

## ENGLISH TRANSLATION

I would like to thank my friends and colleagues who invited me to give this talk.

In the proceedings, you will find my paper which only has historical value: it is the preface to the book in which the contributions to the Introductory Seminar on Computational Linguistics are collected. The seminar was held in 1982 at the Computing Center of Ca' Foscari with the participation of colleagues from Pisa who were interested in CL. The most interesting contribution for what I will say today was that of Zampolli, who did not want to put it on paper - he preferred speaking over writing. I report the essence of his words in the preface: it is necessary to distinguish between those who only do Quantitative Linguistics and those who do Computational Linguistics; the two are not interchangeable. At that time, linguistic resources were not yet discussed. Today I would like to say that we no longer do either: we do language engineering, but I will talk about this later.

What was being done in Venice at that time: both [Quantitative Linguistics and Computational Linguistics] in the '70s and '80s, but mainly Computational Linguistics. In particular, the specialties that distinguished it were Applications for Language Teaching, Applications for Literary and Textual Analysis, and Applications for TTS or text-to-speech. As for CL, apart from PROSODY, Venice implemented generative algorithms for morphological and syntactic analysis - parsers - in the first half of the '80s, and semantic analysis and reasoning algorithms in the second half.

And, here, it is worth going over the lesson one more time:

Definition 1, CL is the scientific study of human language, aiming to provide models of linguistic phenomena through the creation of algorithms that require several stages of trial and error;

then the theoretical purposes of CL, such as the implementation of syntactic and semantic theories in computationally manageable forms, the discovery of linguistic analysis techniques that combine distributional properties with structural ones - paradigmatic to syntagmatic, up to cognitive and neurological models of how the brain functions when using the faculty of language.

But let's go back to Venice and to the first half of the '70s when, during my stay in Australia for my Ph.D., I realized that I had to learn how to use a computer if I wanted to conduct an in-depth study of the complete works of the poet I was working on.

Upon returning to Italy, I learned FORTRAN, which was needed to communicate with the Cyber CDC, the mainframe located at CINECA in Bologna. However, this was not enough: I needed the linguistic resource, i.e., the text of the Collected Poems in digital form, which I discovered consisted of about 62,000 words.

I managed to obtain funding from the CNR with research under the evocative title Stilostatistical Analysis, etc., for two years, 1976/77 and 1977/78. This allowed me to hire people who worked at the Computing Center and the Linguistic Center and thus achieve the much-desired result. I had two texts in digital form: a book of Scientific English readings - which I will discuss later - and the Collected Poems of my poet.

So, I completed the quantitative analysis of my poet's corpus and had the book printed in 1979, which contained a chapter with many statistical tables and a series of Gaussians plotting the distribution of personal pronouns in the six partitions into which the corpus was divided. The quantitative work was then published in 1980 in the LASLA Journal. The plotter was located at the Faculty of Science, or rather Chemistry - as it was called -, and a colleague of mine in Physics made the Gaussians for me.

This is the cover of the second edition of the book that appeared a few months ago with a new chapter on Quantitative Linguistics and another chapter on Computational Analysis, which could not be done at that time.

I taught Scientific English as an Adjunct Professor without compensation - you heard right. Around the same time, I created a grammar of technical English derived from electronic excerpts, concordances, frequency lists, inverse orderings, etc. All examples and the enclosed dictionary were taken from scientific English. The resulting booklet was presented at the Computerized Lemmatization Seminar at the Index Tomisticus, invited by Father Busa, who at that time was in Venice with the Jesuits because there was the IBM Scientific Center in Venice, where he conducted his research and where I went several times. On that occasion, I met Zampolli and other important figures who were present.

The book contained a small chapter explaining how the course was created - here you can see the flowchart with the sequence of operations done by hand... there was no graphic software. Along with the quantitative work, I was also developing my entry into the world of theoretical and computational linguistics through the area most congenial to me, that is, prosody, the science closest to my activity as a musician in the '60s. Prosody encompassed everything that constitutes linguistic science, from phonology to pragmatics. Initially, I was focused on phonology, phonetics, even experimental phonetics with its part of acoustics. I used this information, which I found in journals and books at the CNR Phonetics Center in Padua, where I went several times a week, to build a program for the automatic assignment of word stress in Italian.

My work came to the attention of colleagues in Electronic Engineering, one of whom, Antonio Mian, invited me to the Computational Sonology Center where electronic music was made, but where there was an Italian speech synthesizer that spoke with a female voice - of which I have recordings. Guided by an IBM mainframe, it spoke one word at a time. At that moment, my collaboration with Padua Engineering began, lasting until 1986, for the creation of an Italian TTS and its transfer onto a motherboard.

I was asked to create a program that would allow the synthesizer to pronounce a sentence continuously.

The PROSO program was ready in 1980 and worked perfectly. The collaboration was funded by a CNR project aimed at Biomedical and Health Technologies, with the goal of creating a system on a motherboard for people with hearing loss.

PROSO contained the phonological and prosodic rules it was composed of, which were illustrated in an article published in 1981 in the Accademia della Crusca journal. I presented it

at my first international conference, the FASE, held in Venice. There, I met not only all the legendary figures of acoustic and experimental phonetics like Gunnar Fant, but also researchers from the CSELT in Turin, the research center of STET, now TELECOM.

The director summoned me to Turin, and they were also working on Italian TTS and were interested in my system, now called PROSO. It was decided to initiate engineering theses under my direction, with the aim of making the code more efficient and transforming it into tabular form. One of the thesis students was Maurizio Omologo, who is well-known by our colleagues from IRST - now working at Amazon.

The CL laboratory in Venice was now spread over several locations but was about to be concentrated in Venice. The PROSO system was presented at various international conferences, including the first EACL in Pisa and ICASSP in San Diego. But perhaps the most important moment was the invited presentation as the Italian representative at Doug Applet's conference held at Stanford University.

In Italy, the CL lab became a reference point for creating linguistic resources usable in various contexts. Ca' Foscari established agreements with IRST - where Oliviero Stock was, with the FUB - the Ugo Bordononi Foundation in Rome with Andrea Paoloni, and especially with CSELT in Turin.

Syntax and semantics were entering the PROSO program. In the '80s, there was a revolution in syntactic theories, and West Coast theories were born in California, including LFG by Joan Bresnan, which I immediately started following. I had an epistolary exchange with Bresnan and then met her at XEROX Park to present my theory on syntactic closure. Bresnan was later an invited speaker in Venice, which she loved, at the TAG conference that Joshi asked me to organize in the early 2000s.

The TTS work focused on analyzing texts without vocabulary limits was enriched with a morphological analyzer based on roots and morphemes, a small tagger, and a cascade parser organized around a context-free rule chunker. Giorgio Satta worked on this in his thesis completed in Padua but implemented in Venice at the Computing Center.

The ongoing research results for PROSO were presented at various international conferences and journals and as an invited talks in the UK, where I was invited by Geoffrey Leech and in the USA at Stanford University, where I was invited as the Italian representative of CL by Doug Applet. I tell this anecdote about Stanford because when I arrived, I realized that the event had been organized by Zampolli without he telling me, obviously. Zampolli was impressed by the demo I did at the first EACL conference held in Pisa and came to tell me. However, he then took this initiative [of inviting me in the USA] and, without knowing anything, for Stanford I prepared a talk on syntax entering PROSO. When I finished, he came up to me and whispered that they expected me to talk about the TTS system, not syntax!

The presentations in the USA had an important result: the request from the second largest computer company, Digital EQ, to bring their TTS system, DecTalk, into Italian. This project marked the beginning of the Computational Linguistics Laboratory in the Ca' Garzoni and Moro building.

Ca' Foscari received funding of 200 million, a Vaxstation II, and a paid programmer - one of my engineering graduates from Padua. But most importantly, I was finally granted a room where the linguistics students of that time worked: Cardinaletti, Giusti, Brugè - who are now the linguistics professors at Ca' Foscari - and Paola Merlo, who now works in Geneva. Merlo was working on a parser written in C language, and at that point, we all started using C.

At the LLC, the PROSO system was reworked, now in C language, but most notably, the first international TREEBANK was created, annotating the continuous forms on which the text was printed with the constituent syntactic structure. This structure was then used to create statistics useful for the parser.

The first subcategorized lexicon of Italian was also created at the LLC, annotating the 7,000 most frequent lemmas - nouns, verbs, and adjectives - using an interface written in C.

In 1986, Dario Bianchi from the Faculty of Engineering at the University of Parma came to Venice for a collaboration on a machine translation project. A new chapter of the LLC began thanks to the scientific contribution he brought, ranging from the use of Prolog for the Italian parser - which later became the complete GETARUN system - to the need to combine reasoning with semantics, implemented using the KL-ONE system made available by the University of Berlin.

From 1987 to 1991, the LLC participated in the European EUREKA-PROMETHEUS project, Pro-Art, Man-Machine Interface, which saw for the first time the participation of all Italian units active in the field of Computational Linguistics. The task of the Venice Unit was to produce an algorithm capable of generating spoken dialogues aimed at communicating with the co-pilot, in the domains of traffic, tourism, and weather forecasts. For this purpose, a detailed analysis of a corpus of transcribed spoken language was conducted by Laura Brugè. The language generation algorithm - as well as the Situation Semantics module in the Discourse Model - was created by a student of mine who later became an excellent scholar, Emanuele Pianta, unfortunately, deceased prematurely.

In 1989, the complete system using KL-ONE was operational and was described in an important paper published in 1990 in "Computer and the Humanities." In the early '90s, Ca' Foscari signed international agreements at the European level that allowed researchers from countries like Romania to come to work in Venice with funding. We received four computer science researchers-programmers from Iasi, one of whom was Dan Cristea. As a member of the Board of Directors, I proposed and succeeded in getting funding for a project called SLIM, which aimed to create a system for self-learning English which used speech recognition and synthesis. This was a very innovative system for that time and several demonstrations of it were provided at various international conferences from 1995 onwards. The LLC team now consisted of more than 12 people, 5 native speakers for English and French, 3 programmers, and four linguists.

In the early '90s, publications increased significantly and ranged from speech and experimental phonetics to the analysis of quantifiers and the implementation of an algorithm for their recognition according to the schemes proposed by LFG, to the analysis of pronouns and discourse structure.

To conclude, here is the list of units involved in Computational Linguistics in Italy compiled by AIIA and published in a booklet. Also, here is the list of some people who graduated or conducted research for their thesis at the LLC under my direction. Then a list of the 11 international/national conferences organized by the LLC over time. And now, I return to the fundamental question: does Computational Linguistics still exist? We are witnessing the triumph of LLMs and Deep NNs. But if it is dead, we wish a long life to Computational Linguistics because it is a passion, and I wish you all to be very passionate.