[Originally published in: *MT News International* 14 (June 1996), 9-12.
Reprinted in: *Readings in machine translation*, ed. S.Nirenburg, H.Somers and Y.Wilks
(Cambridge, Mass.: The MIT Press, 2003), 131-135.]

ALPAC: the (in)famous report

John Hutchins

The best known event in the history of machine translation is without doubt the publication thirty years ago in November 1966 of the report by the Automatic Language Processing Advisory Committee (ALPAC 1966). Its effect was to bring to an end the substantial funding of MT research in the United States for some twenty years. More significantly, perhaps, was the clear message to the general public and the rest of the scientific community that MT was hopeless. For years afterwards, an interest in MT was something to keep quiet about; it was almost shameful. To this day, the 'failure' of MT is still repeated by many as an indisputable fact.

The impact of ALPAC is undeniable. Such was the notoriety of its report that from time to time in the next decades researchers would discuss among themselves whether "another ALPAC" might not be inflicted upon MT. At the 1984 ACL conference, for example, Margaret King (1984) introduced a panel session devoted to considering this very possibility. A few years later, the Japanese produced a report (JEIDA 1989) surveying the current situation in their country under the title: *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC.*

While the fame or notoriety of ALPAC is familiar, what the report actually said is now becoming less familiar and often forgotten or misunderstood – and this extensive summary includes therefore substantial extracts.

The report itself is brief – a mere 34 pages – but it is supported by twenty appendices totalling a further 90 pages. Some of these appendices have had an impact as great as the report itself, in particular the evaluation study by John Carroll in Appendix 10.

The first point to note is that the report is entitled: **Languages and machines: computers in translation and linguistics**. It was supposedly concerned, therefore, not just with MT but with the broader field of computational linguistics. In practice, most funded NLP research at the time was devoted to full-scale MT.

The background to the committee is outlined in the Preface:

> The Department of Defense, the National Science Foundation, and the Central Intelligence Agency have supported projects in the automatic processing of foreign languages for about a decade; these have been primarily projects in mechanical translation. In order to provide for a coordinated federal program of research and development in this area, these three agencies established the Joint Automatic Language Processing Group (JALPG).

It was the JALPG which set up ALPAC in April 1964 under the chairmanship of John R.Pierce (at the time, of Bell Telephone Laboratories). Other members of the committee were John B.Carroll (Harvard University), Eric P.Hamp (University of Chicago), David G.Hays (RAND Corporation), Charles F.Hockett (Cornell University, but only briefly until December 1964), Anthony G.Oettinger (Harvard University), and Alan Perlis (Carnegie Institute of Technology). Hays and Oettinger had been MT researchers, although no longer active when ALPAC was meeting (having become disillusioned with progress in recent years); Perlis was a researcher in Artificial Intelligence; Hamp and Hockett were linguists; and Carroll was a psychologist. The committee did, however, hear evidence from active MT researchers such as

Paul Garvin and Jules Mersel (Bunker-Ramo Corporation), Gilbert King (Itek Corporation and previously IBM), and Winfred P.Lehmann (University of Texas).

The committee agreed at the outset that support for research in this area "could be justified on one of two bases: (1) research in an intellectually challenging field that is broadly relevant to the mission of the supporting agency and (2) research and development with a clear promise of effecting early cost reductions, or substantially improving performance, or meeting an operational need." ALPAC rejected (1), deciding that the motivation for MT research was the practical one of (2) alone. For this reason, ALPAC "studied the whole translation problem" and whether MT had a role in it.

The second point to note, therefore, is that the report concentrated exclusively on US government and military needs in the analysis and scanning of Russian-language documents. It was not concerned in any way with other potential uses or users of MT systems or with any other languages.

The first half of the report (pp. 1-18) investigated the translation needs of US scientists and government officials and overall demand and supply of translations from Russian into English. ALPAC began by asking whether, with the overwhelming predominance of English as the language of scientific literature (76% of all articles in 1965), it "might be simpler and more economical for heavy users of Russian translations to learn to read the documents in the original language." Studies indicated that this could be achieved in 200 hours or less, and "an increasing fraction of American scientists and engineers have such a knowledge", and it noted that many of the available opportunities for instruction were underutilized (Appendix 2)

Next it looked at the supply of translations within government agencies (including those sponsoring MT research). They used a combination of contract and in-house translators. The committee was not able to determine the exact number of in-house translators, but it did establish that the average salary of translators was markedly lower than that of government scientists. Nevertheless, it found "a very low rate of turnover among government translators. Indeed, the facts are that the supply exceeds demand." At the time of the report, no post of government translator was vacant while there were over 500 translators registered in the Washington area (statistics in Appendix 8 of the report).

The committee was thus prompted to ask whether there was any shortage of translators. The Joint Publications Research Service, it found, had the capacity to double translation output immediately: out of 4000 translators under contract only 300 on average were being used each month. Likewise, the National Science Foundation's Publication Support Program was prepared to support the cover-to-cover translation of any journal which might be nominated for complete translation by any "responsible" society. Appendix 6 recorded 30 journals being translated from Russian in this way during 1964. Since, some had very low circulations (Appendix 6), ALPAC questioned the justification for this virtually "individual service".

Indeed, ALPAC wondered whether there were not perhaps an excess of translation, on the argument that "translation of material for which there is no definite prospective reader is not only wasteful, but it clogs the channels of translation and information flow." What it found was that many Russian articles were being translated which did not warrant the effort: according to a 1962 evaluation, only some 20 to 30% of Russian articles in some fields would have been accepted for publication in American journals; furthermore the delays in publication of cover-to-cover translations reduced their value. The committee concluded that the main need was for "speed, quality, and economy in supplying such translations as are requested."

At this point, before considering MT as such, the report anticipated its conclusions with the bald statement (p. 16): “There is no emergency in the field of translation. The problem is not to meet some nonexistent need through nonexistent machine translation. There are, however, several crucial problems of translation. These are quality, speed, and cost.”

On quality, ALPAC stressed that it must be appropriate for the needs of requesters: “flawless and polished translation for a user-limited readership is wasteful of both time and money.” But there were no reliable means of measuring quality, and for this reason ALPAC set up an evaluation experiment (reported in Appendix 10). This study by John B.Carroll evaluated both human and machine translations, and it had great influence on many MT evaluations in subsequent years. It was supplemented in Appendix 11 by a study from the Arthur D.Little, Inc. of MT errors, based on the system in use at the time at the Foreign Technology Division, i.e. the system developed by Gilbert King at IBM.

On speed, ALPAC saw much room for improvement: scientists were complaining of delays; the most rapid service (from JPRS) was 15 days for 50 pages; the NSF translation of journals ranged from 15 to 26 weeks; documents sent to outside contractors by the US Foreign Technology Division were taking a minimum of 65 days; and when processed by the FTD's MT system, they were taking 109 days (primarily caused by processes of postediting and production, detailed in Appendix 5).

On cost, ALPAC considered what government agencies were paying to human translators and this varied from $9 to $66 per 1000 words. In Appendix 9 calculations were made of cost per reader of the different forms of translation, including unedited output from the FTD system. These costs included the expenditure of time by readers. Assuming that the average reader took twice as long to read unedited MT documents as good quality human translation (based on the results of Carroll's evaluation in Appendix 10), it concluded that if documents are to be read by more than 20 persons traditional human translation was cheaper than MT. As for the costs of postedited MT, they would include posteditors proficient in Russian; ALPAC concluded that “one might as well hire a few more translators and have the translations done by humans... [or] take part of the money spent on MT and use it either (1) to raise salaries in order to hire bilingual analysts – or, (2) to use the money to teach the analysts Russian.”

At this point, the report turned to “the present state of machine translation” (pp. 19-24). It began with a definition: MT “presumably means going by algorithm from machine-readable source text to useful target text, without recourse to human translation or editing.” And immediately concluded: “In this context, there has been no machine translation of general scientific text, and none is in immediate prospect.”

Support for this contention, ALPAC asserted, came from “the fact that when, after 8 years of work, the Georgetown University MT project tried to produce useful output in 1962, they had to resort to postediting. The postedited translation took slightly longer to do and was more expensive than conventional human translation.” Likewise, ALPAC regarded it as a failure that the MT facility at FTD “postedits the machine output when it produces translations.”

However, the principal basis for its conclusion was the results of Carroll’s evaluation exercise in Appendix 10. “Unedited machine output from scientific text is decipherable for the most part, but it is sometimes misleading and sometimes wrong... and it makes slow and painful reading.” The report then printed (pp. 20-23) what it held to be “typical” samples of the “recent (since November 1964) output of four different MT systems.” These were presumably those used in the evaluation exercise, but this was not stated explicitly. The four systems were from Bunker-Ramo Corporation, from Computer Concepts, Inc., from the USAF Foreign Technology

Division, and from EURATOM. The first would have been the system developed by Paul Garvin after he left Georgetown in 1960. The Euratom system was the Georgetown University system installed in 1963 at Ispra, Italy. The FTD system was, as already mentioned, the one developed by Gilbert King at IBM, using his patented photoscopic store (a precursor of the laser disk). The Computer Concepts company had been set up by Peter Toma after he left the Georgetown project in 1962; the system illustrated was presumably AUTOTRAN, based in many respects on the SERNA version of the Georgetown system, and a precursor of SYSTRAN. Only the Euratom and FTD systems were fully operational at this time, the other two were still experimental prototypes - but this was not mentioned by ALPAC.

After reproducing the MT samples, the report continued: "The reader will find it instructive to compare the samples above with the results obtained on simple, selected, text 10 years earlier (the Georgetown IBM Experiment, January 7, 1954) in that the earlier samples are more readable than the later ones." Twelve sentences from the highly-restricted demonstration model (Hutchins 1994) are then listed, with the comment: "Early machine translations of simple or selected text... were as deceptively encouraging as 'machine translations' of general scientific text have been uniformly discouraging."

There can be no doubt about the deficiencies and inadequacies of the translations illustrated but it was perhaps a major flaw of ALPAC's methodology to compare unfavourably the results of general-purpose MT systems (some still experimental) working from unprepared input (i.e. with no dictionary updating) and the output of a small-scale demonstration system built exclusively to handle and produce a restricted set of sentences.

ALPAC concluded this chapter by stating that it was very unlikely that "we will not suddenly or at least quickly attain machine translation", and it quoted Victor Yngve, head of the MT project at MIT that MT "serves no useful purpose without postediting, and that with postediting the over-all process is slow and probably uneconomical." However, the committee agreed that research should continue "in the name of science, but that the motive for doing so cannot sensibly be any foreseeable improvement in practical translation. Perhaps our attitude might be different if there were some pressing need for machine translation, but we find none."

At this point, ALPAC looked at what it considered the much better prospects of "machine-aided translation" (not, as it stressed, human-aided MT, but what are now referred to as translation tools). It had high praise for the production of text-related glossaries at the Federal Armed Forces Translation Agency in Mannheim (Germany) and for the terminological database at the European Coal and Steel Community, which included terms in sentence contexts - this was the precursor of EURODICAUTOM. (Further details were given in Appendices 12 and 13, pp. 79-90). Its general conclusion was that these aids, primitive as they were, were much more economically effective in the support of translation than any MT systems.

The alternative it saw was postedited MT. However, it admitted that it could not "assess the difficulty and cost of postediting". Appendix 14 (pp. 91-101) reported on a study involving the translation of two excerpts from a Russian book on cybernetics, and the postediting of a an MT version of one of the excerpts. Interestingly, "eight translators found postediting to be more difficult than ordinary translation. Six found it to be about the same, and eight found it easier." Most translators "found postediting tedious and even frustrating", but many found "the output served as an aid... particularly with regard to technical terms." Despite the inconclusiveness of this study, ALPAC decided to emphasise the negative aspects in the body of its report, quoting at length the comments of one translator:

> I found that I spent at least as much time in editing as if I had carried out the entire translation from the start. Even at that, I doubt if the edited translation reads as smoothly as one which I would have started from scratch. I drew the conclusion that the machine today translates from a foreign language to a form of broken English somewhat comparable to pidgin English. But it then remains for the reader to learn this patois in order to understand what the Russian actually wrote. Learning Russian would not be much more difficult.

At the beginning of the next chapter "Automatic language processing and computational linguistics", ALPAC made one of its most often cited statements, namely that "over the past 10 years the government has spent, through various agencies, some $20 million on machine translation and closely related subjects." The statistics provided in Appendix 16 (pp.107-112) reveal that by no means all this sum was spent on MT research in the United States. Firstly, the total includes $35,033 on sponsoring three conferences and $59,000 on ALPAC itself. Secondly, it includes $101,250 in support of research outside the United States (at the Cambridge Language Research Unit) and $1,362,200 in support of research under Zellig Harris at the University of Pennsylvania which even at the time was not considered to be directly related to MT. Thirdly, it lists global sums from the US Air Force, US Navy and US Army (totalling $11,906,600) with no details of the recipients of the grants. Evidence from elsewhere (details in Hutchins 1986:168) suggests that much of the funding was in support of developments in computer equipment rather than MT research (perhaps up to two thirds of the USAF grants). In brief, the funding of US agencies on US research in MT may well have been nearer $12-13 million than the frequently repeated $20 million stated by ALPAC. The sum was still large, of course, and ALPAC was right to emphasise the poor return for the investment.

The main theme of this chapter on "Automatic language processing and computational linguistics" was a consideration of the contribution of MT research to advances of NLP in general. Summarizing the more extensive findings in Appendices 18 and 19, it found that its effect on computer hardware had been insignificant, that it had contributed to advances in "computer software (programming techniques and systems)", but that "by far the most important outcome... has been its effect on linguistics." Here they highlighted insights into syntax and formal grammar, the bringing of "subtler theories into confrontation with richer bodies of data", and concluding that although "the revolution in linguistics has not been solely the result of attempts at machine translation and parsing... it is unlikely that the revolution would have been extensive or significant without these attempts." (This is a view which would certainly be disputed today.) However, despite this favourable influence, ALPAC did not conclude that MT research as such should continue to receive support; rather it felt that what was required was

> basic developmental research in computer methods for handling language, as tools for the linguistic scientist to use as a help to discover and state his generalizations, and ... to state in detail the complex kinds of theories..., so that the theories can be checked in detail.

In the final chapter (pp.32-33), ALPAC underlined once more that "we do not have useful machine translation [and] there is no immediate or predictable prospect of useful machine translation." It repeated the potential opportunities to improve translation quality, particularly in various machine aids: "Machine-aided translation may be an important avenue toward better, quicker, and cheaper translation." But ALPAC did not recommend basic research: "What machine-aided translation needs most is good engineering."

ALPAC's final recommendations (p. 34) were, therefore, that research should be supported on:

1. practical methods for evaluation of translations;

2. means for speeding up the human translation process;
3. evaluation of quality and cost of various sources of translations;
4. investigation of the utilization of translations, to guard against production of translations that are never read;
5. study of delays in the over-all translation process, and means for eliminating them, both in journals and in individual items;
6. evaluation of the relative speed and cost of various sorts of machine-aided translation;
7. adaptation of existing mechanized editing and production processes in translation;
8. the over-all translation process; and
9. production of adequate reference works for the translator, including the adaptation of glossaries that now exist primarily for automatic dictionary look-up in machine translation.

Aware that these recommendations failed to support not just MT but any kind of natural language processing, a statement was inserted in the final report addressed to the president of the National Academy of Sciences from the chairman John R.Pierce in which he stressed the value of supporting "computational linguistics, as distinct from automatic language translation" Elaborating on recommendations in its chapter on NLP, the chairman believed that the National Science Foundation should provide funds for research on a reasonably large scale, "since small-scale experiments and work with miniature models of language have proved seriously deceptive in the past," – obviously alluding to MT experience – "and one can come to grips with real problems only above a certain scale of grammar size, dictionary size, and available corpus."

The ALPAC report was relatively brief; and its direct discussion of MT amounted to just one chapter (pp.19-24) and four appendices (on evaluating translation (pp. 67-75), on errors in MT (pp. 76-78), on postediting MT compared with human translation (pp. 91-101), and on the level of government expenditure on MT (pp. 107-112)). The rest of the report was concerned with the demand for translation in general by US government agencies, the supply of translators, with computer aids for translators, and with the impact of MT on linguistics. However, it was in these few pages that ALPAC condemned MT to ten years of neglect in the United States (longer, as far as government financial support was concerned), and it left the general public and the scientific community (particularly researchers in linguistics and computer science) with the firm conviction that MT had been a failure or, at best, very unlikely to be a useful technology.

In some respects, the impact of ALPAC can be exaggerated. MT research in the US did not come to a complete and sudden halt in 1966. Some projects continued, notably at Wayne State University under Harry Josselson until 1972 and at the University of Texas under Winfred Lehmann and Rolf Stachowitz until 1975 (later revived in 1978 with funding from Siemens). Furthermore, some MT projects supported by government money had ended before ALPAC reported: University of Washington (1962), University of Michigan (1962), Harvard University (1964). In particular, the Georgetown University project, whose system was explicitly criticized by ALPAC, had received no funding after 1963. By this time it had installed operational MT systems at the Oak Ridge National Laboratory and at the Euratom laboratories in Italy.

Furthermore, in hindsight it can, of course, be agreed that ALPAC was quite right to be sceptical about MT: the quality was undoubtedly poor, and did not appear to justify the level of financial support it had been receiving. It was also correct to identify the need to develop machine aids for translators, and to emphasise the need for more basic research in computational linguistics. However, it can be faulted for concentrating too exclusively on the translation needs of US scientists and of US agencies and not recognizing the broader needs of commerce and industry in an already expanding global economy. In this way, ALPAC reinforced an Anglo-

centric insularity in US research which damaged that country's activities in multilingual NLP at a time when progress continued to take place in Europe and Japan. It took two decades for the position to begin to be rectified in government circles, with the report for the Japan Technology Evaluation Center (JTEC 1992) and with ARPA support of US research in this field during the 1990s.

References:

ALPAC (1966) *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council, 1966. (Publication 1416.) 124pp.

Hutchins, W.J. (1986) *Machine translation: past, present, future***.** Chichester: Ellis Horwood.

Hutchins, W.J. (1994) The Georgetown-IBM demonstration, 7th January 1954. *MT News International* no.8 (May 1994), 15-18.

JTEC (1992) *JTEC Panel report on machine translation in Japan*. Jaime Carbonell [et al.] Baltimore, MD: Japanese Technology Evaluation Center, Loyola College in Maryland, January 1992.

JEIDA (1989) *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC, U.S.A.* Tokyo: JEIDA, 1989.

King, M. (1984) When is the next ALPAC report due? In: *10th International conference on computational linguistics...Proceedings of Coling84*, July 1984, Stanford University, Ca. (ACL, 1984), p. 352-353.