

MY FIRST 30 YEARS WITH MT

Peter Toma
President, Voxtran, Inc.
Postfach 101033
53448 Bad Neuenahr-Ahrweiler. 1 Germany
Tel.: +49-2641-203653; fax: +40-2641-203630

1. CalTech, 1956-1957

In 1956 I was working at the California Institute of Technology in Pasadena as an assistant in the Physics Department when one day my colleague Tom Taylor announced, “Peter, we’re going to get a computer!” I instantly knew that this was something for me. I sensed that the arrival of the computer, a Datatron 205, was going to be a turning-point in my life. I would do something with languages, I thought. I got a deep feeling—which has driven me all these years—that I was destined to program the computer to translate languages. At that time I didn’t have the faintest idea that other MT initiatives were already under way.

As I gained knowledge about the computer’s capabilities, I became even more intrigued with the idea of working with text and languages. I volunteered to stay during the night and tend the computer. This enabled me to key in transliterations of Russian text and begin to experiment. The output (on paper tape!) was actually quite good. I had developed an interesting approach. Someone alerted the press and a reporter came. Unfortunately, this publicity was premature. The news reports claimed that I had achieved more than I actually had. I had two choices, either see if I could get support and continue, or join up with another project. I went to Washington, D.C., to attend a meeting at the National Science Foundation. There I met Michael Zarechnak, Gilbert King, and others in the field for the first time.

In 1957 I gave up my job at CalTech and joined Gilbert King at International Telemeter. I hadn’t been there very long, however, when King decided to take the project to the Rome Air Development Center. I preferred not to follow him. I submitted a proposal to the National Science Foundation, but it was not accepted. I joined the project at Georgetown University on March 1, 1958.

2. Georgetown University, 1958-1961

Of the various approaches being tried at Georgetown, the General Analysis Technique (GAT), impressed me the most and I began working with that group. My job was to reduce the linguistic logic to computer representation. The IBM 705 they were using was much better than the Datatron. My first task was to write a morphological program. By then I had read literature in the field and my ideas were more crystallized. After the morphology was completed I proceeded to program the entire GAT system—a program that I baptized SERNA. We presented our work on Friday mornings at our weekly seminars, and I vividly remember the time when it was up to me to present the SERNA system to a visitor—Allen Dulles, then director of the Central Intelligence Agency.

In the fall of 1958 I was given a free hand in the programming with permission to contract part-time programmers. We were paying for computer time by the minute, but then the Pentagon offered us the free use of their IBM 705 on week-ends and helped us to recruit programmers from there. From November 1958 until June 1959 I spent every week-end as well as Christmas and New Year’s Day working at the IBM 705 in the Pentagon with my small but enthusiastic programming staff. Our devotion and hard work culminated in a successful demonstration on June 8, 1959. As I worked, I noticed that the tapes in the tape room were stacked up to the light bulbs and mentioned my concern because I had a strong presentment of danger.

The demonstration went very well. Immediately afterwards I asked Lt. Harry Stevens to copy all our tapes. We then drove them to the sponsoring agency, where I requested that they be put in a safe. Later that same afternoon I flew home to California. A few days later headlines in the newspapers announced that there had been a fire in the tape room at the Pentagon. Luckily, our tapes were backed up.

3. Prelude to SYSTRAN, 1961-1965

In 1961 I left Georgetown and took charge of the machine translation section of Computer Concepts in Los Angeles, where I developed two entirely new MT systems: AUTOTRAN and TECHNOTRAN, both of them oriented toward the IBM 7090 computer. The logic was totally different from what Georgetown had used. In June 1964 I was invited to the University of Bonn to demonstrate the systems. During my flight to Europe I read one of the first IBM 360 manuals and immediately sensed that this line of computers would enable me to realize all my ideas for MT. While demonstrating AUTOTRAN and TECHNOTRAN to European scientists, I devoted my free time to working on a new logic oriented toward the 360 line. Emulating a 360 on the 7090, I was one of the first to write programs for the 360. At that time everyone tried to see how much they could accomplish with the fewest instructions, and in this connection my binary search won particular recognition.

It was around that same time that the Applied Language Processing Advisory Committee (ALPAC) was formed under the aegis of the National Academy of Sciences to assess the state of MT. I was out of the country, but output from AUTOTRAN and TECHNOTRAN was obtained surreptitiously and included in the ALPAC exercise.

In May 1965 the German Science Foundation convened 10 of the country's leading linguists and 10 computer specialists to evaluate my ideas and planned approaches for SYSTRAN. The evaluation was positive and I received a grant. That same year I began studies in Bonn that led to my Ph.D. degree.

4. SYSTRAN, 1965-1987

At the end of 1965 I returned to California and submitted a proposal to the Rome Air Development Center for the development of SYSTRAN and was awarded funding. This and further contacts made it possible to start my own company in La Jolla. One of the most important steps at that time was the development of a special language for linguists which made it possible for them to communicate directly with the computer.

In January 1969 SYSTRAN was installed in the Foreign Technology Division at Wright-Patterson Air Force Base in Dayton, Ohio. From the beginning Systran translated 10 times faster and much better than its predecessor, IBM's Mark II. Yearly contracts during the period 1970-1973 allowed us to introduce many new features—among them, semantic categories and “LS units” to store lexical information for resolving ambiguities.

SYSTRAN's next major project was the Apollo-Soyuz mission. Under a contract from NASA, in two years we improved the Russian-English system and developed English-Russian as well—our first English source combination. In 1975 SYSTRAN had been demonstrated extensively in Europe, and as a result, in 1976 a contract was awarded by the European Commission, which Loll Rolling is going to talk about. Also in 1976 came the XEROX contract for multiple targets from an English source. By that time SYSTRAN was being developed by two La Jolla companies, LATSEC, for the contracts with FTD and NASA, and World Translation Center Inc., for contracts outside the U.S. Government.

While it is beyond the scope of this paper to report my entire history with SYSTRAN, let me just mention that in 1987, after 16 years of pendency, I was granted a U.S. patent for SYSTRAN, the first patent for machine translation as far as I am aware. Two years earlier I had sold both companies to the Gachot family in France.