# CHAPTER 2: The precursors and the pioneers

## 2. 1: Before the computer

The use of mechanical devices to overcome language barriers was suggested first in the seventeenth century. There were two stimulants: the demise of Latin as a universal language for scientific communication, and the supposed inadequacy of natural languages to express thought succinctly and unambiguously. The idea of universal languages arose from a desire both to improve international communication and to create a 'rational' or 'logical' means of scientific communication.

Suggestions for numerical codes to mediate among languages were common. Leibniz's proposals in the context of his monadic theory are perhaps the best known. Another proposal was made by Descartes in comments on the sixteenth proposition of his famous correspondent Anonymous. In a letter to Pierre Mersenne on 20 November 1629 (Mounin 1964:16) Descartes described a proposed universal language in the form of a cipher where the lexical equivalents of all known languages would be given the same code number. Descartes wrote: "Mettant en son dictionnaire un seul chiffre qui se rapporte à *aymer, amare, philein*, et tous les synonymes (of *aimer* in all languages) le livre qui sera écrit avec ces caractères (i.e. the code numbers) pourra être interprété par tous ceux qui auront ce dictionnaire". Actual examples of such mechanical dictionaries were published by Cave Beck in 1657, by Athanasius Kircher in 1663, and by Johann Joachim Becher in 1661. At the height of enthusiasm about machine translation in the early 1960's some writers saw these 17th proposals as genuine forerunners of machine translation. Becher's book, for example, was republished under the title *Zur mechanischen Sprachübersetzung: ein Programmierungversuch aus dem Jahre 1661* (Becher 1962), indicating the conviction of its editor that Becher's ideas foreshadowed certain principles of machine translation. Apart from an ingenious script, Becher's book is distinguished from others of this kind only by the size of the dictionary: 10,000 Latin words (stems and endings) were provided with codings. Like others, however, Becher failed to tackle the real difficulties of providing equivalent entries in other languages (Greek, Hebrew, German, French, Slav, and Arabic were proposed) and the necessary means to cope with syntactic differences.

The vast work by John Wilkins, *An Essay towards a Real Character and a Philosophical Language* (1668), was a more genuine attempt at a universal language in that it sought to provide a logical or rational basis for the establishment of inter-language equivalencies. Wilkins' aim was "a regular enumeration and description of all those things and notions, to which marks or names ought to be assigned according to their respective natures", i.e. a codification which embodied a universal classification of concepts and entities, a genuine interlingua.

All these writers recognised the problems of genuine differences between languages that could not be captured completely in dictionaries, however 'logically' constructed. Many of them like Kircher advised their fellows to write in a simple style and avoid rhetorical flourishes.

Suggestions for mechanical dictionaries on numerical bases continued to be made throughout the following centuries until the middle of the present century. Couturat and Leau in their Histoire *de la langue universelle* (1903) list numerous examples, including one by W.Rieger entitled *Zifferngrammatik, welche mit Hilfe der Wörterbücher ein mechanisches Uebersetzen aus einer Sprache in alle anderen ermöglicht* (Code-grammar, which with the help of dictionaries enables the mechanical translation from one language into all others); a title which links the present mechanical age to the 17th century.

As the reference to Couturat and Leau implies, all these apparent precursors of MT should be regarded more accurately as contributions to the ideal of a 'universal language' and to the development of international auxiliary languages (Large 1985), of which the best known is now Esperanto. Both concepts have in fact inspired many of those engaged in machine translation.

None of these proposals involved the construction of machines; all required the human translator to use the tools provided in a 'mechanical' fashion, i.e. for man to simulate a machine. It was not until the invention of mechanical calculators in the nineteenth and twentieth centuries (starting with the pioneer activities of Charles Babbage) that an automatic device could be envisaged which could perform some translating processes. In fact, the first explicit proposals for 'translating machines' did not appear until 1933, when two patents were issued independently in France and Russia. In both cases, the patents were for mechanical dictionaries.

A French engineer of Armenian extraction, Georges Artsrouni was issued a patent on 22nd July 1933 for a translation machine which he called a "Mechanical Brain" (Corbé 1960).[1] The invention consisted of a mechanical device worked by electric motor for recording and retrieving information on a broad band of paper which passed behind a keyboard. The storage device was capable of several thousand characters, and was envisaged by its inventor in use for railway timetables, bank accounts, commercial records of all sorts, and in particular as a mechanical dictionary. Each line of the broad tape would contain the entry word (SL word) and equivalents in several other languages (TL equivalents); corresponding to each entry were coded perforations on a second band, either paper or metal, which functioned as the selector mechanism. The required entry was punched at the keyboard, the perforations located and the full entry retrieved within, it was claimed, 10 to 15 seconds. A prototype machine was exhibited and demonstrated in 1937; the French railway administration and the post and telegraph services showed considerable interest, and only the start of the Second World War in 1940 prevented installation of Artsrouni's invention.

More important in retrospect was the patent issued in Moscow on 5 September 1933 to Petr Petrovich Smirnov-Troyanskii for the construction of a "machine for the selection and printing of words while translating from one language into another or into several others simultaneously." (quoted from Panov 1960). A brief account by Troyanskii himself written in February 1947 was published in 1959 (Troyanskii 1959). Troyanskii[2] envisaged three stages in the translation process; the machine was involved only in the second stage, performing as an automated dictionary. In the first stage a human editor knowing only the source language was to analyze the input text into a particular 'logical' form: all inflected words were to be replaced by their base forms (e.g. the nominative form of a noun, the infinitive form of a verb) and ascribed their syntactic functions in the sentence. For this process Troyanskii had devised his own 'logical analysis symbols'. In the second stage the machine was designed to transform sequences of base forms and 'logical symbols' of source texts into sequences of base forms and symbols of target languages. In the third stage an editor knowing only the target language was to convert this sequence into the normal forms of his own language. Troyanskii envisaged both bilingual translation and multilingual translation. Although the machine was assigned the task only of automating the dictionary, it is interesting to note that Troyanskii believed that "the process of logical analysis could itself be mechanized, by means of a machine specially constructed for the purpose" (quoted by Panov 1960a). It was this vision of the next steps beyond a simple mechanical dictionary that marks Troyanskii's proposal as a genuine precursor of machine translation.

In the 1933 patent, the technical implementation proposed was a purely mechanical device, a table over which passed a tape listing in vertical columns equivalent words from various languages (not dissimilar to Artsrouni's machine). But, by 1939 he had added an improved 'memory' device operating with photo-elements (Delavenay 1960; Mounin 1964), and by May 1941 it appears that an experimental machine was operational. Troyanskii in fact went further towards the electronic computer; in 1948 he had a project for an electro-mechanical machine

[1] For more on Artsrouni see my unpublished article 'Two precursors of machine translation: Artsrouni and Trojanskij'

[2] For a fuller account of Troyanskii and his proposal see: J. Hutchins and E. Lovtskii 'Petr Petrovich Troyanskii (1894-1950): a forgotten pioneer of machine translation', *Machine Translation* 15 (3), 187-221; and the article cited in footnote 1.

similar to the Harvard Mark I machine developed between 1938 and 1942, and which is regarded as a forerunner of the ENIAC computer (Augarten 1984).

Troyanskii was clearly ahead of his time; Soviet scientists and linguists failed to respond to his proposal when he sought their support in 1939, and later "the Institute of Automation and Telemechanics of the Academy of Sciences was equally unforthcoming in 1944" (Delavenay 1960) In retrospect, there seems to be no doubt that Troyanskii would have been the father of machine translation if the electronic digital calculator had been available and the necessary computer facilities had been ready (Mounin 1964; Panov 1960). History, however, has reserved for Troyanskii the fate of being an unrecognised precursor; his proposal was neglected in Russia and his ideas had no direct influence on later developments; it is only in hindsight that his vision has been recognised.

## 2.2: The first beginnings (1946-1949)[3]

The electronic digital computer was a creation of the Second World War: the ENIAC machine at the Moore School of Electrical Engineering in the University of Pennsylvania was built to calculate ballistic firing tables; the Colossus machine at Bletchley Park in England was built to decipher German military communications. Immediately after the war, projects to develop the new calculating machines were established at numerous centres in the United States and Great Britain (Augarten 1984). The first applications were naturally in the fields of mathematics and physics, but soon the enormous wider potential of the "electronic brain" were realized and nonnumeric applications began to be contemplated.

The first suggestion that electronic computers could be used to translate from one language into another seems to have been made during conversations in New York between Andrew D. Booth and Warren Weaver.

Warren Weaver was at this time vice president of the Rockefeller Foundation. During the war Weaver had served on a scientific mission to investigate Britain's weapons development, and at the Rockefeller Foundation he was closely involved in the sponsorship of computer research and development. Booth had become interested in automatic digital calculation while working at the British Rubber Producers' Association in Welwyn Garden City, and had started to build a machine for crystallographic calculations. In 1945 he was appointed a Nuffield Fellow in the Physics Department at Birkbeck College in the University of London under Professor J.D.Bernal, where he constructed a relay calculator during 1945 and 1946 and initiated plans for computational facilities in the University of London. As a consequence of this work and the efforts of Bernal he obtained funds to visit the United States in 1946 under the auspices of the Rockefeller Foundation. There he visited all the laboratories engaged in computer research and development, at Princeton, MIT, Harvard, and Pennsylvania (Booth 1980).

While he was in the United States he met Weaver for the first time on 20 June 1946 at the Rockefeller Foundation in New York (Weaver 1946). According to Booth (1985): "At that time neither they nor I had any interest in machine translation. The discussions that I had with Warren Weaver were entirely on the subject of coming over to look into the question of acquiring the techniques for building a machine for the University of London based on American experience."

At the end of his US visit in July Booth submitted a report on computer development with particular reference to x-ray crystallography, and he was offered a Rockefeller fellowship to enable him to work at an institution of his own choice in the United States the following year. Booth selected the von Neumann group at the Institute for Advanced Study, Princeton University,

[3] For a fuller account of the early history of MT see: J. Hutchins 'From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology', *Machine Translation* 12 (3), 195-252. A briefer version appered as: 'First steps in mechanical translation', *MT Summit VI: past, present, future, 29 October – 1 November, 1997, San Diego, California, USA: Proceedings*, ed. V.Teller and B.Sundheim (Washington, DC: AMTA, 1997), 14-23

recognised then and now as doing the most advanced research on computers at the time. On his return to the United Kingdom, Booth stopped work on the x-ray calculator and, together with Kathleen Britten, devoted himself to the problem of von Neumann type computers, and in particular to problems of large-scale storage (Booth 1980).

Booth met Weaver again at the Rockefeller Centre in New York on 6th March 1947. According to Booth (1985): "The discussion then was entirely on the question of the Rockefeller Foundation financing a computer for the University of London, and Weaver pointed out that there was very little hope that the Americans would fund a British computer to do number crunching, although they might be interested if we had any additional ideas for using the machine in a non-numerical context, and he suggested I thought about things of that type. I had already thought about non-numerical applications from conversations with A.M.Turing in the mid 1940's... One of these was in fact translation, although at that time I had thought only of using the machine as a dictionary." Weaver suggested treating translation as a cryptography problem.

Weaver had in fact already on 4th March 1947, just before this meeting with Booth, written to Norbert Wiener of the Massachusetts Institute of Technology, one of the pioneers in the mathematical theory of communication, about the possibility of MT. In his letter (Weaver 1947, quoted in Weaver 1949), after commenting on the problem of translation in the post-war world Weaver wrote:

> Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while... Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography... 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."... As a linguist and expert on computers, do you think it is worth thinking about?

Wiener's response on the 30th April disappointed Weaver: "I frankly am afraid the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi mechanical translation scheme very hopeful." (Wiener 1947). Weaver wrote again on 9th May but failed to interest Wiener in the possibilities. He also tried to interest Ivor Richards, the literary critic and linguist, who collaborated with C.K.Ogden on the development of Basic English (Richards 1943, Ogden 1968), and who at this time was professor at Harvard University (Marshall 1947); but with equal lack of success it would appear.

At the Institute for Advanced Study, between March and September 1947, Booth worked with his assistant Kathleen Britten on many tentative ideas for the new London computer, including the recognition of sounds and of printed characters and also the details of a code, which would enable information from a dictionary stored in a computer memory to be retrieved from standard teletype input (Booth et al. 1958:1).

This work was reported in a memorandum sent to the Rockefeller Foundation early the next year, on 12th February 1948. In it, Booth mentioned as "a possible application of the electronic computer... that of translating from one language into another", adding that "We have considered this problem in some detail..." (Booth 1948). Booth admitted later (Booth et al.1958:2) that the program written in Princeton would probably have been of little practical use, nevertheless it was obviously of sufficient encouragement for him to continue to develop his ideas on dictionary lookup procedures.

It is evident that the first serious discussions and investigations of the possibilities of machine translation took place during 1947, beginning with Weaver's letter to Wiener and his meeting with Booth in early March. However, at a later date in 1955 when writing the 'Historical introduction' to the MT collection he edited with Locke (Booth & Locke 1955), Booth recollected the first mention of MT as having occurred during his 1946 visit. This has been generally accepted as the ‘birth’ date of MT; however, in other later publications Booth gives the date 1947 (e.g. Booth 1956, 1958), and he has now confirmed the March 1947 meeting as the one when the MT discussion with Weaver occurred (Booth 1985). On the other hand, in a letter to Bernal on 20 May 1948, written in acknowledgement of Booth’s February report, Weaver himself stated: “I think I may have been the first to suggest this possible use. At any rate I did discuss it two or three years ago, and talked with both you and Booth about it, I believe.” (Weaver 1948)

Alan Turing’s influence in the development of computers and in their possibilities as ‘intelligent machines’ is well known. He may legitimately be regarded as the pioneer of what is now known as Artificial Intelligence. In an essay written during September 1947, he mentions a number of possible ways in which the new computers could demonstrate their ‘intelligence’: “(i) Various games, e.g. chess, noughts and crosses, bridge, poker; (ii) The learning of languages; (iii) Translation of languages; (iv) Cryptography; (v) Mathematics.” (Turing 1947). Evidently, Weaver and Turing were thinking along similar lines independently; and probably, others too.

As there were no facilities available at Birkbeck College, Booth began construction of a small computer at the laboratories of the British Rubber Producers’ Research Association in Welwyn Garden City near London. The machine was operational by 12th May 1948 and a demonstration was given on 25th May to Warren Weaver and Gerard Pomerat, also of the Rockefeller Foundation (Booth 1980, Weaver 1949). On this occasion Weaver met Richard H.Richens, with whom Booth had been collaborating in experiments on mechanical dictionaries.

Richens had first met Booth on the 11th November 1947. His interest in mechanical translation had arisen independently out of experiments with punched cards for storing information at the Commonwealth Bureau of Plant Breeding and Genetics, where he was Assistant Director. “The idea of using punched cards for automatic translation arose as a spin-off, fuelled by my realisation as editor of an abstract journal (*Plant Breeding Abstracts*) that linguists conversant with the grammar of a foreign language and ignorant of the subject matter provided much worse translations than scientists conversant with the subject matter but hazy about the grammar.” (Richens 1984). Richens is to be credited with the first suggestion of the automatic grammatical analysis of word-endings (Delavenay 1960:28). He proposed the segmenting words into their stems (or base forms) and endings (e.g. inflections), both to reduce the size of dictionaries and to introduce grammatical information into a dictionary translation system. For example, in the case of the Latin verb *amat* a search is made for the longest matching stem, i.e. ‘am-’, and for the ending ‘-at’. The stem provides the English translation *love* and the ending gives the grammatical information ‘3rd person singular’. In this way grammatical annotations augment a strict word-by-word dictionary ‘translation’. The validity of the approach was tested by hand and by using punched card machinery on a wide variety of languages; the texts were taken from abstracts in plant genetics. The result of this collaboration with Booth was a memorandum written during 1948.

From a later version of this memorandum, presented at the MIT conference in 1952 (see below), some idea of these early efforts in ‘dictionary translation’ can be gleaned (Richens & Booth 1955).

From the French text:

> Il n’est pas étonn*ant de constat*er que les hormone*s de croissance
> ag*issent sur certain*es espèce*s, alors qu'elles sont in*opér*antes sur
> d’autre*s, si l’on song*e à la grand*e spécificité de ces substance*s.

(Where the stars (*) indicate automatic segmentations.)

The English translation:

*v* not is not/step astonish *v* of establish *v* that/which? *v* hormone m of growth act *m* on certain *m* species *m*, then that/which? *v* not operate *m* on of other *m* if *v* one dream/consider *z* to *v* great *v* specificity of those substance *m*.

(Where v indicates a French word not translated, *m* "multiple, plural or dual", *z* "unspecific", and slashes alternative translations.)

These tentative experiments by Booth and Richens were known to very few. Brief mention was made during the International Conference on Science Abstracting held in Paris, June 20-25, when R. B. Reid of Unesco told some delegates about what he heard of their work from Professor Bernal (Reid 1949). Shortly afterwards, a short account was given by Holmstrom in a report on scientific and technical dictionaries which was submitted to Unesco circulated in mimeograph form in July 1949 (Holmstrom 1951). However, it was undoubtedly the memorandum which Warren Weaver wrote on 15th July 1949 that had most widespread and profound influence (Weaver 1949). The memorandum was distributed to some 200 of Weaver's acquaintances who, it was thought, might have potential interest in the concept of machine translation (Weaver 1949). For probably all recipients this memorandum was literally the first suggestion they had seen that the new electronic computers could be used as translating machines. In effect this memorandum launched machine translation as a scientific enterprise in the United States and subsequently elsewhere. Its historic impact is unquestionable; and it was later published in the Locke & Booth collection (1955) of early contributions to MT.

In his memorandum Weaver dates the origin of his speculations about MT to his wartime experience with electronic computers and to stories of startling achievements in cryptanalysis using computers. Weaver remarks that "it was very natural ...to think, several years ago, of the possibility that such computers be used for translation." He then reproduces the correspondence with Wiener, and refers to Booth's mention of MT in his 1948 memorandum as being the "first instance known to W.W. subsequent to his own notion about it", and outlines the experiments by Richens.

Weaver knew of one other MT experiment at the time of his memorandum. Newspapers had recently reported the use of one of the California computers in a primitive kind of word-for-word translation. These reports had prompted a letter in the *Herald Tribune* of 13th June 1949 from a Max Zeldner demonstrating how ridiculous would be word-for-word translations of literary texts. Mounin (1964: 20) speculated that these reports referred probably to the very tentative experiments mentioned by Olaf Helmer of the RAND Corporation at the 1952 MIT conference (see below).

## 2.3: Weaver's memorandum (1949)[4]

Weaver's memorandum concentrated more on the general strategies and long-term objectives of MT than on the more technical problems Booth and Richens had been tackling. Because of its historic importance it is worth enumerating in some detail the issues and problems raised by Weaver. He raised four points: the problem of multiple meaning, the logical basis of language, the application of communication theory and cryptographic techniques, and the possibilities of language universals.

The problem of deciding which specific meaning an ambiguous word may have in a particular text was, he suggested, solvable in principle if a sufficient amount of the immediate context is taken into account. The practical question of how many contexts are necessary could be answered by a statistical study of different types of texts on a variety of subject matters. Weaver

---

[4] See also: J. Hutchins 'Warren Weaver memorandum: 50th anniversary of machine translation', *MT News International* 22 (July 1999), 5-6, 15; and 'Warren Weaver and the launching of MT: brief biographical note', *Early years in machine translation: memoirs and biographies of pioneers*, ed. W.J.Hutchins (Amsterdam: John Benjamins, 2000), 17-20.

explicitly rejected the idea of actually storing in dictionaries long sequences of words for this purpose, but did suggest that "some reasonable way could be found of using the micro context to settle the difficult cases of ambiguity."

He expressed optimism about finding logical aspects in languages. In connection with a remark by Hans Reichenbach that he "was amazed to discover that, for (apparently) widely varying languages, the basic logical structures have important common features", Weaver commented that Reichenbach's observation seems to be confirmed in the linguistics literature and he mentioned work by Erwin Reifler (below) on the comparative semantics of Chinese and English. Secondly, Weaver expressed the conviction that "insofar as written language is an expression of logical character" then the theorem of McCulloch and Pitts (1943) on the mathematical possibility of computing logical proofs appeared to offer a demonstration of the logical possibility of implementing the translation process on a computer, i.e. that the problem of translating with a computer is formally solvable.

On the applicability of communication theory Weaver pointed to the important work of Shannon (which he was instrumental in popularizing, e.g. Shannon & Weaver 1949) and its relation to cryptography. The success of computers in cryptanalysis had prompted Weaver's speculations about MT, as the letter to Wiener in 1947 illustrates. In the memorandum he recounted an anecdote about the deciphering of a short 100-word coded text in Turkish. The cryptographer did not know Turkish, nor did he know that the message was in Turkish. (Before distributing his memorandum, Weaver checked the veracity of the story with his informant, Professor W.Prager of Brown University: the decipherment had been done by R.E.Gilman of the Mathematics Department (Prager 1948).) For Weaver this achievement was sufficient demonstration of the power of cryptanalytic techniques independent of the languages involved. Allied to the probabilistic foundation of communication theory, Weaver believed that the translation problem could be largely solved by "statistical semantic studies".

For Weaver the "most promising approach of all" was the investigation of language invariants or universals. He linked this again with logical structures of language and with probabilistic uniformities. The analogy he suggested was "of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower... communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication... Thus may it be true that the way to translate ... is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language..."

## 2.4: From Weaver to the first MT conference (1950-1952)

Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time. But, it did more. It indicated potentially fruitful lines of research in statistical analyses of language, on the logical bases of language, and on semantic universals of language. In addition, it pointed to some actual, even if tentative, achievements in the work of Booth and Richens. It was, however, received with considerable scepticism by many linguists who rejected it for its naivety in linguistic matters and for its unfounded assumptions on the logicality of language, and they were naturally sceptical about the possibility of formalising language and translation processes.

The memorandum had also been noticed by the press. An account appeared in *Scientific American* in December 1949, which reported on the experiment by Booth and Richens. This in turn was picked up by the British newspaper the *News Chronicle* in the spring of 1950, and so appeared the first of what in coming years were to be frequent misunderstandings and exaggerations. Booth's

APEXC computer program was described as an “electronic translator”, at which an operator “could select which of a dozen or more languages he desired to translate. As fast as he could type the words, say, in French, the equivalent in Hungarian or Russian would issue on the tape.”

**2.4.1: First MT studies**.

Weaver’s own favoured approach, the application of cryptanalytic techniques, was immediately recognised as mistaken (see Mounin (1964: 31-39) for a detailed discussion). Confusion between the activities of deciphering and translation arise whenever the same person does both. Obviously, no translating is involved when an English message is deciphered into English by an English-speaking recipient. Likewise, the decipherment of the highly complex Enigma code used by Germany in the Second World War, with its immensely complex sequences of manipulations and transpositions, was not translation; it was only after the German texts had been deciphered that they were translated. The Colossus computers at Bletchley Park were applied to cracking the cipher, not to translating the German text into English. In practice, the cryptanalyst generally knows what the language is of the texts to be deciphered and often what their content is likely to be and the circumstances in which the message was transmitted. All this helps him to guess which letters and words are likely to be most frequent in the text. In the case cited by Weaver, the decipherment was based on “the frequencies of the letters, pairs of letters, etc. in English” (Prager 1948); fortunately they were much the same in Turkish and the original could be interpreted.

Though the cryptanalytic approach was mistaken, there were sufficient stimulating ideas in Weaver’s paper to launch MT as a serious line of research in the United States. During the next two years, individuals and groups began MT studies at a number of locations, the Massachusetts Institute of Technology (MIT), the University of Washington in Seattle, the University of California at Los Angeles (UCLA), the National Bureau of Standards (NBS) also in Los Angeles and the RAND Corporation nearby at Santa Monica.

On 10th January 1950, Erwin Reifler circulated privately the first of a series of studies on MT (Reifler 1950). Reifler was a Sinologist of German origin, head of the Department of Far Eastern and Slavic Languages and Literature at the University of Washington in Seattle. Recognising the problem of multiple meanings as an obstacle to word-for-word translation of the kind attempted by Booth and Richens, Reifler introduced the concepts of ‘pre-editor’ and ‘post-editor’. The human ‘pre-editor’ would prepare the text for input to the computer and the ‘post-editor’ would resolve residual problems and tidy up the style of the translation. One suggestion was that the pre-editor should indicate the grammatical category of each word in the source language (SL) text by adding symbols or diacritic marks, e.g. to distinguish between the noun *cónvict* and the verb *convíct*. The post-editor’s role was to select the correct translation from the possibilities found by the computer dictionary and to rearrange the word order to suit the target language. As we shall see, the concepts of pre-editor and post-editor recur in one form or another throughout the development of MT research.

Following Weaver’s suggestion for statistical studies of microcontext for resolving problems of multiple meaning, Abraham Kaplan at the RAND Corporation investigated polysemy in mathematics texts. (The study was completed on 30th November 1950.) A group of test subjects were presented with a set of words, each with a number of possible meanings, and asked to select the most applicable sense. Kaplan limited the test to nouns, verbs and adjectives on the assumption that “these are the major carriers of the content of any discourse, and probably more markedly exhibit ambiguities”. Each word was presented first in isolation, then together with preceding and following words (up to two before and after), and finally the whole sentence. It was found that the “most practical context is ... one word on each side, increased to two if one of the context words is a particle”, i.e. an article, preposition or conjunction (Kaplan 1955). Despite its limitations and deficiencies (Kaplan excluded homographs: words of different grammatical categories having the same form) and the tentativeness of the conclusions, this study encouraged hopes that problems of ambiguity were tractable and that statistical analyses could contribute useful linguistic data for MT.

In the latter half of 1950, a survey was conducted by W.F.Loomis on behalf of Weaver to find out all those who were interested in MT and what research was underway. The survey revealed a surprising amount of activity already (Loomis 1951): apart from Booth, Richens and Reifler, two groups had been set up in California. One was at the RAND Corporation in Santa Monica under J.D.Williams, and Kaplan's paper was to be the first of a series of MT studies. The other had been formed by Harry D.Huskey of the National Bureau of Standards in Los Angeles, with the intention of using the SWAC (Standards Western Automatic Computer) for MT research. The group included Victor A. Oswald of the German Department at UCLA and William E.Bull of the UCLA Spanish Department, and was soon joined by Kenneth E.Harper of the UCLA Slavic Languages Department. In support of its work, the group received some funds from the Rockefeller Foundation in July 1951.

From this group appeared during 1951 the first published journal article devoted to MT (Oswald and Fletcher 1951). Up to this time papers on MT had been mimeographed memoranda and reports of limited circulation. (It is true to say that to the present day information about MT research is largely contained in report literature of limited distribution; journal and book publication has on the whole been considered secondary.) The article by Victor Oswald and Stuart L. Fletcher, of UCLA and NBS respectively, was also the first attempt to investigate syntactic aspects of MT. Arguing that word-for-word translation of a language like German would produce obviously unsatisfactory results, Oswald and Fletcher proposed a detailed grammatical coding of German sentences indicating syntactic functions of nouns and verb forms in clauses and enabling the identification of ‘noun blocks’ and ‘verb blocks’. On the basis of the codes, certain sequences were identifiable as candidates for rearrangement when the output was to be in English. The procedures were framed as ‘instructions’ for the SWAC computer at NBS, but were not in fact implemented in any way. In an appendix, the authors illustrated how a mathematics text by Cantor would be coded, reordered and then translated word by word into English, e.g. the original:

> Bevor wir diese Definition im Einzelnen zergliedern, wollen wir einige Beispiele von Mengen betrachten, die uns anschauliches Material zum Verständnis der Definition liefern sollen.

reordered:

> Bevor wir zergliedern diese Definition im Einzelnen, wir wollen betrachten einige Beispiele von Mengen, die sollen liefern uns anschauliches Material zum Verständnis der Definition.

English:

> Before we analyze this definition in detail we want-to regard some examples of sets, which shall furnish us perceptible material for-the understanding of-the definition.

Oswald and Fletcher concluded that syntax “does not constitute, as had been thought by some, a barrier to mechanical translations”; they stressed the problems of solving the “lexicographic difficulties” of MT.

### 2.4.2: Bar-Hillel’s survey (1951)

At the Massachusetts Institute of Technology a meeting on MT was held soon after the appearance of Weaver's memorandum. According to Locke and Yngve (1958: 511): “In January 1950 Dr. Weaver met at M.I.T. with a dozen men from nearly as many different fields, including the heads of our Research Laboratory of Electronics, of our Digital Computer Laboratory, and of the Department of Modern Languages... The conclusion was cautious: the possibility of translation by machine was worth examining”. In the next year, 1951, Yehoshua Bar-Hillel became the first person to be appointed full-time specifically for research on MT; the appointment in the Research

Laboratory of Electronics was made possible by a grant form the National Science Foundation (quite possibly with the influence of Weaver who was a director of NSF at the time.) Bar-Hillel's task was to study the possibilities of MT and to plan future research at MIT.

At the end of 1951 he produced a survey of the current position (Bar-Hillel 1951). The paper[5] raised many of the issues which dominated discussion of MT in the following years: the feasibility of fully automatic MT, the role of post-editing, the objectives of syntactic analysis, the role of statistical information, the possibility of universal grammar, the logical foundations of language, and restricted vocabularies.

Bar-Hillel argued that since problems of semantic ambiguities could not be resolved at present, "high-accuracy, fully-automatic MT is not achievable in the foreseeable future". This fact, however, should not discourage MT research: a less ambitious target is feasible, "a mixed MT... in which a human brain intervenes... either at the beginning of the translation process or the end, perhaps at both, but preferably not somewhere in the midst of it." Bar-Hillel saw no alternative to the post-editor whose task would be the "elimination of semantical ambiguities, in addition, of course, to stylistic smoothing."

As for the machine processes, Bar-Hillel saw these as comprising the morphological analysis of each word into its stem and grammatical category, the "mechanical identification of small syntactical units" and the "transformation of the given sentence into another that is logically equivalent to it." The second stage required what Bar-Hillel called an "operational syntax", an explicit programmable method for syntactic analysis, the rudiments of which Bar-Hillel recognised in Oswald and Fletcher's paper.

While recognising some value in statistical analyses, he was sceptical of proposals to include only high frequency vocabulary in dictionaries since the residue are likely to be those "highly loaded with information" in a given text. In any case, he felt that the current limitations of computer storage would be solved by hardware developments.

At the end of the paper he considered the possibilities of constructing a "universal, or at least general grammar, perhaps even ... a whole artificial exchange-language" on the basis of work by logicians such as Ajdukiewicz, Carnap, Reichenbach and himself, and by the linguist Zellig S.Harris (cf.ch.3.4-5 below), or, less ambitiously, "transfer-grammars... in which the grammar of one language is stated in categories appropriate to some other language". Finally, he proposed the application of MT to situations where restricted vocabularies are used, e.g. by pilots and meteorologists, and where "the theoretical difficulties of such a type of MT are clearly less formidable". To these he added regularized languages such as Basic English and auxiliary languages such as Esperanto and Interlingua.

As we shall see, not only had Bar-Hillel raised many of the major MT issues recurring in the following years and to some extent to the present day, but he also stated views which he was to repeat at greater length and with great impact in the early 1960's (ch.8.3)

### 2.4.3: First MT conference (1952)[6]

By 1952 interest in MT had grown enough for the Rockefeller Foundation to sponsor the first conference on MT. It was held at the Massachusetts Institute of Technology from 17th to 20th June 1952, and was organised by Bar-Hillel. Eighteen individuals interested in MT attended. As might be expected, MIT was well represented, by seven participants: Yehoshua Bar-Hillel; Jay W.Forrester, head of the Computer Laboratory; William N.Locke, Dept. of Modern Languages; James W.Perry, Center of International Studies; Vernon Tate, Director of Libraries; Jerome

---

[5] See also: J. Hutchins 'Bar-Hillel's survey, 1951', *Language Today* 8 (May 1998), 22-23.

[6] For more on the conference see: J. Hutchins 'Looking back to 1952: the first MT conference', *TMI-97: proceedings of the 7th Interantional Conference on Theoretical and Methodological Issues in Machine Translation, July 23-25, 1997, St.John's College, Santa Fe, New Mexico, USA* (Las Cruces: Computing Research Laboratory, New Mexico State University), 19-30.

B.Wiesner, director of the Research Laboratory of Electronics; and Dudley A.Buck, Electrical Engineering Dept. The Los Angeles area was represented by four participants: Victor Oswald and William E.Bull from UCLA; Olaf Helmer from RAND; and Harry D.Huskey from NBS. Erwin Reifler came from the University of Washington, as did Stuart C.Dodd, the director of the Washington Public Opinion Laboratory. The remainder were: Andrew D.Booth from Birkbeck College (the sole British participant), Leon Dostert from the Institute of Languages and Linguistics of Georgetown University, Duncan Harkin from the Department of Defense, Victor H.Yngve from the University of Chicago, and A.Craig Reynolds from the Endicott Laboratories of IBM. The backgrounds of participants reflect already some of the principal sources of interest in MT: electronic engineering and computing, linguistics, librarianship and information science, military and governmental bodies. Although the proceedings were not published at the time, accounts of the conference appeared subsequently by Reynolds (1954) and Reifler (1954), both compiled immediately after the conference; and, somewhat later, a number of the papers were printed in the collection by Locke and Booth (1955).

The conference opened with a public session. Bar-Hillel enunciated the need and possibilities for MT, particularly to cover the immense and growing volume of scientific research and popular periodical literature of a country. Leon Dostert spoke on his experience in setting up the simultaneous translation systems at the Nuremberg trials, at the United Nations and other international conferences. He was sceptical of the contribution of MT except for the processing of material currently not touched in specialized fields. Olaf Helmer mentioned the tentative experiments at RAND, and Perry spoke of the relationship of MT to automatic indexing and retrieval systems.

The presentations on the following days included Reifler and Bar-Hillel on pre-editing and post-editing respectively, and Booth on his work with Richens on mechanizing a dictionary. For Reifler (1954) the low intelligibility of the output from the latter supported his argument for pre- and post-editing. Oswald presented his proposals for treating German syntax, and then with Bull suggested that problems of ambiguity could be overcome by constructing micro-glossaries for particular subject fields; the glossaries could be established from statistical analyses of the relevant literature, and an investigation of the vocabulary of brain surgery had already been undertaken (later published in Oswald and Lawson, 1953). Bull warned that no scientific vocabulary constituted a closed domain and the lexical ambiguities in general vocabulary would still remain. Dodd outlined his ideas of regularizing English syntax and morphology for use in MT, e.g. “I will send he to she”, “he have”, “she did be loved” (Dodd 1955); and Reifler suggested encouraging writers to write with MT in mind (i.e. to write in MTese, as later researchers put it). Dostert put forward the advantages of a “pivot language” in the context of multilingual MT. Reifler (1954) thought this would most likely be English and he also thought that MT output should be regularized on the lines suggested by Dodd; studies of language universals could contribute to this regularization in the MT context.

According to Reynolds (1954), Bar-Hillel's operational syntax was “a completely new concept to the linguists of the conference who had intuitively felt that such a structure did exist but without the tools of symbolic logic had been unable to isolate the essential features that lead to the exceedingly simple arithmetic operations.” There was naturally some discussion on the problems of writing computer programs for MT, on computer costing and on the use of punched card machinery. The conference concluded with statements from participants about what research they hoped to do. The most challenging proposal came from Dostert, who had been converted from his original scepticism in the course of the conference. He suggested “the early creation of a pilot machine or of pilot machines proving to the world not only the possibility, but also the practicality of MT” (Reifler 1954).

## 2. 5: From the MIT conference to the Georgetown-IBM demonstration (1952-1954)

The conference was an undoubted success, the participants were enthusiastic about the prospects and the general public had been made aware of the possibilities. Although no formal conclusions were drawn, it seems there was general agreement on what the next stages of MT research should be (Booth & Locke 1955: 6-7): “word frequency and word translation studies” on micro-glossaries and investigation of suitable storage, input and output methods for “an operating automatic dictionary at the end of approximately 12 months”; immediate studies on “operational analysis of syntax”; later longer-term work would come on multilingual MT, universal grammar and interlinguas. In retrospect, the expectations were far too optimistic, but they provided a stimulating framework for further MT research.

In the following year, articles presenting MT to the general public began to appear, e.g. Booth (1953) and Bar-Hillel (1953) in Computers *and Automation*, and MT appeared in a textbook for the first time, in a chapter on ‘Some applications of computing machines’ in the book by Booth and his wife, the former Kathleen Britten (Booth & Booth 1953).

During 1953 Bar-Hillel left MIT to return to the Hebrew University of Jerusalem; he was replaced by Victor Yngve, who had participated at the 1952 conference. At this time, Yngve shared the view that since “word-for-word translations are surprisingly good, it seems reasonable to accept a word-for-word translation as a first approximation and then see what can be done to improve it” (Yngve 1955). A partial translation of a German mathematics text was simulated in which function words and endings were left untranslated:

> Die CONVINCINGe CRITIQUE des CLASSICALen IDEA-OF-PROBABILITY IS eine der REMARKABLEen WORKS des AUTHORs. Er HAS BOTHen LAWe der GREATen NUMBERen ein DOUBLEes TO SHOWen: (1) wie sie IN seinem SYSTEM TO INTERPRETen ARE, (2) THAT sie THROUGH THISe INTERPRETATION NOT den CHARACTER von NOT-TRIVIALen DEMONSTRABLE PROPOSITIONen LOSEen...

Although readers with some familiarity with German could work out the general drift it was clear to Yngve that syntactic analysis was essential for better MT output; and this was to be the central emphasis of the MIT group which he formed in the coming years  (ch.4.7 below)

One of the most significant outcomes of the 1952 conference was the establishment of a MT research team at Georgetown University by Leon Dostert to start work towards the pilot experiment he had advocated to demonstrate the practical feasibility of machine translation. Dostert was fully aware of the considerable linguistic problems of MT but had concluded that “rather than attempt to resolve theoretically a rather vast segment of the problem, it would be more fruitful to make an actual experiment, limited in scope but significant in terms of broader implications” and that the test should serve as the basis for a series of progressively larger and more complex tests. In brief, Dostert advocated what was later to be generally known as the ‘empirical’ approach to MT research. Collaboration was agreed with IBM, under the aegis of the company's programme of endowed research in computation; Paul Garvin was appointed at Georgetown to work on the linguistic procedures for the trial translation of Russian texts; and Peter Sheridan of the IBM Scientific Computing Service was given the task of implementing the procedures on the IBM 701 machine.

The program was ready by the end of 1953 and on the 7th January 1954 a public demonstration of the Georgetown-I.B.M. experiment took place at IBM's Technical Computing Bureau in New York.[7] This small-scale experiment in Russian-English translation was one of the

[7] For more details of the demonstration see: J. Hutchins ‘The Georgetown-IBM demonstration, 7th January 1954’, *MT News International* 8, 15-18.

most important events in the early history of MT. It was the first real demonstration of MT on a computer; earlier experiments had either been hand simulations or had used punched card machinery. No pre-editing of text was required and the output appeared to be fully intelligible. Furthermore, it was the first implementation which went beyond word-for-word translation. Nevertheless, its limitations were clearly recognised: a vocabulary of just 250 Russian words, only six rules of grammar and a carefully selected sample of Russian sentences (ch. 4.3 below). For the general public and the scientific community at large, however, the experiment showed that MT was a feasible objective, and it undoubtedly helped to stimulate the funding of MT research by U.S. governmental agencies in the following decade.

The demonstration received wide publicity; Sheridan and Dostert lectured widely on the system and the potential future of MT. Further demonstrations were given during the year, for example on the 14th September at a New York meeting of the American Chemical Society (*MT* 1(3) Dec 1954). Another demonstration was given to D.Y. Panov from the USSR Academy of Sciences, who was visiting computer installations in the United States. On his return, research on MT began in the Soviet Union.

## 2. 6: From the Georgetown-IBM experiment to the first international conference (1954-1956)

During 1954 two further MT groups were founded, an informal group at Cambridge in Great Britain by Margaret Masterman, and a research team at Harvard University by Anthony Oettinger who also in the same year Oettinger presented the first doctoral dissertation on MT (ch.4.9 below). In March appeared the first issue of the journal *Mechanical Translation* issued from M.I.T. under the editorship of Victor Yngve and supported by a grant from the Rockefeller Foundation. This journal was to carry in subsequent years many of the most important articles on MT.

During 1955 the first news of Russian activity became known. As a result of his visit to the demonstration at IBM, Professor Panov began MT experiments in January 1955 on the BESM computer at the Institute of Precision Mechanics and Computer Technology in Moscow. Within the next two years other MT groups were formed in the USSR at the Steklov Mathematical Institute, at the Institute of Linguistics and at the Laboratory of Electrical Modelling, all in Moscow, and at the University of Leningrad (ch.6 below).

In August 1955 the Cambridge Language Research Group held a meeting at King’s College, Cambridge. Participants included R.H.Richens, the mathematicians A.F.Parker-Rhodes and E.W.Bastin and the linguists J.R.Firth, M.A.K.Halliday and R.A.Crossland. Already in the abstracts of the proceedings (CLRU 1956) there is evidence of the distinctive emphasis of CLRU on interlinguas, pidgin languages, logical and semantic foundations, and lattice theory (ch.5.2 below). At about this time also, Silvio Ceccato started a MT project in Milan, adopting a distinctive ‘philosophical’ approach to MT grammar and semantics (ch.5.3 below)

Publications on MT were now beginning to appear in larger numbers, shortly voluminous. A major publication of 1955 was the collection of articles edited by Locke and Booth (1955), the first book to be published devoted entirely to MT research. In it appear many of the most significant papers of the period up to 1955, including Weaver's memorandum, the report by Richens and Booth on their early work, and contributions to the 1952 conference. It includes also a valuable ‘Historical introduction’ and early papers by Oettinger (1955) on automatic dictionaries (predating his dissertation), by Yngve (1955) on syntactic analysis, by Booth (1955) on available computer hardware, and by Bar-Hillel (1955) on the problems of idiomatic usage.

It was clear that MT was growing fast. In October 1956 the first international conference was organised by MIT attended by 30 workers from the United Sates, Great Britain and Canada, and papers were received from the Soviet Union. The presentations by MT groups of their current research activity revealed the general pattern of approaches which was to dominate the next five years or so. On the one hand there was the emphasis on dictionary, lexicographic and semantic problems, e.g. at Washington University, at Los Angeles (UCLA and RAND), at Michigan, at Harvard and at Cambridge; and on the other hand, there was the greater emphasis on syntactic problems at Georgetown and MIT. Another major division was between the 'empirical' approach, typified by RAND and Georgetown, and the 'theoretical' approach, typified by MIT and Cambridge. A third division was between those groups aiming to produce operational systems in the near future, however limited (e.g. Birkbeck, Washington and to some extent Georgetown), and those with the longer term objective of producing a good quality product (e.g. Harvard and MIT).

From 1956 research on MT was pursued with great vigour all over the world by numerous groups. In the United States MT research received increasingly large grants from governmental, military and private sources. By 1964, however, the promise of operational MT systems still seemed distant and the sponsors set up a committee, which recommended in 1966 that funding of MT should be reduced. It brought to an end a decade of intensive MT research activity. The work of the various groups during these years will be described in chapters 4 to 7, after a preliminary chapter outlining some of the basic MT methods and strategies.

# Chapter 8: Expectations and criticisms: the decade from 1956 to 1966

The decade between the Georgetown-IBM demonstration in 1954 and the publication of the ALPAC report in 1966 may be characterized as one of initial widespread optimism of 'imminent' success followed by nearly equally widespread disillusionment at the slowness of progress. The turning point came roughly in 1960, marked by Bar-Hillel's influential survey (1960) and by changes of emphasis in the research strategies of many MT groups. In very broad terms, research in the earlier period from 1954 to 1960 concentrated primarily on semantic and lexicographic problems and research in the later period from 1960 to 1965 tended to concentrate more on syntactic problems and parsing strategies.

## 8. 1: Period of optimism

In the first five years after the 1954 Georgetown-IBM demonstration (ch.2.5), MT research was pursued with great enthusiasm and optimism. Numerous groups were formed in many countries. In the United States, the five early centres of Washington, Georgetown, MIT, Harvard and Los Angeles (RAND), were joined during the next five years by groups at Michigan, IBM, Ramo-Wooldridge, NBS, Texas, Berkeley (University of California), Wayne State, and numerous shorter lived projects. In the Soviet Union large scale projects were begun at three Moscow centres and in Leningrad. Major projects were set up in a number of centres in Japan, and research was beginning in Czechoslovakia and France.

The mood of optimism is well captured in the book by Emile Delavenay (1960) completed in December 1958. This introduction of MT to the general public provided a brief history of the progress of MT, and a survey of the different approaches, the problems and the methods. From it we now have an invaluable picture of the way MT research was perceived at the time. "The translation machine... is now on our doorstep. In order to set it to work, it remains to complete the exploration of linguistic data..." Delavenay was aware of considerable problems; the complexities of syntax were only just becoming apparent. For Delavenay MT research in semantics and syntax was directed towards the refinement of basically word for word transpositions. Crude though they were, the achievements so far were believed to provide the foundations for future progress. "While a great deal remains to be done, it can be stated without hesitation that the essential has already been accomplished." Delavenay could even contemplate the translation of literature; it was just a matter of compiling the appropriate dictionaries to convey the local colour of the original. He went further. "Will the machine translate poetry? To this there is only one possible reply - why not?" Anything seemed possible with the awesome power of the new electronic machines. For Delavenay it was just a matter of time. "Translating machines will soon take their place beside gramophone records and colour reproductions in the first rank of modern techniques for the spread of culture and of science."

It is easy to smile indulgently at the naive optimism of Delavenay in 1958. He was not alone by any means. Bel'skaya in the Soviet Union, as we have seen (ch.6.1), shared his belief in the possibility of translating literary works. While most MT researchers did not agree on this point, they all believed in the ultimate success of MT, even though they differed, sometimes vehemently, about the best methods to be pursued.

## 8. 2: Variety of approaches and methods.

There were first the differences between those groups who sought to produce a working system as soon as possible and those who held that fundamental research must be done before operational systems could be contemplated. Principal among the former were the Washington and IBM groups under Reifler and King. Also belonging to the pragmatists were Dostert and his Georgetown team, Booth at Birkbeck College, the Ramo-Wooldridge group, the ITMVT group in Moscow, and the ETL group in Tokyo. Those taking longer perspectives included the groups at MIT, Harvard, RAND, Cambridge, Milan, and in the Soviet groups at MIAN in Moscow and at the University of Leningrad.

Within these broad divisions there were further differences. While the Washington and IBM groups developed essentially the word-for-word approach with refinements introduced through lexicographic information, the Georgetown and Ramo-Wooldridge approaches concentrated more on problems of syntax and structural manipulation. The IBM group and the Japanese ETL group were committed to the development of special-purpose machines, and Booth at Birkbeck inclined also towards hardware solutions.

Among the groups engaged in fundamental research there were differences between the ‘empiricists’ and the ‘theorists’. Most committed to the ‘empirical’ approach was the RAND group, which distrusted traditional grammars and dictionaries and believed that MT must be based on actual usage.

The RAND group was also the strongest advocate of the ‘cyclic’ approach to MT system development, i.e. devising rules and programs for a particular text, testing them on another text, making amendments, testing on another text, and so forth, with the hope that eventually no further modifications will be necessary. Progressive modification was, of course, assumed to be necessary in the development of any operational system; what was particular to the ‘cyclic’ method was the explicit concentration on the corpus alone and the exclusion of any information from other sources, e.g. the analyst’s own knowledge of the language and its structure. The cyclic method was adopted by the groups at Birkbeck and Ramo-Wooldridge, and in an extreme form by Brown at Georgetown (ch.4.3)

The Harvard group was also committed to fundamental empirical research, and, like the RAND group, believed in the need for careful preparatory work in order to achieve eventually MT of a high quality. As at RAND, the Harvard group concentrated in this period on the compilation of a large Russian dictionary.

The most theory-oriented groups were those at MIT, Cambridge, Milan, and Leningrad. The group at MIT stressed the need for fundamental linguistic studies, particularly in the area of syntax and was greatly influenced by the formal linguistics of Chomsky and his followers. The Cambridge and Leningrad groups were more interested in semantic approaches, both concerned with the construction of interlinguas, and at Cambridge investigating the thesaurus concept for semantic organization. At Milan the emphasis was on a conceptual (i.e. non-linguistic) analysis of lexical and structural relationships.

Cutting across these various divisions were differences of MT system design. Most of the groups adopted the ‘direct translation’ strategy (ch.3.9), particularly those aiming for practical operating systems in the near future, e.g. Washington, IBM, Georgetown, Ramo-Wooldridge, Birkbeck. The most popular strategy among those with long term perspectives was the ‘interlingual’ strategy, e.g. Cambridge, Milan, and Leningrad. During this period emerged also the first versions of the ‘transfer’ strategy in the ideas of Yngve at MIT on ‘syntactic transfer’.

To a large extent, however, MT research in this five year period was often dominated by problems of computer hardware. Many of the groups, particularly in the Soviet Union but also, for example, the Cambridge group, had no access to computer facilities and much of their programming was simulated on punched card equipment. Even for those groups which did have computer equipment there were perennial problems of insufficient storage. Internal core storage was often very small, and external storage had to be on punched card or magnetic tape. Dictionary searching was therefore a major problem: for most the only real option at the time was sequential access to dictionaries on tape, and therefore preliminary alphabetisation of text words. The slowness of serial access prompted the development of the random-access photoscopic disk storage device by IBM (ch.4.2). Booth advocated the binary partition technique, but he had few followers.

It was the hope of most MT researchers that problems of text input would be solved in the near future by optical character readers. It was recognized that without much faster means of converting texts into machine-readable form the prospects of operational MT systems being economical were greatly reduced.

## 8. 3: Doubts and criticisms, Bar-Hillel's report

While optimism remained the prevailing mood, there were signs of some loss of confidence. Many groups had begun in the expectation of relatively quick success. There were not only problems of technical facilities and hardware, but also the complexities of the linguistic problems. These were becoming more and more apparent. The mood of optimism was now by no means universal within the MT community. Critics of MT were growing and becoming more vociferous year by year. There had always been those who were highly sceptical of any attempts to 'mechanise' language. Norbert Wiener (ch.2.2) was only the first of many who doubted that semantic problems could ever be resolved. For a long time those involved in MT research could ignore such objections, they could always claim that they were only at the very start of the enterprise, and that in any case most objectors did not understand and probably did not want to understand what MT research was really about. However, when criticism came from within, from one intimately familiar with MT research and its practitioners then it could not be ignored so easily.

In February 1959, Yehoshua Bar-Hillel published his *Report on the state of machine translation in the United States and Great Britain*, prepared on the basis of personal visits to most MT groups in the US during October 1958 and on information received during the following months in response to a circular letter (Bar-Hillel 1959).

One year later the report[1] received wider distribution when it appeared in a revised form in the annual review journal *Advances in Computers* (Bar-Hillel 1960). The main addition in the later version was a survey of Soviet research based principally on a report by Rozentsveig (1958), the book by Panov (1960), and unpublished accounts by Oettinger and John W.Carr III of their visits to Soviet MT groups during 1958. (Probably Bar-Hillel also saw the survey by Oettinger (1958) before publication.) As far as the US projects were concerned, the later version did incorporate some updated information, but in essence it described the situation as Bar-Hillel saw it in late 1958. In particular, it should be noted that the wording of his forceful criticisms was not revised in any way.

The basic argument was that MT research was, with few exceptions, mistakenly pursuing an unattainable goal: fully automatic translation of a quality equal to that of a good human translator. Instead, it should be less ambitious and work towards the realisation of systems involving human collaboration.

In view of the later impact of this review it is well to keep in mind the stage at which MT research had reached in late 1958. Hardly any of the projects had been engaged on full scale research for more than two years. As we have seen (ch.4), the large Georgetown team was formed only in June 1956, the teams at Harvard and Cambridge received their first National Science Foundation grants in 1956, the RAND group was set up in March 1957, and research at NBS, IBM, Berkeley, Wayne State and Texas did not start until 1958. Funded research had begun a little earlier, in 1955, by the relatively small-scale teams at Birkbeck College and Michigan University, but the only really long established large projects were those at MIT and Washington, and even there it may be noted that large scale funding started only in 1954 and 1956 respectively. The situation was much the same in the Soviet Union: only two of the Russian projects had been active for more than two years, and then only since 1955; others did not begin until 1956 or later (ch.6).

It is significant that Bar-Hillel did not cite any actual examples of translations produced by any of the projects, nor even allude to any particular linguistic or computational problems of their systems. Instead he concentrated his remarks explicitly on general questions of methodology. The thrust of his argument was that current methods could not conceivably produce fully automatic systems providing reasonable quality translations, either in the short or long term. The argument was based on highlighting the methodological shortcomings of individual projects and an abstract 'demonstration' of the impossibility of what he called 'fully automatic high quality translation' (FAHQT).

Bar-Hillel had become convinced that FAHQT was unattainable "not only in the near future but altogether". It was in fact a view he had expressed in his 1951 review (ch.2.4.2), before most MT projects had even been thought of. Now in 1959 he felt able to give a 'proof' in 'A demonstration of the non-feasibility of fully automatic, high quality translation' (Appendix III in Bar-Hillel 1960). His argument was based on the short sentence:

The box was in the pen

in the context:

[1] For more on Bar-Hillel and this report see: J.Hutchins 'Yehoshua Bar-Hillel: a philosopher's contribution to machine translation', *Early years in machine translation: memoirs and biographies of pioneers*, ed. W.J.Hutchins (Amsterdam: John Benjamins, 2000), 299-312.

> Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.

On the assumption that *pen* can have two meanings, a 'writing utensil' and an 'enclosure where small children can play', Bar-Hillel claimed that "no existing or imaginable program will enable an electronic computer to determine that the word pen in the given sentence within the given context has the second of the above meanings." This amounted to the very strong claim that in certain (not infrequent) cases no amount of context will suffice to resolve this type of homonymy. The reason is that we as intelligent human readers know "the relative sizes of pens, in the sense of writing implements, toy boxes, and pens, in the sense of playpens" and that this knowledge is "not at the disposal of the electronic computer". To put such information in a MT system would mean that "a translation machine should not only be supplied with a dictionary but also with a universal encyclopedia". For Bar-Hillel such a requirement was "utterly chimerical and hardly deserves any further discussion". Not only is human knowledge a vast store of facts, but it is also the infinite set of inferences which may be drawn from facts.

Bar-Hillel conceded that ambiguity can be resolved by the use of microglossaries and contextual clues. However, he thought that use of microglossaries increased the risk of mistranslation (meaningful but erroneous in a particular instance), and that contextual analysis can have only limited effectiveness. Resolution of some but not all ambiguities would not be good enough if the aim is 'high quality' translation.

Much of the point of his argument has now been somewhat blunted by achievements in computational linguistics and in AI semantic analysis (ch.15 below), but at the time Bar-Hillel's case against the FAHQT goal convinced many not involved in MT research that MT as such was doomed to failure, and it has continued to represent a challenge and point of departure for arguments about MT to the present day.

Bar-Hillel attributed the adherence of some MT groups to the FAHQT aim as a residue of the early initial successes in MT. In the first few years of MT there had been "a considerable amount of progress" in solving a large number of linguistic and computational problems and producing crude translations which expert readers could make some sense of. This progress had convinced many that "a working system (was) just around the corner". It had been realized only gradually that although "the problems solved until then were indeed many" they were "just the simplest ones" and "the 'few' remaining problems were the harder ones – very hard indeed." However, he did not condemn basic theoretical research as such, even if FAHQT was the distant aim, because it might be justified by "interesting theoretical insights", whether of benefit to practical MT or not.

As for operational MT he contended that researchers had either to sacrifice quality (low quality products were acceptable in many circumstances) or to acknowledge the necessity for post-editing. He advocated the latter aim, "high quality translation by a machine-post-editor partnership", as the most fruitful area of future MT development. The goal then should be partially automatic MT, commercially competitive with human translation, which could be gradually improved and refined with more and more of the post-editing operations carried out automatically. This goal required, however, the development of more reliable and flexible optical character readers, more attention to dictionary compilation, research on the efficiencies of different dictionary formats (full forms vs. stems and endings), and investigation of the need for pre-editing of input.

Not only did Bar-Hillel hold strong convictions about the aims of MT but he was also highly critical of two particular approaches: the ‘empirical’ approach, and the ‘interlingual’ approach. Adherents of the former distrusted traditional grammars and dictionaries and held that MT grammars and dictionaries must be built from scratch (often on the basis of statistical analyses of large text corpora) Bar-Hillel condemned it as both “wasteful in practice and not sufficiently justified in theory”. Faith in statistics derived from earlier overestimations of the power of statistical theory of communication (Shannon-Weaver information theory), and there was no reason to reject normative grammars so completely, as they “are already based... upon actual texts of incomparably larger extension than those that serve as a basis for the new compilers”.

As for the ‘interlingual’ approach, while he admitted that achievements in mathematical logic might reinforce the hope that the 17th century idea need not fail in the 20th century, he dismissed the idea for two reasons: one was the fallacy of the economic argument for multilingual systems, the other was the fallacy of what he saw as the basic assumption of adherents of the approach, namely that translation into a ‘logical’ interlingua was simpler than translation into a natural language. (The arguments for and against interlinguas will be taken up more fully below.)

With such a collection of strongly held views, it is not surprising that very few of the current MT systems escaped criticism, often harsh. In one way or another nearly all the US groups were found wanting. The only one to escape was that of Rhodes at NBS, praised for its efficient parser and its “practical aims” involving “no attempt... to obtain a FAHQT output”. Specific criticisms were often harsh. For example, the work at MIT on the programming language COMIT was considered unnecessary, and the MIT group was censured for “reluctance to publish incomplete results”. In similar vein, Harvard was accused of “strong distrust of the achievements of other groups”. The Cambridge (CLRU) group’s applications of lattice-theory were dismissed as “only reformulations... of things that were said and done many times before”, and its conception of the thesaurus was too obscure for Bar-Hillel. (“Since I could not persuade myself that I really understood the Cambridge group's conception (or conceptions?) of the thesaurus (or thesaurus-lattice) approach to MT, I shall say nothing about it”, adding acidly: “Perhaps the reader will be luckier.”)

In general, Bar-Hillel’s opinions were not based on a careful evaluation of the actual achievements of MT projects but they were already formed before the review was undertaken. He can be criticised for bias, prejudices, inaccuracies, and antagonisms, but the basic thrust of his main argument had some validity: “Fully automatic, high quality translation is not a reasonable goal, not even for scientific texts”. The only reasonable goals were “either fully automatic, low quality translation or partly automatic, high quality translation”. Both were considered to be technically (although perhaps not commercially) feasible at the time. Bar-Hillel did not think “great linguistic sophistication” to be either “requisite” or “helpful” for developing practical MT systems; basic linguistic research should continue with the long term improvement of MT in aim. However, there was a considerable overlap in linguistic research, resulting in “costly repetitions of achievements as well as, and even more often so, of failures”. It was his opinion that much of this wastefulness was attributable to MT researchers’ too long-lived adherence to FAHQT goals.

As one of the best known pioneers of MT, Bar-Hillel had written a report which was bound to influence public opinion, and it did. There is no doubt that it contributed to the disillusionment which steadily grew in the following years, and that it was held up as 'proof' of the impossibility of MT. To this day, Bar-Hillel's article is still cited as an indictment of MT research (not only for this early period, but in general). There can be few other areas of research activity in which one publication has had such an impact.

## 8. 4: The state of the art in 1960, and new directions.

Bar-Hillel's report had most effect on the external perception of MT; in most respects, it did not greatly affect the internal development of MT. Those groups which were not concerned with high quality MT (however defined) continued the development of systems to produce output needing a greater or lesser degree of post-editing; they would agree with Bar-Hillel that FAHQT was not achievable and so they would not attempt to achieve it. This was the essence of the Georgetown view (and of course also of the IBM group). On the other hand, those groups which did have FAHQT aims were spurred on to disprove Bar-Hillel's contention that it was impossible. They would probably agree with the Cambridge group that Bar-Hillel's criteria for high quality was too absolute and that his demonstration was not a 'proof'. (For CLRU the 'pen' problem could be resolved by thesaural methods, cf.ch.5.2) Bar-Hillel's arguments had, therefore, the unfortunate effect of polarizing MT research between those who saw the need for more basic research in order to achieve successful MT and those who were concerned with the solution of practical MT problems.

Since the 1956 MIT conference (ch.2.6) there had been a number of important conferences at which contributions on MT had been made. For example, the international conference on 'Information Processing' organised by Unesco in June 1959 had included an important session on MT (Unesco 1960). Reports were made by Bel'skaya from the ITMVT in Moscow, Japanese speakers from the Electrotechnical Laboratory in Tokyo, as well as speakers from Cambridge, Harvard, MIT, and RAND. It was at this conference also that A.F.R. Brown of Georgetown gave the first public demonstration of an 'operational' system. Whether intended or not, the impression given was that Georgetown group was about to launch a commercial MT system. Some of those engaged in MT were highly critical; it was obvious that the quality of output was not good enough, and that promises of improvements could not possibly be fulfilled.

Although the Russians were making important contributions and there were also the groups in Britain, elsewhere in Europe and in Japan, there is no doubt that at this time the main impetus for MT research was in the United States. The National Symposium on Machine Translation held in February 1960 at the University of California at Los Angeles (Edmundson 1961) brought together all the major (and most minor) active US groups: RAND, Ramo-Wooldridge, NBS, IBM, Georgetown, Texas, MIT, Berkeley, Washington, Wayne State, Harvard (cf. relevant sections in ch.4.). Some had only recently been formed (Texas, Berkeley, Wayne State) and were outlining future plans, but most had already a number of years' experience. This meeting was the first occasion at which a number of the new approaches and methods were publicly aired which were to characterize the next five years of MT research.

Principal among these were the descriptions of new methods of syntactic analysis: Hays (1961) gave the first public description of dependency grammar (ch.3.4); Garvin

(1961) described the development of his ‘fulcrum’ theory, which he had begun at Georgetown and was to continue at Ramo-Wooldridge (ch.4.6); and Rhodes (1961) outlined the ‘predictive analyzer’ under development at NBS, and which had previously been described only in an internal report (Rhodes 1959).

Up to this time, syntactic analysis had been seen primarily as providing data for the local manipulation of lexical items where SL and TL word orders did not match; for this all that was required was the determination of grammatical categories and sequences; the identification of groupings (in phrases and clauses) was not always considered necessary, since lexical information was sufficient to determine relationships of this kind. The main exponents of this attitude were the Washington group (and IBM) and, in a less extreme form, the Georgetown group. In general, syntax was considered less important than the problems of dictionary compilation, microglossaries and lexical ambiguity. The only group to tackle syntax at an early stage was at MIT. For many researchers, however, the MIT concentration on ‘abstract’ formal syntax was too remote from the practicalities of MT; indeed, Bar-Hillel himself in his report (1960) had censured MIT commitment to Chomskyan grammar as “premature” (despite his own formalistic inclinations, as witness his categorial grammar, ch.3.4) From 1960, however, syntax became the dominant theme of much MT research. Symptomatic of the change was the switch at Harvard from dictionary work to the development of the NBS ‘predictive analyzer’ (announced at this 1960 meeting by Oettinger & Sherry (1961)) and the attention devoted to syntax by groups which had previously tended to ignore it (e.g. the Cambridge group, ch.5.2). With the formulation of alternative methods of syntactic analysis there was confidence that this problem area, at least, could be conquered, and from this time on, the formalisation of syntactic theories and development of efficient parsers advanced rapidly.

Further evidence of future changes came in the presentation by Sydney Lamb of Berkeley of the first version of his ‘stratificational’ conception of language structure (ch.3.10 and 4.10). It marked the beginning of multi-level approaches, i.e. the separation of morphological, syntactic and semantic stages of SL analysis and TL synthesis, which though perhaps implicit in Yngve’s ‘syntactic transfer’ approach (ch.4.7) had not before been formulated clearly. Lamb’s model was itself to influence other MT groups (notably CETA at Grenoble, ch.10.1). For the present, however, the dominant strategy was to remain the ‘direct’ approach.

The new methods of syntactic analysis, the linguistic modelling of Lamb and the increasing formalism were characteristic, of course, primarily of the more theoretical inclinations of MT research. There was clear evidence at the conference of antagonism between those groups aiming for ‘production systems’ and those concentrating on basic research. Oettinger was particularly insistent that the Georgetown group should make its intentions plain: “In the last issue of the Georgetown Newsletter a statement was made that an automatic translation system would be operating as a production system within a year. I would like to know whether that is the system that was described today, in which the mean number of errors is 13.4%...” Dostert (of Georgetown) replied that all they intended was to have within a year a lexicon adequate for translating Russian texts in the narrow field of organic chemistry; he saw nothing wrong in developing systems which “produce inelegant but reasonably reliable text”. But for Oettinger (and many others), the quality was just not good enough – little better than word-for-word translation – and it was considered intolerable that the public should be offered such error-ridden systems.

But how long would it be before MT research produced good quality translations? There were no clear ideas on how quality could be evaluated, but many knew there was much more linguistic work to be done – not only in syntax, but even more in semantics (which, as King (IBM) commented, "we haven't even begun to talk or think about seriously".) Oettinger was certainly not sanguine (prophetically, he commented "All of us who are dedicated to research have to face the prospects that our efforts may have been in vain"); but even he was surprised that Hays and Yngve should estimate "10 years' worth of problems" ahead. Although in retrospect even this was an underestimate, it was clear evidence that some MT researchers were in danger of becoming 'perfectionists' (Bar-Hillel thought some were already); and at the conference, Edmundson said "it is apparent to all of us that some of the MT investigators are not really concerned with production translation."

## 8. 5: Official support for US research

In the same year, MT research in the US received what was in effect an official seal of approval. It came in a report which was compiled by the Committee on Science and Astronautics of the U.S. House of Representatives in June 1960 (U.S. House of Representatives, 1960) The committee received reports and testimony from the funding agencies (Central Intelligence Agency, US Air Force, US Army, and US Navy) on the importance of MT to "the overall intelligence and scientific effort of our Nation" and for translations from English "for the exchange of cultural, economic, agricultural, technical, and scientific documents that will present the American way of life to people throughout the world." A survey was included to indicate US and foreign activity in MT (particularly the strong Soviet interest), and details were provided of the possible systems and of the current research methodologies. While recognizing the dangers of duplicating effort, the committee believed that "all approaches are valid and should be pursued so that the Nation will benefit from an early interim capability while waiting for the long-term research to provide a highly accurate system." The committee was particularly impressed by the NBS research on syntactic analysis. In the "near future" it foresaw the establishment of a "national center for machine translation", and eventually a "national machine translation production program... operating on a full-time basis." In the long term it considered desirable the development of a "special-purpose computer, designed for translation." In the short term it approved the development "at an early date, of a limited machine translation with postediting (which) will provide the scientific community with a sample of things to come."

The committee's recommendations would have pleased all active US groups. Dostert of Georgetown would have found approval for his suggested national centres and early production prototypes; the theoretically inclined groups would have been encouraged to pursue the aim of high quality MT; those, like IBM, developing special-purpose machines would have been greatly reassured of continued support; and all would have liked the 'official' acknowledgement of the national importance of their efforts.

## 8. 6: National and international conferences.

In the few years since the 1956, MT research had become, in Bar-Hillel's words (1960), "a multimillion dollar affair", with the major proportion of the effort in the United States. With so many groups active in the US, many researchers were becoming worried about duplication and lack of information about what others were doing. All the groups except MIT and Texas were engaged on investigations of Russian-English systems; and many were compiling large Russian dictionaries and lexical databases. Some groups were already cooperating, such as Washington and IBM, Harvard and NBS, Ramo and Wayne State, and RAND had long been making its dictionary information available; yet it was clearly felt that closer formal links were desirable.

As a result of informal discussions at the UCLA conference, a meeting was arranged later in the same year (in July 1960) at Princeton for all MT groups sponsored by US Federal agencies (NSF, CIA, USAF, US Army, US Navy). The participants included not only most of those US groups present in UCLA, with the exception of IBM, Ramo-Wooldridge and Harvard, but also the European groups at Cambridge and Milan, which were receiving US grants at the time (Wayne State University 1960). The success of the meeting in promoting frank exchanges of views led to a series of similar working conferences to be known as 'Princeton-type' meetings, all organized by the Wayne State group (Josselson 1970). The second conference of the series was held at Georgetown in 1961, and was devoted to 'grammar coding'; the third, again at Princeton, in 1962, was on the theme of 'syntactic analysis' (participants now included IBM and Ohio State); and the fourth devoted to 'computer-based semantic analysis' was held at Las Vegas in 1965 (where participation was extended to non-Federally funded groups). There may be some dispute about the amount of real cooperation at these meetings – Zarechnak of Georgetown believed little tangible of benefit resulted (Zarechnak 1979: 42) – but they did mark genuine steps towards the fully-fledged discipline that MT was seen to be. As a further indicator of 'scientific maturity', it was at the third conference in 1962 that the Association for Machine Translation and Computational Linguistics was formally constituted.

Until about 1960, MT research had been concentrated in the large projects of the United States and the Soviet Union, with a significant contribution also from Great Britain. The period 1959-61 saw the appearance of numerous MT groups in many other countries: France (particularly the important CETA group), Belgium, Mexico, Czechoslovakia, Hungary, Rumania, East Germany, West Germany, China, and Japan (ch.5-7). In addition there was a new group in Britain at the National Physical Laboratory (ch.5.4), and among numerous new US projects the one at Ohio State (ch.4.13).

It was an opportune time for an international conference. In November 1961, the newly formed group at the National Physical Laboratory arranged a conference at its headquarters in Teddington, near London. Participants included representatives of all the major US and British groups, a French contingent, members of the Milan group, and a Japanese, Itiroo Sakai. There was unfortunately no Russian representation. The proceedings (NPL 1962) confirm increasing interest in syntactic analysis and in formal linguistic studies. Nearly all the contributions were concerned with some aspect of syntax; only the presentations from the Milan group, the Cambridge group, Wayne State, and IBM dealt with semantic and lexical problems in any depth. The contrast with the 1956 international conference, where only MIT and Georgetown were interested in syntactic questions, could not be more marked.

In the following year (in June 1962), the NATO Advanced Summer Institute on Automatic Translation of Languages was held in Venice (NATO 1966). Substantial contributions were made by Bar-Hillel, Brown, Ceccato, Vauquois, and others. Bar-

Hillel's 'Four lectures' (also in Bar-Hillel 1964) were basically restatements of his views on the impossibility of FAHQT, backed up by demonstrations of the failure of contemporary AI experiments in 'learning' machines. The other contributors were all confident in their various ways of eventual success: Brown by empirical means, Ceccato by conceptual modelling, Vauquois by formal linguistics.

The NATO link was a sign of the growing international stature of MT. However, there were increasing doubts among outside observers. Basically, there were still no actual operative systems. There had been numerous promises, e.g. by Dostert (Georgetown) and Reifler (Washington), and there had been accounts of significant progress from numerous groups, which had sometimes been interpreted as promises by eager journalists. Why? The answer, many concluded, must be that MT was inherently impossible. A spokesman for this view, which he argued passionately, was Mortimer Taube.

## 8. 7: Taube's Computers and Common Sense.

In 1961 appeared the book by Mortimer Taube entitled *Computers and Common Sense* (Taube 1961). Taube gave expression to a prevalent anti-computer view of the time, seeing mechanization of quintessentially human processes as 'dehumanising' and ultimately and necessarily doomed to failure. Understandably and justifiably, Taube began by looking for actual achievements in MT, and he found none. In support he quoted Oettinger (1960: 346) on the absence of working systems: "While several rules for producing smooth Russian-English automatic translations have been proposed in the literature, published experimental results have been conspicuously absent." However, Taube spoilt his case by exaggeration: "it can be stated categorically that twelve years after the Warren Weaver memorandum no practical, usable MT computer program exists... there does not even exist a practical or usable mechanical dictionary." In so far as there were no fully operational MT systems in evidence at this time he was, of course, correct. But he was wrong about their being no usable, working mechanical dictionaries; he had after all read about the most substantial one in Oettinger's book.

Taube's main argument was that MT is formally impossible. It is impossible because computers demand precise formalisation, and language cannot be formalised. Part of the argument rested on Gödel's theorem that consistent axiomatization of mathematics is impossible: "Since, at the very least, language must include all mathematics and since there seems to be a proof that all mathematics is not susceptible to formal treatment, it would follow that natural language is not susceptible to formal treatment." Specifically in relation to translation, Taube denied the possibility of strict synonymy, echoing the contemporary arguments of the philosopher Quine (1960) on the indeterminacy of translational synonymy. Human translation is intuitive and "we assume that machines are not capable of intuition. Hence, if they are to translate at all, they must translate formally." The conclusion was inescapable: "mechanical translation in the formal sense is impossible because translation in the formal sense from one natural language to another is impossible."

Formal linguistic analysis, as exemplified particularly in the work of Noam Chomsky and Zellig Harris, was attacked by Taube as a misguided aberration which "has cast a mystique over the whole field of MT." His basic point was that attempts to define transformation rules and develop formal grammars without reference to notions of synonymy and significance were logically impossible. Taube's objection then was to formalism which neglected semantics. For Taube there was just no point in continuing MT research. He noted that even practitioners could find no sound economic reasons. The University of Washington study (1959) had concluded that without optical character readers MT would be twice as expensive as human translation. Taube added sarcastically: "It seems that the main area of research should be print readers and not translating machines." He had no time for justifications based on prospects of the potential spin-offs (to linguistics or information retrieval), which were being mentioned increasingly by MT researchers, e.g. during the hearing for the House of Representatives report (5.3.2 above). Taube could find "nowhere in the literature of MT... a systems engineering study of its feasibility or practicality. In the absence of such a study and in the light of the known informality of language and meaning, research in MT takes on the character not of genuine scientific investigation, but of a romantic quest such as the search for the Holy Grail."

It matters little whether, in retrospect, Taube was right or wrong in some of his specific criticisms. McCorduck (1979), for example, points out that in his discussion of learning machines, Taube's "insistence on limiting what computers could do to a highly restricted sense of mathematical formalism meant he had to ignore those programs which did seem to exhibit learning, in particular, Samuel's checkers program." On MT Taube assumed that MT could only mean fully automatic systems producing translations as good as any human translator could do. Unlike Bar-Hillel, he did not even discuss the practical value of lower quality post-edited MT. Taube was also wrong in assuming that all MT research was based on the linguistic theories of Chomsky and Harris. As we have seen, many groups explicitly rejected such approaches: Washington, IBM, Milan, Cambridge, and Georgetown (to a large extent). Only at MIT had Chomskyan ideas been taken up at this time. On the other hand, he was right to reject an excessive attention to syntax which excluded semantics. Hidden in the polemics it is possible to read an argument for semantic as well as syntactic analysis in MT (McCorduck 1979); whether Taube would have agreed is doubtful. He would probably have dismissed the formalisation of semantics as yet another logical impossibility. His concern was to expose the formalist fallacies of determinism and to expose the mistaken idealism (or in some cases what he saw as wilful deceit) of those who were attempting the mechanization of human thought processes, and MT was just one example.

In the MT community Taube's book seems to have been ignored as an irrelevant 'curiosity', as it was in AI circles (McCorduck 1979). But it must surely have had an impact on the public perception of MT research. Together with Bar-Hillel's article, it must have contributed substantially to the growing impatience about the evidently slow progress of MT. Public perception would also have been influenced, at a more trivial level, by the frequently repeated stories of supposed MT howlers, such as the well known *invisible idiot* and *The liquor is alright but the meat is spoiled* versions (ch.1). It did not any longer appear surprising that the repeated promises of imminent working systems had failed to be fulfilled, and Taube had seemingly showed why MT would never come.

## 8. 8: Operational systems, and the 'semantic barrier'

As it happens, in the next few years operational systems were put into full time service. In 1964 the Mark II version of the IBM system developed by King was installed at the Foreign Technology Division of the US Air Force (ch.4.2). Apparently, the earlier version Mark I, had been used to translate Russian newspaper articles since 1959; but King did not reveal this until much later (King 1963). The Georgetown system for Russian-English translation was demonstrated during 1962 with some measure of success and, as a result, was delivered in 1963 to the EURATOM centre in Ispra (Italy) and in 1964 to the Atomic Energy Commission at the Oak Ridge National Laboratory. The output of both the Georgetown and the IBM systems was admitted to be of poor quality and usually in need of extensive post-editing, as the examples given in ch.4.2-3 above show. Nevertheless, the systems served a real need, many users expressing satisfaction with even unedited texts.

By this time, however, it was clear that research by more 'theoretical' groups was not succeeding. Intensive research on syntactic analysis had served only to show the intractability of syntactic ambiguity. Oettinger (1963) concluded from his own experience at Harvard with the predictive analyzer that "the outlook is grim for those who still cherish hopes for fully automatic high-quality mechanical translation." Likewise Yngve (1964) confessed that the intensive research at MIT had shown that "Work in mechanical translation has come up against what we will call the semantic barrier... We have come face to face with the realization that we will only have adequate mechanical translations when the machine can 'understand' what it is translating and this will be a very difficult task indeed."

It was already clear that sponsors were becoming less willing to support MT. In 1963, research at Georgetown was terminated, in circumstances which remain somewhat unclear (ch.4.3). In same year, in October, the director of the National Science Foundation requested that the National Academy of Sciences set up an independent committee to advise the Department of Defense, the Central Intelligence Agency, and the National Science Foundation itself on the future funding of MT research.

## 8. 9: The ALPAC report

The National Academy of Sciences formed the Automatic Language Processing Advisory Committee (ALPAC) in April 1964 under the chairmanship of John R. Pierce of Bell Telephone Laboratories. The other members included two linguists, Eric P. Hamp (University of Chicago) and Charles F. Hockett (Cornell University); a psychologist, John B. Carroll (Harvard University); two MT specialists, David G. Hays (RAND Corporation) and Anthony G. Oettinger (Harvard University); and one AI researcher, Alan Perlis (Carnegie Institute of Technology). The constitution is not without significance: neither of the MT specialists believed in the continuation of MT research as such; at RAND, the emphasis had shifted since the early 1960's towards basic research in computational linguistics (ch.4.4), and at Harvard disillusion with the practicality of MT had been growing for many years, so that by 1964 active work in this area had virtually ceased (ch.4.9). Eric Hamp was a linguist of the Bloomfieldian school generally sceptical of mathematical and computational linguistics. On the other hand, Charles Hockett had been for some time an enthusiast for mathematical linguistics, and wrote a basic text on the topic (Hockett 1967); however before its publication Hockett underwent a "radical shift in point of view" between 1964 and 1965, and he became convinced that formal grammar of the Chomskyan kind was utterly misguided (Hockett 1968). Finally, Alan

Perlis represented the emerging view of AI that linguistics-based (specifically syntax-based) approaches to language analysis were inevitably inadequate.

The committee[2] undertook studies of the existing demand, supply and costs of translations, the demand and availability of translators, an evaluation of some MT output and the costs of post-editing. ALPAC was concerned almost exclusively with translation from Russian and with economic considerations. From its surveys of US government translators and the provision of translations, it concluded that although poorly paid the "supply of translators greatly exceeds the demand", and that "all the Soviet literature for which there is any obvious demand is being translated". Indeed, it was concerned about "a possible excess of translation"; it thought that cover-to-cover translations of Soviet scientific journals "contain, in addition to much valuable information, many uninspired reports that the U.S. scientist could have been mercifully spared". ALPAC concluded that "the emphasis should be on speed, quality, and economy in supplying such translations as are requested."

One solution was the provision of machine aids for translators. The committee had been impressed by developments of text-related glossaries by the Germany Army's translation service (later LEXIS), and of the automatic dictionary by the University of Brussels for the European Coal and Steel Community (DICAUTOM). The value of mechanized dictionaries had, of course, been recognized from the beginning of MT research, e.g. by Booth, and later by Oettinger, and machine aids of many kinds have now been developed, from the national terminology databanks to the personal glossaries. (They will be treated briefly in ch.17.6.) Increased support for the development of aids for translators was the least controversial of the committee's recommendations. Machine translation, however, was most definitely not a solution. From its survey of the state of MT, ALPAC concluded that "there has been no machine translation of general scientific text, and none is in immediate prospect." In support, it drew attention to the fact that all MT output had to be post-edited. This was seen a 'failure': "when, after 8 years of work, the Georgetown University MT project tried to produce useful output in 1962, they had to resort to postediting". For some reason, ALPAC failed to acknowledge that most human translation, particularly when produced in translation agencies, is also revised ('post-edited') before submission to clients. The committee appears to have assumed that 'raw' MT output ought to be acceptable without editing.

The committee sponsored evaluations of three (unspecified) MT systems by John B. Carroll, reported briefly in Appendix 10, and in greater detail by Carroll (1966), which showed conclusively that on ratings for intelligibility and informativeness all were significantly poorer than three human translations. It also received evaluation studies by the IBM Research Center and by Arthur D. Little Inc. of the IBM system at FTD (Appendix 11). The latter revealed that in a sample of 200 pages there were 7,573 errors, of which 35% were omitted words, 26% wrong words, 12% incorrect choices and 13% wrong word orders. The committee reached the general conclusion that although "unedited machine output from scientific text is decipherable for the most part... it is sometimes misleading and sometimes wrong... and it makes slow and painful reading". As examples of "recent (since November 1964) output of four different MT systems", the ALPAC report included passages from a Russian article on space biology translated by

[2] For more on ALPAC see: J.Hutchins 'ALPAC: the (in)famous report', *MT News International* 14: 9-12.

the systems at Bunker-Ramo, Computer Concepts, FTD and EURATOM. The EURATOM system was, of course, as the report indicated, the Georgetown system (ch.4.3), and the Computer Concepts system was Toma's AUTOTRAN prototype of the later Systran (ch.4.13), also at this stage in most essentials the Georgetown system. The FTD system was the IBM Mark II (ch.4.2), and the Bunker-Ramo translations seem to have been produced by one of the earlier experimental versions (ch.4.6). Only the FTD and the EURATOM (Georgetown) systems were in fact operational systems; the others were still experimental, but this was not mentioned by ALPAC.

The MT output from these systems was obviously inadequate and unsatisfactory. Nevertheless, it was unfair of ALPAC to compare them unfavourably with the output of the 1954 Georgetown-IBM experiment (ch.4.3). For ALPAC, the work at Georgetown was typical of the progression from the "deceptively encouraging" early achievements to the current "uniformly discouraging" results. The committee failed to distinguish between a small-scale demonstration program working on prepared text and a large-scale working system dealing with unexamined texts.

More seriously, the committee failed to examine the theoretical research of other MT groups. The concentration on the Georgetown and IBM systems was probably understandable in view of the publicity these projects attracted, but it was surely amiss in neglecting to evaluate the projects at MIT, the University of California at Berkeley, the University of Texas, Bunker-Ramo, Wayne State University, and even Harvard University. It is true that the committee heard the testimony of Paul Garvin (Bunker-Ramo) and received the views of Victor Yngve (MIT). It is not known what these researchers had to say in substance, but the ALPAC report chose to emphasise Yngve's opinion that MT research had "come up against a semantic barrier" and that progress in MT required fundamental research in text understanding. This confirmed their belief that while support of MT research should be reduced, since "there is no immediate or predictable prospect of useful machine translation", there should be support of basic research in computational linguistics.

The committee recognized the contribution of MT to the development of computer software and to theoretical linguistics. It acknowledged that computational linguistics had grown out of MT research, and it believed that "The advent of computational linguistics promises to work a revolution in linguistics", with implications for language teaching, psycholinguistics, and computer aids in information retrieval and translation. Given the recognition of the fruitful interaction between MT and computational linguistics it appeared perverse to many at the time and subsequently (e.g. Zarechnak 1979) that ALPAC should recommend increased funding for computational linguistics but no more funds for MT.

The ALPAC report was widely condemned as narrow, biased and shortsighted. It was criticized strongly by Pankowicz of the Rome Air Development Center (one of the biggest sponsors) for its "factual inaccuracies... hostile and vindictive attitude... use of obsolete and invalid figures... distortion of quality, speed and cost estimates... concealment of data reflecting credit on MT... wilful omission of dissenting statements" (quoted by Josselson, 1970). MT researchers protested that improvements were imminent and that, in short, ALPAC's dismissal of MT was premature. But, whether the criticisms were valid or not, the damage had been done; MT research in the United States suffered immediate reductions and a loss of status which it has still not fully recovered. Whereas in 1963 there had been ten US groups (Georgetown, MIT, Harvard, NBS, Berkeley, Ohio State, Wayne State, Texas, Bunker-Ramo, and IBM) by 1968 there were just three (at Berkeley, Texas and Wayne) and two of these suffered interruptions and reductions in funding (at Berkeley between 1965 and 1968 (ch.4.10 and 11.2), and at Texas between 1968 and 1970 (ch.10.3) The effect of ALPAC was also felt in other countries where quite different conditions prevailed: none of the three British groups were engaged in active MT research after 1968 (although the Birkbeck College group had stopped some time before ALPAC, in 1962); research in Japan and the Soviet Union continued at much reduced levels; only the French group CETA (ch.5.5 and 10.1) appears to have been relatively unaffected.

## 8. 10: Expenditure on MT research.

Probably the most persuasive argument of ALPAC was that so little had been achieved despite huge investments of public money by the US government. As Roberts and Zarechnak (1974) put it, "from 1956 to 1965" MT research was being supported at "17 institutions to the tune of almost $20,000,000." These figures are frequently repeated, but they are in fact misleading. Appendix 16 of the ALPAC report lists expenditure under three headings: National Science Foundation (NSF) grants totalling $6,585,227, Central Intelligence Agency (CIA) grants totalling $1,314,869, and Department of Defense (DOD) grants totalling $11,906,600. The DOD grants are broken down into grants by the USAF (mainly through the Rome Air Development Center), totalling $9,613,000, by the US Navy ($971,600) and by the US Army ($1,322,000); but no details are given of the recipients of the grants. The CIA grants went all to the Georgetown University project.

The NSF grants are listed under 17 recipients. This is the source of the "17 institutions" but it is an incorrect interpretation: three headings refer to grants for MT conferences (to MIT in 1956 and 1960, and to Wayne State in 1960 and 1962), and one to grants for ALPAC itself. The institutional recipients include the Cambridge Language Research Unit (a non-US group), the University of Pennsylvania (the group under Zellig Harris which was not engaged in MT research as such, cf.4.13 above), and the University of Chicago (a grant to Victor Yngve to continue his MT work begun at MIT). There remain therefore ten US groups receiving NSF grants for MT research in the period 1956-65: Georgetown ($106,600, plus $305,000 transferred from the CIA, and in addition to the direct CIA grants), Harvard ($1,054,960, plus $206,500 transferred from RADC), MIT ($911,700), Berkeley ($722,400), Ohio Sate ($154,700), Wayne State ($444,000), Texas ($613,200), NBS ($221,200), Ramo-Wooldridge and Bunker-Ramo ($561,784), and Washington ($54,700). The total of direct NSF grants in the period is thus about $5,000,000.

Many of these institutions were also receiving grants from DOD sources, e.g. Washington from the USAF, Texas from US Army, and Wayne State from US Navy. It is known that the USAF supported the IBM project on a large scale, and that the Cambridge and Milan groups also received grants. However, there are good reasons to suspect that by no means all the huge DOD expenditure of nearly $12 million went to MT research.

In the report of the 1960 US House of Representatives mentioned earlier (8.5), details of USAF grants for 1960-61 are provided. Some support was being given to known MT research: the Milan project ($124,000), the Cambridge and Harvard projects ($125,000, via NSF), and the Ramo-Wooldridge project ($130,000). Other grants went to

MT-related research: to the MT evaluation project at the University of Washington (ch.4.1), and to Indiana University ($99,000, although this semantics research was only distantly relevant, ch.4.13). By far the greatest proportion of the total $3 million was going, however, to IBM for development of its photoscopic disk 'translator' ($1,787,000), to Baird Atomic Inc. for the development of an optical character reader to be used in conjunction with the IBM equipment ($381,000), and to Intelligent Machines Research Corp. for the development of ancillary computer equipment ($267,000). If we assume that, say, a third of the IBM grant went in fact towards the development of MT research as opposed to technical refinement of the Mark II equipment, then this still means that nearly two thirds of the 1960 USAF grants for "mechanical translation" were in fact going on hardware development. It would be too great an assumption to extrapolate these proportions to the USAF total for the period 1956-65, but it may not be too inaccurate to suppose that the figure for MT research should be nearer $5 million than the $9,613,000 given in the ALPAC report.

A similar reduction should probably be made for the US Navy grants: the 1960 House of Representatives report again indicates that most US Navy contracts for MT activities were in character readers, pattern recognition, high density storage). Of the total $532,500 for the period 1953-60 less than a quarter ($115,900) went to 'pure' MT research (Wayne State). Only the US Army seems to have concentrated exclusively on MT projects as such in 1960 (supporting Texas and NBS). If these proportions are cautiously extrapolated for the whole period 1953-65 then perhaps the figure for US Navy and US Army MT research contracts should be nearer $1,500,000 than the $2,300,000 in the report.

These adjustments result in a total expenditure of approximately $13 million by the US government and military agencies on MT research at 11 US institutions and 2 foreign institutions during the period 1956-65. Although the figures are substantially smaller (by a third) than those given by ALPAC and repeated frequently in subsequent years, the level of financial support was nevertheless immense. There was a good deal of justification for sponsors to expect practical returns for their support.

## 8. 11: The situation in 1966.

The ALPAC report may have dismissed solid achievements too readily, but it was probably quite right to call a halt to indiscriminate and wasteful expenditure on unrealistic projects. Unfortunately, it destroyed at the same time much of the credibility of MT research. After ALPAC few American researchers were willing to be associated with MT; indicative of the change was the deletion of Machine Translation from the title of the Association of Computational Linguistics in 1968.

In the decade of research since 1956, a considerable amount had in fact been achieved, not only in the United States but also in Britain, the Soviet Union and elsewhere. As far as operating systems were concerned, there were after all two now in regular use: the IBM Mark II at the USAF's Foreign Technology Division at the Wright-Patterson Air Force Base, and the Georgetown systems installed for the Atomic Energy Commission at the Oak Ridge Laboratory and for the EURATOM Centre in Italy. Operating systems had also been tested at the National Physical Laboratory and in the Soviet Union. Admittedly, the achievement in this respect was far less than had been anticipated at the beginning of the decade; and furthermore, the quality of output was much poorer than would have been hoped for.

These operating systems were all products of the ‘engineering’ approach to MT, systems which started from fairly rudimentary foundations (often word-for-word systems) and were progressively refined on a trial and error basis. In the case of the IBM system the refinements were by lexicographic means (ch.4.2); in the case of the Georgetown and NPL systems, refinements also included addition of rudimentary syntactic analyses (ch.4.3 and 5.4). The system at Ramo-Wooldridge was also progressively refined by improved parsing and dictionary information (ch.4.6) although by 1967 it was not yet operational.

As opposed to the engineering approaches there were the numerous ‘theoretical’ approaches, many of which sought to perfect procedures and methods before implementation. Few of the groups taking the ‘perfectionist’ attitude succeeded in going beyond preliminary small-scale experimental tests. Prominent among the theoretical groups were MIT, Harvard, Berkeley, Cambridge, and the Soviet groups at MIAN and Leningrad. Typical of the attitude of these groups were the remarks of Lamb (Berkeley), who thought it “a very curious thing that so much of the work in the field of Russian-to-English MT has been devoted to writing translation programs instead of investigating the structure of Russian” (Lamb 1961).

There were, of course, many differences between the theoretical groups, as we have seen (8.2). Some concentrated almost exclusively on semantic problems (e.g. Cambridge, Leningrad and the Moscow Pedagogical Institute), but increasingly projects were concentrating primarily on syntactic approaches (8.4): phrase structure grammar (MIT and Texas), dependency grammar (RAND and CETA), the fulcrum approach (Ramo-Wooldridge, Wayne State), and the extensive testing of the ‘predictive syntactic analyzer’ (NBS and Harvard). Initial high hopes for syntactic approaches were in the end deflated by the problems of structural ambiguity, e.g. the prepositional phrase problem (ch.3.6), and by unexpected multiple analyses and parsing failures (ch.4.9). Nevertheless, foundations were laid for future advances in parsing techniques.

Problems of dictionary techniques were throughout subject to close examination. As we have seen (8.2) the most popular method of dictionary access was the serial technique, but there were now alternatives: the binary partition method of Booth (5.1) and the tree structure technique of Lamb (ch.4.10). In large part, dictionary searching was constrained by hardware deficiencies. There was therefore, considerable interest in special-purpose equipment; not only the IBM photoscopic store, but also the Yamato machine in Japan (ch.7.1) and, according to Bar-Hillel (1960), they were also proposed in the Soviet Union. In the course of time, computer storage became less problematic and the notion of special ‘translating machines’ became largely irrelevant. Nevertheless, computer facilities were often unsatisfactory and many groups had virtually no chance to test their ideas on actual computers (ch.8.2 above).

The most characteristic MT system design of the period was the 'direct translation' approach, one-directional systems where SL analysis is directed explicitly towards the specific features of the TL, and where analysis and synthesis procedures are not separated (ch.3.9) This was the most common form, and is seen in the Washington, IBM, Georgetown, Birkbeck, NPL, and a number of the Soviet and Japanese systems. However, the advantages of separating SL and TL analysis were also recognized from the earliest days, e.g. by Panov in the Soviet Union (rudimentary 'transfer' systems can be seen in some of the Soviet systems), and most specifically in the 'syntactic transfer' approach at MIT (ch.4.7). This exploration of bilingual transfer mechanism converting SL phrase structure analyses into TL phrase structure equivalents was to influence the development of the CETA and Texas approaches (ch.5.5 and 10.1; 4.11 and 10.3). Concurrently there emerged in both the Soviet Union and the United States the concept of multi-level analysis, the separation of stages of morphological, syntactic and semantic analysis (ch.3.9) Its most explicit formulations appeared in the 'stratificational' theory of Lamb at Berkeley (ch.4.10) and in the theories of Mel'chuk (ch.6.3), later elaborated as the 'meaning-text' model (ch.10.2).

Most ambitious of all were the various proposals for interlingual systems. The idea of interlingual MT was put forward by Weaver in his 1949 memorandum. Researchers were both attracted by the centuries-old notion of a 'universal language' (ch.2.1) and by the practical advantages of translation via an 'intermediary language' in multilingual environments (ch.3.9). Richens was a strong advocate and at the Cambridge Language Research Unit his ideas of a basic universal set of primitive sense elements were tested in conjunction with the thesaurus approach (ch.5.2) The Milan group investigated in depth an interlingua which was intended to be a direct representation of 'conceptual' relations, independent of any languages, and was explicitly not based on universal or common linguistic features. Different conceptions of an interlingua were put forward by the Soviet researchers Andreev and Mel'chuk (ch.6.3-4). Mel'chuk proposed that an interlingua should be the sum of all features common to two or more languages. Andreev proposed an interlingua which would be a complete artificial language with its own lexicon and syntax, based on the most frequently common features in the languages under consideration. Other suggestions were the use of one of the artificial auxiliary languages, such as Esperanto or Interlingua (e.g. Gode 1955) – but, rather surprisingly, it is only recently that the idea has been taken seriously (ch.16.3). At the same time, Booth argued that the case for an 'intermediary language' was fallacious, firstly because in a multilingual environment of *n* languages, the number of programs can be reduced even more than the *2n* with an interlingua if one of the *n* languages itself is the mediating language (Booth et al. 1958). In fact Reifler (1954) and Panov (1960) had made the same proposal, suggesting English and Russian respectively as the best mediating languages. More pertinently, however, Booth argued that all SL-TL conversion involved abstract representations which could be regarded as 'intermediary'. In these remarks we may perhaps see the germs of the later ('deep syntactic') conceptions of the 'pivot language' of CETA (ch.5.5 and 10.1) and of the interlingual representations of the Texas group (ch.4.11 and 10.3).

In the face of the semantic complexities of MT, a number of researchers suggested limitations of vocabulary or grammar of input or output text. One line led to the development of the microglossary approach (Oswald 1957), the construction of specialised bilingual dictionaries to reduce the incidence of multiple meanings. The idea was taken up by many US projects (Washington, IBM, Georgetown, RAND), and in the Soviet Union. A second related line was the restriction of the MT system as a whole to one particular scientific field, which may be expected to have its own particular 'sublanguage': a popular choice was mathematics (e.g. Wayne State) Even more radical were Dodd's and Reifler's suggestions that writers should use 'regularized' language in texts (ch.2.4.3); but these ideas were not to be taken up in this form until later (ch.17.1-4) However, the notion of 'pidgin' MT languages did receive attention, by Reifler (ch.4.1) and in particular by the Cambridge group, which argued that 'low level' MT output could be made more understandable by the regular useful of 'pidgin' variables (ch.5.2)

These were proposals prompted by the amount of post-editing that MT output evidently needed. Initially post-editing was seen as part of the feedback process of improving systems (e.g. at RAND, ch.4.4) Increasingly, it was realised that in the foreseeable future revision of MT text would be a necessity in operational systems, although many still hoped that eventually post-editing would wither away with 'higher quality' MT. However, there was surprisingly little discussion of what was meant by translation quality.

Miller & Beebe-Center (1958) were the first to suggest measures of evaluation. Judges were asked to rate human translations and simulated MT versions of a set of abstracts on a percentage scale (0%= "no translation at all", 100%= "best imaginable translation"). The results indicated that subjective scaling was a poor measure. More successful was a second evaluation on the basis of identifiable 'errors' of vocabulary and syntax. Finally the authors suggested comprehension tests as evaluative measures. Nothing more was attempted until output of actual MT systems became available in the late 1960's. In evaluations of the IBM system (ch.4.2), tests of 'reading comprehension' were used by Pfafflin (1965) and by Orr & Small (1967) and Pfafflin also used a measure of 'clarity' in preference to an error analysis. In his evaluation studies for ALPAC, Carroll (1966) criticised the unreliability of comprehension scores, and used instead correlations of 'intelligibility' and 'informativeness' measures.

It is remarkable that quality evaluation should have been neglected, but it is perhaps symptomatic of the surprising failure of the US funding agencies to monitor the research they were sponsoring. One explanation is the wish to encourage as many different approaches as possible. A feature of most groups was the single-minded concentration on a particular method or technique. While MT was in these early stages (as it was still essentially at the time of ALPAC), this commitment was not necessarily bad; it meant, for example, that some approaches were tested to their limits. Examples are the thesaurus technique at Cambridge and the predictive analyzer at Harvard. What was perhaps less excusable was the sponsors failure to prevent duplication with so many groups doing basic investigations of Russian.

It is understandable in the 'cold war' climate of the 1950's and early 1960's (cf. the comments in the House of Representatives report above) that US research should have concentrated so heavily on Russian-English translation (all groups except MIT were involved), but the neglect of other languages is nevertheless remarkable. It is true that German-English MT was the main focus at MIT and Texas, and that Chinese-English translation was investigated at Washington, IBM, Georgetown, and (towards the end of the period) more extensively at Berkeley, Ohio State and Ramo-Wooldridge. There were also studies of French-English (principally by Brown at Georgetown, but also at IBM) and of Arabic (at MIT and Georgetown). But Japanese, for example, was virtually neglected, except for a small study by Kuno at Harvard; and there was no MT research on Spanish within the US – a surprising omission in view of the US involvement in Latin America.

By contrast, Soviet research was far more diversified. There was again, for obvious reasons, a preponderance of work on English-Russian systems; but equal importance was attached to research on French, Hungarian, and languages of Soviet republics. In the rest of Europe also, the range of languages was more diverse: French, German, Italian, Latin, Czech, Rumanian. In Japan, research naturally concentrated on Japanese and English.

In 1966, research on MT had reached a low ebb. After a slow start, the decade since 1956 had seen a vigorous growth of activity. Initial encouraging results, based primarily on word-for-word experiments and early trials of ‘direct translation’ systems, were followed after roughly 1960 by a gradual realisation of the immense complexities. The solutions looked for in increasing sophistication of structural analyses and syntactic formalism did not come readily; and the ‘semantic barrier’ seemed insurmountable. There were MT systems in operation, but the output was not satisfactory. It seemed that MT had been an expensive failure.

# Chapter 19: Present developments and future prospects.

## 19. 1: The general scene in the 1980s.

When the Commission of the European Communities announced the installation of Systran in their translation service, it marked the end of MT's quiet period in the scientific and technical doldrums. MT emerged from 'academic' obscurity into public view, and further technological advances brought it on to the market place.

In the 1960s MT was primarily the concern of university linguists and computer engineers. Commercial interests were not absent, but they did not flourish. The dramatic miniaturisation and the increasing sophistication of computer equipment have changed the picture in the 1980s. Now MT is also the concern of translators and of those selling translation aids and systems. Academic interests remain, but are now not dominant.

In the 1970s MT activity was concentrated on two main areas. Firstly, progressive improvements were being made to systems of the 'direct translation' type; the most important being Systran. The second main area was the development of 'indirect' approaches, initially the interlingual systems at CETA and LRC and the ambitious Soviet 'meaning-text' model, and then the development of the advanced transfer approaches (TAUM, GETA, SUSY), which emerged as the most realistic and feasible foundation for MT systems of the 'second generation' (where the 'first generation' was considered to be the direct approach favoured in the previous decade). At the same time, the 1970s saw the beginning of semantics-based approaches, primarily associated with AI, and the development of commercial interactive systems.

In the 1980s these lines have continued to be investigated and developed. There is now a variety of MT systems which almost defies any neat classification. It is still often legitimate to apply the labels of the 1960s: practical vs. theoretical, empirical vs. perfectionist, direct vs. indirect, interlingual and transfer. But now there are new labels and new perspectives: interactive vs. fully automatic, 'try-anything' systems vs. 'restricted language' systems, mainframe systems vs. microcomputer or word-processor systems, AI-based systems vs. linguistics-oriented systems.

In overall MT design strategy it is clear that, while the transfer approach has now been adopted by most (linguistically) advanced systems, the 'direct' approach is still attractive as a well-tested strategy, and the interlingual approach continues to appeal to the idealists and 'perfectionists'. It may noted, however, that occasionally these familiar categories no longer seem completely appropriate: Systran has features of a 'transfer' approach in a basically 'direct' design; and the differences between the 'interlingual' and the 'transfer' approach have become blurred in Logos, Eurotra, and METAL, for example.

For large scale systems there is now a broad measure of agreement in basic design: the flexible transfer strategy of GETA and SUSY is to be found in its essentials in the Eurotra project and in the newer Kyoto project. There is now a sound foundation of well-tested linguistic and computational techniques of analysis, transfer and synthesis. The systems represent a continuation of the earlier 'indirect' systems (MIT, CETA, LRC, TAUM). At the same time, there are revivals of older approaches. There are again, for example, projects with 'empirical' approaches (the statistical and contextual models in the Soviet Union); and there are word-centred or lexicon-driven approaches (as in a number of AI systems).

In the past decade, however, the innovation with the most immediate impact has been the emergence of the interactive systems (ALPS, Weidner, etc.); as Vauquois (1976) forecast one of the most fruitful way forward in MT research has been the development of on-line interrogation and editing techniques; there are now a number of systems available, more will no doubt appear in the next few years, and nearly all systems at present under development, whether designed for mainframe of microcomputer, envisage some kind of interactive component.

Finally, there remain attractions in 'restricted language' systems. Most of the earlier systems were designed primarily for particular subject areas (chemistry, mathematics, etc.); the most successful recent example has been METEO for meteorological reports. A variant is the restriction to language which computer programs can handle relatively easily (TITUS, Smart), or the pre-editing of text input.

All these various types of MT systems have each their own places and functions. The MT field is a mixture of practice and research. Practical operational systems apply well tested methods, both in large scale systems for mainframe computers (e.g. Systran and PAHO) and in smaller systems for microcomputers (e.g. Weidner and ALPS). Experimental systems test new ideas, such as AI methods, new parsers, logico-semantic representations. Some innovative approaches are to be found in the sophisticated and complex linguistics-oriented 'transfer' systems (GETA, Eurotra, Kyoto); others are to be found in the smaller scale AI-inspired knowledge-based systems (Yale, LUTE and ATLAS); and others in the new interlingual approaches (DLT, Rosetta). From these projects will no doubt emerge the techniques of future fully automatic systems.

It is evident that the earlier divisions between pragmatists (e.g. IBM, Georgetown) and perfectionists (e.g. MIT, Berkeley) remain to the present. Now the pragmatists are mainly the designers of interactive systems, and the perfectionists are among those who experiment with AI approaches.

This opposition of practicality and idealism is a sign of health. MT is not just a special sort of data processing, just another application of computers. This lesson was learnt in the early days of MT when computer engineers found that translation was rather more than manipulating a dictionary. Nor is MT just another branch of linguistics, of computational linguistics, or of artificial intelligence. It is not an 'academic' quest for theories of language or for models of human intellectual behaviour. The primary stimulus for MT research has been the urgent needs of scientists, engineers, technologists, economists, administrators, etc. to cope with an ever increasing volume of materials in foreign languages. In the 1950s and 1960s most demand was for access to scientific literature (in the United States, mainly from Russian); now the demand is for technical, legal, commercial and administrative translations prompted by the 'linguistic equality' of the European Communities and the biculturalism of Canada, by the needs of multinational companies and the exigencies of global telecommunications.[1]

MT is a marriage of practical needs and theoretical idealism. Wilks referred to MT as a 'testbed' for AI theories. MT has been and continues to be a test for linguistic theories, although linguists have been reluctant to acknowledge the fact. MT provides a touchstone for the realism of theory; small scale models applied to carefully selected or specially composed texts are no proof, large scale application to actual texts outside the laboratory are the real test. As yet, the more sophisticated approaches (of GETA, Eurotra, and AI systems) have not demonstrated their capabilities; only the earlier cruder 'direct' systems and the interactive systems have shown their practicality.

---

[1] Developments in MT research and usage since the book was written in 1984/85 are recorded in: W.J.Hutchins, 'Recent developments in machine translation: a review of the last five years', *New directions in machine translation: international conference, Budapest 18-19 August 1988*, ed. D.Maxwell, K.Schubert, T.Witkam (Dordrecht: Foris, 1988), pp. 7-63; J.Hutchins, 'Latest developments in machine translation technology: beginning a new era in MT research', *The Fourth Machine Translation Summit: MT Summit IV*. Proceedings: International Cooperation for Global Communication, July 20-22, 1993, Kobe, Japan. [Tokyo: AAMT, 1993], pp.11-34; J.Hutchins, 'Research methods and system designs in machine translation: a ten-year review, 1984-1994', *Machine Translation: Ten Years On*, 12-14 November 1994. [Proceedings of conference at] Cranfield University ... [London: British Computer Society, 1994]; J.Hutchins, 'A new era in machine translation', *Aslib Proceedings* 47(1), 1995, pp. 211-219; J.Hutchins, 'The state of machine translation in Europe', *Expanding MT horizons*: proceedings of the Second Conference of the Association for Machine Translation in the Americas, 2-5 October 1996, Montreal, Quebec, Canada. [AMTA, 1996], pp. 198-205; and J. Hutchins, 'The development and use of machine translation systems and computer-based translation tools', *International Conference on Machine Translation & Computer Language Information Processing*, 26-28 June 1999, Beijing, China, ed. Chen Zhaoxiong [Beijing: Research Center of Computer & Language Engineering, Chinese Academy of Sciences, 1999], pp.1-16. All available at: http://ourworld.compuserve.com/homepages/WJHutchins.

## 19. 2: Linguistic problems and methods

MT research is a mixture of solid achievement and speculative experiment. There are now few problems with morphological analysis and with dictionary searching routines. There is broad agreement among MT researchers on general strategies for these processes, and the same techniques are found repeatedly in one system after another. This has been true since the mid-1960s and it is surprising, therefore, that there seem still to be no standard programming 'packages' available (Damerau 1976).

Syntactic analysis techniques are likewise well tested and relatively efficient. There is now a wide variety of computational techniques and aids to choose from (cf. King 1983): context free parsers, ATN parsers, tree transducers, charts, etc. There are, of course, still disagreements about the relative merits of different parsing strategies: bottom-up vs. top-down, etc. (Ch.3). A more recent issue is the question of production systems vs. procedural systems (Rosner 1983, Johnson 1983): production systems are sets of unordered generalised rewrite rules, which may be applied as appropriate to input strings or trees (as in Q-systems and charts, cf. Ch. 9.14-15 above); procedural systems by contrast specify sequences of rewriting rules (e.g. ATN parsers), they represent a deterministic approach to parsing. As yet, choice of syntactic analysis strategies would seem to be largely a matter of individual preferences (Johnson 1983)

Two recent developments in formal linguistics may well influence future developments in parsing, since both deny the need for transformational rules. One is the theory of 'lexical-functional grammar' (e.g. Bresnan 1982), in which relationships between semantically related surface structures are handled by lexical rules. Thus, whereas in the standard transformational approach (Chomsky 1965) active and passive forms are derived from the same underlying 'deep' phrase structure, in lexical-functional theory the equation is expressed through correlations of functional structures assigned to the active and passive verb forms, i.e. that the object of the active form of a verb has the same thematic function as the subject of the verb's passive form. In this way language-specific features are removed from the syntactic (transformational) component into the lexicon. As a parser the model supports a lexicon-driven approach.

The other development is the theory of 'generalized phrase structure grammar' by Gazdar and others (1985). Chomsky claimed to have proved that phrase structure grammars (and hence all context-free grammars) were mathematically inadequate for describing natural language sentence structures. Gazdar argues that Chomsky imposed unnecessary constraints on the formal properties of phrase structure grammars, in particular by excluding the use of complex symbols. Gazdar proposes formalisms which capture the generalisations and generative power of transformations without going outside the limitations of context-free grammar.

While syntax and parsing are thoroughly researched and well founded, the greatest problems remain, as they always have been, in the area of semantics. The use of semantic features for resolving certain types of syntactic ambiguity and the use of case frame analyses are now well established in very many systems of all kinds. The problems arise whenever information is required from outside the sentence being analysed. Pronominal reference is one area of difficulty; in practical systems a simple rule of taking the most recent noun phrase as the antecedent may work by and large, and corrections can be made during post-editing. Satisfactory treatment, however, clearly requires often 'real world' knowledge. Likewise, very many problems of lexical ambiguity and of TL lexical selection demand some 'understanding' of texts. Much can be achieved with restrictions of systems to particular sublanguages, but unfortunately few actual texts fall neatly into one particular subject field; most refer to other areas and disciplines.

There has been considerable activity in linguistics in all these areas. The work on sublanguages arose directly from problems of MT and related spheres of computational linguistics, e.g. Kittredge & Lehrberger (1982), both members at one time of the TAUM group, and Sager (1981) of the Pennsylvania group. There has been a growing substantial interest by linguists in recent years in discourse analysis and text structures (e.g. Halliday & Hasan 1976, van Dijk 1977, 1980, de Beaugrande & Dressler 1981). Petr Sgall of the Prague group has made particular study of topic-comment structures and presupposition, both traditional specialities of Czech linguists (Ch.13.5 above). Although this linguistics research has aroused considerable interest there is little evidence yet of its direct effects on MT research.

It is an embarrassing fact for those who believe there can only be better quality MT with more linguistically sophisticated systems that the most successful operational MT systems so far (based essentially on the 'direct translation' approach) owe almost nothing to linguistic theory, or rather to be precise, to syntactic theory. What is the reason for this apparent 'failure' of syntactic theory to provide or suggest models appropriate to MT? One may be the distinction between 'competence' and 'performance'. Linguistic theory (largely inspired by Chomsky's theoretical position) has concentrated on the formal definition of language systems and has generally neglected the investigation of language behaviour in social contexts; it has pursued the goal of 'scientific' rigour, idealisation and abstraction without checking its hypotheses and theoretical models against empirical observations of actual linguistic usage. It is paradoxical that the primary impetus for the formalisation of grammars which made the automation of linguistic processes appear feasible should have itself encouraged the dissociation of theory and practical reality which has led to be adoption of unrealisable models.

The influence of linguistic theory in the past is to be seen in the strict separation of stages of analysis (morphological, syntactic, semantic), the adoption of the 'filtering' approach. This rigidity was found to be a major cause of the failures of the 'interlingual' systems (Ch.10). It was also a mistake, as Ljudskanov (1968) pointed out, to assume that translation processes necessarily took place at the 'deepest' levels of analysis; surface information was often sufficient, and was usually needed in any case. As we have seen, the newer linguistics-oriented systems have abandoned filtering, retain surface information and mix semantic, syntactic and lexical information in representations and procedures (GETA, Eurotra, METAL). These systems are no longer implementations of any particular linguistic theories; they are independent MT models, albeit founded on linguistics approaches. As such, they are still essentially developments of syntax-based MT strategies, and have problems whenever analysis has to go beyond sentence boundaries and whenever disambiguation needs to take account of 'real world' knowledge. Nevertheless, these systems have such a high degree of flexibility that there is no reason to doubt their capacity to incorporate new methods of analysis, including AI methods, without detrimental effects on existing successful procedures.

## 19. 3: Artificial intelligence and MT

For AI researchers, the difficulties and past 'failures' of linguistics oriented MT point to the need for AI semantics-based approaches: semantic parsers, preference semantics, knowledge databases, inference routines, expert systems, and the rest of the AI techniques.

There is no denying the basic AI argument that at some stage translation involves the 'understanding' of a SL text in order to convey its 'meaning' in a TL text. For AI researchers this means that texts must be given semantic or 'conceptual' representations (independent of particular languages), that parsing must be based on semantic criteria, that 'knowledge' of the world must be called upon to interpret relationships and resolve ambiguities, etc. The result is an impressive battery of AI techniques, some of which have been already described (Ch 15 above).

The concept of knowledge databases as sources of information necessary for disambiguation and inference mechanisms has considerable attractions. Many of the newer MT projects intend to incorporate such data in their systems. However, there have been no estimations of the size or complexity required for the knowledge base of a MT system. Therefore, despite assurances from AI researchers, there are many MT researchers who doubt the capacity of such approaches in large scale systems.

The degree of understanding required for translation is, in any case, still an open question. Does the translator of a biochemistry document need to be a biochemist? He must have some knowledge, it is true, but does he need to 'understand' as fully as the biochemist? It is surely evident that much translation can take place at a fairly superficial level (Slocum 1984a, Ljudskanov 1968). If this were not so, then 'word-for-word' translations would not convey any meaning at all. For practical MT purposes, it has yet to be shown that complete understanding in the AI sense is necessary. It is therefore not surprising that many MT projects still aim to translate at lexical and phrase structure levels (e.g. the Kyoto projects, Ch.18.7).

When considering the applicability of AI methods to MT, there need be no commitment to AI philosophy. That is to say, it does not matter whether we believe that AI programs map texts onto 'universal' conceptual representations or whether we take these 'representations' as just useful computer data structures and nothing more. What matters is whether the programs work, i.e. produce paraphrases, answer questions, translate. It may be dangerous, as Weizenbaum (1976) argues, to imagine that the machine really 'understands' and that it is simulating the way we as humans understand, since this has social and moral implications. But there is no need to see the computer as anything other than a symbol manipulator.

The most successful application of AI methods has been the development of expert systems, i.e. knowledge bases incorporating the 'lore' of experts in a particular field (e.g. medical diagnosis or geology) and an 'inference engine' to work out conclusions from data submitted by users. Most rules are conditional and probabilistic ('if X and Y are true then Z is likely with a probability of p'). An important feature is the ability of the computer to describe the steps leading to a conclusion or explain the logical reasoning being pursued, thus enabling the user to assess the validity of the output.

What could be the role of expert systems in MT? The most obvious one would be to assist in SL text disambiguation (as proposed for the GETA, SALAT, TRANSLATOR, DLT and LUTE systems). MT methods of semantic analysis have tended to apply 'selection restrictions' and measures of semantic compatibility (Ch.3.6). An expert-system approach would be less rigid, more probabilistic, more inferential - more, it would be hoped, like a human translator. The idea is certainly attractive, but the complexities of codifying the knowledge of an expert translator would appear to be particularly daunting.

Most AI systems have been experimental projects on highly restricted domains. The results in many cases seem promising, but there is no guarantee that if extended to larger scale system the methods would still work. AI researchers are confident: "Knowledge-based machine translation... has been implemented in a pilot system" which "established the technical feasibility of KBMT... Newer and more robust semantic-based parsing techniques and better natural language generators argue in favor of converting the KBMT approach from a laboratory exercise to production-quality translation systems in the very near future" (Carbonell & Tomita 1985). There were similar pronouncements by MT researchers in the 1950s. The history of MT has numerous examples of methods which promised much at first but proved disappointing failures.

Such exaggerations should not be allowed to disguise the real promise of AI techniques. There is no doubt that AI methods of semantic analysis will become standard components of MT systems in the future. But they will be components; AI approaches are in themselves not the whole solution. In the 1960s there was a tendency for a number of MT groups to concentrate on a single approach (particularly in syntax). The lesson of the 1970s was that integrated flexible approaches were needed. A number of AI researchers are recognising that efficient parsing cannot be exclusively semantics-based, and that some use of syntactic information must be made (Lytinen 1985). Similar conclusions were reached much earlier by the Grenoble and Saarbrücken groups, and flexibility of this nature has been built into the Eurotra system.

In this respect, it should be noted that a number of AI researchers have developed 'lexicon-driven' systems. A example is the 'word-expert' parser of Small (1983), where each word has associated with it (in the lexicon) a set of predictions about preceding and succeeding lexical items and creates tentative ‘concept structures’ (or frames). Even word endings such as the plural *-s* invoke predictions. Although similar techniques have been applied in many AI experiments (e.g. the Yale group), this is the most extensive exploration. It may be noted in passing that both the word-expert approach of Small and the lexical-functional theory of Bresnan (above) are examples of the recent revival of ‘lexis-oriented’ grammars (which is also to be seen in the adoption by a number of MT systems of ‘lexicon-driven’ methods of analysis and synthesis, in general philosophy similar to some systems of the 1960s.)

There are now many MT groups already applying AI techniques; and many recent projects are to investigate the possibilities in greater depth. American researchers have been impressed by the achievements of AI techniques, and the same is true for many European MT researchers. A particular stimulus in Japan has been the launch of the ‘Fifth Generation Computer Systems’ project in October 1981 by the Japanese Ministry of International Trade and Industry. The project has political and social motives, to give Japan a leading, perhaps dominant, role in the future world economy. The Japanese see information technology as the key. “The goal is to develop computers for the 1990s and beyond – *intelligent* computers that will be able to converse with humans in natural language and understand speech and pictures... that can learn, associate, make inferences, make decisions...” (Feigenbaum & McCorduck 1984). MT has a central role in the project, as one of the “basic application systems” along with “consultation systems, intelligent programming systems and an intelligent VLSI-CAD system.” The aim is a multilingual system with a vocabulary of 100,000 words, guaranteeing “a 90% accuracy, with the remaining 10% to be processed through intervention by man, and with total costs... at 30% or lower than those of translation by man” (Moto-Oka 1983)

In many respects, these immediate objectives are not unrealistic – the Eurotra project has similar aims and they are expected to be achievable with present MT expertise in linguistics oriented approaches and with present computer technology (Ch.14.2 above). However, more distant objectives of the Fifth Generation appear rather more utopian. Yasuo Kato, general manager of systems research at the Nippon Electric Corporation, one of the consortium backing the project, predicts that “in 20 years you will have (an interpreting) machine that you can put in your pocket. The system will recognize your voice, translate what you say, and read it out in another language.” (quoted in *Business Week*, Sept.16, 1985)

There is no obligation to accept such prognostications when acknowledging the future impact of AI in MT systems. A more sober assessment would predict a gradual integration of AI approaches in MT systems, leading to definite (perhaps highly significant) improvements over the next twenty years, but with the ultimate completely automatic system still a distant goal.

## 19. 4: MT quality and revision.

The goal of the MT ‘perfectionists’ of the 1960s was fully automatic high quality translation. While it may be true that some AI-inspired researchers are still aiming for FAHQT, this goal has long been abandoned by the great majority of those working in the MT field. It is recognised that revision is normal and expected for all translations, whether done by humans or by computers. Debates about what is meant by ‘high quality’ or ‘fully automatic’ are largely irrelevant. What matters is whether the MT output is satisfactory for its intended use (revised or not) and whether the operation is cost-effective. The question of translation quality is entirely relative to the needs of recipients, as Bar-Hillel (1971) acknowledged some years after his influential FAHQT pronouncements (Ch. 8.3). There can be valid uses of poor quality output in unedited form if it is produced quickly, cheaply and not intended for publication. If better quality is required then collaboration of man and machine is essential. As we have seen, this means in practice either post-edited or interactive MT.

What is meant by good quality MT output? It is a surprisingly elusive concept. Judgements of quality are primarily subjective. MT output is ranked lower than human translation generally because MT systems still make gross grammatical 'mistakes', select the wrong words and produce stylistic barbarisms. Objective criteria are difficult to establish. Van Slype (1982) has attempted to assess the value of different yardsticks: intelligibility (of output text, e.g. via readability scales, cloze tests), fidelity (to SL original, e.g. via measures of information transfer), acceptability (to recipient of translation), time spent in revision (post-editing), number of 'errors' corrected and type. Costing of MT in comparison with human translation requires careful accounting of all stages of the translation processes from receipt of SL text to final despatch of revised TL version. Evaluation of MT is made that much more difficult by the lack of any objective measures of human translation. As Van Slype (1982) comments, human translation has never been the subject of quality controls of the kind common in industrial enterprises. It has only been evaluated 'objectively' since comparisons began to be made with MT output.

There is now considerable practical experience of post-editing 'raw' MT output. Reports of 'professional' MT revisers have been presented at the annual series of Aslib conferences (e.g. Lawson 1982, 1985) The work rate of post-editors depends largely on their attitude to mechanisation and word processors and on their motivation; translators are trained to aim for high quality products, they find it difficult initially to accept low quality MT with equanimity. Translation is a creative activity; many find MT revision inhibiting: it is "not an aid, but a hindrance, because it limits their freedom of expression" (Wagner 1985). Others, however, rise to the challenge. Such differences in attitude can easily account for the wide ranges in productivity which have been reported, from 5 pages a day to 25 pages or more a day. But, the amount of revision is clearly very much dependent on the quality and difficulty of the original text. Nevertheless, it is the conviction of many translation services that MT with post-editing is a viable economic proposition. The experience of translation bureaux such as ITT, which has installed the Weidner system, is generally favourable: they claim increased productivity with no degradation of quality (Magnusson-Murray 1985).

Initially post-editors tend to be overzealous in their correction of MT texts, changing not only grammatical and lexical mistakes but also altering the style of texts. They have to learn to edit without complete rewriting. Some become very fast: the idea of 'rapid post-editing' has developed (e.g. in the translation service of the Commission of the European Communities) in recognition that not all users want 'purple prose' versions. There is now a growing realisation that for many recipients stylistic refinements are not necessary; it appears that on the whole users are more content with low quality texts than translators and post-editors. There would seem to be a growing market for rapidly produced low quality translation, which the advent of MT has opened up. There is undoubtedly a considerable latent demand for texts to be translated for information purposes which would otherwise not be translated at all.[2] But, of course, there is a natural fear by many translators that standards will be lowered and with a consequent detrimental impact on their professional image and social standing.

## 19. 5: Future prospects

Taube expressed the fears of an earlier generation about the insidious mechanistic and dehumanising influence of the computer (Ch.8.7). Weizenbaum (1976) has expressed the continuing fears of a later generation. It is not surprising that many translators should have similar fears about the potential influence of MT.

Technical advances in on-line interactive facilities may well dispel such fears. There is no need to assume that computerisation means either that translators will be no longer needed or that the translator will be a slave to a machine. There is ample evidence to show that there will always

[2] The need is now being met (at various levels of quality) by MT services on the Internet since the mid 1990s.

be a demand for the high quality product which only the human translator can achieve; not only in the translation of literary works where sensitivity to cultural and linguistic nuances is crucial, but also in such areas as diplomatic and legal translation where meticulous accuracy is of paramount importance.

The most likely development in the near future is a translator's workstation, on the lines envisaged by Melby (Ch.17.10), and exemplified by the recent SUSANNAH project at Saarbrücken (Ch.13.2). A workstation would combine a word processor, access to machine aids (not only dictionaries but other reference materials), a multi-level interactive MT system (of the ALPS kind), and a fully automatic MT system. The translator would select whichever method might be the most appropriate for a text to be translated. The technology is already available, and translators are likely to be attracted by a 'translating machine' which they can control.[3]

It is for such reasons that interactive MT systems can be expected to remain popular for a long time. There is also no sign that fully automatic systems will be able in the near future to resolve the semantic problems that prevent good quality output; interactive disambiguation and stylistic improvement will continue to offer an obvious solution, and one which satisfies the professional integrity of the translator.

At the same time, it is unlikely that the fully automatic systems which are now operative or under development will disappear in the near or even distant future. Large scale systems on mainframe computers will continue to retain a place in the larger translation services and in service bureaux. Systran is already available through bureaux to smaller translation agencies and individual clients. Free-lance translators already have access over public telecommunication networks to remote term banks (e.g. EURODICATOM). Remote access to mainframe MT systems is surely the next stage.

By then, it is to be hoped or expected that 'direct' systems such as Systran will be superseded by more sophisticated 'transfer' systems like Eurotra. There are signs already of quality degradation in Systran (at least in the Russian-English version, Ch.12.1); and there are good grounds for optimism that transfer systems (Eurotra, METAL, Kyoto) will produce significantly better quality output, even though the high quality anticipated by AI researchers may not be forthcoming in the near future.

Many of the interactive systems (ALPS, Weidner, Logos) are already available on microcomputers (mainly Wang and IBM PC). With progressive miniaturisation there is the future prospect of even larger translation program packages for personal computers, perhaps approaching the size and complexity of the present mainframe systems. Eventually perhaps, there will be genuine hand-held 'personal translators', which will be so useful for businessmen, travellers and students, and which should be substantially more satisfactory than the rudimentary 'phrase dictionary' types introduced abortively in the late 1970s by Nixdorf, Craig and other manufacturers (Smith 1984). In the more immediate future there is most likely the development of programs which combine the writing and composing of business correspondence and simultaneous translation; initially they will be 'restricted language' interactive systems on the lines suggested by the researchers at UMIST and Carnegie-Mellon (Ch.17.13), but later no doubt there will be systems with almost unrestricted input.

In the immediate future, a major problem still to be solved is that of cheap input to MT systems. The need for optical character readers has been recognised from the earliest days, and yet there are still no inexpensive machines capable of reading a wide variety of typefaces and fonts (Roman and non-Roman). Until they appear the costs of all types of MT will remain relatively high. Voice input is, of course, often mentioned as a future possibility; but the problems of phonetic

[3] The first commercial translator's workstations appeared in the early 1990s. Their rapid adoption was due not only to the facilities mentioned but also to the development of means for storing and accessing large databases of previously translated texts ('translation memories'). See: J.Hutchins 'The origins of the translator's workstation', *Machine Translation* 13(4), 1998, pp.287-307.

analysis and transcription, for more than highly restricted vocabularies, are almost equal to the problems of 'high-quality' semantic analysis.

Development of MT systems is no longer an exclusively 'academic' pursuit; it is in the hands of commercial companies (and as a consequence, much technical information is unavailable): Siemens, Philips, Fujitsu, Hitachi, Weidner, etc. But more importantly for the future, it is increasingly in the hands of translators themselves: as post-editors they provide feedback for system improvements (e.g. in the CEC Systran service, Ch. 14) and they can influence directly the direction of commercial MT development. It can be argued that one of the failures of earlier MT research was the neglect of what translators actually did and what they actually wanted from computer systems (Masterman 1979).

The centres of MT research since the mid 1960s have been in Europe (predominantly France, Germany and the Soviet Union), Japan, Canada and the United States. The pattern reflects the general concentration of scientific research in the major industrialised nations. It is not then surprising that there has apparently been no MT research in South America, Africa and India. But there have also been other omissions: Australia and Scandinavia (with the exception of the Danish contribution to Eurotra), Poland and East Germany. There is evidence of some activity in China again since the 1950s (cf. Ch. 15.5), although little is known of what has been achieved; and interest is growing in the Middle East, where it can be expected that MT projects will be established in the future. A model for the future might be the international collaboration of the Eurotra project; the political complexities of funding the research teams have been daunting, but if the project proves successful there will surely be others to come.

Although in comparison with the 1960s the range of languages involved in MT has greatly expanded, there are still notable gaps. The earlier neglect of Spanish (Ch.8.11) has been rectified by Systran, PAHO and a number of interactive systems; but there is still a relative neglect of Arabic, Portuguese and (since the mid 1960s) Italian. Some major languages have yet to be treated in MT systems; in particular, the languages of India and Africa (e.g. Hindustani, Bengali, Panjabi, Swahili) and most Southeast Asian languages (Vietnamese was a short-lived exception). On the other hand even some European languages have not yet been tackled in MT systems: Polish, Swedish, Greek. Involvement in the Eurotra project will entail MT research at some future date on some of these 'neglected' languages: Portuguese, Italian, Greek, and Danish. For other languages, however, it would seem that what is required is some economic or political incentive; for good or ill, international understanding and cooperation have not in the past been sufficient motivations for financing substantial MT projects.

Advances in computer technology and programming have made possible the proliferation of MT projects. Whereas until the mid 1970s MT research demanded large scale computer facilities, it is now possible for anyone with a microcomputer to experiment. There is evidence of this happening already, and it will no doubt increase. In the past, MT programming was almost exclusively in assembly code for specific machines; now increasingly, MT programs are written in high level languages (at present, the most popular are LISP, Prolog, and Pascal) and this itself will encourage more interest by many who otherwise have no opportunity to do MT research. It is notable that many AI projects have involved small numbers of researchers, often no more than two or three.

While small projects have made valuable contributions and will no doubt continue to do so in the future, it is equally indisputable that major advances in MT will probably come primarily from the larger projects of the size of GETA, Eurotra, Kyoto, and so forth. Only these have the capacity to test new linguistic and AI techniques on a sufficiently large scale. The lesson from the past ought to be clear enough: the complexity of MT derives from the complexity of language and the huge variety of text types. Small scale projects often seem initially of great promise, but there are frequently disappointing results when methods are tested on larger systems. The history of MT reveals how difficult it is to achieve success: of the many projects in the past thirty years, only a handful have resulted in operational systems. However, much has been learned and there is genuine and well-founded optimism for future progress.

## 19. 6: Generations and periods.

In computer science it is common practice to refer to generations of computers: vacuum tube computers, transistorised computers, integrated circuit computers, and very large-scale integrated computers (VLSI). In the anticipated 'fifth generation', the basic serial von Neumann design of computers until now is expected to be replaced by parallel architectures, new memory organisations and new programming languages.

There has been an inclination among MT researchers to refer also to generations of systems. For some the first generation is represented by the simple word-for-word systems, the second generation added syntactic analysis and the third incorporated semantics of some kind (Locke 1975, Toma 1977). For others, the first generation is represented by the 'direct translation' systems, the second by the 'indirect' systems and the third by systems based on AI approaches (Vauquois 1976, Hutchins 1978). As a result Systran, for example, is sometimes classified as a 'third generation' system because it incorporates some semantic analysis, and sometimes as a 'first generation' systems because it adopts the direct translation approach. In addition, however, there is no place in either classification for the interactive systems, unless they are regarded as 'transitional' stages between generations, as does Melby (1980) with the Brigham Young system, or as 'hybrid' forms - i.e. CULT would belong to the first generation as an interactive 'direct' system.

It may be more appropriate to see the development of MT in terms of evolutionary periods. The first period extended from the end of the Second World War until the mid 1950s (the Georgetown-IBM demonstration and the MIT conference in 1956). The second period, which lasted until the ALPAC report in the mid 1960s, was characterised by vast US governmental and military support, great enthusiasm and considerable disappointments. The third period was MT's 'quiet' period when research concentrated on 'indirect' systems and when the first operational systems became well established. The fourth period began in the mid 1970s with the interest of the Commission of the European Communities in the possibilities of MT (Systran), the first public MT system (METEO), the reorganisation of MT activity in the Soviet Union, and shortly afterwards, the appearance of commercial systems, increasing research on AI approaches and the revival of Japanese interest.

The periods are roughly decades, and so it is tempting to believe that a new decade is now beginning. If so, it may possibly be marked by a burst of activity associated with the Japanese 'fifth generation' project and by a resurgence of MT research activity in the United States, for which there are already signs (conferences at Georgetown and Colgate University in 1985, and the foundation of a new specialist journal *Computers and Translation*.) Whether entering a new period or not,[4] the future of MT is secure: it satisfies a genuine urgent need, it is the subject of worldwide research and development, and it is becoming a commercial product like other technical aids and office equipment; the application of the computer to translation is a reality, for many it is already as much a part of life as the computer itself.

[4] A new era began in 1989/1990 with the appearance of corpus-based approaches to MT (statistics-based and example-based machine translation), the first translator workstations, and, shortly afterwards, the beginnings of MT on the Internet. See references in footnote 1.