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.