Chapter 6

Machine translation: History of Research and Applications

W. John Hutchins

Formerly University of East Anglia, the United Kingdom

From 1949 to 1970

Within a few years of the first appearance of the 'electronic calculators' research had begun on using computers as aids for translating natural languages. The major stimulus was a memorandum in July 1949 by Warren Weaver, who put forward possible lines of research. One was a statistical approach expressed as a dictum that '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 referred also to war-time success in code-breaking, from developments by Shannon in information theory and from speculations about universal principles underlying natural languages (Weaver 1949). Within a few years research had begun at many US universities, and in 1954 the first public demonstration of the feasibility of translation by computer was given, in a collaboration of IBM and Georgetown University (Hutchins 2004). Although using a very restricted vocabulary and grammar it was sufficiently impressive to stimulate massive funding of what became known since that date as 'machine translation' (MT).

This first decade saw the beginnings of the three basic approaches to MT. The first was the 'direct translation' model, where programming rules were developed for translation specifically from one source language (SL) into one particular target language (TL) with a minimal amount of analysis and syntactic reorganisation. The second approach was the 'interlingua' model, based on abstract language-neutral representations (codes or symbols independent of both SL and TL), where translation would then be in two stages, from SL to the interlingua and from interlingua to TL. The third approach was less ambitious: the 'transfer approach', where conversion was through a transfer stage from abstract (i.e. disambiguated) representations of SL texts to equivalent TL representations; in this case, translation comprised three stages: analysis, transfer, and generation (or synthesis). (For a general historical survey of MT see Hutchins 1986.)

At the University of Washington Erwin Reifler led a team on German-English and Russian-English translation, which later led to the IBM system developed by Gilbert King on a special memory device (the 'photoscopic disk') developed for the US Air Force and in operation from 1958. The largest MT group in the US was at Georgetown University, which

did not continue with the method used in the 1954 experiment but based its system on rules derived from traditional grammars. There were three levels of analysis: morphological (including identification of idioms), syntagmatic (agreement of nouns and adjectives, government of verbs, modification of adjectives, etc.), and syntactic (subjects and predicates, clause relationships, etc.) Much of the linguistic research for the Russian-English system was undertaken by Michael Zarechnak; the program was based on work by Petr Toma (later designer of Systran) and by Antony Brown (his SLC program for French-English). In this form it was successfully demonstrated in 1961 and 1962, and as a result Russian-English systems were installed at Euratom in Ispra (Italy) in 1963 and at the Oak Ridge National Laboratory of the US Atomic Energy Commission in 1964.

Anthony Oettinger at Harvard University adopted in a gradualist approach. From 1954 to 1960 his group concentrated on the compilation of a massive Russian-English dictionary, to serve as an aid for translators (a forerunner of the now common computer-based dictionary aids), to produce crude word-for-word translations for scientists familiar with the subject, and as the basis for more advanced experimental work. From 1959 research turned to a method of syntactic analysis originally developed at the National Bureau of Standards under Ida Rhodes. This 'predictive syntactic analyzer' sought to identify permissible sequences of grammatical categories (nouns, verbs, adjectives, etc.) and to predict the probabilities of the following categories. Multiple parsings were generated to examine all possible predictions, but the results were often unsatisfactory, and by 1965 the Harvard group had effectively ceased MT research.

Research at the Massachusetts Institute of Technology, started by Bar-Hillel in 1951, was directed by Victor Yngve from 1953 until its end in 1965. Whereas other groups saw syntax as an adjunct to lexicographic transfer, as a means of resolving ambiguities and rearranging TL output, Yngve placed syntax at the centre: translation was a three-stage process, a SL grammar analyzed input sentences as phrase structure representations, a 'structure transfer routine' converted them into equivalent TL phrase structures, and the TL grammar rules produced output text. An important contribution of MIT was the development of the first string-handling programming language (COMIT). Eventually the limitations of the 'syntactic transfer' approach became obvious, and in 1964 Yngve acknowledged that MT research had come up against 'the semantic barrier... and that we will only have adequate mechanical translations when the machine can "understand" what it is translating' (Yngve 1964: 279).

There were other US groups at the University of Texas led by Winfried Lehmann, and at the University of California led by Sydney Lamb (who developed his 'stratificational'

model of language), both linguistics-based models. There were, however, no American groups taking the interlingua approach. This was the focus of projects elsewhere. At the Cambridge Language Research Unit, Margaret Masterman and her colleagues adopted two basic lines of research: the development of a prototype interlingua producing crude 'pidgin' (essentially word-for-word) translations, and the development of tools for improving and refining MT output, primarily by means of the rich semantic networks of a thesaurus (conceived as lattices of interlocking meanings.) At Milan, Silvio Ceccato concentrated on the development of an interlingua based on conceptual analysis of words (species, genus, activity type, physical properties, etc.) and their possible correlations with other words in texts.

In the Soviet Union research was as vigorous as in the United States and showed a similar mix of empirical and basic theoretical approaches. At the Institute of Precision Mechanics the research under D.Y. Panov on English-Russian translation was on lines similar to that at Georgetown, but with less practical success – primarily from lack of adequate computer facilities. More basic research was undertaken at the Steklov Mathematical Institute by Aleksej A. Ljapunov, Olga S. Kulagina and Igor A. Mel'čuk (of the Institute of Linguistics) – the latter working on an interlingua approach that led eventually to his 'meaning-text' model. This combined a stratificational dependency approach (six strata: phonetic, phonemic, morphemic, surface syntactic, deep syntactic, semantic) with a strong emphasis on lexicographic aspects of an interlingua. Fifty universal 'lexical functions' were identified at the deep syntactic stratum covering paradigmatic relations (e.g. synonyms, antonyms, verbs and their corresponding agentive nouns, etc.) and a great variety of syntagmatic relations (e.g. inceptive verbs associated with given nouns, *conference: open, war: break out*; idiomatic causatives, *compile: dictionary, lay: foundations*, etc.) Interlingua investigations were consonant with the multilingual needs of the Soviet Union and undertaken at a number of other centres. The principal one was at Leningrad State University, where a team under Nikolaj Andreev conceived an interlingua not as an abstract intermediary representation but as an artificial language complete in itself with its own morphology and syntax, and having only those features statistically most common to a large number of languages.

By the mid-1960s MT research groups had been established in many countries throughout the world, including most European countries (Hungary, Czechoslovakia, Bulgaria, Belgium, Germany, France, etc.), China, Mexico, and Japan. Many of these were short-lived; an exception was the project which begun in 1960 at Grenoble University (see next section).

In the 1950s optimism had been high; developments in computing and in formal linguistics, particularly in the area of syntax, seemed to promise great improvement in quality. There were many predictions of imminent breakthroughs and of fully automatic systems operating within a few years. However, disillusion grew as the complexity of the linguistic problems became more and more apparent. In a review of MT progress, Bar-Hillel (1960) criticized the prevailing assumption that the goal of MT research should be the creation of fully automatic high quality translation (FAHQT) systems producing results indistinguishable from those of human translators. He argued that it was not merely unrealistic, given the current state of linguistic knowledge and computer systems, but impossible in principle. He demonstrated his argument with the word *pen.* It can have at least two meanings (a container for animals or children, and a writing implement). In the sentence *The box was in the pen* we know that only the first meaning is plausible; the second meaning is excluded by our knowledge of the normal sizes of (writing) pens and boxes. Bar-Hillel contended that no computer program could conceivably deal with such 'real world' knowledge without recourse to a vast encyclopedic store.

By 1964, the US government sponsors had become increasingly concerned at the lack of progress; they set up the Automatic Language Processing Advisory Committee (ALPAC), which concluded in its report (ALPAC 1966) that MT was slower, less accurate and twice as expensive as human translation and that 'there is no immediate or predictable prospect of useful machine translation'. It saw no need in the United States for further investment in MT research; instead it recommended the development of machine aids for translators, such as automatic dictionaries, and continued support in basic research in computational linguistics.

The ALPAC report brought a virtual end to MT research in the United States for over a decade and it had great impact elsewhere in the Soviet Union and in Europe. However, MT research did continue in Canada, in France and in Germany. Within a few years Peter Toma, one of the members of the Georgetown University project, had developed Systran for operational use by the USAF (1970) and by NASA (in 1974/5), and shortly afterwards Systran was installed by the Commission of the European Communities for translating from English into French (1976) and later between other Community languages.

Throughout this period, research on MT became an 'umbrella' for much contemporary work in structural and formal linguistics (particularly in the Soviet Union), semiotics, logical semantics, mathematical linguistics, quantitative linguistics, and nearly all of what would now be called computational linguistics and language engineering (terms already in use since early 1960s). Initially, there were also close ties with cybernetics and information theory. In general, throughout the early period, work on MT (both theoretical and

practical) was seen to be of wide relevance in many fields concerned with the application of computers to ‘intellectual’ tasks; this was true in particular for the research on ‘interlingua’ aspects of MT, regarded as significant for the development of ‘information languages’ to be used in document retrieval systems.

From 1970 to 1989

Research did not stop completely, however, after ALPAC. Even in the United States groups continued for a few more years, at the University of Texas and at Wayne State University. But there was a change of direction. Where ‘first generation’ research of the pre-ALPAC period (1956-1966) had been dominated by mainly ‘direct translation’ approaches, the ‘second generation’ post-ALPAC was to be dominated by ‘indirect’ models, both interlingua and transfer based.

In the 1960s in the US and the Soviet Union MT activity had concentrated on Russian-English and English-Russian translation of scientific and technical documents for a relatively small number of potential users, most of whom were prepared to overlook mistakes of terminology, grammar and style in order to be able to read something which they would have otherwise not known about. Since the mid-1970s the demand for MT has come from quite different sources with different needs and different languages. The administrative and commercial demands of multi-lingual communities and multinational trade have stimulated the demand for translation in Europe, Canada and Japan beyond the capacity of the traditional translation services. The demand is now for cost-effective machine-aided translation systems which can deal with commercial and technical documentation in the principal languages of international commerce.

At Montreal, research began in 1970 on a syntactic transfer system for English-French translation. The TAUM project (Traduction Automatique de l’Université de Montréal) had two major achievements: firstly, the Q-system formalism for manipulating linguistic strings and trees (later developed as the Prolog programming language), and secondly, the Météo system for translating weather forecasts. Designed specifically for the restricted vocabulary and limited syntax of meteorological reports, Météo has been successfully operating since 1976 (since 1984 in a new version). The TAUM group attempted to repeat this success in another field, that of aviation manuals, but failed to overcome the problems of complex noun compounds and phrases, and the project ended in 1981.

A similar fate met the ITS system at Brigham Young University. This was a transfer-based interactive multilingual system based on Eldon G. Lytle’s junction grammar. The aim was a commercial system but an internal evaluation in 1979 – a decade after the project had

begun – concluded that the system had become too complex, and recommended the development of practical computer aids for translators (cf. ALPS, below).

Throughout the 1980s research on more advanced methods and techniques continued. For most of the decade, the dominant strategy was that of 'indirect' translation via intermediary representations, sometimes interlingual in nature, involving semantic as well as morphological and syntactic analysis and sometimes non-linguistic 'knowledge bases'. There was an increasing emphasis on devising systems for particular subject domains and for particular specific purposes, for monolingual users as well as bilingual users (translators), and for interactive operation rather than batch processing.

The most notable research projects were the GETA-Ariane system at Grenoble, SUSY and ASCOF at Saarbrücken, Mu at Kyoto, DLT at Utrecht, Rosetta at Eindhoven, the knowledge-based MT project at Carnegie-Mellon University (Pittsburgh), and two ambitious international multilingual projects: Eurotra, supported by the European Communities, involving teams in each member country; and the Japanese CICC project with participants in China, Indonesia and Thailand.

Between 1960 and 1971 the group established by Bernard Vauquois at Grenoble University developed an interlingua system for translating Russian mathematics and physics texts into French. The 'pivot language' of CETA (Centre d'Etudes pour la Traduction Automatique) was a formalism for representing the logical properties of syntactic relationships. It was not a pure interlingua as it did not provide interlingual expressions for lexical items; these were translated by a bilingual transfer mechanism. Syntactic analysis produced first a phrase-structure (context-free) representation, then added dependency relations, and finally a 'pivot language' representation in terms of predicates and arguments. After substitution of TL lexemes (French), the 'pivot language' tree was converted first into a dependency representation and then into a phrase structure for generating French sentences. A similar model was adopted by the group at Texas during the 1970s in its METAL system: sentences were analyzed into 'normal forms', semantic propositional dependency structures with no interlingual lexical elements.

By the mid-1970s, the future of the interlingua approach was in doubt. The main problems identified were attributed to the rigidity of the levels of analysis (failure at any one stage meant failure to produce any output at all), the inefficiency of parsers (too many partial analyses which had to be 'filtered' out), and in particular loss of information about surface forms of the SL input which might have been used to guide the selection of TL forms and the construction of acceptable TL sentence structures.

After the disappointing results of its interlingua system, the Grenoble group (GETA, Groupe d'Etudes pour la Traduction Automatique) began development of its influential Ariane system. Regarded as the paradigm of the 'second generation' linguistics-based transfer systems, Ariane influenced projects throughout the world in the 1980s. Of particular note were its flexibility and modularity, its algorithms for manipulating tree representations, and its conception of static and dynamic grammars. However, like many experimental MT systems, Ariane did not become an operational system, and active research on the system ceased in the late 1980s.

Similar in conception to the GETA-Ariane design was the Mu system developed at the University of Kyoto under Makoto Nagao. Prominent features of Mu were the use of case grammar analysis and dependency tree representations, and the development of a programming environment for grammar writing (GRADE). Another experimental system was developed at Saarbrücken (Germany), a multilingual transfer system SUSY (Saarbrücker Übersetzungssystem), displaying a heterogeneity of techniques: phrase structure rules, transformational rules, case grammar and valency frames, dependency grammar, the use of statistical data, etc.

The best known project of the 1980s was the Eurotra project of the European Communities. Its aim was the construction of an advanced multilingual transfer system for translation among all the Community languages – on the assumption that the 'direct translation' approach of the Communities' Systran system was inherently limited. Like GETA-Ariane and SUSY the design combined lexical, logico-syntactic and semantic information in multilevel interfaces at a high degree of abstractness. No direct use of extra-linguistic knowledge bases or of inference mechanisms was made, and no facilities for human assistance or intervention during translation processes were to be incorporated. A major defect was the failure to tackle problems of the lexicon, both theoretically and practically; by the end of the 1980s no operational system was in prospect and the project ended.

During the latter half of the 1980s there was a general revival of interest in interlingua systems, motivated in part by contemporary research in artificial intelligence and cognitive linguistics. The DLT (Distributed Language Translation) system at the BSO software company in Utrecht (The Netherlands), under the direction of Toon Witkam, was intended as a multilingual interactive system operating over computer networks, where each terminal was to be a translating machine from and into one language only. Texts were to be transmitted between terminals in an intermediary language, a modified form of Esperanto. A second interlingua project in the Netherlands was the Rosetta project at Philips (Eindhoven) directed

by Jan Landsbergen. The aim was to explore the use of Montague grammar in interlingual representations, and as a secondary goal, the exploration of the reversibility of grammars, i.e. grammatical rules and transformations that could work in both directions between languages.

In the latter half of the 1980s Japan witnessed a substantial increase in MT research activity. Most of the computer companies (Fujitsu, Toshiba, Hitachi, etc.) began to invest large sums into an area which government and industry saw as fundamental to the coming 'fifth generation' of the information society. The research, initially greatly influenced by the Mu project at Kyoto University, showed a wide variety of approaches. While transfer systems predominated there were also interlingua systems, e.g. the PIVOT system at NEC and the Japanese funded multilingual multinational project, from the mid 1980s to the mid 1990s, already mentioned above.

As in the previous decade, many research projects were established in the 1980s outside North America, Western Europe, and Japan – in Korea (sometimes in collaborative projects with Japanese and American groups), in Taiwan (e.g. the ArchTran system), in mainland China at a number of institutions, and in Southeast Asia, particularly in Malaysia.

There was also an increase in activity in the Soviet Union. From 1976 most research was concentrated at the All-Union Centre for Translation in Moscow. Systems for English-Russian (AMPAR) and German-Russian translation (NERPA) were developed based on the direct approach, but there was also work under the direction of Yurij Apres'jan based on Mel'čuk's 'meaning-text' model – Mel'čuk himself had been obliged to leave the Soviet Union in 1977. This led to the advanced transfer systems FRAP (for French-Russian), and ETAP (for English-Russian). Apart from this group, however, most activity in the Soviet Union focused on the production of relatively low-level operational systems, often involving the use of statistical analyses – where the influence of the 'Speech Statistics' group under Raimund Piotrowski (Leningrad State University) has been particularly significant for the development of many later commercial MT systems in Russia.

During the 1980s, many researchers believed that the most likely means for improving MT quality would come from natural language processing research within the context of artificial intelligence (AI). Investigations of AI methods in MT began in the mid-1970s with Yorick Wilks' work on 'preference semantics' and 'semantic templates'. A number of projects applied knowledge-based approaches – some in Japan (e.g. the LUTE project at NTT, and the ETL research for the Japanese multilingual project), others in Europe (e.g. at Saarbrücken and Stuttgart), and many in North America. The most important group was at Carnegie-Mellon University in Pittsburgh under Jaime Carbonell and Sergei Nirenburg,

which experimented with a number of knowledge-based MT systems (Goodman and Nirenburg 1991).

The 1980s witnessed the emergence of a variety of operational MT systems. First there were a number of mainframe systems. Best known is Systran, operating in many pairs of languages; others were: Logos for German-English translation and for English-French in Canada; the internally developed systems for Spanish-English and English-Spanish translation at the Pan American Health Organization; systems developed by the Smart Corporation for large organizations in North America; and the Metal system from Siemens for German-English translation; and major systems for English-Japanese and Japanese-English translation came from Japanese computer companies, Fujitsu, Hitachi and Toshiba.

The wide availability of microcomputers and of text-processing software led to a commercial market for cheaper MT systems, exploited in North America and Europe by companies such as ALPS, Weidner, Linguistic Products, Tovna and Globalink, and by many Japanese companies, e.g. Sharp, NEC, Oki, Mitsubishi, Sanyo. Other microcomputer-based systems came from China, Taiwan, Korea, Bolivia, Eastern and Central Europe, e.g. PROMT from Russia.

Finally, not least, there was the beginning of systems offering some kind of translation for spoken language. These were the phrase-book and PC-based systems which included the option of voice output from written text – it seems that Globalink in 1995 was the earliest. But automatic speech synthesis of text-to-text translation is not at all the same as genuine 'speech-to-speech translation'. Research on speech translation did not start until the late 1980s [see below].

Applications of MT up to 2000: Translation Tools

Until the middle of the 1990s there were just two basic ways in which machine translation systems were used. The first was the traditional large-scale system mounted on mainframe computers in large companies. The purpose was to use MT in order to produce publishable translations. The output of MT systems were thus revised (post-edited) by human translators or editors familiar with both source and target languages. Revision for MT differs from the revision of traditionally produced translations; the computer program is regular and consistent with terminology, unlike the human translator, but typically it contains grammatical and stylistic errors which no human translator would commit. Hence, there was opposition from translators (particularly those with the task of post-editing) but the advantages of fast and consistent output has made large-scale MT cost-effective. In order to

improve the quality of the raw MT output many large companies included methods of 'controlling' the input language (by restricting vocabulary and syntactic structures) in order to minimise problems of disambiguation and alternative interpretations of structure and thus improve the quality. Companies such as the Xerox Corporation used the Systran systems with a 'controlled language' from the late 1970s (Elliston 1978) for the translation of English language documents into Scandinavian languages. Many companies followed their example, and the Smart Corporation specializes to this day in setting up 'controlled language' MT systems for large companies in North America. In a few cases, it was possible to develop systems specifically for the particular 'sublanguage' of the texts to be translated, as in the Météo system mentioned above. Indeed, nearly all systems operating in large organisations are in some way 'adapted' to the subject areas they operate in: earth moving machines, job applications, health reports, patents, police data, and many more.

Personal Computers became widely marketed since the early 1980s and software for translation became available soon afterwards: ALPS (later Alpnet) in 1983, Weidner in 1984 (later acquired by the Japanese company Bravis). They were followed from the mid-1980s onwards by many companies marketing PCs – including most of the Japanese manufacturers of PCs – and covering an increasingly wider range of language pairs and on an increasingly wide range of operating systems. Since the mid 1990s a huge range of translation software has been available (Hutchins 2003).

What has always been uncertain is how purchasers have been using these PC systems. In the case of large-scale (mainframe) 'enterprise' systems it is clear that MT is used to produce drafts which are then edited by bilingual personnel. This may also be the case for PC systems, i.e. it may be that they have been and are used to create 'drafts' which are edited to a higher quality. On the other hand, it seems more likely that users want just to get some idea of the contents (the basic 'message') of foreign texts and are not concerned about the quality of translations. This usage is generally referred to as 'assimilation' (in contrast to the use for publishable translations: 'dissemination'). We know (anecdotally) that some users of PC systems have trusted them too much and have sent 'raw' (unedited) MT translations as if they were as good as human translations.

The same comments apply to the marketing since the early 1990s of hand-held translation devices or 'pocket translators'. Many, such as the Ectaco range of special devices, are in effect computerized versions of the familiar phrase-book or pocket dictionary, and are clearly marketed primarily to the tourist and business traveller. The small dictionary sizes are obviously limited. Although sold in large numbers, there is no indication of how successful in actual use they may be. Recently, since the end of the 1990s they have been largely replaced

by online MT services. (see below).

Mainframe, client-server and PC systems are overwhelmingly ‘general purpose’ systems, i.e. they are built to deal with texts in any subject domain. Of course, ‘enterprise’ systems (particularly controlled language systems) are over time focused on particular subject areas, and adaptation to new areas is offered by most large MT systems (such as Systran). A few PC-based systems are available for texts in specific subject areas, e.g. medical texts and patents (the English/Japanese Transer systems). On the whole, however, PC systems deal with specific subjects by the provision of subject glossaries. For some systems the range of dictionaries is very wide, embracing most engineering topics, computer science, business and marketing, law, sports, cookery, music, etc.

Few translators have been happy with fully automatic translation. In particular they do not want to be post-editors of poor quality output. They prefer dedicated computer-based aids, in particular since the early 1990s the availability of ‘translation memories’. An early advocate of translation aids was Martin Kay (1980), who criticised the current approaches to MT as technology-driven rather than user-driven. He argued that the real need was assistance in translation tasks. These aids include facilities for multilingual word-processing, for creating in-house glossaries and termbanks, for receiving and sending texts over telecommunication networks, for accessing remote sources of information, for publishing quality documents, and for using interactive or batch MT systems when appropriate. Above all, translators need access to previous translations in ‘translation memories’, i.e. bilingual corpora of aligned sentences and text segments. Translators can find examples of existing translations of text which match or are similar to those in hand. Not only is consistency improved and quality maintained, but sections of repetitive texts are not translated again unnecessarily. Ideas for translation memories date back to proposals by Arthern (1978) and Kay (1980), but it was not until the early 1990s that they came onto the market with systems from Trados, SDL, Atril, Champollion, etc. Systems which integrate a variety of aids are known as translator's workstations or workbenches and have been commercially available from a number of vendors (Trados, STAR, IBM). (For a historical survey see Hutchins 1998.)

A special application of MT since the early 1990s has been the localisation of software products. (For a survey see Esselink 2003). Software producers seek to market versions of their systems in other languages, simultaneously or very closely following the launch of the version in the original language (usually English), and so localisation has become a necessity in the global markets of today. The repetitive nature of the documentation (e.g. software manuals), changing little from one product to another and from one edition to the next, made the use translation memories and the development of ‘controlled’

terminologies for MT systems particularly attractive. But, localisation involves more than just translation of texts. It means the adaptation of products (and their documentation) to particular cultural conditions, ranging from correct expression of dates (day-month-year vs. month-day-year), times (12-hour vs. 24-hour), address conventions and abbreviations, to the reformatting (re-paragraphing) and re-arranging of complete texts to suit expectations of recipients.

Corpus-based MT Research – 1989 to the Present

The dominant framework of MT research until the end of the 1980s was based on essentially linguistic rules of various kinds: rules for syntactic analysis, lexical rules, rules for lexical transfer, rules for syntactic generation, rules for morphology, etc. The rule-based approach was most obvious in the dominant transfer systems of the 1980s (Ariane, Metal, SUSY, Mu and Eurotra), but it was also the basis of all the various interlingua systems - both those which were essentially linguistics-oriented (DLT and Rosetta), and those which were knowledge-based (KANT). Rule-based methods continued into the 1990s: the CAT2 system (a by-product of Eurotra) at Saarbrücken, the Catalyst project at Carnegie-Mellon University (a domain-specific knowledge-based system) for the Caterpillar company, a project at the University of Maryland based on the linguistic theory of 'principles and parameters', and Pangloss, an ARPA-funded research at Carnegie-Mellon, Southern California, and New Mexico State University.

Since 1989, however, the dominance of the rule-based approach has been broken by the emergence of new methods and strategies which are now loosely called ‘corpus-based’ methods. The most dramatic development was the revival of a purely statistics-based approach to MT in the Candide project at IBM, first reported in 1988 (Brown *et al*. 1988, 1990), and developed to its definitive form in 1993 (Brown *et al*. 1993). Statistical methods were common in the earliest period of MT research (such as the distributional analysis of texts at the RAND Corporation), but the results had been generally disappointing. With the success of newer stochastic techniques in speech recognition, the IBM team at Yorktown Heights began to look again at their application to MT. The distinctive feature of Candide was that statistical methods were used as the sole means of analysis and generation; no linguistic rules were applied. The researchers at IBM acknowledged that their approach was in effect a return to the statistical approach suggested by Warren Weaver (1949). The system was tested on the large corpus of French and English texts contained in the reports of Canadian parliamentary debates (the Canadian Hansard). What surprised most researchers (particularly those involved in rule-based approaches) was that the results were so acceptable: almost half the phrases translated either matched exactly the translations in the corpus, or

expressed the same sense in slightly different words, or offered other equally legitimate translations.

Stages of translation in statistical machine translation (SMT) systems are: first alignment of bilingual corpora (i.e. texts in original language and texts in target language, or texts in comparable corpora which are not directly alignable), either by word or phrase; then, frequency matching of input words against words in the corpus, extraction of most probable equivalents in the target language ('decoding'); reordering of the output according to most common word sequences using a 'language model', a monolingual corpus providing word frequencies of the TL; and finally production of the output in the target language. In broad terms the process was in effect a revival of the 'direct translation' approach of some MT pioneers (see the quote from Weaver above), but refined of course by sophisticated statistical techniques.

Since this time, statistical machine translation (SMT) has become the major focus of most MT research groups, based primarily on the IBM model, but with many subsequent refinements (Ney 2005). The original emphasis on word correlations between source and target languages has been replaced by correlations between 'phrases' (i.e. sequences of words, not necessarily 'traditional' noun phrases, verb phrases or prepositional phrases), by the inclusion of morphological and syntactic information, and by the use of dictionary and thesaurus resources. Subsequent refinements have been the inclusion of structural information (usually dependency relations) in hierarchical trees similar to some earlier rule-based systems. For transfer from source to target, SMT systems incorporate string-to-string (or phrase-to-string) transfer relations based on the bilingual corpora, and the output is revised (corrected) via frequency information from monolingual corpora ('language models'). The SMT approach has been applied to an ever widening range of language pairs. The main centres for SMT research are the universities of Aachen, Edinburgh, and Southern California, and they have been recently joined by the Google Corporation. There are a number of ambitious SMT projects. Within Europe and funded by the European Union is the Euromatrix project involving many European researchers in an 'open' network under the general leadership of the Edinburgh centre. The project began in 2006 (Koehn 2007) with the aim of developing SMT systems between all the languages of the European Union. Some language pairs already exist, many in different versions, particularly between languages such as English, French, German, Spanish. A major effort of the project has been the development of SMT for 'minor' languages not previously found in MT systems, such as Estonian, Latvian, Slovenian, Macedonian, etc. The project does not exclude rule-based methods when appropriate (i.e. as hybrid systems – see below); and given the complexity of translation and

the range of types of languages is it presumed that multiple approaches will be essential. (An insightful summary of achievements in SMT systems for translation of European languages is found in Koehn *et al*. (2009) Apart from the Euromatrix project, groups active in Europe include researchers at many German and Spanish universities, researchers at the Charles University Prague, who have made fundamental contributions to the SMT of morphologically rich languages (Czech and others), and researchers in the Baltic countries.

The second major 'corpus-based' approach – benefiting likewise from improved rapid access to large databanks of text corpora – was what is known as the 'example-based' (or 'memory-based') approach (Carl and Way 2003). Although first proposed in 1981 by Makoto Nagao (1984), it was only towards the end of the 1980s that experiments began, initially in some Japanese groups and during the DLT project mentioned above. The underlying hypothesis of example-based machine translation (EBMT) is that translation by humans often involves the finding or recalling of analogous examples, i.e. how a particular expression or some similar phrase has been translated before. The EBMT approach is founded on processes of extracting and selecting equivalent phrases or word groups from a databank of parallel bilingual texts, which have been aligned either by statistical methods (similar perhaps to those used in SMT) or by more traditional rule-based methods. For calculating matches, some research groups use semantic methods, e.g. a semantic network or a hierarchy (thesaurus) of domain terms, other groups use statistical information about lexical frequencies in the target language. A major problem is the re-combination of selected target language examples (generally short phrases) in order to produce fluent and grammatical output. Nevertheless, the main advantage of the approach (in comparison with rule-based approaches) is that since the texts have been extracted from databanks of actual translations produced by professional translators there is an assurance that the results should be idiomatic. Unlike SMT, there is little agreement on what might be a 'typical' EBMT model (cf. Turcato and Popowich 2003), and most research is devoted to example-based methods which might be applicable to any MT system (rule-based or statistical.).

Although SMT is now the dominant framework for MT research, it is recognised that the two corpus-based approaches are converging in many respects: SMT systems are making more use of phrase-based alignments and of linguistic data, and EBMT systems are making wider use of statistical analysis techniques.

Increasingly, resources for MT (components, algorithms, corpora, etc.) are widely available as 'open source' materials. For SMT well known examples are: *GIZA++* for alignment, and the *Moses* basic translation engine. For rule-based MT there is the *Apertium* system from Spain which has been the basis of MT systems for Spanish, Portuguese, Galician,

Catalan, Welsh, Swedish, Danish, Slovenian, etc.

Many researchers believe that the future for MT lies in the development of hybrid systems combining the best of the statistical and rule-based approaches. In the meantime, however, until a viable framework for hybrid MT appears, experiments are being made with multi-engine systems and with adopting statistical techniques with rule-based (and example-based) systems. The multi-engine approach involves the translation of a given text by two or more different MT architectures (SMT and RBMT, for example) and the integration or combination of outputs for the selection of the 'best' output – for which statistical techniques can be used (in what are called 'combination systems'). An example of appending statistical techniques to rule-based MT is 'statistical post-editing,' i.e. the submission of the output of an RBMT system to a 'language model' of the kind found in SMT systems.

Evaluation

Evaluations of MT systems date back to the earliest years of research: Miller and Beebe-Center (1956) were the first; Henisz-Dostert evaluated the Georgetown Russian-English system (Henisz-Dostert 1967) and John Carroll (1966) did the study influenced the negative conclusions of ALPAC, – all were based on human judgments of comprehensibility, fluency, fidelity, etc. and all were evaluations of Russian-English systems. In the years from 1970 to 1990 the European Commission undertook in-depth evaluations of the Systran English-French and English-Italian systems before they were adopted (Van Slype 1979). In the 1990s there were numerous workshops dedicated specifically to the problems of evaluating MT, e.g. Falkedal 1991, Vasconcellos 1992, and the workshops attached to many MT conferences. The methodologies developed by Japan Electronic Industry Development Association (Nomura and Isahara 1992) and those designed for the evaluation of ARPA (later DARPA) supported projects were particularly influential (ARPA 1994), and MT evaluation proved to have significant implications for evaluation in other areas of computational linguistics and other applications of natural language processing. Initially, most measures of MT quality were performed by human assessments of such factors as comprehensibility, intelligibility, fluency, accuracy and appropriateness – for such evaluation methods the research group at ISSCO has been particularly important – e.g. King *et al*. (2003). However, human evaluation is expensive in time and effort and so efforts have been made, particularly since 2000, to develop automatic (or semi-automatic) methods.

One important consequence of the development of the statistics-based MT models (SMT, above) has in fact been the application of statistical analysis to the automatic evaluation of MT systems. The first metric was BLEU from the IBM group, followed later by

the NIST (National Institute for Standards and Techniques); for BLEU see Papineni *et al*. (2002); for NIST see Doddington (2002). Both have been applied by (D)ARPA in its evaluations of MT projects supported by US research funds.

BLEU and NIST (and other subsequently developed metrics such as METEOR) are based on the availability of human produced translations (called 'reference texts'). The output from an MT system is compared with one of more 'reference texts'; MT texts which are identical or very close to the 'reference' in terms of word sequences score highly, MT texts which differ greatly either in individual word occurrences or in word sequences score poorly. The metrics tend to rank rule-based systems lower than SMT systems even though the former are often more acceptable to human readers. Nevertheless, current automatic evaluation is undeniably valuable for monitoring whether a particular system (SMT or EBMT) has or has not improved over time. Many researchers are currently seeking metrics which produce results more closely matching human judgments; or indeed, metrics based directly on collaborative human evaluations from 'crowd sourcing' (e.g. using *Mechanical Turk*, as in Callison-Burch 2009).

A consequence of the change from rule-based approaches to statistics-based methods has been that MT researchers do not need any longer to have considerable knowledge of the source and target languages of their systems; they can rely upon metrics based on human produced 'reference texts' to suggest improvements; furthermore, the use of statistics-based methods means that researchers can produce systems much more quickly than with the previous laborious rule-based methods.

Speech Translation since 1990

Reports of the speech translation research in Japan appeared from 1988 onwards (e.g. the research at ATR, by Tomita *et al*. 1988). Reports of the JANUS system at Carnegie-Mellon came in 1993 (Woszczyna *et al*. 1993) and in the same year news of the Verbmobil project based in Germany (Wahlster 1993) and of the SLT project in the SRI group in Cambridge (Rayner *et al*. 1993). The NESPOLE research project came in 2001 (Lavie *et al*. 2001)

The research in speech translation is faced with numerous problems, not just variability of voice input but also the nature of spoken language. By contrast with written language, spoken language is colloquial, elliptical, context-dependent, interpersonal, and frequently in the form of dialogues. MT has focused primarily on well-formed, technical and scientific language and has tended to neglect informal modes of communication. Speech translation therefore represents a radical departure from traditional MT. Some of the problems of spoken

language translation may be reduced by restricting communication to relatively narrow domains. Business communication was the focus of the government funded research at a number of German universities (the Verbmobil project), where the aim was the development of a system for three-way negotiation between English, German and Japanese (Wahlster 2000). The focus of the ATR research in Japan has been telephone communication between English and Japanese primarily in the area of booking hotel accommodation and registration for conferences. The potentialities of speech translation in the area of health-communication are obvious. Communication may be from doctor to patient or interactive, or may be via a screen displaying possible 'health' conditions. Examples are the MedSLT project from SRI where voice input locates potential phrases and the translation is output by speech synthesis (Rayner and Bouillon 2002), and the interactive multimodal assistance provided by the Converser system (Seligman and Dillinger 2006). A somewhat similar 'phrasebook' approach is found in the DIPLOMAT system from Carnegie-Mellon (Frederking *et al*. 1997). The system was developed for the US Army for communication from English to Serbo-Croat, Haitian Creole and Korean: spoken input is matched against fixed phrases in the database and translations of the phrases are output by a speech synthesizer. Nearly all the systems were somewhat inflexible and limited in range – the weakest point continues to be speech recognition.

One of the most obvious applications of speech translation is the assistance of tourists in foreign countries. In most cases, translation is restricted to 'standard' phrases extracted from corpora of dialogues and interactions in tourist situations, although, in recent years, researchers have turned to systems capable of dealing with 'spontaneous speech'. Despite the amount of research in an apparently highly-restricted domain it is clear that commercially viable products are still some way in the future.

Usage and Applications since 1990

Since the early 1990 the use of unrevised MT output has grown greatly, such that now it may well be true that 'raw' unedited MT is the principal form in which people encounter translation from any source.

For the general public, the main source of translation since the mid-1990s has been the availability of free MT services on the Internet (Gaspari and Hutchins 2007). Initially, online MT services in the early 1990s were not free. In 1988 Systran in France offered a subscription to its translation software using the French postal services Minitel network. At about the same time, Fujitsu made its Atlas English-Japanese and Japanese-English systems available through the online service Niftyserve. Then in 1992 CompuServe launched its MT

service (based on the Intergraph DP/Translator), initially restricted to selected forums, but which proved highly popular, and in 1994 Globalink offered an online subscription service – texts were submitted on line and translations returned by email. A similar service was provided by Systran Express. However, it was the launch of AltaVista's Babelfish free MT service in 1997 (based on the various Systran MT systems) that attracted the greatest publicity. Not only was it free but results were (virtually) immediate. Within the next few years, the Babelfish service was joined by FreeTranslation (using the Intergraph system), Gist-in-Time, ProMT, PARS, and many others; in most cases, these were online versions of already existing PC-based (or mainframe) systems. The great attraction of these services was (and is) that they are free to users – it is evidently the expectation of the developers is that free online use will lead either to sales of PC translation software, although the evidence for this has not been shown, or to the use of fee-based 'valued-added' post-editing services offered by providers such as FreeTranslation. While online MT has undoubtedly raised the profile of MT for the general public, there have, of course, been drawbacks.

To most users the idea of automatic translation was something completely new – many users 'tested' the services by inputting sentences containing idiomatic phrases, ambiguous words and complex structures, and even proverbs and deliberately opaque sayings, and not surprisingly the results were unsatisfactory. A favorite method of 'evaluation' was back translation: i.e. translation into another language and then back into the original language (Somers 2007). Not surprisingly, users discovered that MT suffered from many limitations – all well-known to company users and to purchasers of PC software. Numerous commentators have enjoyed finding fault with online MT and, by implication with MT itself. On the other hand, there is no doubt that the less knowledge users have of the language of the original texts the more value they attach to the MT output; and some users must have found that online MT enabled them to read texts which they would have previously had to pass over.

Largely unknown by the general public is the use of MT systems by the intelligence services. The languages of most interest are, for obvious reasons, Arabic, Chinese, Persian (Farsi). The older demand for translation from Russian (see above) has almost disappeared. The need is for the translation of huge volumes of text. The coming of statistical machine translation has answered this need to a great extent: SMT systems are based on large corpora, often concentrating on specific topics (politics, economics, etc.), and the systems can be delivered quickly. As a sideline we may mention one intriguing application of SMT methods to the decipherment of ancient languages (Ravi and Knight 2011) – reviving the cryptographic speculations of Weaver in 1949 (section 1 above).

Collaboration in the acquisition of lexical resources dates from the beginning of MT

research (e.g. the Harvard Russian-English dictionary was used by the MT project at the National Physical Laboratory). A notable effort in the late 1980s was the Electronic Dictionary Research project in the late 1980s, supported by several Japanese computer manufacturing companies. The need grew with the coming of corpus-based systems (see above). Since the latter part of the 1990s large lexical resources have been collected and made available in the United States through the Linguistic Data Consortium and in Europe through the European Language Resources Association (ELRA), which in 1998 inaugurated its major biennial series of conferences devoted to the topic – the Language Resources and Evaluation Conferences (LREC). The Internet itself is now a source for lexical data, such as Wikipedia. One of the earliest examples of 'mining' bilingual texts from the World Wide Web was described by Resnick (1999).

The languages most often in demand and available commercially are those from and to English. The most frequently used pairs (for online MT services and apparently for PC systems) are English/Spanish and English/Japanese. These are followed by (in no particular order) English/French, English/German, English/Italian, English/Chinese, English/Korean, and French/German. Other European languages such as Catalan, Czech, Polish, Bulgarian, Romanian, Latvian, Lithuanian, Estonian, and Finnish, Bulgarian, Romanian were more rarely found in the commercial PC market or online until the last decade. Until the middle of the 1990s, Arabic/English, Arabic/French and Chinese/English were also rare, but this situation has now changed for obvious political reasons. Other Asian languages have been relatively neglected: Malay, Indonesian, Thai, Vietnamese and even major languages of India: Hindu, Urdu, Bengali, Punjabi, Tamil, etc. And African languages (except Arabic dialects) are virtually invisible. In terms of population these are not 'minor' languages – many are among the world's most spoken languages. The reason for neglect is a combination of low commercial viability and lack of language resources (whether for rule-based lexicons and grammars or for statistical MT corpora). There is often no word-processing software (indeed some languages lack scripts), no spellcheckers (sometime languages lack standard spelling conventions), no dictionaries (monolingual or bilingual), indeed a general lack of language resources (e.g. corpora of translations) and of qualified and experienced researcherso (For an overview see Somers 2003a).

Summary

Machine translation has come a long way from its tentative and speculative beginnings in the 1950s. We can see three stages of development, each spanning two decades. The first 20 years include the pioneering period (1949-1966) when numerous different approaches were investigated: dictionary-based word-for-word systems, experiments with

interlinguas, syntax-based systems with multiple levels of analysis, and the first operational systems (IBM and Georgetown). The period ended with the influential ALPAC report of 1966. The next two decades (1967-1989) saw the development of linguistic rule-based systems, mainly in the framework of transfer grammars, and experiments with sophisticated interlingua and artificial intelligence systems; in the same decade there was increased application of MT for commercial users, including the use of controlled languages and sublanguages, and applications such as localization; and there was also the first computer-based translation aids. The third period, since the early 1990s, has seen the domination of corpus-based approaches, translation memories, example–based MT, and in particular statistical MT; but there has also been much greater attention to evaluation methods; lastly, applications and usages of MT have widened markedly, most significantly by the access to and use of MT and resources over the Internet.

References

ALPAC (1966) *Languages and Machines: Computers in Translation and Linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, D.C.: National Academy of Sciences, National Research Council, 1966.

Arnold, Doug J., Lorna Balkan, R. Lee Humphreys, Seity Meijer, and Louisa Sadler (1994) *Machine Translation: An Introductory Guide*, Manchester and Oxford: NCC Blackwell.

ARPA (1994) *ARPA Workshop on Machine Translation*, 17-18 March 1994, Sheraton Premier Hotel at Tyson's Corner, Vienna, Austria.

Arthern, Peter J. (1979) 'Machine Translation and Computerized Terminology Systems: A Translator's Viewpoint', in Barbara M. Snell (ed.) *Translating and the Computer: Proceedings of a Seminar*, London, 14 November 1978, Amsterdam: North-Holland Publishing Company, 77-108.

Bar-Hillel, Yehoshua (1960) 'The Present Status of Automatic Translation of Languages', *Advances in Computers* 1: 91-163.

Booth, Andrew D. (ed.) (1967) *Machine Translation*, Amsterdam: North-Holland Publishing Company.

Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, Robert L. Mercer, and Paul S. Roossin (1988) 'A Statistical Approach to French / English Translation', in *Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages,* 12-14 June 1988, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, Pennsylvania, the United States of America.

Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin (1990) 'A Statistical Approach to Machine Translation', *Computational Linguistics* 16(2): 79-85.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer (1993) 'The Mathematics of Statistical Machine Translation: Parameter Estimation', *Computational Linguistics* 19(2): 263-311.

Callison-Burch, Chris (2009) 'Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk', in *EMNLP-2009: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 6-7 August 2009, Singapore, 286-295.

Carl, Michael and Andy Way (eds.) (2003) *Recent Advances in Example-based Machine Translation*, Dordrecht: Kluwer Academic Publishers.

Carroll, John B. (1966) 'An Experiment in Evaluating the Quality of Translations', *Mechanical Translation and Computational Linguistics* 9(3-4): 55-66.

Doddington, George (2002) 'Automatic Evaluation of Machine Translation Quality Using N-gram Co-occurrence Statistics', in *HLT 2002: Human Language Technology Conference: Proceedings of the 2nd International Conference on Human Language Technology Research*, 24-27 March 2002, San Diego, California, the United States of America, 138-145.

Edmundson, Harold P. (ed.) (1961) *Proceedings of the National Symposium on Machine Translation*, 2-5 February 1960, the University of California, Los Angeles, the United States of America, London: Prentice-Hall.

Elliston, John S.G. (1978) 'Computer-aided Translation: A Business Viewpoint', in Barbara M. Snell (ed.) *Translating and the Computer: Proceedings of a Seminar*, 14 November 1978, London, Amsterdam: North-Holland Publishing Company, 149-158.

Esselink, Bert (2003) 'Localisation and Translation', in Harold L. Somers (ed.) *Computers and Translation: A Translator's Guide*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 67-86.

Falkedal, Kirsten (ed.) (1991) *Proceedings of the Evaluators' Forum*, 21-24 April 1991, Les Rasses, Vaud, Switzerland.

Frederking, Robert E., Ralf D. Brown, and Christopher Hogan (1997) 'The DIPLOMAT Rapid-deployment Speech MT System', in *Proceedings of the MT Summit VI: Machine Translation: Past, Present, Future*, 29 October – 1 November 1997, San Diego, California, the United States of America, 261-262.

Gaspari, Federico and W. John Hutchins (2007) 'Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects', in *Proceeding of the MT Summit XI*, 10-14 September 2007, Copenhagen, Denmark, 199-206.

Goodman, Kenneth and Sergei Nirenburg (eds.) (1991) *The KBMT Project: A Case Study in Knowledge-based Machine Translation*, San Mateo, CA: Morgan Kaufmann Publishers.

Goutte, Cyril, Nicola Cancedda, Marc Dymetman, and George Foster (eds.) (2009) *Learning Machine Translation*, Cambridge, Massachusetts: MIT Press.

Henisz-Dostert, Bozena (1967) 'Experimental Machine Translation', in William M. Austin (ed.) *Papers in Linguistics in Honor of Léon Dostert*, The Hague: Mouton, 57-91.

Hutchins, W. John (1986) *Machine Translation: Past, Present, Future*, Chichester: Ellis Horwood and New York: Halsted Press.

Hutchins, W. John and Harold L. Somers (1992) *An Introduction to Machine Translation*, London: Academic Press.

Hutchins, W. John (1998) 'The Origins of the Translator's Workstation', *Machine Translation* 13(4): 287-307.

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

Hutchins, W. John (2003) 'Commercial Systems: The State of the Art', in Harold L. Somers (ed.) *Computers and Translation: A Translator's Guide*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 161-174.

Hutchins, W. John (2004) 'The Georgetown-IBM Experiment Demonstrated in January 1954', in Robert E. Frederking and Kathryn B. Taylor (eds.) *Proceedings of Machine Translation: From Real Users to Research: Proceedings of the 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004*, 28 September – 2 October 2004, Washington, D.C., the United States of America, Berlin: Springer Verlag, 102-114.

Kay, Martin (1980) 'The Proper Place of Men and Machines in Language Translation', Research Report CSL-80-11, Xerox Palo Alto Research Center, Palo Alto, CA, the United States of America.

King, Margaret (ed.) (1987) *Machine Translation Today: The State of the Art*, Edinburgh: Edinburgh University Press.

King, Margaret, Andrei Popescu-Belis, and Eduard Hovy (2003) 'FEMTI: Creating and Using a Framework for MT Evaluation', in *Proceedings of the MT Summit IX*, 23-27 September 2003, New Orleans, the United States of America, 224-231.

Koehn, Philipp (2007) 'EuroMatrix – Machine Translation for all European Languages', Invited Talk at *MT Summit XI*, 10-14 September 2007, Copenhagen, Denmark.

Koehn, Philipp (2009) *Statistical Machine Translation*, Cambridge: Cambridge University Press.

Koehn, Philipp, Alexandra Birch, and Ralf Steinberger (2009) '462 Machine Translation Systems for Europe', in *MT Summit XII: Proceedings of the 12th Machine Translation Summit*, 26-30 August 2009, Ottawa, Ontario, Canada, 65-72.

Lavie, Alon, Chad Langley, Alex Waibel, Fabio Pianesi, Gianni Lazzari, Paolo Coletti, Loredana Taddei, and Franco Balducci (2001) 'Architecture and Design Considerations in NESPOLE!: A Speech Translation System for E-commerce Applications', in *HLT-2001: Proceedings of the 1st International Conference on Human Language Technology Research*, 18-21 March 2001, San Diego, California, the United States of America.

Locke, William N. and Andrew D. Booth (eds.) (1955) *Machine Translation of Languages: Fourteen Essays*, Cambridge, Massachusetts: The Technology Press of the Massachusetts Institute of Technology and New York: John Wiley.

Miller, George A. and J.G. Beebe-Center (1956) 'Some Psychological Methods for Evaluating the Quality of Translations', *Mechanical Translation* 3(3): 73-80.

Nagao, Makoto (1984) 'A Framework of a Mechanical Translation between Japanese and English by Analogy Principle', in Alick Elithorn and Ranan Banerji (eds.) *Artificial and Human Intelligence*, Amsterdam: North-Holland Publishing Company, 173-180.

Newton, John (ed.) (1992) *Computers in Translation: A Practical Appraisal*, London and New York: Routledge.

Ney, Hermann (2005) 'One Decade of Statistical Machine Translation', in *Proceedings of the MT Summit IX: The 10th Machine Translation Summit,* 12-16 September 2005, Phuket, Thailand, i-12-i-17.

Nirenburg, Sergei, Harold L. Somers, and Yorick Wilks (eds.) (2003) *Readings in Machine Translation*, Cambridge, Massachusetts: MIT Press.

Nomura, Hirosato and Hitoshi Isahara (1992) 'The JEIDA Methodology and Survey', in Muriel Vasconcellos (ed.) *MT Evaluation: Basis for Future Directions: Proceedings of a Workshop Sponsored by the National Science Foundation*, 2-3 November 1992, San Diego, California, the United States of America, 11-12.

Papineni, Kishore A., Salim Roukos, Todd Ward, and Wei-Jing Zhu (2002) 'BLEU: A Method for Automatic Evaluation of Machine Translation', in *ACL-2002*: *40th Annual Meeting of the Association for Computational Linguistics*, 6-12 July 2002, Philadelphia, the United States of America, 311-318.

Ravi, Sujith and Kevin Knight (2011) 'Deciphering Foreign Language', in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 19-24 June 2011, Portland, Oregon, the United States of America, 12-21.

Rayner, Manny, Hiyan Alshawi, Ivan Bretan, David Carter, Vassilios Digalakis, Björn Gambäck, Jaan Kaja, Jussi Karlgren, Bertil Lyberg, Steve Pulman, Patti Price, and Christer Samuelsson (1993) 'A Speech to Speech Translation System Built from Standard Components', in *HLT '93: Proceedings of the Workshop on Human Language Technology*, 21-24 March 1993, Plainsboro, New Jersey, the United States of America, 217-222.

Rayner, Manny and Pierrette Bouillon (2002) 'Flexible Speech to Speech Phrasebook Translator', in *Proceedings of the ACL-2002 Workshop on Speech-to-speech Translation*, 11 July 2002, Philadelphia, the United States of America, 69-76.

Resnick, Philip (1999) 'Mining the Web for Bilingual Text', in *ACL-1999: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 20-26 June 1999, University of Maryland, College Park, Maryland, the United States of America, 527-534.

Seligman, Mark and Mike Dillinger (2006) 'Usability Issues in an Interactive Speech-to-speech Translation System for Healthcare', in *HLT-NAACL 2006: Proceedings of the Workshop*

on Medical Speech Translation, 9 June 2006, New York, the United States of America,1-8.

Slocum, Jonathan (ed.) (1988) *Machine Translation Systems*, Cambridge: Cambridge University Press.

Somers, Harold L. (2003a) 'Translation Technologies and Minority Languages', in Harold L. Somers (ed.) *Computers and Translation: A Translator's Guide*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 87-103.

Somers, Harold L. (ed.) (2003b) *Computers and Translation: A Translator's Guide*, Amsterdam and Philadelphia: John Benjamins Publishing Company.

Somers, Harold L. (2007) 'Machine Translation and the World Wide Web', in Ahmad Kurshid, Chistopher Brewster, and Mark Stevenson (eds.) *Words and Intelligence II: Essays in Honor of Yorick Wilks*, Dordrecht: Springer Verlag, 209-233.

Tomita, Masaru, Marion Kee, Hiroaki Saito, Teruko Mitamura, and Hideto Tomabechi (1988) 'Towards a Speech-to-speech Translation System', *Interface: Journal of Applied Linguistics* 3(1): 57-77.

Turcato, Davide and Fred Popowich (2003) 'What Is Example-based Machine Translation?' in Michael Carl and Andy Way (eds.) *Recent Advances in Example-based Machine Translation*, Dordrecht: Kluwer Academic Publishers, 59-81.

van Slype, Georges (1979) *Critical Study of Methods for Evaluating the Quality of Machine Translation*, Final Report BR19142, Brussels: Bureau Marcel van Dijk [for] European Commission.

Vasconcellos, Muriel (ed.) (1992) *MT Evaluation: Basis for Future Directions: Proceedings of a Workshop Sponsored by the National Science Foundation*, 2-3 November 1992, San Diego, California, the United States of America.

Wahlster, Wolfgang (1993) 'Verbmobil: Translation of Face-to-face Dialogs', in *Proceedings of the MT Summit IV: International Cooperation for Global Communication*, 20-22 July 1993, Kobe, Japan, 127-135.

Wahlster, Wolfgang (ed.) (2000) *Verbmobil: Foundations of Speech-to-speech Translation*, Berlin: Springer Verlag.

Weaver, Warren (1949) 'Translation'. Reprinted in William N. Locke and Andrew D. Booth (eds.) *Machine Translation of Languages: Fourteen Essays*, Cambridge, Massachusetts: Technology Press of the Massachusetts Institute of Technology, 15-33.

Woszczyna, Monika, Noah Coccaro, Andreas Eisele, Alon Lavie, Arthur E. McNair, Thomas Polzin, Ivica Rogina, Carolyn P. Rose, Tilo Sloboda, Masaru Tomita, Junya Tsutsumi, Naomi Aoki-Waibel, Alex H. Waibel, and Wayne Ward (1993) 'Recent Advances in JANUS: A Speech Translation System', in *Proceedings of the 5th International*

Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: *MT in the Next Generation* (*TMI-93*), 14-16 July 1993, Kyoto, Japan, 195-200.

Yngve, Victor H. (1964) 'Implications of Mechanical Translation Research', in *Proceedings of the American Philosophical Society*, 108(4): 275-281.