Congressional inquiry into MT

On September 11, 1990, the US House of Representatives sub-committee on science, research and technology held a hearing to review progress in machine translation, and set the stage for future policy in this area. Muriel Vasconcellos, chairman of the American Translators Association committee on trans-lation and computers, reported in the *ATA Chronicle* November 1990 issue.

Two of the witnesses at the hearing spoke about the problem of US Government requirements, particularly in the field of translation from Japanese to English.

Martha Caldwell Harris, director of the National Research Council's Office of Japan Affairs, expressed concern over what she called the "asymmetries in access to technical information between the United States and Japan". Despite significant efforts over the last five years, the fundamental problem persists, she said, and she urged that policymakers, concerned with US competitiveness and security, look for new ways to improve access to

developments in Japanese science and technology. She felt that machine translation can be a useful tool for addressing this problem, and argued for the importance of developing a machine translation technology in North America that would be geared directly to US interests.

Deborah Wince-Smith, Assistant Secretary of Commerce for technology policy, referred to the call, in the 1988 agreement between the United States and Japan on cooperation in research and technology, for improved access and dissemination of scientific and technical information.

Dale Bostad reported to the hearing on the experience of the last 20 years in using Systran to translate Russian, and more recently German and French, into English for the US Air Force. The proof that consumers are satisfied with machine translation, he said, is that they are now accepting raw output directly, retrieving it via desktop PCs linked to the US Air Force's mainframe computer, and that the monthly number of such inquiries is rising steadily.

Scott Bennett reported on METAL, another operational machine translation system, which was developed at the University of Texas, and which is now being used to translate the language pairs German to English, English to German, French to English, German to Spanish, German to Danish, Dutch to French and French to Dutch.

Muriel Vasconcellos described the ten years' experience of machine translation production at the Pan American Health Organisation using SPANAM (Spanish to English) since January 1980 and ENGSPAN (English to Spanish) since mid-1985. Research systems not yet in operational use were represented by *Jaime Carnonell* of Carnegie Mellon University and *David Johnson* of IBM.

The panellists pointed out that developing general purpose machine translation systems, as opposed to domain-specific ones, is a costly, long-term undertaking. Moreover the

analysis of new source languages, needed for tapping into foreign technology, is three times more expensive than the addition of new target languages to an existing source language component.

Non-Roman alphabets, they made clear, escalate the amount of investment required, and it is calculated that this, with the difficulties of interpreting word and sentence boundaries, and the need for an immense lexicon, adds to the problem by a factor of three. Accordingly, it was estimated that the building of a robust Japanese to English system will cost at least $15 million.

Congressman *Doug Walgren* of Pennsylvania asked penetrating questions about the respective roles of government, industry and the universities in seeing that an adequate effort is mounted to gain access to critical languages and the development of the machine translation source component languages required. Dr Vasconcellos emphasised that industry cannot do the job alone; that the long-term nature of the challenge cannot be supported by the typical customer base, for which the requirements are usually quite different. It was felt that the United States government can be most effective if it helps to pinpoint needs, set priorities, reduce duplication, and identify, mobilise, coordinate, and pool resources. In this connection it was made clear that the National Translations Center, whose appropriation of $127,000 is under threat of removal from the federal budget, is in a position to play a major coordinating role of the type that is needed.

The Sub-committee asked for a set of recommendations, especially in regard to the further development of English machine translation. They were told that the necessary background on the state of the art would be gathered by a team of machine translation experts due to make a visit to Japan. Congressman Walgren called for a specific plan, with a detailed outline of what the United States needs to do and to spend in order to get from "here" to "there".

Writing for the *ATA Chronicle*, Dr Vasconcellos recalled that 30 years ago, in May 1960, the House of Representatives committee on science and astronautics held similar hearings for a similar purpose. Could history be repeating itself? she asked:

"What is repeating itself is that once again the United States is slipping behind. Thirty-odd years ago Sputnik caught us napping, and suddenly there was an urgent need to get up to date on Soviet technology. We eventually managed to do so, in part with the help of Georgetown University's Russian to English machine translation system (GAT), installed at the Atomic Energy Commission in Oak Ridge, Tennessee, and Systran's Russian to English system at the Air Force Foreign Technology Division in Dayton, Ohio — at first a mere fledgling, but by 1978 a mature partner in the business of scanning Soviet literature.

During much of the post-Sputnik era, English remained the principle language for the publication of research results throughout the world. Gradually, however, the picture began to change, and today we have been rudely awakened once again, this time upon learning that 50% of all scientific research is now being reported in languages other than English — principally, Russian, German and French, but also Chinese, Korean, Farsi and Arabic.

At the same time, we are faced with the fact that our pool of translators in these languages, especially the non-European ones, is not large enough to deal with the tons of information that needs to be perused, and we are forced to recognise that unless this information is translated and assimilated within months from the time it appears, we can no longer hope to maintain a competitive edge in many of the areas that are critical to the national interest."