-technical audience, describing both how language and translation models are trained, how decoding works, and how systems are tuned.)

Koehn, P. 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press. (This textbook is the standard reference for SMT. While its primary audience is students of computer science, some chapters are accessible to non-mathematical audiences. The book can be supplemented by a draft chapter on neural machine translation, updated in September 2017, and available at: https://arxiv.org/pdf/1709.07809.pdf.)

Hutchins, J. (ed.) 2000. *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*. Amsterdam/Philadelphia: John Benjamins Publishing. (This edited volume contains contributions by or about the earliest researchers in machine translation, providing personal recollections not just of scientific endeavours, but also of the political and social contexts in which MT research was conducted in the 1950s and 1960s.)

References

ALPAC. 1966. *Language and Computers in Translation and Linguistics. A Report by the Automatic Language Processing Advisory Committee*. Washington, DC: National Academy of Sciences, National Research Council Publication 1416. http://www.mt-archive.info/ALPAC-1966.pdf

Bar-Hillel, Y. 1960. ‘The present status of automatic translation of languages’ *Advances in Computers* 1, 91–163.

Bateman, R. 1985. Introduction to interactive translation. *Tools for the Trade: Translating and the Computer 5*. Proceedings of a conference… 10-11 November 1983, London Press Centre, ed. Veronica Lawson. (London: Aslib, 1985); pp.193-7 http://www.mt-archive.info/70/Aslib-1983-Bateman.pdf

Battersby, M. 2015. ‘Inside Out: Pixar makes crucial change for Japanese audiences by editing out broccoli’. *The Independent* 24 July 2015.

Bédard, C. 1989. ‘The return of low-linguistics MT. PC-Translator: can a low cost machine translator do the job?’ *Language Technology/Electric Word* 14, July-August 1989: 53, 57.

Bentivogli, L., A. Bisazza, M. Cettolo and M. Federico. 2016b. ‘Neural versus Phrase-Based Machine Translation Quality: a Case Study’. In *EMNLP 2016*. arXiv:1608.04631v1 [cs.CL] 16 Aug 2016.

Bentivogli, L., Bertoldi N., Cettolo, M., Federico, M,. Negri M. and Turchi, M. 2016a. ‘On the Evaluation of Adaptive Machine Translation for Human Post-Editing’. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24(2): 388-99.

Bentivogli, L., M. Federico, S. Stüker, M. Cettolo and J. Niehues. 2016c. ‘The IWSLT Evaluation Campaign: Challenges, Achievements, Future Directions’. In G. Rehm, A. Burchardt et al. (eds) *Proceedings of the LREC 2016 Workshop “Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem”*, 14-19.

Boitet, C., H. Blanchon, M. Seligman and V. Bellynck. 2010. ‘MT on and for the Web’. Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010). http://www-clips.imag.fr/geta/herve.blanchon/Pdfs/NLP-KE-10.pdf

Bojar, O., C. Federmann, B. Haddow, P. Koehn, M. Post and L. Specia. 2016. ‘Ten Years of WMT Evaluation Campaigns: Lessons Learnt’, In G. Rehm, A. Burchardt et al. (eds) *Proceedings of the LREC 2016 Workshop “Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem”*, 27-34.

Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin. 1988a. ‘A statistical approach to French/English translation’. In *Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (TMI 1988)*, June 12-14.

Pittsburgh, PA, USA. No page numbers.

Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin. 1988b. ‘A statistical approach to language translation’. In *Proceedings of the 12th International Conference on Computational Linguistics* Vol.1. Budapest, Hungary, August 22-27, 1988. Budapest: John von Neumann Society for Computing Sciences, 71-6.

Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R. Mercer and P. Roossin. 1990. ‘A Statistical Approach to Machine Translation’. *Computational Linguistics* 16(2): 79-85.

Brown, P., V. Della Pietra, S. Della Pietra and R. Mercer. 1993. *Computational Linguistics - Special issue on using large corpora: II*. 19(2): 263-311.

Brynjolfsson, E. and A. McAfee. 2011. *Race Against the Machine*. Lexington, MA: Digital Frontier Press.

Bureau of Labor Statistics. 2016. *Occupational outlook handbook*, 2016-17 edition: Interpreters and translators. Washington, DC: U.S. Department of Labor. http://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm

Burrell, Jenna. 2016. ‘How the machine ‘thinks’: Understanding opacity in machine learning algorithms’. *Big Data & Society*, January–June 2016: 1–12.

Castilho S., J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley and A. Way. 2017. ‘Is neural machine translation the new state of the art?’ *The Prague Bulletin of Mathematical Linguistics* 108(1): 109–20.

Cettolo, M. et al. 2016. The IWSLT 2016 Evaluation Campaign https://workshop2016.iwslt.org/downloads/IWSLT_2016_evaluation_overview.pdf

Chan, S. 2017. *The Future of Translation Technology: Towards a World Without Babel*. London/New York: Routledge.

Chang, H., B. Bederson, P. Resnik and Y. Kronrod. 2011. ‘MonoTrans2: a new human computation system to support monolingual translation’. In *CHI '11*

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1133-6.

Common Sense Advisory. 2016. ‘Global Market Research Firm Common Sense Advisory Finds Post-edited Machine Translation (PEMT) Among Fastest-growing Segments of the Language Industry’. https://www.commonsenseadvisory.com/default.aspx?Contenttype=ArticleDet&tabID=64&moduleId=392&Aid=36546&PR=PR

Cronin, M. 2003. *Translation and Globalization*. London/New York: Routledge.

Cronin, M. 2013. *Translation in the Digital Age*. London/New York: Routledge.

DGT. 2016. *Translation Tools and Workflow*. Brochure. https://publications.europa.eu/en/publication-detail/-/publication/e0770e72-afa1-4971-8824-6190512537dc/language-en

Doherty, S. 2016. ‘The Impact of Translation Technologies on the Process and Product of Translation’. *International Journal of Communication* 10: 947–969.

Domingos, P. 2017. *The Master Algorithm*. London: Penguin.

ECMT. 2013. https://webgate.ec.europa.eu/fpfis/mwikis/thinktank/index.php/European_Commission_Machine_Translation

Elliott, D., S. Frank and E. Hasler. 2015. ‘Multi-language image description with neural sequence models’. *CoRR*, abs/1510.04709.

Elliott, D., S. Frank, L. Barrault, F. Bougares and L. Specia 2017. ‘Findings of the Second Shared Task on Multimodal Machine Translation and Multilingual Image Description’. *Proceedings of the Conference on Machine Translation (WMT), Volume 2: Shared Task Papers*, 215–33.

European Commission. 2017. https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-mtec_en

Farajian M. A., M. Turchi, M. Negri and M. Federico. 2017. ‘Multi-Domain Neural Machine Translation through Unsupervised Adaptation’. *Proceedings of the Conference on Machine Translation (WMT), Volume 1: Research Papers*, 127–37.

Firth, J. R. 1957. ‘A synopsis of linguistic theory 1930-1955’. In *Studies in Linguistic Analysis*. Oxford: Philological Society, 1-32.

Forcada, M. 2017. Making Sense of Neural Translation. *Translation Spaces*. 6(2): 291-309.

Forcada, M. L., M. Ginestí-Rosell, J. Nordfalk, J. O’Regan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez and F.M. Tyers. 2011. ‘Apertium: a free/opensource platform for rule-based machine translation’. *Machine Translation* 24(1): 1–18.

Frey, C. B. and M. A. Osborne. 2017. ‘The Future of Employment: How susceptible are jobs to computerisation’. *Technological Forecasting and Social Change*. 114: 254-80.

Gardner, D. and M. Davies. 2007. ‘Pointing out frequent phrasal verbs: A corpus-based analysis’. *TESOL Quarterly* 41: 339-59.

Goltz, N. 2017. ‘Linked Democracy 3.0 - Global machine translated legislation and compliance in the age of artificial intelligence’. *SSRN Electronic Journal*. https://researchcommons.waikato.ac.nz/handle/10289/11197

Harnad, S. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-46.

Hearne, M. and A. Way. 2011. ‘Statistical Machine Translation: A Guide for Linguists and Translators’. *Language and Linguistics Compass* 5(5): 205–26.

Henderson, J.B. 2010. ‘Artificial Neural Networks’. In A. Clark, C. Fox and S. Lappin (eds) *Handbook of Computational Linguistics and Natural Language Processing*. Chicester: John Wiley & Sons, 397d-5a.

Hutchins, W. J. 1986. *Machine Translation: past, present, future*. Chichester: Ellis Horwood.

Hutchins, W. J. (ed.) 2000. *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*. Amsterdam/Philadelphia: John Benjamins Publishing.

Hutchins, W. J. 2004. 'Two precursors of machine translation: Artsrouni and Trojanskij'. *International Journal of Translation* 16 (1): 11–31.

Hutchins, W. J. 2006. *The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954*. Updated version of paper originally presented at the AMTA Conference, September 2004. http://www.hutchinsweb.me.uk/GU-IBM-2005.pdf

Hutchins, W. J. and H. L. Somers. 1992. *An Introduction to Machine Translation.* London: Academic Press.

Internetlivestats. 2018. http://www.internetlivestats.com/internet-users/#trend

Jakobson, R. 1959. 'On Linguistic Aspects of Translation'. In R. Brower (ed.) *On Translation*. Cambridge, MA: The Harvard University Press, 232-9.

Johnson, M., M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat. 2017. 'Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation'. *Transactions of the Association for Computational Linguistics*, 5: 339–51.

Joscelyne, A. 1998 'AltaVista translates in real time' *Language International* 10 (1): 1998: 6-7.

Karamanis, N., S. Luz and G. Doherty. 2011. 'Translation practice in the workplace: contextual analysis and implications for machine translation'. *Machine Translation* 25(1): 35–52.

Kenny, D. 2011. 'Electronic Tools and Resources for Translators'. In K. Malmkjær and K. Windle (eds) *The Oxford Handbook of Translation Studies*. Oxford: Oxford University Press, 455-72.

Kenny, D. 2012a. ‘The Ethics of Machine Translation’. In S. Ferner (ed.) *Proceedings of the NZSTI Annual Conference 2011*, 121-31.

Kenny, D. 2012b. ‘Linguistic Approaches to Translation’. In C.A. Chapelle (general editor) *The Encyclopedia of Applied Linguistics*. Hoboken, NJ: Wiley-Blackwell, 3439-46.

Kenny, D. and S. Doherty. 2014. ‘Statistical Machine Translation in the Translation Curriculum: overcoming obstacles and empowering translators’. *The Interpreter and Translator Trainer* 8(2): 276-94.

Koehn, P. 2010a. *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, P. 2010b. ‘Enabling Monolingual Translators: Post-Editing vs. Options’. In *NAACL HLT 2010 Proceedings*. Los Angeles, CA: ACL, 537-45.

Koehn, P. 2016. ‘The State of Neural Machine Translation (NMT)’. *Omniscien Technologies Blog*. 2016-11-30. https://omniscien.com/category/blog/ Last accessed 30 Jan 2018.

Koehn, P. 2017. ‘Neural Machine Translation’. Draft chapter for the book *Statistical Machine Translation*. https://arxiv.org/pdf/1709.07809.pdf

Koehn, P. and R. Knowles. 2017. ‘Six Challenges for Neural Machine Translation’. *First Workshop on Neural Machine Translation*, https://arxiv.org/pdf/1706.03872.pdf

Kuhn, T. 2014. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics* 40(1): 121-70

Langlais, P., S. Gandrabur, T. Leplus and G. Lapalme. 2005. ‘The Long-Term Forecast for Weather Bulletin Translation’. *Machine Translation* 19(1): 83-112.

Lommel, A. 2017. ‘How Augmented Translation Affects the Language Services Industry’. Common Sense Advisory Blog, July 5, 2017. http://www.commonsenseadvisory.com/default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=39818&moduleId=390

Malmkjær, K. 2011. ‘Meaning and Translation’. In K. Malmkjær and K. Windle (eds)*The Oxford Handbook of Translation Studies*. Oxford: Oxford University Press, 108-22.

Manning, C. D. and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.

Melby, A. 1995. *The Possibility of Language.* Amsterdam/Philadelphia: John Benjamins Publishing.

Microsoft. 2018. https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/

Mitchell, C. 2010. ‘Translation and Materiality. The Paradox of Visible Translation’. *Translating Media* 30(1): 23-9.

Moorkens, J. 2017. ‘Under pressure: translation in times of austerity’. *Perspectives,* 25(3): 464-77.

Nagao, M. 1984. ‘A framework of a mechanical translation between Japanese and English by analogy principle’. In A. Elithorn and R. Banerji (eds) *Artificial and human intelligence*. Amsterdam: Elsevier Science Publishers, 173-80.

O'Hagan, M. (ed.). 2011. *Translation as a Social Activity. Community Translation 2.0. Special issue of Linguistica Antverpiensia* 10/2011.

Poibeau, T. 2017. *Machine Translation*. Cambridge, MA: The MIT Press.

Pontin, J. 2018. ‘Greedy, brittle, opaque, and shallow: the downsides to deep learning’. *Wired* 02 February 2018. https://www.wired.com/story/greedy-brittle-opaque-and-shallow-the-downsides-to-deep-learning/

Raley, R. 2003. ‘Machine Translation and Global English’. *The Yale Journal of Criticism* 16(2): 291-313.

Reynolds, M. 2017. ‘Google uses neural networks to translate without transcribing’. *New Scientist*. https://www.newscientist.com/article/2126738-google-uses-neural-networks-to-translate-without-transcribing/ 04 April 2017.

Sereda, S. P. 1982. ‘Practical experience of machine translation’. In V. Lawson (ed.) *Conference Proceedings: Practical experience of machine translation*, Amsterdam/New York/Oxford: North-Holland Publishing Company, 119-27.

Somers, H. 1993. ‘Current Research in Machine Translation’. *Machine Translation*, 7: 231-46.

Somers, H. 1999. ‘Review article: Example-based Machine Translation’. *Machine Translation*, 14(2): 113-57.

Turovsky, B. 2016. ‘Ten years of Google Translate’. https://blog.google/products/translate/ten-years-of-google-translate/

Way, A. 2009. ‘A Critique of Statistical Machine Translation’. *Linguistica Antverpiensia.* 8: 17-41.

Weaver, W. 1949. *Translation*. http://www.mt-archive.info/Weaver-1949.pdf

Wiggins, D. 2017. *Riding the Machine Translation Hype Cycle – From SMT to NMT to Deep NMT*. Omniscien Blog. https://omniscien.com/riding-machine-translation-hype-cycle/ last accessed 01 December 2017.

Wittgenstein, L. 1968. *Philosophical Investigations.* 3rd edition. Oxford: Basil Blackwell. Translated by G. E. M. Anscombe.