28
Translation, artificial intelligence and cognition

Michael Carl

28.1 Introduction

Artificial intelligence (AI) and machine translation (MT) have a closely entangled history. Both fields of research originated in the United States in the 1950s and were ignited during the Cold War by the attempt to use computers to automatically translate Russian scientific journals into English (Camburn, 2013). While the early success of MT triggered an enthusiasm for fully automatic high-quality translation, it was quickly recognized that—in order to scale up MT—one had to understand the grammar of the source and the target language, the morphology (the grammar of word forms), and the syntax (the grammar of sentence structure). But to understand syntax, one also had to understand semantics (the meaning of words and sentences) and the pragmatics of language use. The attempt to translate texts automatically from one language into another thus quickly developed into a much bigger endeavour, which involved all areas of Computational Linguistics. According to the Stanford Encyclopedia of Philosophy (2018, n.p.), Computational Linguistics is “the scientific and engineering discipline concerned with understanding written and spoken language from a computational perspective […] [and] a computational understanding of language [that] provides insight into thinking and intelligence”.

While the idea of using computers to simulate intelligent human behaviour (i.e. to provide insight into thinking and intelligence) goes back to Turing (see Section 28.2), the term “Artificial Intelligence” was coined at a workshop in 1956, when a small group of ten scientists met for six weeks to discuss how “to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves” (McCarthy et al., 1955, p. 1). The underlying conviction was that “every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” (McCarthy et al., 1955, p.1). Based on this assumption, Simon and Newell subsequently developed a theory of the general problem solver in 1963 and the physical symbol systems hypothesis (PSSH) in 1972, which is the basis of several computational cognitive models and brought about a huge push in AI (Newell & Simon, 1963, 1972). The human mind was conceived as a machine that manipulates symbols, and the manipulation operations were identified as computational processes of a program which could equally well run in the human brain or in a computer. It was believed that “[a] physical symbol system has the necessary and sufficient means for general intelligent action” (Newell & Simon, 1976, p. 116).
Accordingly, in the 1970s and the early 1980s, natural language processing (NLP) approached language explicitly as a cognitive process. These NLP systems were based on intuitions about how people “understand”, and linguistic theories were thought to constitute the building blocks for mental representations and models of human thinking. For instance, Nagao (1984, p.173) proposed to model MT in line with human behaviour, stating that “we have to think about the mechanism of human translation, and have to build a model based on the fundamental function of language processing in the human brain”. Nagao suggested a cognitive approach to translation, which became known as example-based machine translation (EBMT), deemed to mimic the human translation process. Based on this model, a large number of EBMT systems have been developed (see e.g. Carl & Way, 2003, for an overview); however, none of them were based on actual psycholinguistic investigation taking into account how humans really translate.

Until the late 1980s, most work in MT was concerned with symbolic systems, involving large collections of handwritten rules to model linguistic phenomena. These kinds of “knowledge-based” systems had a number of bottlenecks, mainly due to the unexpected complexity and ambiguity of natural languages. Since the late 1980s, there has been a significant shift towards statistical methods, whereby rules and probabilistic generalizations are learned from data rather than being produced manually. This “data-driven” development became possible due to the availability of texts in electronic form and later from the World Wide Web and increased computer processing power, which is required to derive the information with necessary precision from large data sets. In addition, in his Chinese room experiment, Searle (1980) showed that intelligent behaviour (e.g. translation) can be simulated by machines without proper understanding of the meaning. Thus, intelligent behaviour is possible by manipulating symbols without insights into the underlying causal relations and implications. Together, this has led to a path-breaking rise of statistical MT, and more recently to machine learning and Big Data technologies with unprecedented translation quality, which have been developed without reference to human–machine analogy or psycholinguistic insights, hence decoupling intelligent behaviour from deeper understanding. However, very recently—within the discussion of human–machine parity in translation (Läubli et al., 2020)—we see a resurgence of interest in psycholinguistics in the context of interactive translation, translation reception and evaluation.

Crocker (2013) argues that the divergence between computational linguistics and computational psycholinguistics (and thus computational cognitive modelling) already started in the 1970s. He sees the reasons for this separation in divergent goals. While psycholinguistics is concerned with questions of linguistic competence (such as the processes and representations that allow humans to arrive at incremental sentence understanding), computational linguistics, and in particular its engineering branch, NLP, became more concerned with linguistic performance. This includes how linguistic knowledge can be used—among other things—for spell checking, keyword and information extraction, text summarization, document classification, question answering, MT, etc. Questions that would address how humans process and understand language in real time became increasingly uninteresting for NLP applications, so that MT and computational cognitive modelling are now only marginally overlapping fields of research. Current NLP applications are limited tasks that need to be performed accurately and robustly, often without real “understanding” (e.g. spam filters, information retrieval, document clustering, summarization, etc.). In its present form, modern NLP has shifted emphasis onto coverage and efficiency with little concern about cognitive plausibility.

However, with the increasing quality and usability of MT output, Cognitive Translation Studies has recently become concerned with human–computer interaction as a form of extended cognition (Muñoz Martín, 2010). As O’Brien (2017, p. 313) points out, “[w]ith the current state of the art, machine translation (MT) cannot show us how translators work at a cognitive level. We can, however, investigate how translators and end users […] interact
with MT through cognitive process studies”. Despite increased translation quality, MT output is in many cases not good enough for at least some end users. Recent years have witnessed increased research into how cognitive effort in translation can be assessed and measured, and how humans and machines can work in tandem as a hybrid intelligence system, driven by the aim of minimizing human translation effort while maintaining the required translation quality. These studies reveal that the success of human–computer interaction—for instance interactive MT (e.g. Daems & Macken, 2019; Knowles et al., 2019) and MT post-editing—depends on various parameters, such as the quality of the MT output, the type and domain of the texts, the expertise of the translators, the availability of suitable interfaces and visualizations of translation options, etc. These studies are indicative of a renewed interest in the conditions of the human translating mind, which often have the goal of designing “machines to augment human expertise (intelligence augmentation)” so as to “leverage the best of both machine intelligence and human expertise rather than exclude one or the other out of hubris or habit”1 (Dillinger, 2018, emphasis in original).

Section 28.2 traces some historical aspects in the relation between artificial intelligence and cognition. It outlines the PSSH, which claims that humans and computers are symbol-manipulating machines. The PSSH was challenged in the 1980s by a connectionist alternative, which today is experiencing a powerful, ubiquitous revival. Section 28.3 traces the development of interactive machine translation (IMT) in relation to embedded, extended and situated forms of cognition (Clarke, 2008; Muñoz Martin, 2010, 2017), where the translator engages in a dialogue with the MT system during translation production. We point to the particularly innovative and disruptive potential of IMT—as it models a tight integration of humans and machines—within different incarnations, presupposing different conceptions of AI, the functioning of the human mind, and technological frameworks. Section 28.4 summarizes the discussion with a view on dual-process/dual-system theory as a possible framework for integrating the various approaches, and the potential of Cognitive Translation Studies to advance AI approaches and MT.

28.2 Core topics at the interface between artificial intelligence and cognition

In 1936, Alan Turing developed a (theoretical) machine to formalize the notion of computation. Turing showed how this machine was capable of manipulating symbols and could thus solve problems that would up till then have required human intelligence. Turing’s formalization of the notion of computation implied that machines could mimic the human mind—thus providing the technology and conceptual backbone of what would later be called “artificial intelligence”.

Starting from the late 19th century, behaviourism was the dominant method for understanding human behaviour. Behaviourism postulates that only measures of observable behaviours and events, and in later incarnations also feelings, states of mind and introspection, can be objects of scientific investigation. From the 1950s, behaviourism was heavily criticized on the basis that it would not examine the role of mental processes, and by the 1970s there was a consensus that the processes and mechanisms of the mind could not be understood purely on the basis of behavioural experiments. It became obvious that behaviourism could not explain the intricate details of the human mind and its manifestation in behavioural flexibility. As there are very many parameters that influence the results in any experiment involving the human mind, observations could not be explained with sufficient precision.
28.2.1 The physical symbol systems hypothesis

Computational cognitive sciences emerged under the assumption that computational models are necessary for understanding a system as complex and as diverse as the human mind. According to Sun (2008), “[a] cognitive architecture helps to narrow down possibilities, provides scaffolding structures, and embodies fundamental theoretical postulates.” (Sun, 2008, p. 9). Newell and Simon (1976), as well as Marr (1982), suggest cognitive architectures in which cognitive processes can be formulated on three more or less independent levels of description: (1) at a computational or knowledge level, to determine which are the goals and purposes of the cognitive processes; (2) at the symbol representation level, to identify the tasks, symbol structures and manipulation procedures; and (3) at the physical, implementation level, by any available physical means. The idea was that these levels of description are to a large degree independent and compositional. Cognition (thinking, perception, decision making, etc.) consists of manipulating the syntax of symbols, and their analysis would enable the prediction of behaviour. In this view, the mind is a syntactically driven machine in which meaning derives from the manipulation of symbols. A number of cognitive architectures, such as Soar (Laird, 2012), ACT-R (Anderson & Bower, 1973) and CLARION (Sun, 2017), have been developed, which provide a framework for implementing and testing the PSSH, aiming to provide conceptual clarity and precision of the predictions.

28.2.2 Strong vs. weak AI

The PSSH has generated controversial discussions. It stipulates the strong AI hypothesis, by which machines can be built that represent the human mind and the activities humans perform. The weak AI hypothesis, in contrast, states that machines can be built that simulate (but do not represent) human behaviour. Searle (1980, p. 417) states that “[i]n strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs themselves are the explanations”. In this view, the brain is a computer, and the mind is the result of the program that the brain runs. Searle further explains that “instantiating a formal program with the right input and output is a sufficient condition of, indeed is constitutive of, intentionality […] the essence of the mental is the operation of a physical symbol system” (Searle, 1980, p. 421) Strong AI protagonists thus maintain that Marr’s (1982) level (1) and level (2) descriptions are sufficient to represent the human mind.

In his Gedankenexperiment, the Chinese Room, Searle (1980) attempts to reject the strong AI hypothesis through a paradox. He argues that perfect answers (and thus e.g. perfect translations) are possible without an understanding of the questions (i.e. a source text), just by following a sufficiently precise set of formal rules. Searle uses the Chinese Room experiment to show that the human mind is more than a syntactically driven machine which runs in the brain, and that no Turing machine (i.e. no usual computer) can ever achieve human-like understanding. For Searle (1980, p. 446), “the whole idea of strong AI is that we don’t need to know how the brain works to know how the mind works”; and for weak AI it is not important to know how the brain works. While there was—and still is—a debate about under what conditions understanding can emerge in a computer, there is little doubt that perfect automatic translation is possible without proper understanding. For instance, an MT system may produce perfect translations for some sentences or texts, but we do not assume that the system “understands” what is being translated. Accordingly, the development of MT systems—and of computer linguistics or AI applications in general—has not paid much attention to psycholinguistic research.
in recent years. As Crocker (2013, p. 484) points out, computational linguistics has “a greater interest in optimizing the computational properties of parsing algorithms, such as their time and space complexity. Computational psycholinguistics, in contrast, places particular emphasis on the incremental processing behaviour of the parser”, which is crucial in the understanding of how humans process language.

28.2.3 Connectionist modelling

Artificial neural networks (NNs) are the most common type of connectionist networks, which provide an alternative computational framework for computational cognitive modelling. There is a wide variety of NN architectures, which are derived from an abstraction of how the brain works. The earliest implementation of a simple artificial neuron goes back to McCulloch & Pitts (1943) and the perceptron (Rosenblatt, 1958). However, it was not until the mid-1980s, due to increased computer processing power, that the usage of multi-layer perceptrons (which include one or more hidden layers) and the discovery of back-propagation algorithms (Rumelhart et al., 1986) resulted in its breakthrough. NNs are interconnected simple processing units that operate in parallel and that spread activation through layers of connected nodes. A learning algorithm changes the behaviour of NNs over time (i.e. back-propagation adjusts activation weights between the individual nodes), but the numeric properties of the connections and node activation make an interpretation of the inner working of NNs notoriously difficult.

There was strong opposition to connectionism from proponents of the classical computational theory of mind, due to the fact that it is usually impossible to know what activity patterns inside NNs actually mean. Both connectionists and defenders of the PSSH maintain that the brain implements a computer. However, the latter state in addition that the mind operates with symbols, which have representational and syntactic properties, and that mental processes (e.g. thinking) manipulate sequences of those representations defined by the combinatorial structure of the representations, while the former argue that symbols cannot (easily) be isolated in distributed NNs—and usually not without loss of information—which undermines the basic assumptions of the PSSH.

In defence of the language of thought (LOT), Fodor and Pylyshyn (1988) launched a debate stating that connectionists cannot provide a scientific explanation of cognition, as connectionism lacks a representational system and a combinatorial syntax and semantics. Connectionist models tend to focus on a single cognitive phenomenon or process, while cognitive architectures (such as ACT-R; e.g. Anderson & Bower, 1973) cover a large range of cognitive activities. In connectionist systems, nodes are labelled, but “the operation of the machine is unaffected by the syntactic and semantic relations that hold among the expressions that are used as labels” (Fodor & Pylyshyn, 1988, p. 296). In addition, some connectionists claim that NNs can learn representations that have a combinatorial syntax and semantics and produce structure-sensitive transformations on those representations (e.g. Chorowski & Zurada, 2011). These structures, they claim, can be extracted in the form of rules, in the following manner: “Given a trained neural network and the examples used to train it, produce a concise and accurate symbolic description of the network” (Roy, 2002, p. 269). However, it is unlikely that any human could understand the “symbolic description” of a moderately intelligent system (in a reasonable amount of time), which undermines the notion of symbol altogether.

Most connectionists follow Chalmers (1993), for whom the “deepest philosophical commitment of the connectionist movement is […] the rejection of the atomic symbol as the bearer of meaning”, as symbolic descriptions “do not carry enough information with them to be useful in modeling human cognition”. Other attempts develop two-factor theories, which “explain certain ‘low-level’ aspects of cognition without resort to representations, but […] representational hypotheses will still be needed to account for the intentionality-based features of
cognition and [...] higher level processes” (Von Eckardt, 2012, p. 45; see also Kahneman, 2011, later).

28.2.4 Neo-connectionism

While connectionist architectures in the 1980s were developed and discussed in the light of how they simulate human cognition, this aspect is largely lacking in the discussion of newer NN architectures. The recent success of Neural Machine Translation (NMT) systems starting around 2014 and, to a large extent, is due to (1) increased computer processing power (memory and speed) and (2) a technique that encodes words in vector spaces (word embeddings). Vector models capture the semantic similarities between words based on the distribution of collocational properties in large corpora words that are used in the same contexts tend to carry similar meanings. Earlier vector models include Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). More recently, NNs have been used to embed words as vectors into a lower dimensional space. For instance, an NMT system with a vocabulary of 50,000 words would initially encode each word as a “one-hot” vector of 50,000 bits, in which only one bit is set. This one-hot vector is then mapped into a continuous vector space of 100 to 500 or more nodes, which represent the contextualized meaning of the encoded word. The main benefit of vector space models and word embeddings is that they can be learned as an unsupervised task, which does not require pricey annotation of the data. Word embeddings make it possible to measure the semantic similarity between words, e.g. using the cosine similarity. As Koehn (2017, p. 36) points out, they also allow for semantic inference such as:

queen = king + (woman − man)  
queens = queen + (kings − king)

The “meaning” of the nodes that make up word embeddings is, however, very different from the micro-features in a sub-symbolic system, as conceived by Fodor and Pylyshyn (1988). Fodor and Pylyshyn assume that concepts can be automatically decomposed into micro-features from the samples in the training set, so that, e.g., BACHELOR corresponds to a vector in a space of features that includes (+ADULT, +HUMAN, +MALE, −MARRIED) or CUP consists of a set of nodes which contains (+has-a-handle), etc. Accordingly, they argue, sub-symbolic connectionist states do have semantics, though it is not the semantics of representations at the “conceptual level”. Hence, symbol-based cognition emerges from the sub-symbols, so that connectionist systems are merely an implementation of the classical symbol systems hypothesis.

However, despite observations that in current vector space encodings “countries are clustered close together and syntactically similar words occupy similar locations in the vector space”, it is more than questionable whether it will be possible to assign a compositional meaning to each of the 500 or so nodes (i.e. sub-symbolic micro-features) of a word embedding. Current NMT systems are first of all technological solutions, and the architecture of the networks is inspired by pragmatic considerations, not through computational cognitive insights.

NMT systems consist of an input layer (the word embedding), an output layer and several hidden layers. The nodes in the output layer can be decoded into symbols (e.g. words), for instance with support of the softmax transformation, by which representations of words emerge on the surface of the output. It is thus possible to compose and decompose connectionist input and output representations into words (i.e. symbols at the conceptual level), but the mapping mechanisms that are responsible for reordering and re-phrasing the input vectors into an
output representation—which are learned during the training phase of the NMT system, and which are encoded in the hidden layers—are implicit (non-concatenative) realizations, which may be difficult or impossible to tokenize, extract and represent in some meaningful compositional way.

28.2.5 Connectionism and cognitive theories

Clarke (2008) has suggested a similar view to understand human language processing: words (and languages) are “inputs [to the mind] (whether externally or internally generated) that drive, sculpt, and discipline the internal representational scheme” (Clarke, 2008, p. 54) They are “basic tools to discipline and stabilize dynamic processes of reason and recall”, which give us and others the “ability to reliably follow trajectories in representational space” and create the “stable attendable structure to which subsequent thinking can attach” (Clarke, 2008, pp. 53–59) However, the encoding and processing of words and language in the human mind does not amount to an LOT, as suggested by Fodor and Pylyshyn (1988).

In order to create an attendable and attachable structure in NMT systems, and to make them flexible and better adaptable to new domains or text types, methods need to be developed that in some way interfere in the translation process to adjust systematic shortcomings of the NMT systems. For instance, the infusion of specific translations (e.g. terminology) or fixed, idiomatic translations into an NMT system during the translation phase or the controlled production of preferred target structures would probably require some sort of online access into the internal structure of an NMT system. Research efforts are being made to infuse higher-order monitoring and intervention processes as extensions in the NN architecture, which allow targeted manipulations of the NMT processes that may lead to reliable output in a more predictive manner.

Dual-process and dual-system theories (e.g. Kahneman, 2011) provide a conceptual framework, which accounts for similar processes in the human cognitive system. According to Kahneman (2011), while one system (System 1) is automatic, fast and realizes non-conscious processes, the other system (System 2) is controlled, slow and processes conscious forms of thinking. These two processes are implemented in two different but connected systems. The integration of higher-order monitoring processes and lower-order automatized routines has also been the architecture in human–computer interaction in translation. In the next section, we will discuss how MT systems have been embedded in translation workflows and used as an instrument for embedded and extended cognition.

28.3 Recent developments concerning machine translation and cognition

Numerous models have been elaborated to describe cognitive processes in translation, and several of these models provide evidence of a dual process/dual system of reasoning in translation (e.g. Hönig, 1991; Ivir, 1981; Schaeffer & Carl, 2013; Tirkkonen-Condit, 2005). The dual-process/dual-system theory assumes that there are two types of systems (system 1 and system 2), which result in two different, overlapping processes which can be disentangled to some extent. “The implicit system [system 1] is non-conscious or pre-conscious, rapid, parallel, low effort, high capacity and shaped by biologically constrained, domain-specific learning. The explicit system [system 2], by contrast, is conscious, slow, serial, high effort, limited capacity and responsive to verbal instruction” (Frankish, 2010, p. 920). For Frankish (2010, p. 919), the implicit system 1 is “associated with parallel, connectionist processing; the [explicit system 2] […] with serial, rule-governed processing”.
The dual-system approach has also been extended and criticized. Instead of a dual system, Muñoz Martín (2017, p. 564), for instance, presupposes a mesh of interconnected processes, which are “conscious and unconscious, logical and analogical, rational and emotional to different degrees, often at once”. Some incarnations of the dual-system theory and its one-system alternative could be understood as existing on a continuum that could plausibly be characterized as either one system or a dual system (Mugg, 2015). However, none of these theories is in principle incompatible with the assumption that the human mind is embedded within a social-cultural milieu and extended into the body and the environment. The mind is embedded, as it is “designed to function in tandem with the environment” (Rowlands, 2010), and it is also extended, as it offloads tasks onto the external environment to free up limited cognitive resources (e.g. using a notepad as external memory storage).

MT and IMT document workflows and their technological solutions are part of the socio-cultural milieu in which translators operate. Since the 1950s, attempts have been made to integrate the human translator and automated assistance (e.g. an MT system) into one hybrid cognitive system, in which a translator is either embedded—to give advice whenever a decision cannot be taken automatically by the MT system with high enough precision—or extended with automated help to free up her/his limited cognitive resources. The aim of any hybrid integration is to eliminate the weakness of each component by the (relative) strength of the other component so as to increase the performance of the integrated human–machine interactive system. This is possible if the weaknesses and strengths are unevenly distributed in the two (or more) sub-components (e.g. in system 1 and system 2). According to the conceptualization of the human mind as a device that manipulates symbols and the connectionist approach, these strengths have been differently defined. This becomes particularly apparent in the changing design of IMT over the past decades, which will be traced in this section.

Bruderer (1978) suggests four principles by which MT systems can cooperate with humans:

1. fully automatic MT, which is possible for easy texts, restricted domains and/or limited quality expectations
2. pre-edition, mainly to reduce ambiguities and to simplify the source text for better MT results
3. post-edition, the process to amend grammar and style of the MT output
4. IMT, where the translator intervenes in dialogue with the MT system

In contrast to the other forms of human–MT cooperation, IMT presumes MT operations preceding and following human intervention, thus embedding or extending human translation activities. While the concept of pre- and post-editing has remained—to a large extent—similar over the past decades, irrespectively of whether the underlying MT technology has been rule-based, statistical or neural concepts of IMT have dramatically changed with evolving translation technology in the past 30 years and accordingly have taken different views on embedded/extended cognition.

### 28.3.1 Interactive translation aids

For Hutchins and Somers (1992), the antecedents of MT are, in fact, mechanical dictionaries, which would be used interactively to support translators in their work and reduce their mechanical and cognitive workload. They report that in 1933 two independent teams, led by George Artsrouni and Petr Smirnov-Troyanskii, worked on two different interactive translation devices
that can be seen as predecessors of MT. Artsrouni proposed to store a dictionary on a paper-
tape-based general-purpose storage device that would be queried automatically to show word 
translations that could be used by a translator. Smirnov-Troyanskii, in turn, suggested a multi-
step MT system in which the machine and a human would collaborate interactively. A human 
operator would first transform each source word into its basic form using a specific editor. The 
machine would then translate each word into a target-language basic word form, and finally a 
reviser would generate the final sentence form in the target language.

Such simple hybrid systems can be represented as simple box-and-arrow models, which make 
explicit the cognitive structure of the integrated system. The different processing steps can be 
represented by boxes and the relationships between them by arrows. Such diagrams—similar to 
a computer flowchart—facilitate a detailed understanding of an extended/embedded human 
translator and her/his machine interaction and make it possible to design the workflow in a way 
that compensates their mutual weaknesses and accumulates their strengths.

### 28.3.2 Interactive embedded translation

A next step in the development of IMT systems was related to the introduction of rule-based 
MT (RBMT) systems in the 1970s. RBMT systems typically first analyse the syntactic struc-
ture of the source text (ST). The ST structure and the lexical items are then transferred and 
mapped onto the target language, and a target-language surface form is generated. Due to the 
large number of possible analyses for an ST string, it was considered the most difficult task in 
early RBMT systems to generate (or select) a correct ST structure. In the case a decision was 
too difficult for the machine to take (such as prepositional phrase (PP) attachment), interactive 
RBMT systems would allow the MT system to “consult a human when it does not know what 
to do” (Bisbey & Kay, 1972, p. 9). Bisbey and Kay developed the MIND system, which embeds 
a “monolingual consultant to resolve ambiguities in the [machine] translation process” (Bisbey 
& Kay, 1972, p. 9). The human component was considered an additional embedded resource for 
the MT system with easy access to additional (world-)knowledge resources that are required to 
resolve linguistic ambiguities that the computer could not (yet) solve. In line with the predom-
inant view at the time that cognitive processes in humans and machines were symbolic by nature, 
the human–machine communication was assumed to take place on a symbolic metalanguage 
level. It was “hoped that interaction could be restricted to analysis […] [but] it was found that 
some interaction was also required in transfer” (Melby et al., 1980 p. 424).

According to Weaver (1988), the aim of interaction in RBMT systems was then considered 
to be helping the computer produce better draft translations, which would reduce the amount of 
successive post-editing. Weaver mentions several forms of interaction, the most common being 
lexical interaction by which an ST parser is provided with additional (semantic) constraints that 
help the system find better translations. A disadvantage of early RBMT interaction was that the 
“on-line interaction requires specially trained operators” (Melby et al., 1980). To overcome this 
shortcoming, Boitet et al. (1995) suggested a dialogue-based RBMT system whereby untrained 
editors would choose a disambiguating item from a pre-processed list in order to disambiguate 
the source linguistic analysis of the text to be translated.

### 28.3.3 Interactive extended translation

An alternative perspective on human-computer interaction evolved as a consequence of 
these difficulties. Instead of embedding the translator into a computational system to resolve 
remaining ambiguities that RBMT systems cannot solve alone, Martin Kay (1980) called for
a re-examination of the role of humans and machines (including his own MIND system) with his Translator Amanuensis and envisioned a gradual extension of the translator, i.e. a human–machine partnership to assist translators. With his Translator Amanuensis, Kay suggests an incremental, pragmatic approach to human–machine interaction to produce not only a translation of an ST but also a device that constantly accumulates experiences that have been agreed upon between the human translator and the machine. By doing so, it extends the translator’s memory, as it memorizes the translator’s previous solutions, and it reduces motor activity (typing effort), thus enhancing its translation performance. Kay’s vision of interactive extended translation can be considered a precursor of computer-assisted translation (CAT tools) and translation memory systems (TMS) as we know them today, being incrementally built through translation activities into which translators “offload” their translation solutions, which become the basis of future translations.

28.3.4 Interactive statistical translation

A further step in extending translators with AI systems arrived with data-driven technology. By the 1990s, statistical MT (SMT) systems were developed, which would focus on target-text generation, turning the focus away from SL analysis. In their seminal paper, Brown et al. (1990, p. 79) proposed SMT in which “every sentence in one language is a possible translation of any sentence in the other language”, and they subsequently elaborated a mathematical framework to assign the (hypothetical) probabilities with which a translator would produce the various different target sentences for a given SL sentence (Brown et al., 1993). The emphasis in SMT shifted from selecting the best (or appropriate) ST analysis to the question of how to filter the best out of a large number of possible translations.

SMT systems have been further developed into interactive SMT systems, e.g. TransType (Foster et al., 2002) and TransType2 (Macklovitch, 2006), which offer an interaction with the translator. TransType proposes translation completion at the translator’s cursor position, thus providing situated assistance at the time when a translator needs it. Not only the translator’s memory, but also the human translation process, is extended into the translation environment, which provides the translator with potentially new translation solutions, but “the translator retains full control over their output” (Macklovitch & Valderrábanos, 2001) rather than being an appendix to an MT system. Foster et al. (1997) show how TransType can predict up to 70% of the characters in the translator’s intended target text. Experiments with TransType (Foster et al., 2002) show a 10% gain in translation time. Later experiments with TransType2 (Macklovitch, 2006) showed a productivity increase from 15% to 55%, depending on the text to be translated, with professional translators in a translation agency.

28.3.5 Interactive cloud-based translation

A continuation of TransType2 was carried out in the context of the CASMACAT project (Alabau et al., 2013). The objective of CASMACAT was to devise computer interfaces based on the outcome of cognitive studies and translators’ behaviour. CASMACAT implemented a browser-based system, featuring a number of innovative techniques, such as online learning and various visualization options.

CASMACAT makes use of cloud-based translation servers, which is by now a common practice in the translation industry. Cloud-based solutions process computationally heavy translation tasks with remote, powerful hardware and communicate the results to a client, usually in real time. Sophisticated software solutions allow real-time updating of the translation...
models, and enhanced translation solutions are immediately available for successive interactive translations, e.g. of the same document. While CASMACAT is a non-commercial prototype, Lilt\(^6\) is the first commercial solution that uses browser-based interface and online learning (Green et al., 2014).

### 28.3.6 Interactive crowd-based translation

Cloud-based techniques enable novel business models whereby source texts are fragmented into many small translation jobs, which can be distributed over a large crowd of translators, while the shared resources on the MT server help ensure consistency of the translated text segments within a distributed translation project. These techniques may be complemented by active learning strategies, in which the MT system re-orders the source-language segments to be translated and interactively edited in order to maximize its expected learning effect. Thus, the text is presented in an order that suits the machine, so that it can learn most quickly or efficiently from the human corrections. Interactive crowd-based MT supports a further move from “content being rolled out in a static, sequential manner” to translated content being “integrated into a dynamic system of ubiquitous delivery” (Cronin, 2013, p. 498).

Companies such as Unbabel\(^7\) are experimenting with interactive translation editing tools for hand-held devices in which segments of limited length are post-edited out of context by the crowd and re-assembled into coherent text in the server back-end. Similarly, MotaWord\(^8\) produces translations on a collaborative cloud platform “coordinated efficiently through a smart back end” in which, according to their website, over 5,600 translators participate. Such distributed translation platforms allow a translator to request new segments on the fly so that they can assign and change their workload at will, while simultaneous co-sharing of translation resources and sophisticated quality assessment routines make it possible to maintain the minimum requested translation quality.

### 28.3.7 Interactive machine translation teaching

An extension of this trend toward decontextualization of the translation job is put forward by Dillinger (2018). For Dillinger,\(^9\) there are (at least) 50 shades of collaboration in which IMT systems are designed as augmented intelligence, where translators become “machine teachers”. Dillinger promotes a “New Paradigm for Building Machine Learning Systems” (Simard et al., 2017), in which the role of a domain expert, becomes that of a “teacher […] who transfers concept knowledge to a learning machine”. According to this paradigm, a “concept is a mapping from an example to a label value”, whereby the teacher selects examples that exemplify useful aspects of the concept under consideration. Concepts can be decomposed and re-factored, sub-concepts can be created, and each concept might provide its own mechanisms for example selection, labels, schemas (i.e. hierarchies of concept relations) and features, etc. Such concepts with their examples—Simard claims—are important in AI systems to train a learning system for analysing pictures, speech, text, or the relationship between people, things and texts; they are necessary for named entity recognition, document classification, sentiment or intent detection, etc., etc.

The augmented intelligence approach assumes that the “teacher has access to feature concepts not available to the training set”; i.e. experts have access to knowledge that the learning agent (the machine) has not, and teachers can “use their domain knowledge to pick the right examples and counterexamples for a concept and explain why they differ”.

510
For instance, according to Dion Wiggins, the word “run” is the most ambiguous English word, with 197 meanings, including “I ran for office”, “I went for a run”, “I ate bad food and got the runs”, “I scored a home run”, “The dry run went well”, etc., etc. Such sentences are likely to produce translation errors, which could be addressed in a systematic manner. In an MT teacher scenario, the MT system could first produce all translations that it is uncertain of (for instance, translations that include the word “run”) and ask a teacher—through an interactive information exchange with a learning system—to provide a disambiguating label, probably best in the form of a correct translation. The learning agent would then store and generalize these examples and re-use them in similar contexts.

Note the similarity of this scenario to the role of the translator in the MIND and Dialogue systems discussed in Section 28.3.3. However, in contrast to embedded IMT in the 1970s, the actual interactive interface might just look like a well-known translation workbench. In line with such a vision, Alonso and Vieira (2017) project a vision of the Translator’s Amanuensis into the year 2020 (TA2020), where they anticipate a multimodal (text, speech and picture) translation workstation, to be used by experts and by non-experts, that would allow ubiquitous translation production and translation retrieval, embedded in any kind of hand-held or other device.

28.4 Summary and outlook

From the late 1950s until about the 1980s, the predominant framework of AI was based on the Physical Symbol Systems Hypothesis (PSSH) which postulates that intelligent computer programs represent human cognition, information processing, memory, learning and decision processes. According to this hypothesis, the human mind was thought to process sequences of symbols, and computational models were built not (only) to explain but instead to represent intelligent behaviour, such as translation. A number of techniques were suggested (see Section 28.3.2) to interactively embed a translator into an RBMT system to provide additional knowledge resources at run time.

By the mid-1980s, the PSSH was challenged by connectionist architectures of massively parallel cooperative and competitive neuron-like units. Inspired by the functioning of biological systems, sets of neurons would accumulate potential and emit spikes of activation, and connections between them could be learned and the activation adjusted based on a training corpus and a learning algorithm. The PSSH was increasingly undermined as artificial NN (and other machine learning algorithms) were able to simulate intelligent behaviour without manipulating symbols but rather reacting directly to the input signal.

In addition, it was shown that human understanding is more than just the ability to manipulate symbols. Instead of aiming at fully automatic, all-purpose MT solutions, the ALPAC report (1966) suggested that tools should be developed that would support translators in their work. Such computer-assisted translation (CAT) tools would support and extend the translators’ mechanical and cognitive capacities; humans and machines would collaborate in the production of translations in novel workflows.

Clarke (2008, pp. 39–42) makes the distinction between “a user in command of a new tool” and a “reconfigured user”, in which a user and the tool form a brand-new systemic whole. In the former case, the agent “would always use tools the way we typically begin to use them”, in a way that requires conscious allocation of cognitive resources and effort. In the latter case, the tool character disappears and becomes a new (cognitive or bodily) extension of the human agent. The agent “learns a complex problem-solving routine that makes a variety of deep implicit...
commitments to the robust bioexternal availability of certain operations”. He argues that our “plastic neural resources become recalibrated [...] so as to automatically take account of new [...] opportunities”. This recalibration is possible because our minds are “forever testing and exploring the possibilities for incorporating new resources and structures deep into [our] embodied acting and problem-solving regimes”. A translator will thus recalibrate, automatize and offload translation tasks and translation decisions in line with the opportunities offered by the Computer Assisted Translation environment.

Kahneman (2011) makes a distinction between more automatized and more conscious processes within the dual-process/dual-system theory. An implicit system [system 1] is pre-conscious, rapid, parallel, with high capacity and low effort, whereas the explicit system [system 2] is slow, serial, conscious and involves high effort. Tirkkonen-Condit (2005) adapts this theory in a monitor model of translation production, which consists of a “literal translation automaton” (i.e. system 1) that operates on a lexical as well as a syntactic level during the production of translations. This literal translation automaton is quick with a high capacity, it does not require working memory and it “generates literal or formally corresponding linguistic material as long as the material thus produced is semantically and syntactically acceptable” (Tirkkonen-Condit, 2005, p. 412). System 2 processes consist of a self-aware monitoring mechanism, which controls and intervenes in unwanted literal rendering.

As discussed in Section 28.3, different IMT frameworks address the two systems in a different way: The interactive RBMT paradigm (see Section 28.3.2), addresses the conscious system 2, in which an embedded translation assistant is asked to disambiguate the ST analysis (e.g. whether an item is animated, countable or liquid, how a PP should be attached in a syntactic analysis, etc.). While the human translation assistant adjusts to the environment (gets used to the graphical user interface and the kinds of question to be solved), there is no intention (and maybe no chance) for the translator to automatize the decision process and thus to reduce the cognitive load.

The interactive SMT paradigm, by contrast, together with its cloud- and crowd-based translation scenarios (Sections 28.3.4–28.3.6), starts out with the attempt to extend the translators’ cognitive process. It provides the translator with bilingual terminologies, translation memories (Section 28.3.3), and automatically generated translation proposals and translation completions. This IMT framework aims at mapping otherwise cognitively effortful operations (i.e. system 2 operations, such as internal or external search for translation alternatives or terminology, etc.) onto automatized system 1 processes in which the translator is invited to accept or select translation suggestions. The interactive SMT (or more recently also NMT) paradigm is thus likely to reduce the translators’ workload, allowing “genuine deep implicit commitments” (see Clarke, 2008), in which the translator and the MT system become a brand-new systemic whole.

Risku (2014, p. 339) defines cognition in translation as consisting of interconnected and self-organizing processes, which “includes all operations that work on internal and external representations with the aim of creating translations”. Working with CAT and IMT is thus a form of extended cognition, because the mind offloads tasks into the environment. As Muñoz Martín (2017, p. 564) puts it, “the brain uses (parts of) the environment as a tool for thought—such as rereading the original on the screen instead of memorizing it [...] thereby rendering the distinction internal/external irrelevant”. In an extended and situated translation environment— as discussed e.g. in Section 28.3—it is crucial to understand what pieces of information can (or should) be offloaded, how and which translation processes can be best supported by the IMT system, and how the offloaded information should be retrieved and/or textually integrated at the right point in time.

An underlying prerequisite for this to work is the complementarity of the human and machine component in the integrated IMT process. However, this complementarity
Translation, AI and cognition

is undermined as MT errors become increasingly “human-like” and therefore difficult to detect and eventually impossible or unnecessary for humans to improve. Läubli et al. (2020) discuss the possibility of human–machine parity in translation and show that only with advanced evaluation methods and/or professional raters a significant difference can be measured between human and machine translation. Yamada (2019, p. 87), finds that NMT errors are similar to those of human translators, making them more difficult to spot and correct. He concludes that “translation training is necessary for students to be able to shift their attention to the right problems (such as mistranslation) and be effective post-editors.” Novel methods and visualization tools might be required that highlight potential IMT errors or highly translation-ambiguous words so as to notify the human agent about the doubts of the (N)MT. By post-editing and even more so through IMT, the translator has already become a “machine teacher” who instructs the machine as to which translation(s) should be preferred under what conditions and in which context. Cognitive research in translation studies may need to address this new role of translation and the translator and investigate the impact of AI on subliminal translation and translation reception processes. Muñoz Martín (2017, p. 563) notes that cognition “works in tandem with the environment and [the various sub-systems] cannot be analyzed in isolation”. This suggests that cognitive translation studies ought to develop methods to determine and assess the factors of interaction and co-variation between these subsystems aiming at better understanding the nature and functions of the various internal and environmental components and to cope with the biases that new AI technologies (re)produce.

Notes

1 www.linkedin.com/pulse/50-shades-ai-mike-dillinger-phd/
2 Currently almost all major MT providers (Google, Bing, Systran, etc.) make use of some sort of NMT system.
3 http://ruder.io/word-embeddings-1/
6 http://labs.lilt.com/
7 https://unbabel.com/
8 www.motaword.com
9 www.linkedin.com/pulse/50-shades-ai-mike-dillinger-phd/

Further reading

Up-to-date interactive machine translation system.
On techniques of IMT as well as multimodal interfaces and adaptive learning in IMT.
An earlier handbook of cognitive aspects of translation.

Overview of research topics and applications of interactive machine translation:
References


