

## LET ME TELL YOU HOW IT REALLY WAS

Michael Zarechnak  
Georgetown University  
Washington, DC 20057  
mzarechnak@aol.com

*Bliss was it in that dawn to be alive,  
But to be young was very heaven.*

— Wordsworth

I would like to share with you, the younger generation, some memories that may be worth examining within the context of the past, present, and future of machine translation.

In the history of MT, the year 1947 is associated with the names of A.D. Booth and Warren Weaver and their correspondence in which they speculated about using a computer to translate messages from one natural language into another.

Those were the days of many fast-moving events. The ensuing decade was in fact the dawn of the Space Age. The launch of Sputnik I, the first man-made satellite to orbit the Earth, by the Soviet Union in 1957, was a shock to Americans. Fueled by the Cold War, MT researchers in the 1950s and 1960s were doing pioneering scientific work, paving the way for later commercial development.

In 1954 I was at Harvard University studying under Professor Roman Jakobson. The October issue of *Voprosy Yazykoznanija* reported the world's first demonstration of machine translation by Georgetown University and IBM in the United States. In that demonstration a computer was programmed to translate selected sentences from Russian into English using a restricted dictionary of 250 words. Professor Jakobson gave me an excellent recommendation to Professor Léon Dostert, founder and head of the Georgetown MT project. I soon joined the group of linguistic researchers led by Professor Paul Garvin who had provided the linguistic formulation for the Georgetown-IBM experiment.

I believe that nobody can detract from Professor Garvin's pioneering contribution to solving, on a computer, the two basic problems involved in the translation process: (1) the decision steps connected with the selection of lexical units, and (2) the arrangement units in the target language. These two problems reflect the two axes in natural language structure referred to by de Saussure as the associative and syntagmatic axes.

As a true pioneer, Professor Dostert combined vision and practice. The goal of the Russian-English MT project was to produce output in English such that the outside evaluator, a specialist in the field, would find the quality of the text acceptable as a source for gleaning the information contained in the original language. The dictionary entries containing the Russian words, their English equivalents, and their morphological, syntagmatic, syntactic, and semantic codes were keypunched and introduced into the computer memory.

Léon Dostert had a practical vision of machine translation and believed that it was feasible. He was always pushing for it to become a reality. He knew how to motivate his co-workers by letting them conduct research independently, provided only that, at the end of the day, they would be willing to come down from the level of theory and try to embody the theory in some practical system.

Not all the heads of other MT projects thought that MT was feasible, and many of them concentrated on limited studies of this or that phenomenon. There were exceptions, of course—most notably, the research of Professors Victor Yngve and Winfred Lehmann.

At MIT, Victor Yngve developed a special programming language to help linguists write their own rules and pursue the MT goal as a long-range project with possible insights on the way. He devoted his attention to the left-branching of English. He also produced the journal *Mechanical Translation*, which published an article of mine on semantic support for the resolution of syntactic ambiguities.

At the University of Texas, Winfred Lehmann was theoretically oriented and produced a tripartite system for German-English. Later the system was further expanded and became known as METAL, the basis for one of today's most sophisticated MT systems.

Léon Dostert set himself apart from the others by stating that one should work the way Descartes did: tackle one part of a problem at a time, learn from it, take on another part, and so on. Eventually, through these cumulative efforts, insights would grow and general comprehension might be closer at hand.

My contribution to machine translation is reflected in the original name of the Georgetown system, General Analysis Technique (GAT), which stressed that the analysis of the source and target language should be done independently of one another and that the transfer of meaning should be done between a pair of languages. To accomplish these tasks the members of the MT team should know both languages and be linguistically trained. Statistical considerations should be invoked when the structural (distributional and contrastive) information does not provide a clear solution.

The GAT system's lexical entries and translation rules were expressed in a machine-independent form known as the Simulated Linguistic Computer (SLC). SLC was developed by Dr. A.F. Brown on the basis of the special programming language that he created for French-English MT. It was demonstrated at a Unesco conference in 1959. He began translating a single sentence and then added another sentence, correcting or adding more information to translate these sentences into English. This approach resulted in the insight that one should make a distinction between "global" operations, whose domain was the entire sentence (e.g., a "subject routine"), and "local" operations, triggered by individual words (e.g., a dictionary entry with multiple equivalents in English). SLC permitted us to replace the fifteen passes through the computer (based on the implementation developed by Peter Toma) with only two: one for dictionary lookup and the other for everything else. In 1993 Dr. Brown ported the GAT-SLC system to the C programming language and today it runs on PCs.

The researchers who did not believe in the feasibility of MT succeeded in persuading a government panel in 1966 that "high-quality" MT could not be achieved and that research money should instead be devoted to computationally oriented theoretical linguistic studies. Their conclusions were published by the National Science Foundation under the name of the Automatic Language Processing Advisory Committee (ALPAC). It was not until after a conference organized by Professor Lehmann in 1971, at which the earlier critic Yehoshua Bar-Hillel no longer defended "fully automatic high-quality translation," that it became clear that something had to be done to return to a pragmatic approach to MT development.

That experience should be our lesson for the future. At Georgetown we worked with bilingual people who were linguistically trained. They were our models for capturing intuitions. We should pay more attention to semantic and pragmatic situations and try to formalize them. The intersecting pragmatic observations for a given set of languages might provide some clues for building an intermediary language.