The History and Promise of Machine Translation

Lane Schwartz

Department of Linguistics

University of Illinois at Urbana-Champaign

Urbana IL, USA

lanes@illinois.edu

Abstract

This work examines the history of machine translation (MT), from its intellectual roots in the 17th century search for universal language through its practical realization in the late 20th and early 21st centuries. We survey the major MT paradigms, including transfer-based, interlingua, and statistical approaches. We examine the current state of human-machine partnership in translation, and consider the substantial, yet largely unfulfilled, promise that MT technology has for human translation professionals.

1 Introduction

With the possible exception of the calculation of artillery trajectory tables (Goldstine and Goldstine 1946), machine translation has a strong claim to be the oldest established research discipline within computer science. Machine translation is a modern discipline, one whose success and very existence is predicated on the existence of modern computing hardware. Yet certain core ideas which would eventually serve as the foundations for this new discipline have roots which predate the development of electronic digital computers.

This work examines the history of machine translation (MT), from its intellectual roots in the 17th century search for universal language through its practical realization in the late 20th and early 21st centuries. Beginning with the development of the first general-purpose electronic digital computers in the late 1940s, substantial effort was invested in the development of computer software capable of performing fully-automatic high quality machine translation. We survey the major MT paradigms, including transfer-based, interlingua, and statistical approaches.

As we survey the history of machine translation, we find it useful to recall the following quote from Mark Twain: “History never repeats itself, but the Kaleidoscopic combinations of the pictured present often seem to be constructed out of the broken fragments of antique legends”

(Twain and Warner 1874). This excerpt from Mark Twain's *The Gilded Age: A Tale of Today* may be the origin of the modern proverb apocryphally ascribed to Twain, "History never repeats itself, but it rhymes." In this work, we will examine how machine translation in the 20th and 21st centuries has been shaped and reshaped over the decades by diverse ideas from semiotics, information theory, linguistics, computation complexity theory, and human-computer interaction, and how the relative impact of these various influences have respectively waxed and waned and re-emerged throughout the history of machine translation.

In 1980, Martin Kay called for a re-examination of the role of humans and machines in the translation process, away from the lofty goal sought by many during the mid-20th century of fully automatic high-quality machine translation, and towards a gradual partnership between human translators and well-designed computer programs built to modestly assist their human counterparts (Kay 1980). Kay's dream in many ways presaged the development of computer-aided translation (CAT) tools that are in wide use today. We conclude by examining the current state of human-machine partnership in translation, and consider the substantial, yet largely unfulfilled, promise that MT technology has for human translation professionals.

2 The Road to Machine Translation

2.1 The Search for Universal Language

Our examination of the history of machine translation begins in 17th century Europe. For the scholars of that era, the search for meaning in language was intimately tied with two accounts found early in the Book of Genesis.

> Out of the ground the Lord God formed every beast of the field and every bird of the air, and brought them to Adam to see what he would call them. And whatever Adam called each living creature, that was its name. (Genesis 2:19 NKJV)

Many scholars of 17th century Europe believed that the original human language, created by Adam in the Garden of Eden, perfectly linked the form of human language with the meaning of each thing and concept. In the words of the 17th century German mystic Jakob Böhme (1623), Adam "knew the property of all creatures, and gave names to all creatures, from their essence, form and property. He understood the Language of Nature, the manifested and formed Word in everyone's essence, for thence the name of every creature is arisen."

The prominent scholars of 17th century Europe believed that the many languages found in the world in their day no longer possessed the perfect qualities of the language of Adam. For them, that perfection was lost at the Tower of Babel:

> Now the whole earth had one language and one speech. And they said, “Come, let us build ourselves a city, and a tower whose top is in the heavens.” But the Lord came down to see the city and the tower which the sons of men had built. And the Lord said, “Indeed the people are one and they all have one language, and this is what they begin to do; now nothing that they propose to do will be withheld from them. Come, let Us go down and there confuse their language, that they may not understand one another’s speech.” So the Lord scattered them abroad from there over the face of all the earth, and they ceased building the city. Therefore its name is called Babel, because there the Lord confused the language of all the earth; and from there the Lord scattered them abroad over the face of all the earth. (Genesis 11:1,4-9)

At Babel, “the perfect bonds between words and things were shattered” (Stillman 1995). It was proposed that the restoration of a pansophic language, where words and meaning were one, could resolve the political and moral crises of the day. The Czech philosopher John Comenius (1642) exemplified this position, writing that “in this way alone can the commonwealth of men, now torn to pieces, be restored.”

Many of the great thinkers of the 17th century would search for a universal language that could recover the lost language spoken by the biblical Adam. A common theme in this search was the search for a system of representing each unique concept with a corresponding language-independent symbol. In 1605, Francis Bacon wrote about his understanding of the Chinese writing system. Bacon believed, mistakenly, that Chinese characters represented “neither letters nor words” but rather “things or notions”:

> And we understand farther, that it is the use of China, and the kingdoms of the high Levant, to write in characters real, which express neither letters nor words in gross, but things or notions; insomuch as countries and provinces, which understand not one another’s language, can nevertheless read one another’s writings, because the characters are accepted more generally than the languages do extend; and therefore they have a vast multitude of characters, as many, I suppose, as radical words. (Bacon 1605)

In 1629, the French philosopher and mathematician Marin Mersenne wrote to René Descartes discussing one universal language which he had recently seen proposed. Descartes (1629) replied to Mersenne, critical of some details of the particular proposal, but very much intrigued by the idea. Throughout the 17th century, many of the era’s most prominent thinkers considered the problem of universal language, leading to numerous proposals, some more developed (Dalgarno 1661; Leibniz 1666) than others (Newton 1661). One of the most well developed proposals was published by the English philosopher John Wilkins; in his 600 page book, Wilkins (1668) attempts to lay out a detailed ontology on which to build a universal language.

In retrospect, it may appear obvious that this search for a language in which words and their denotata are perfectly unified was doomed to fail. Slaughter (1982) wrote "As linguistic enterprises, we know the schemes to be nonsense. Neither man nor language works in ways that are compatible with an artificial language." This perspective was indeed recognized by some at the time. Both Descartes and Hobbes urged caution in the construction of detailed scientific semantic ontologies as a prerequisite step for pansophic language, cautioning against attempts "(as some have done ridiculously) to prove that the kinds of things are not infinite" (Hobbes 1655). Hobbes argued that a perfect reunion of words and concepts "cannot be performed as long as philosophy remains imperfect". In the 18th century, the search for universal language was parodied in the novels *Gulliver's Travels* (Swift 1726) and *Candide* (Voltaire 1759).

Yet, this century of semiotic exploration succeeded in laying foundations for later semantic schemas. "The focal point in the language schemes is, clearly *order* of some sort. The early universal languages attempt to bring order into the linguistic code; the more developed, philosophical languages tackle the order in and of nature. The tables or the taxonomies of the universal languages become the most crucial aspect of the universal language movement" (Slaughter 1982). Attempts in the 20th century to develop a practical interlingua for machine translation (see §4.1.3) strongly reflected 17th century attempts to develop a pansophic language.

2.2 Mechanical Translating Devices before 1940

In their critical bibliography, Mel'čuk and Ravič (1967) refer to the earliest known surviving reference to an attempted mechanical translation system. According to Mel'čuk and Ravič, the system, ostensibly a prototype mechanical translating typewriter, was reported in the Estonian newspaper *Vaba Maa* on 24 February 1924.

The earliest systems for which detailed records exist were developed independently in the early 1930s in France and in Russia. The French system, developed by Georges Artsrouni (Corbé 1960; Hutchins 2004), appears to have been a general-purpose mechanical lookup and retrieval system based on a paper tape. The device, patented in 1933 and publicly demonstrated at the Paris Universal Expo in 1937, was not a full-fledged machine translation system, but could be configured as a mechanical bilingual dictionary; in this regard Arstrouni's machine could be viewed as a very preliminary precursor to the subsequent direct translation paradigm (see §4.1.1).

The Russian proposal by Petr Petrovič Trojanskij envisioned a special purpose device that directly foreshadowed ideas that would be developed more fully in the 1950s and 1960s (Trojanskij 1933; Hutchins and Lovtskii 2000). Trojanskij proposed a three-part translation process broadly similar to the subsequent analysis-transfer-generation paradigm (see §4.1.2). Trojanskij's proposal, which was never built, also was the first to envision the use of monolingual pre- and post-editors.

2.3 Warren Weaver's 1949 Memorandum

In 1947, Warren Weaver of the Rockefeller Foundation began a correspondence with Norbert Wiener, a professor at MIT, regarding the possibility of using computers to perform translation (Weaver 1947). Weaver posited that translation of human languages could be posed as a problem in cryptography. Inspired by the success of code-breaking during World War II, and by contemporary advances in information theory (Shannon 1948), Weaver distributed a memorandum that expanded the ideas from his prior correspondence (Weaver 1949).

Weaver's memorandum was the first publication of the 20th century known to researchers in America and western Europe to suggest "the possibility that [modern electronic computers might] be used for translation," and in particular, that methods from cryptography might be useful for machine translation. Weaver acknowledged that mechanical translation of literary texts may be too challenging, but that basic machine translation could still be useful for translation of technical documents.

Weaver began his memo noting the negative potential for languages to "impede cultural interchange" and serve as a "serious deterrent to international understanding." Weaver speculated that certain cryptographic successes may be in part due to possible statistical properties that are "to some significant degree independent of the language used," and which may be "to some statistically useful degree, common to all languages." In particular, Weaver proposed that ambiguity in text might be resolvable by looking at a window of n words around the ambiguous word. This insight, and his discussion of information theory (Shannon 1948), directly foreshadowed later use of n-gram language models and the noisy channel model in statistical machine translation (Brown et al. 1988 — see §4.2). Weaver believed that while "'perfect' translation is almost surely unattainable," machine translation capable of translating with "only X percent 'error' are almost certainly attainable."

Weaver concluded with an analogy of human language as a "series of tall closed towers, all erected over a common foundation." Weaver compared direct machine translation to individuals attempting to communicate between two towers by shouting, suggesting that a more fruitful approach may be for these individuals to descend the towers into a common basement, where communication may be easier; this serves as an analogy to translation approaches which make use of deep semantics and linguistic universals. This appeal to "the common base of human communication" using "the real but as yet undiscovered universal language" is very similar in spirit to 17th century searches for a universal language (§2.1), and directly foreshadows deep semantic transfer and interlingual systems (§4.1.3) of subsequent decades.

Weaver (1949) stated his hope that his memorandum "might possibly serve in some way as a stimulus to someone else, who would have the techniques, the knowledge, and the imagination to do something about it." The subsequent decades of research and development in machine translation built Weaver's hope into reality. Weaver's 1949 memorandum on translation is widely recognized as the intellectual starting point of machine translation in the mid-20th century. Written at the same time as the development of the first general-purpose electronic digital computers, that memorandum (Weaver 1949) served to kick-start the research field of machine translation. Less than two years after Weaver's memorandum, four research groups in Britain and the U.S. had begun preliminary MT research, with at least seven more groups interested (Loomis 1951).

3 The First Wave: A Decade of Optimism

3.1 Bar-Hillel's 1951 survey

In 1951, Yehoshua Bar-Hillel began a new position, at the Massachusetts Institute of Technology (MIT), as the world's first full-time machine translation researcher (Hutchins 1998). Bar-Hillel began by visiting the few labs in the U.S. which had conducted MT research to date (Rand Corporation, UCLA, and the University of Washington), writing up his thoughts and findings in a survey of the state-of-the-art in machine translation research (Bar-Hillel 1951).

Bar-Hillel focused primarily on machine translation for dissemination, and posited that research into machine translation was likely to "provide valuable insights into linguistic communication." Bar-Hillel discussed the state of the art in machine translation, as he observed it in systems at the three institutions he visited. In this discussion, Bar-Hillel criticized the idea that a system which covered 90% of the required vocabulary entries would be sufficient, largely because "the remaining few percent of words …will be least predictable and highly loaded with information"; his conclusion with respect to this was that technology capable of storing large vocabularies would be required.

In his survey, Bar-Hillel foresaw several important directions that MT research would follow in the coming years, as well as several of the important problems that would be encountered. Bar-Hillel claimed that because no contemporary machine was capable of fully resolving semantic ambiguity, “fully automatic MT …is achievable only at the price of inaccuracy.” To alleviate this problem, Bar-Hillel foresaw various possible human-machine partnerships, wherein humans could serve as pre-editors or post-editors to MT systems.

Bar-Hillel stated that certain processes were fundamental to any MT system, among them morphological analysis, basic syntactic analysis, and transformation and reordering; these ideas directly foreshadowed later systems which used the analysis-transfer-generation paradigm (§4.1.2). Bar-Hillel also claimed that long term, machine translation’s best hope would be the development of a truly language-independent universal grammar; Bar-Hillel believed that modern technology could make such a system, hoped for since the 17th century, finally a reality. Nevertheless, Bar-Hillel insisted that work on MT should not wait for the development of such a universal grammar.

3.2 The First MT Conference

Bar-Hillel's paper was distributed in advance, along with Weaver's 1949 memorandum, as background reading to all participants at the first Conference on Mechanical Translation, hosted by Bar-Hillel at MIT on 17-20 June 1952. Summaries of the conference (Reifler 1954; Reynolds 1954) were published in the newly created journal *Mechanical Translation*.

In the introductory address to participants, Bar-Hillel framed the problem of machine translation in very human terms:

> But it is … a very difficult job. To get a good look on it, imagine yourself on the task of translating from one language you do not understand into another you do not understand either, and this without the benefit of dictionaries from the source language into English and from English into the target language. This looks quite formidable but would still be manageable to a certain degree for an intelligent translator, provided he were given - what? What would you require to be given before undertaking such a task? Remember the restriction put upon you: you will be forever denied the understanding of the text you are going to translate. (Bar-Hillel 1952)

For present-day users of machine translation, this thought experiment remains extremely apt.

From a modern perspective, the ideas presented at this first conference were by turns highly insightful and utterly unrealistic. Reifler (1952), for example, correctly foresaw two MT paradigms that would actually emerge decades later; these predictions include spoken language translation and the method of interactive disambiguation attempted by Melby (1995). In the latter case, Reifler correctly predicted claimed that such a method would in practice "remain academic" in large part because it would be "too slow."

With regard to MT designed for dissemination, Reifler stated that if such MT is to be "ready for publication, then MT, like all translations, will require a post-editor." This is an extremely insightful statement; even today, nearly all discussion of MT quality in comparison to human translation quality neglects the fact that high-quality human translation is typically predicated on a human editor verifying the quality of the human translator's work.

The main proposal of Reifler (1952) is, however, extremely naive and unrealistic, both from a technical standpoint and from a societal perspective. Channeling the most unrealistic aspects of 17th century pansophic language proposals, Reifler proposed that source text authors simply adopt the use of a special-purpose orthography designed to fully disambiguate such features as part of speech and semantic roles in order to ease the job of the

MT system. These assumptions are wildly unrealistic, both technologically and in human terms. In human terms, it is utterly unrealistic to impose such artificial orthographic requirements on authors of texts. And technologically, while the availability of part-of-speech tags and semantic role labels may in principle certainly make MT easier than when such information is not available, its presence by no means solves all problems in MT.

3.3 A decade of optimism

Overall, the mood of the conference was upbeat; armed with the newly developed electronic digital computers, and with the promise of (relatively) high-density magnetic storage media on the horizon, many researchers expected that substantial progress could be made in a relatively short amount of time. Hutchins (2014) described the period following this first MT conference as a "decade of optimism."

In 1954, researchers at IBM and Georgetown collaborated on the first public demonstration of machine translation (Dostert 1955). This demonstration system used a set of hand-crafted rules and a small vocabulary to translate a fixed set of sentences from Russian to English. Despite the limitations of the system, the demonstration was widely (and credulously) reported as a resounding success in the press (Hutchins 1999a). This demonstration greatly increased the profile of machine translation, and also helped lead the way for substantial U.S. federal funding of machine translation research.

Later that same year, the first Ph.D in machine translation was awarded to Anthony Oettinger by the department of Applied Mathematics at Harvard University (Oettinger 2000). Oettinger (1954) examined how to design a program for storing a bilingual Russian-English technical dictionary on an electronic digital computer with a magnetic storage device. Oettinger also reported preliminary success with simulated direct word-for-word machine translation, post-edited by monolingual domain experts.

In 1955, many of the presentations from the 1951 MT Conference were collected and published (along with a handful of new articles) in the first book on MT, edited by Locke and Booth (1955). Warren Weaver contributed the foreword to the book. In the foreword, Weaver wrote:

> Students of languages and of the structures of languages, the logicians who design computers, the electronic engineers who build and run them — and specially the rare individuals who share all of these talents and insights — are now engaged in erecting a new Tower of Anti-Babel. … The hopes for this new development are, one can believe, so reasonable and limited that this new tower will not fail through arrogance. No reasonable person thinks that a machine translation can ever achieve elegance and style. Pushkin need not shudder.

> (Weaver 1955)

Weaver concluded by comparing machine translation to cargo trucks loaded with information content:

> This, in fact, is the reasonable purpose of this effort. Not to charm or delight, not to contribute to elegance or beauty; but to be of wide service in the work-a-day task of making available the essential content of documents in languages which are foreign to the reader.

The same volume published the first known concrete algorithm for performing machine translation (Richens and Booth 1955) — on punch-card tabulating machines. The description by Richens and Booth describes two variants: an optimized variant for those with access to the latest punch-card hardware, and a slower variant for those with access to only the standard older punch-card hardware. The description is notable for its high level of detail — sufficient details are provided that a motivated researcher today should be able to emulate the Richens and Booth (1955) algorithms in software with relative ease. Three years later, the same research group (Booth 1958) published source code (in assembly language) for performing machine translation on an electronic digital computer with a magnetic drum.

3.4 The first wave crashes

In the final years of the 1950s, at the behest of the U.S. Office of Naval Research, Bar-Hillel was charged with conducting a second survey on the state of machine translation. Bar-Hillel noted that in 1952, at the time of the first MT conference, the total number of people working on MT, "[r]educed to full-time workers … could not at that time have been much more than three, and the amount of money spent that year not much more than ten thousand dollars." In his report, Bar-Hillel (1960) estimated that worldwide investment in MT research in 1958 had reached $3 million, with between 200 and 250 full-time equivalent MT researchers worldwide. The pessimism towards MT present in Bar-Hillel's 1960 report stood in stark contrast to his earlier optimism:

> During the first years of the research in MT, a considerable amount of progress was made which sufficed to convince many people, who originally were highly skeptical, that MT was not just a wild idea. It did more than that. It created among many of the workers actively engaged in this field the strong feeling that a working system is just around the corner. Though it is understandable that such an illusion should have been formed at the time, it was an illusion. (Bar-Hillel 1960)

Bar-Hillel argued that fully-automatic high quality machine translation, a goal which many contemporaneous MT research groups enthusiastically espoused, was unattainable “not only in the near future but altogether.” He claimed that the problems which had been solved to date were “but just the simplest ones, whereas the ‘few’ remaining problems were the harder ones—very hard indeed.” While Bar-Hillel’s 1960 report triggered concern, it did not result in widespread changes in research direction or technique among other MT researchers (Hutchins 1999b).

In 1964, the Automatic Language Processing Advisory Committee (ALPAC) was established under the aegis of the National Research Council, in large part to provide independent advice to U.S. federal agencies which had been providing funding support to MT research for the previous ten years. The committee examined the motivations for federal funding of both computational linguistics and machine translation, and considered whether additional federal funding for these disciplines was warranted.

One notable point identified by the committee was the need for robust evaluation metrics to judge the quality of both machine and human translations. To this end, the committee commissioned a study (Carroll 1966) which developed 9-point evaluation metrics for intelligibility and informativeness, and applied those scales to three human and three machine translations of a Russian scientific article.

The committee’s final report (ALPAC 1966) concluded that the current U.S. government demand for translations did not justify the use of MT, and that the poor quality of MT meant that MT was in practice usable only after post-editing. The final report recommended that any future government funding be focused on basic research in computational linguistics, and in the development of computational aids for human translators, rather than on machine translation. While MT research would continue in Europe and to a lesser extent in the U.S. after the publication of the ALPAC report, the report led many to view MT as a failed endeavor and U.S. federal funding for machine translation was virtually non-existent for the subsequent twenty years (Hutchins 1996).

4 Survey of Machine Translation Paradigms

By the late 1960s, the broad outlines (if not the actual implementations) of the major approaches to MT had been proposed. Bernard Vauquois (1968) encapsulated the major approaches in the diagram that became widely known in the machine translation field as the Vauquois triangle (reproduced here in Figure 1). The Vauquois triangle depicts machine translation as a process involving three major processes: analysis of the source text, transfer from a source representation to a target representation, and generation of the target text.

Three major paradigms (presented in §4.1.1–4.1.3 below) broadly

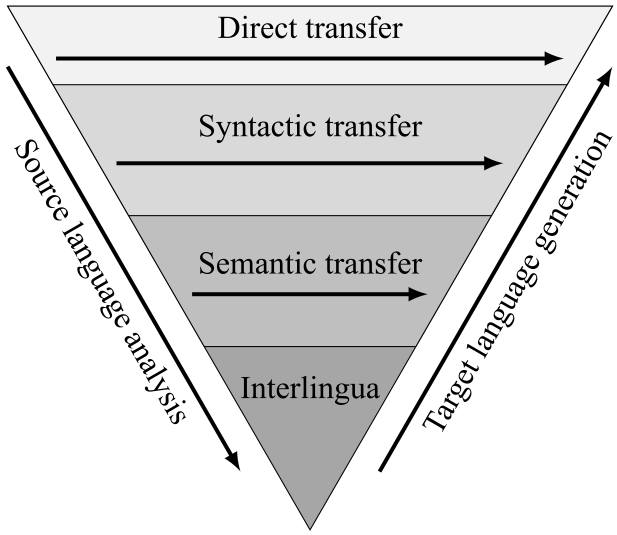

Figure 1: Machine translation triangle, adapted from Vauquois (1968).

"We can imagine as many levels as we wish from the zero level (level of the text considered as a string of characters), asymptotically towards a level of understanding. At each level a formalization of the input sentence can be defined. Then, it may be assumed that the deeper the level chosen, the easier the transfer is. At the limit, if the ideal level of understanding could be reached for a given sentence in one language, the same structural specifier would represent all the paraphrases of this sentence in all languages." (Vauquois 1976) encompass the dominant approaches to machine translation in the 20th century, in which machine translation software was built around programmed rules that directly encoded linguistic insights with regard to the translation process for a given language pair. An alternative family of methods (presented in §4.2 below) examine techniques by which the rules governing the translation process can be obtained automatically from corpora and applied in a stochastic process to obtain translations of previously unseen source texts. These statistical methods, first proposed by Weaver (1949) and first implemented by Brown et al. (1988), now represent the dominant approaches to machine translation in the early 21st century.

4.1 Rule-based Approaches

4.1.1 Direct Transfer

The earliest MT systems implemented various forms of direct transfer. In this approach, little or no source analysis was performed, and source tokens (or sometimes phrases) were directly transformed into their equivalent target language translations, usually through the use of encoded dictionaries. While some direct transfer systems implemented basic processes for applying morphological inflection and syntactic agreement on the target side, the direct transfer paradigm typically involved relatively little in the way of target language generation or processing. This direct transfer phase corresponds to the horizontal arrow at the top of the Vauquois triangle (see Figure 1).

Proposals for direct word-for-word translation can be traced back at least as far as Georges Artsrouni's 1933 patent (§2.2) for a mechanical lookup and retrieval system (Corbé 1960). Practical research attempting to implement direct word-for-word translation using encoded bilingual dictionaries began using punch-cards as early as 1948 (Booth and Richens 1952) and using an electronic digital computer as early as 1950 (Oswald and Fletcher 1951).

A consensus quickly emerged that at least some handling of multi-word expressions must be performed. Less than two years after conducting his initial direct word-for-word experiments, Victor Oswald, an early MT researcher and professor of Germanic studies at UCLA, was invited to discuss his work at the first MT conference. Addressing the conference attendees, Oswald said:

> When I learned that I had been summoned to address myself to the topic of word-by-word translation I felt like a geographer invited to discuss the utility of the conception that the world is flat. In short, I can only say that word-by-word translation is not possible, if we are to understand by the term a wordwise transverbalization from one language into another, particularly

from German into English. (Oswald 1952)

More sophisticated direct transfer systems were proposed and developed throughout the 1950s and into the 1960s. In addition to basic word-for-word dictionary lookup, these more advanced direct transfer systems typically included basic morphological analysis (splitting each word into a stem and affixes), and processes for matching and translating multi-word expressions, prioritized rule tables for resolving word order, grammatical agreement, and lexical and semantic ambiguities (Vauquois 1976). Two of the more well developed direct transfer systems were the Russian-English punch-card system developed for the U.S. Air Force by the University of Washington (Reifler et al. 1958) and the Chinese-English system developed later by the same group (Reifler 1967).

4.1.2 Analysis-Transfer-Generation

A subsequent major advance in MT methodology was the development of the analysis-transfer-generation paradigm. This paradigm explicitly arranges the machine translation process into three distinct phases. In the first phase, the raw source text is analyzed, resulting in a intermediate representation such as a source language syntax tree or semantic representation. In the second phase, the source language intermediate representation is transformed into an equivalent target language intermediate representation. Finally, in the third phase, target language text is generated from the target language intermediate representation. These three phases respectively correspond to the downward, horizontal, and upward arrows in the Vauquois triangle (see Figure 1).

The first known proposal for transfer-based mechanical translation can be traced to a 1933 patent a special purpose mechanical translation device. Trojanskij (1933) proposed a three-part translation process broadly similar to the analysis-transfer-generation paradigm. Trojanskij's system was never built. Independently, Bar-Hillel (1951) proposed that certain processes were fundamental to any MT system, among them morphological analysis, basic syntactic analysis, and transformation and reordering; these ideas directly foreshadowed later systems which used the analysis-transfer-generation paradigm.

Concrete development of syntactic transfer systems largely began at MIT under Victor Yngve (see, for example, Yngve 1960). This approach is typified by the Arabic-English system of Satterthwait (1965). Many of the dominant research and commercial systems of the 1960s through the 1990s were variants of the analysis-transfer-generation paradigm; examples of such systems include Logos, METAL, and Systran.

4.1.3 Interlingua

The third major paradigm in machine translation took the motivation behind the analysis-transfer-generation methods to its logical conclusion, attempting remove the need for a transfer phase by proposing the use of truly interlingual intermediate representations. This paradigm models the MT process as two phases. The source text is first subjected to deep analysis, resulting in a language-independent intermediate representation. Because the intermediate representation is language-independent, no transfer phase is needed, and the second phase directly generates the target language text from the intermediate representation. These two phases respectively correspond to the downward and upward arrows in the Vauquois triangle (see Figure 1).

The idea of using an interlingua for machine translation can be found as early as Weaver's famous 1949 memorandum (see §2.3) that kick-started 20th century research in machine translation. Two years later, Bar-Hillel (1951) claimed that long term, machine translation's best hope would be the development of a truly language-independent universal linguistic representation; Bar-Hillel argued that modern technology could make such a representation, hoped for since the 17th century (see §2.1), finally a reality.

Serious attempts at developing interlingual machine translation were typified by the University of Grenoble CETA system (Vauquois 1975), as well as the system developed at the University of Texas in the early 1970s. Research into interlingual machine translation continued through the 1990s (see, for example, Dorr 1992), but interlingual approaches remained less popular and generally less successful than shallower transfer-based approaches.

While in principle an interlingual MT system should use an intermediate representation that is truly language independent, in practice the actual intermediate representations used in these systems fell short of this ideal. Attempts to develop interlingual representations in the 20th century in many cases fell victim to the same shortcomings as attempts to develop universal semantic representations of language in the 17th century (Melby 1995). One of the only current attempts to develop an interlingua has centered around the Universal Networking Language initially developed by the United Nations University (UNL 1996; Cardeñosa et al. 2005).

4.2 Corpus-based Approaches

4.2.1 The cryptographic metaphor

In the years immediately following World War II, Warren Weaver, director of the Natural Sciences Division at the Rockefeller Foundation, began to contemplate whether recent advances in cryptographic analysis and the development of electronic digital computers might together be brought to bear on the problem of translation. In 1947, Weaver wrote to a colleague:

> Also knowing nothing official about, but having guessed
> and inferred considerable about, powerful new mechanized

> methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode." (Weaver 1947)

The idea of treating translation as a problem in cryptography was considered in greater detail two years later when Weaver wrote and distributed a memorandum on the topic to researchers interested in the nascent field of machine translation. Weaver (1949) took additional inspiration from recent work in information theory by Shannon (1948). Weaver argued that Shannon's approach for a statistical model of human language based on a limited window of context could be used to disambiguate words with multiple possible meanings during machine translation.

> If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words. "Fast" may mean "rapid"; or it may mean "motionless"; and there is no way of telling which. But if one lengthens the slit in the opaque mask, until one can see not only the central

word in question, but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word. (Weaver 1949)

Weaver's memorandum sparked substantial interest and research in machine translation, but virtually none of his contemporaries followed through with the proposed cryptographic approach to MT. Weaver's cryptographic metaphor was attacked in a detailed point-by-point critique by Mounin (1964). Mounin's rejection of the cryptographic approach was considered definitive well into the 1980s (see, for example Hutchins's 1986 MT textbook).

4.2.2 Example-based MT

In the rule-based MT paradigms presented in §4.1, translation rules were crafted by human linguistic analysts. In the mid-1980s, researchers in Japan began to explore the question of whether some of these rules could be obtained automatically from a parallel corpus of translated data. Nagao (1984) proposed that when a new sentence to be translated is encountered, segments of the sentence could be translated by finding existing translations of those segments or of similar segments. However, such MT based on translating by example still required substantial heuristics for deciding which translations of a segment to use, and how to glue translating segments together. Example-based MT systems thus represent a preliminary corpus-based paradigm part way between rule-based systems and statistical corpus-based approaches to MT. Operating on fuzzy matching principles, example-based MT bears some similarity to the translation memories later developed for use by human translators.

4.2.3 Word-based statistical MT

Throughout the 1970s and 1980s, research into automatic speech recognition (ASR) made substantial progress, in large part through the use of the noisy channel approach proposed by Shannon (1948). Buoyed by their success in ASR and the related task of automatic spelling correction, a research group at IBM under the direction of Fred Jelinek considered how to apply these same statistical techniques to machine translation, a field nearly entirely dominated by rule-based paradigms.

Substituting a probabilistic translation dictionary for the acoustic model used in ASR, the IBM team (Brown et al. 1988) presented the first concrete proposal for machine translation based on the information-theoretic ideas of Weaver and Shannon. Over the next five years, the IBM team developed five increasingly sophisticated models for probabilistic word-for-word translation (Brown et al. 1993). IBM's statistical machine translation (SMT) models learned probabilistic translation rules directly from a parallel corpus of translated data, without requiring linguistic insights or human intervention. Just as Voltaire's classic novel criticized the idealistic philosophy and universal language proposal of Leibniz in the 17th century, the IBM statistical machine translation system called Candide represented a stark alternative to the manually crafted rules of the dominant transfer-based and interlingual MT systems of the 20th century.

In 1995, the Center for Language and Speech Processing at Johns Hopkins University began an annual series of 6-week summer workshops designed to advance the state of the art in speech and language processing. In 1999, for the first time the workshop included a team focusing on machine translation. The team constructed a word-based statistical MT toolkit called Egypt, the first statistical MT system built outside of IBM. The Egypt toolkit included a training module (called GIZA) capable of training word-based models from a parallel corpus and a decoder[1] (called Weaver) capable of translating using those trained models (Al-Onaizan et al. 1999).

4.2.4 Phrase-based SMT

MT research in the years after the introduction of word-based SMT proceeded along two distinct tracks, as some research groups continued investing in traditional rule-based approaches, while others began tentative explorations

[1] In statistical machine translation, the software component that performs the translation is called the *decoder*. This term follows from Warren Weaver's cryptographic analogy: "When I look at an article in Russian, I say 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'" (Weaver 1947)

into the statistical corpus-based approach pioneered by IBM. In the late 1990s, a number of research groups began to examine how IBM's word-based translation models could be extended to allow models that captured translations of multiple adjacent words (Melamed 1997; Wang and Waibel 1998; Och et al. 1999), and began to develop algorithms for translating with these new models (Zens et al. 2002; Watanabe et al. 2003; Koehn et al. 2003).

Following Koehn et al. (2003), this new paradigm became known as phrase-based statistical machine translation. From a linguistic perspective, this name is somewhat misleading, as the so-called "phrases" in phrase-based translation need not represent linguistic constituents. Phrase-based translation models typically do contain many phrase pairs that correspond to linguistic constituents (such as *das Buch* → *the book*), but such models also contain many more phrase pairs that do not (such as *spaß am* → *fun with the*).

By the middle years of the '00s, phrase-based machine translation had almost completely replaced word-based MT as the dominant statistical MT paradigm. The first widely available statistical phrase-based decoder was Pharaoh (Koehn 2004). The Pharaoh system included a decoder, but no software to train models, and was distributed under a non-commercial license and in binary form only, which meant that other research groups could not modify and extend the system. Two years later, another JHU summer workshop team (Koehn et al. 2006) was organized with the goal of developing an open source phrase-based SMT system. Over the subsequent decade, the resulting system, called Moses,[2] has been widely extended and widely adopted in research, government, and industry as the *de facto* standard for statistical phrase-based machine translation. Since 2006, popular online MT systems such as Google Translate have used phrase-based SMT for most language pairs (Och 2006).

4.2.5 Syntactic SMT

For several decades in the late 20th century, rule-based transfer systems represented the state-of-the-art in machine translation. As word-based and later phrase-based statistical machine translation developed, the question arose whether the best of statistical and rule-based techniques might be combined in a statistical transfer-based paradigm that used syntactic rules extracted automatically from parallel data. Beginning in the late 1990s, various statistical MT techniques were developed (Wu 1997; Alshawi et al. 1998; Yamada and Knight 2001) that made use of syntax on either the source language side, the target language side, or both. By the end

[2] The choice of names for both Pharaoh and Moses continued the Egypt-themed naming convention for SMT tools established at the 1999 JHU summer workshop. The name Moses, in particular, was chosen for the open source MT tool to succeed the closed-source Pharaoh because Moses set the people free from Pharaoh.

of the first decade of the 21st century, most research groups working on statistical machine translation had experimented with various methods of incorporating syntax into a statistical machine translation framework.

In the early years of syntactic SMT development, such systems were typically outperformed by phrase-based SMT alternatives. Over the past decade, various improvements in syntactic SMT grammar formalisms (for example, Zollmann and Venugopal 2006) and feature handling (for example, Chiang et al. 2009) have combined to make syntactic SMT approaches more competitive, at least for certain language pairs. Currently, syntactic SMT methods are known to provide state-of-the-art performance in terms of translation quality for some language pairs (Baker et al. 2009; Bojar et al. 2015) such as Urdu-English and English-German.

4.2.6 Current trends in SMT

Statistical machine translation remains an extremely active research field. While phrase-based SMT has been the dominant MT paradigm of the past decade, it seems likely that new paradigms will emerge and old paradigms will be reinvigorated. Research is ongoing into how to best develop and integrate increasingly sophisticated syntactic, semantic, and even discourse-level knowledge into statistical machine translation frameworks (see, for example, Banarescu et al. 2013). Separately, substantial recent effort has examined how to perform machine translation using neural networks (see, for example, Kalchbrenner and Blunsom 2013) in which word meaning is indirectly represented via a potentially language-independent high-dimensional vector.

5 Human-Machine Partnerships

Throughout the history of machine translation, one major question has dominated: What roles should humans and machines respectively play in the translation process? To make sense of this question, it is necessary to consider the various use cases which necessitate translation.

5.1 MT and lay end-users

Most laypersons familiar with machine translation today have used MT in the context of information assimilation; in a typical example, a person encounters a news article online written in a foreign language, and uses functionality built into their web browser to translate the content using Google Translate. For many language pairs, fully automated machine translation today is of sufficiently high quality that the hypothetical layperson in the preceding example can likely ascertain the broad outlines of the news content, and quite possibly most or even all of the salient details.

A similarly empowering use case is the use of MT for real-time long-distance communication by end-users who do not share a common language. While text-to-text chat translation services have existed in some form for some time (see, for example Flournoy and Callison-Burch 2000), the widespread adoption of MT through such mainstream internet communication modalities as Facebook and Skype seems likely to substantially magnify the impact of such services. In particular, the advent of free online voice-to-voice MT through Skype Translator has enormous potential for positively impacting lay-user to lay-user communication in areas of life outside the official settings where professional interpreting services are typically encountered.

In each of these use cases, MT has begun to fulfill the long-sought goal of improving "communication between peoples" (Weaver 1947). The widespread availability of MT has enlarged the circle of persons able to access foreign language information content. Many end-users in these use cases, perhaps even the vast majority, would not seek and pay for professional human services were MT unavailable; rather they would simply forgo access to the foreign-language content or (if possible) seek another source for that content. The lay users here are active agents who make use of MT as an empowering tool, enabling access to otherwise inaccessible information content.

5.2 MT and subject matter experts

A related use case involves end-users who are subject matter experts in a given domain. In the simplest embodiment of this use case, the end-user uses MT to simply access foreign language information content (exactly as in §5.1 above). In a somewhat more involved use case, MT is used to enable (potentially large-scale) automated searches through foreign language content (see, for example, Nuutila 2005 and Onyshkevych 2014). In its most complex form, the (potentially monolingual) subject matter expert attempts a more ambitious undertaking: post-editing a machine translation into a usable and valid translation of the source text.

In the 1940s, before the development of any actual MT systems, it was observed that monolingual subject matter experts armed with minimal resources (such as a bilingual dictionary) could in some cases perform better translations than bilingual translators not familiar with the subject matter of a specialized text (Richens 1984). This hypothesis was independently tested in the context of post-editing MT (with preliminary success) by Oettinger (1954) and Schwartz (2014). Both studies indicated the ability of a monolingual subject matter expert to successfully construct a high quality post-edited translation given machine translations of a document in the user's area of expertise.

MT can be effectively used by subject matter experts to enable more thorough access to a multilingual range of research sources (Ross 2015). But this use case presents an even larger potential for certain language service providers (LSPs) and freelance translators with specialized subject matter expertise. In the case of LSPs who specialize in certain subject areas, this use case lays out the possibility for an entirely new category of language workers: subject matter experts who are monolingual in the target language or who are bilingual, but not in the language of a particular source document. Such workers could provide first-pass post-editing and triage, passing only the most challenging segments on to translators or post-editors fluent in the source language. Similarly, this workflow could be utilized by freelance translators who are expert in a particular subject area to broaden the base of source languages from which they translate.

5.3 MT and professional end-users

The use case for MT most commonly considered in conjunction with translation professionals involves the use of MT as a starting point for a human constructing a translation; this idea has a long history. Trojanskij (1933) proposed the use of monolingual source language pre-editors to prepare a text for translation, in conjunction with a mechanical dictionary for performing the translation, monolingual target language post-editors to correct the output, and bilingual literary editors to polish and revise the final product (Hutchins and Lovtskii 2000). Pre- and post-editors were proposed independently by later MT researchers as research began in the 1940s and 1950s (Bar-Hillel 1951).

Many MT researchers during the 20th century assumed that post-editing represented only a temporary waypoint, the need for which would be obviated once fully automatic high quality MT was developed. Others, such as Bar-Hillel (1952), argued that in order for MT to be successful, it must necessarily be attempted in the context of a human-machine partnership. Even when used in partnership with humans, the question of MT efficacy for human professionals has a long and controversial history. The ALPAC report famously claimed that MT was neither necessary to satisfy demand for translation nor effective in speeding up the translation process when used in conjunction with post-editing. Other more recent studies have shown productivity gains when MT was used with post-editing (Plitt and Masselot 2010). As MT research began, Bar-Hillel predicted that even if human-machine MT partnerships were not economically feasible at that time, such partnerships would make sense in the future:

> Electronic machines will doubtless become cheaper, human brains probably more expensive. A partnership that could not stand free competition today may well outfit its human competitors in some not too remote future. (Bar-Hillel 1952)

MT today has unquestionably research the point where it is sufficiently cheap for Bar-Hillel's prediction to come true; the questions that remain are how such MT systems can be made accessible to human translators in ways that are sensitive to professional requirements (such as client confidentiality), and what roles the human and machine partners should respectively play in the translation process.

5.3.1 Machine-centric translation

Some of the earliest ideas for human-machine partnership in MT (dating back to Trojanskij 1933) proposed that the human partner ensure all incoming source texts be simple and unambiguous, in order to make the machine's task of translation as easy as possible. In these proposed approaches, pre-editors would annotate source texts to clarify all ambiguities, either manually (Reifler 1952) or via a special-purpose user interface (Boitet and Blanchon 1994). Alternatively, authors of texts intended for translation would write using an specially-designed artificial auxiliary language (Dodd 1952) or in a simplified variant of a natural language, such as Basic English (Ogden 1930). While in general, such approaches were not widely adopted, successful use of controlled language has been adopted in certain industrial environments in which documents are authored with the intention of being translated. Kamprath et al. (1998), for example, describe Caterpillar's use of controlled language for multilingual localization of specialized documentation.

As MT researchers throughout the latter half of the 20th century adopted the analysis-transfer-generation approach to MT (see §4.1.2), a recurring problem was commonly encountered: How should an MT system proceed when a source sentence could be legitimately analyzed in more than one way? Where early machine-centric ideas suggested that all ambiguity be resolved by a pre-editor, later proposals in the 1970s suggested that MT systems be "aided by human translator in a conversational way" (Vauquois 1976) when challenging decision points in the translation process were encountered. A prominent example of this approach can be seen in research at Brigham Young University beginning with Melby (1973).

The third mode of machine-centric translation is by far the simplest and most widespread: post-editing. In post-editing scenarios, a translation professional is presented with the raw output of an MT system and tasked with transforming this output into a fluent and adequate translation of the source text. Whether such a post-editing process increases the speed or quality of the overall process is almost certainly dependent on numerous factors, not least of which is the quality of the MT system being used. But regardless of the actual and perceived value of the process, in typical post-editing partnerships the machine partner is responsible for propounding the structure and vocabulary of each target segment, and the human partner simply corrects any errors in translation. While post-editing scenarios can in principle be configured such that the MT system learns from such human feedback (Denkowski et al. 2014), most post-editing scenarios are not.

5.3.2 Human-centric translation

In 1980, Martin Kay called for a re-examination of the role of humans and machines in the translation process, away from machine-centric translation paradigms, and towards a gradual and human-centric partnership in which well-designed computer programs modestly assist their human counterparts. Kay's dream in many ways presaged the development of computer-aided translation (CAT) tools that are in wide use today. One major component of such a human-centric translation system had been first proposed two years earlier:

> the system would be instructed to compare the new text, probably sentence by sentence, with all the previously recorded texts … and to print out the nearest available equivalent for each sentence (Arthern 1978)

This proposed component, today known as a translation memory (TM), is now an essential component for many translation professionals, and is integrated into all major CAT tools.

Some CAT tools also allow for integration with specific MT systems, presenting MT output as an additional information source alongside TM fuzzy matches. The adoption and use of MT in CAT tools, in a framework where the human partner maintains agency over the translation process, is not widespread to date; and the use of MT that adapts to human feedback (as proposed by Melby 1989 and implemented by Denkowski et al. 2014) is virtually non-existent.

Recent research systems have proposed an even more advanced mode of human-centric interactive translation. In such systems, MT output is presented to the human partner, but the translation process is driven by the human, not by the machine. As the human partner begins typing a translation, the machine translation system comes up with a newly hypothesized translation of the sentence that begins with the words typed by the human partner. At any point, the human partner can either accept the proposed MT completion of the sentence, or continue to type in their own translation of the source sentence. While the first such systems were only research prototypes (Langlais et al. 2000; Barrachina et al. 2009; Koehn 2009), recent open source (Sanchis-Trilles et al. 2014) and commercial offerings (Green 2015) have begun to support this interactive human-driven translation modality.

6 Conclusion

The history of machine translation in the 20th and 21st centuries was shaped over the decades by diverse ideas from semiotics, information theory, linguistics, computation complexity theory, and human-computer interaction. The influence of 17th century notions of universal language surfaced in early discussions of MT in the late 1940s and early 1950s. The reflections of this centuries-old search for a universal semantic representation can be clearly seen throughout the remainder of the 20th century in serious efforts to develop MT based on deep syntactic transfer and interlingua.

Various developers of machine translation through the years have shown both great hubris and great humility. Beginning in the 1950s, certain MT researchers made unsubstantiated claims that fully automated high quality MT would be easily achieved within about five years time. The failure of these wildly optimistic claims to materialize was a contributing factor in the ALPAC report, and the subsequent de-prioritization of MT in the U.S. for a time. Other researchers, most notably Warren Weaver and Martin Kay, argued for a more modest approach to MT, developing capabilities modestly and incrementally, and building human-centric interfaces in which MT serves as one tool among many for the humans whom it was built to serve.

References

Al-Onaizan, Yaser, Jan Curin, Michael Jahr, Kevin Knight, John Lafferty, Dan Melamed, Franz Josef Och, David Purdy, Noah A. Smith, and David Yarowsky (1999). Statistical machine translation. JHU Summer Workshop Final Report.

ALPAC, Automatic Language Processing Advisory Committee (1966). *Language and Machines: Computers in Translation and Linguistics*. Publication 1416. Washington, D.C.: National Research Council.

Alshawi, Hiyan, Srinivas Bangalore, and Shona Douglas (1998, August). Automatic acquisition of hierarchical transduction models for machine translation. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, Montréal, Québec, pp. 41–47. Association for Computational Linguistics.

Arthern, Peter J. (1978, November). Machine translation and computerized terminology systems: a translator's viewpoint. In *Translating and the Computer: proceedings of a seminar*, London, pp. 77–108. North-Holland. Published 1979.

Bacon, Francis (1605). *Of the Proficience and Advancement of Learning, Divine and Human*. Oxford: Leonard Lichfield.

Baker, Kathy, Steven Bethard, Michael Bloodgood, Ralf Brown, Chris Callison-Burch, Glen Coppersmith, Bonnie Dorr, Wes Filardo, Kendall Giles, Anni Irvine, Mike Kayser, Lori Levin, Justin Martineau, Jim Mayfield, Scott Miller, Aaron Phillips, Andrew Philpot, Christine Piatko, Lane Schwartz, and David Zajic (2009). Semantically informed machine translation (SIMT). SCALE Summer Workshop Final Report 2, JHU Human Language Technology Center Of Excellence.

Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider (2013, August). Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, Sofia, Bulgaria, pp. 178–186.

Bar-Hillel, Yehoshua (1951). The present state of research on mechanical translation. *American Documentation 2*(4), 229–237.

Bar-Hillel, Yehoshua (1952, June). Mechanical translation: needs and possibilities. In *Proceedings of the 1st Conference on Mechanical Translation*, Cambridge, Massachusetts. Massachusetts Institute of

Technology.

Bar-Hillel, Yehoshua (1960). The present status of automatic translation of languages. *Advances in Computers 1*, 91–163.

Barrachina, Sergio, Oliver Bender, Francisco Casacuberta, Jorge Civera, Elsa Cubel, Shahram Khadivi, Antonio Lagarda, Hermann Ney, Jesús Tomás, Enrique Vidal, and Juan-Miguel Vilar (2009, March). Statistical approaches to computer-assisted translation. *Computational Linguistics 35*(1), 3–28.

Boitet, Christian and Hervé Blanchon (1994, June). Multilingual dialogue-based MT for monolingual authors: the LIDIA project and a first mockup. *Machine Translation 9*(2), 99–132.

Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi (2015, September). Findings of the 2015 workshop on statistical machine translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, Lisbon, Portugal, pp. 1–46.

Booth, A. Donald and Richard Richens (1952, June). Some methods of mechanized translation. In *Proceedings of the 1st Conference on Mechanical Translation*, Cambridge, Massachusetts. Massachusetts

Institute of Technology.

Booth, Kathleen H. V. (1958). *Programming for an Automatic Digital Calculator*, Chapter 10: Mechanical Translation. New York: Academic Press Inc.

Brown, Peter F., John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, Robert L. Mercer, and Paul S. Roossin (1988, August). A statistical approach to language translation. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*, Budapest, Hungary, pp. 71–76.

Brown, Peter F., Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer (1993, June). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics 19*(2), 263–311.

Böhme, Jakob (1623). *Mysterium Magnum*.

Cardeñosa, Jesús, Alexander Gelbukh, and Edmundo Tovar (Eds.) (2005, February). *Universal Networking Language: Advances in Theory and Applications*, Volume 12 of *Research on Computing Science*.

Carroll, John B. (1966, September, December). An experiment in evaluating the quality of translations. *Mechanical Translation and*

Computational Linguistics 9(3-4), 55–66.

Chiang, David, Kevin Knight, and Wei Wang (2009, June). 11,001 new features for statistical machine translation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'09)*, Boulder, Colorado, pp. 218–226.

Comenius, John (1642). *The Way of Light*.

Corbé, Michael (1960, March). La machine à traduire française aura bientôt trente ans. *Automatisme 5*(3), 87–91.

Dalgarno, George (1661). *The Art of Signs: or A Universal Character and Philosophical Language*. J. Hayes.

Denkowski, Michael, Chris Dyer, and Alon Lavie (2014, April). Learning from post-editing: Online model adaptation for statistical machine translation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL'14)*, Gothenburg, Sweden, pp. 395–404.

Descartes, René (1629, November). Letter to Mersenne.

Dodd, Stuart (1952, June). Model english for mechanical translation. In *Proceedings of the 1st Conference on Mechanical Translation*, Cambridge, Massachusetts. Massachusetts Institute of Technology.

Dorr, Bonnie (1992, August). Parameterization of the interlingua in machine translation. In *Proceedings of the Fifteenth International Conference on Computational Linguistics (COLING'92)*, Nantes, France, pp. 624–630.

Dostert, Leon E. (1955). The Georgetown-I.B.M. experiment. In William N. Locke and A. Donald Booth (Eds.), *Machine Translation of Languages*. Westport, Connecticut: Greenwood Press.

Flournoy, Raymond S. and Chris Callison-Burch (2000, November). Reconciling user expectations and translation technology to create a useful real-world application. In *Translating and the Computer 22: Proceedings of the Twenty-second International Conference on Translating*.

Goldstine, Herman H. and Adele Goldstine (1946). The electronic numerical integrator and computer (ENIAC). *Mathematical Tables and Other Aids to Computation 2*(15), 97–110.

Green, Spence (2015, October–November). Beyond post-editing: Interactive and adaptive translation technologies. In *Tutorial Sessions of MT Summit XV*, Miami, Florida.

Hobbes, Thomas (1655). *Elements of Philosophy. The First Section: Concerning Body*.

Hutchins, John (1986). *Machine Translation: Past, Present, Future*. Ellis Horwood Series in Computers and their Applications. Chichester, England: Ellis Horwood.

Hutchins, John (1996, June). ALPAC: the (in)famous report. *MT News International* (14), 9–12.

Hutchins, John (1998, May). Milestones in machine translation no.3: Bar-Hillel's survey 1951. *Language Today* (8), 12–13.

Hutchins, John (1999a, January). Milestones in machine translation. no 5: The IBM-Georgetown demonstration, January 1954. *Language Today* (16), 19–20.

Hutchins, John (1999b, September). Milestones in machine translation. no 6: Bar-Hillel and the nonfeasibility of FAHQT. *International Journal for Language and Documentation*, 20–21.

Hutchins, John (2004, Jan-June). Two precursors of machine translation: Artsrouni and Trojanskij. *International Journal of Translation 16*(1), 11–31.

Hutchins, John (2014). The history of machine translation in a nutshell.

Hutchins, John and Evgenii Lovtskii (2000). Petr Petrovich Troyanskii (1894–1950): A forgotten pioneer of mechanical translation. *Machine Translation 15*(3), 187–221.

Kalchbrenner, Nal and Phil Blunsom (2013, October). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP'13)*, Seattle, Washington, pp. 1700–1709.

Kamprath, Christine, Eric Adolphson, Teruko Mitamura, and Eric Nyberg (1998, May). Controlled language for multilingual document production: Experience with caterpillar technical english. In *Proceedings of the Second International Workshop on Controlled Language Applications (CLAW'98)*, Pittsburgh, Pennsylvania.

Kay, Martin (1980, October). The proper place of men and machines in language translation. Technical Report CSL-80-11, Xerox PARC, Palo Alto, California. Reprinted in *Machine Translation*, 12(1–2):3–23, 1997.

Koehn, Philipp (2004, September–October). Pharaoh: a beam search decoder for phrase-based statistical machine translation models. In *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA'04)*, Washington, D.C., pp. 115–124.

Koehn, Philipp (2009, August). A web-based interactive computer aided translation tool. In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, Suntec, Singapore, pp. 17–20.

Koehn, Philipp, Marcello Federico, Wade Shen, Nicola Bertoldi, Ondřej Bojar, Chris Callison-Burch, Brooke Cowan, Chris Dyer, Hieu Hoang, Richard Zens, Alexandra Constantin, Christine Corbett Moran, and Evan Herbst (2006). Open source toolkit for statistical machine translation: Factored translation models and confusion network decoding. JHU Summer Workshop Final Report.

Koehn, Philipp, Franz Joseph Och, and Daniel Marcu (2003, May–June). Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL '13)*, Edmonton, Canada, pp. 48–54.

Langlais, Philippe, George Foster, and Guy Lapalme (2000, May). TransType: A computer-aided translation typing system. In *Proceedings of the ANLP/NAACL 2000 Workshop on Embedded Machine Translation Systems*, Seattle, Washington, pp. 46–51.

Leibniz, Gottfried (1666). *On the Art of Combinations*.

Locke, William N. and A. Donald Booth (Eds.) (1955). Westport, Connecticut: Greenwood Press.

Loomis, W. F. (1951, January). Present interest in mechanical translation. Memorandum.

Melamed, I. Dan (1997, July). A portable algorithm for mapping bitext correspondence. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97)*, Madrid, Spain, pp. 305–312.

Melby, Alan (1973, August–September). Junction grammar and machine assisted translation. In *Computational and mathematical linguistics: Proceedings of the International Conference on Computational Linguistics (COLING'73)*, Volume 2, Pisa, Italy, pp. 785–824.

Melby, Alan (1989, April). Translation support. In *IFTT'89: Harmonizing human beings and computers in translation. International Forum for Translation Technology*, pp. 26–28.

Melby, Alan (1995). *The Possibility of Language*, Volume 14 of *Benjamins Translation Library*. Amsterdam: John Benjamins. With C. Terry Warner.

Mel'čuk, I.A. and R.D. Ravič (1967). *Автоматический перевод: 1949-1963 критико-библиографический справочник (Automatic translation: 1949-1963, A critical bibliographic reference)*, Chapter История и современное состояние АП (The history and present state of MT). Moscow: Всесоюзный Институт Научной и Технической Информации : Институт Языкознания.

Mounin, Georges (1964). *La Machine à Traduire*. Number 32 in Janua

Linguarum. The Hague: Mouton.

Nagao, Makoto (1984). A framework of a mechanical translation between Japanese and English by analogy principle. In *Proceedings of the International NATO Symposium on Artificial and Human Intelligence*, Lyon, France, pp. 173–180.

Newton, Isaac (1661). Of an universall language. In Ralph W. V. Elliot (Ed.), *The Modern Language Review*, Volume 52:1, pp. 1–18. Published Jan. 1957.

Nuutila, Pertti (2005, September). *Rough Machine Translation in the Communication Process*. Licentiate thesis, University of Tampere.

Och, Franz Joseph (2006, 28 April). Statistical machine translation live. Google Research Blog.

Och, Franz Josef, Christoph Tillmann, and Hermann Ney (1999, June). Improved alignment models for statistical machine translation. In *Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP'99)*, College Park, Maryland.

Oettinger, Anthony G. (1954). *A Study for the Design of an Automatic Dictionary*. Ph. D. thesis, Harvard University, Cambridge, Massachusetts.

Oettinger, Anthony G. (2000). Machine translation at Harvard. In John

Hutchins (Ed.), *Early Years in Machine Translation: Memoirs and Biographies of Pioneers*, Volume 97 of *Studies in the History of the Language Sciences*. John Benjamins.

Ogden, Charles K. (1930). *Basic English: A General Introduction with Rules and Grammar*. London: Paul Treber.

Onyshkevych, Boyan (2014, October). Low resource languages for emergent incidents (LORELEI). DARPA Broad Agency Announcement (DARPA-BAA-15-04).

Oswald, Victor (1952, June). Word-by-word translation. In *Proceedings of the 1st Conference on Mechanical Translation*, Cambridge, Massachusetts. Massachusetts Institute of Technology.

Oswald, Victor and Stuart Fletcher (1951). Proposals for the mechanical resolution of German syntax patterns. *Modern Language Forum 36*(3–4), 1–24.

Plitt, Mirko and François Masselot (2010, January). A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics* (93), 7–16.

Reifler, Erwin (1952, June). Mechanical translation with a pre-editor, and writing for MT. In *Proceedings of the 1st Conference on Mechanical Translation*, Cambridge, Massachusetts. Massachusetts

Institute of Technology.

Reifler, Erwin (1954, August). The first conference on mechanical translation. *Mechanical Translation 1*(2).

Reifler, Erwin (1967). Chinese-English machine translation, its lexicographic and linguistic problems. In Andrew D. Booth (Ed.), *Machine Translation*. Amsterdam: North-Holland Publishing Company.

Reifler, Erwin, W. Ryland Hill, David L. Johnson, Lew R. Micklesen, Robert E. Wall, Jr., Udo K. Niehaus, Aristotelis D. Stathacopoulos, Donald Culbertson, and Donald W. Moore (1958, June). Linguistic and engineering studies in automatic language translation of scientific Russian into English. Technical report to the Air Force Intelligence Laboratory, Rome Air Development Center, under contracts AF 30(602)-1566 and AF 30(602)-1827, University of Washington, Seattle, Washington.

Reynolds, Craig (1954, December). The conference on mechanical translation held at M.I.T., June 17-20, 1952. *Mechanical Translation 1*(3), 47–55.

Richens, Richard H. (1984, May). Letter to John Hutchins, partially reproduced in *Machine Translation: Past, Present, Future* (Hutchins 1986).

Richens, Richard H. and Andrew D. Booth (1955). Some methods of mechanized translation. In William N. Locke and A. Donald Booth (Eds.), *Machine Translation of Languages*. Westport, Connecticut: Greenwood Press.

Ross, Daniel (2015, May). Using machine translation to enable highly multilingual scholarship. Presented at the First Machine Translation Marathon in the Americas (MTMA'15).

Sanchis-Trilles, Germán, Vicent Alabau, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Ulrich Germann, Jesús González-Rubio, Robin L. Hill, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Herve Saint-Amand, Chara Tsoukala, and Enrique Vidal (2014, December). Interactive translation prediction versus conventional post-editing in practice: a study with the CasMaCat workbench. *Machine Translation 28*(3), 217–235.

Satterthwait, Arnold C. (1965, February). Sentence-for-sentence translation: An example. *Mechanical Translation and Computational Linguistics 8*(2).

Schwartz, Lane (2014, October). Monolingual post-editing by a domain expert is highly effective for translation triage. In *Proceedings of the The Third Workshop on Post-editing Technology and Practice*, Vancouver, Canada.

Shannon, Claude E. (1948, July, October). A mathematical theory of communication. *The Bell System Technical Journal 27*, 379–423,623–656.

Slaughter, Mary M. (1982). *Universal languages and scientific taxonomy in the seventeenth century*. Cambridge: Cambridge University Press.

Stillman, Robert E. (1995). *The New Philosophy and Universal Languages in Seventeenth-Century England: Bacon, Hobbes, and Wilkins*. Lewisburg: Bucknell University Press.

Swift, Jonathan (1726, October). *Gulliver's Travels*. London: Benjamin Motte.

Trojanskij, Petr Petrovič (1933). Description of a machine for selecting and typing words when translating from one language into another or several others simultaneously. In John Hutchins and Evgenii Lovtskii (Eds.), *Machine Translation*, Volume 15:3, pp. 187–221. Published Sept. 2000.

Twain, Mark and Charles Dudley Warner (1874). *The Gilded Age: A Tale of Today*. San Francisco: American Publishing Company.

UNL (1996). Universal networking language: an electronic language for communication, understanding and collaboration.

Vaba Maa (1924, February 24). Kirjutusmasin-tõlk (typewriter-interpreter). *Vaba Maa* (46), p. 4.

Vauquois, Bernard (1968, August). A survey of formal grammars and algorithms for recognition and transformation in mechanical translation. In J.H. Morrell (Ed.), *Proceedings of the International Federation for Information Processing Congress (IFIP-68)*, Volume 2, Edinburgh, Scotland, pp. 1114–1122.

Vauquois, Bernard (1975). *La traduction automatique à Grenoble*. Paris: Dunod.

Vauquois, Bernard (1976). Automatic translation - a survey of different approaches. *Statistical Methods in Linguistics*.

Voltaire (1759). *Candide*.

Wang, Ye-Yi and Alex Waibel (1998, August). Modeling with structures in statistical machine translation. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)*, Montréal, Québec, pp. 1357–1363.

Watanabe, Taro, Eiichiro Sumita, and Hiroshi G. Okuno (2003, July). Chunk-based statistical translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL'03)*, Sapporo, Japan, pp. 303–310.

Weaver, Warren (1947, March). Letter to Norbert Wiener.

Weaver, Warren (1949, July). Translation. Memorandum.

Weaver, Warren (1955). The new tower. In William N. Locke and A. Donald Booth (Eds.), *Machine Translation of Languages*. Westport, Connecticut: Greenwood Press.

Wilkins, John (1668). *An Essay towards a Real Character and a Philosophical Language*. London: Royal Society.

Wu, Dekai (1997, September). Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics 23*(3), 377–403.

Yamada, Kenji and Kevin Knight (2001, July). A syntax-based statistical translation model. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics (ACL'01)*, Toulouse, France, pp. 523–530.

Yngve, Victor H. (1960, October). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society 104*(5), 444–466.

Zens, Richard, Franz Josef Och, and Hermann Ney (2002, September). Phrase-based statistical machine translation. In *Proceedings of Advances in Artificial Intelligence, 25th Annual German Conference on AI, (KI'02)*, Aachen, Germany, pp. 18–32.

Zollmann, Andreas and Ashish Venugopal (2006, June). Syntax augmented machine translation via chart parsing. In *Proceedings on the Workshop on Statistical Machine Translation (WMT'06)*, New York City, pp. 138–141.