ANTONIO ZAMPOLLI

A life for Computational Linguistics

Antonio Zampolli has been one of the pioneers of Computational Linguistics at the international level and thanks to his countless initiatives Computational Linguistics was born, developed and grounded in Italy.

After graduating in Statistical Linguistics, in 1960 Zampolli started working with Father R. Busa S.J. at the "Centro per l'Automazione dell'Analisi Linguistica (CAAL)" of Gallarate. The work consisted in the electronic processing of St. Thomas of Aquino's works (about 10 millions of occurrences), indeed an exceptional exploit for that time.

In 1969 the Direction of CNUCE founded a Linguistic Division and Zampolli became its Director. At the same time he started teaching his first course of Computational Linguistics at the University of Pisa (the first and for many years the only one in Italy). In the same year, a project in collaboration with the Italian Parliament was launched: its aim was developing *information retrieval* procedures on legal texts, by means of creating and using a very innovative and futuristic "machine dictionary" with phonetic, morpho-syntactic and semantic information. This prototype has later been the starting point for important European projects, such as ACQUILEX and ItalWordNet.

In the '70s, many international activities came into being. The famous Pisa Summer Schools (*International Summer Schools in Computational and Mathematical Linguistics*) have been 'invented' and organised by Zampolli, and at that time they were not as fashionable as nowadays. He gathered in Pisa well-know international scholars representing the most innovative directions of research: among the others, M. Kay, B. Partee, B. Quemada, J. Lyons, C. Fillmore, E. Keenan, M. Gross, D.G. Hays, Ch. Muller, T. Winograd, I. Mel'cuk, Y. Wilks, E. Bach, W.A. Woods, M.A.K. Halliday. These Summer Schools got together European and American researchers and contributed to crucial developments in the state-of-the-art of Computational Linguistics, Linguistics, Artificial Intelligence, Lexicology and Computational Lexicography. Zampolli remembered with pleasure that the *Lexical Functional Grammar* was born on a tower in San Gimignano, with J. Bresnan and R. Kaplan drawing trees (syntactic ones, obviously) on a piece of paper rather than enjoying the wonderful view! A whole generation of computational linguists is in practice a product of these Summer Schools, which are enthusiastically remembered still today.

In 1980 Zampolli founded in Pisa the Institute of Computational Linguistics (ILC) of the National Research Council (CNR) and he has always been its Director. Thanks to Zampolli and to ILC, the CNR has been the first (and for a long period, the sole) public body

supporting Computational Linguistics in Italy. Zampolli played an essential role in promoting and participating to national and international "infrastructures for co-ordination and planning" (ELRA, ELSNET, ICCL, TAL Forum, ICCWLRE, TEI, EURALEX, ALLC, etc.), by means of which he often guided the international community towards those he considered the main priorities. ILC always played a crucial role in these strategic activities and it is now one of the major centers of the field at an international level.

He was a man of great 'visions', as few have been: he was able of anticipating and creating the future that other would have seen only much later, he always had new intuitions and fighted for them to become a reality. His exceptional ability in mixing together people from different fields and having different interests (see the famous Grosseto Workshop of 1986, where some of the most important initiatives for recommendations and best practices in the field of LRs were defined in front of a breakfast table) and his overwhelming capacity to convey his enthusiasm and his passion, will always be remembered.

For all of us he has been a guide, a 'maestro', a *living legend* (as he has been defined once), an extraordinary man not only at the professional level but especially as a person, for his sympathy and sense of humour, his love for the mountains, for music…

Just a few examples of his ideas, chosen among those he loved the most.

He 'invented' the field of Language Resources (LRs) and supported the idea of reusable LRs, in times when many linguists considered as a shame to talk about data. Nowadays the so-called *data-driven approach* is the paradigm which underlies Computational Linguistics and it is based on the use of large amounts of linguistic data with its relevant descriptions. He introduced the term "language resources", feeling the need to stress the infrastructural role of these components, similar to the role of basic resources (i.e. aqueducts, electricity energy, roads) for the industrial development of a country. Nowadays, talking about LRs is taken for granted but someone's endeavour for founding and promoting the field has been necessary.

The Workshop *On Automating the Lexicon* that Zampolli organised with D. Walker, J. Sager, L. Rolling and N. Calzolari in May 1986, is by now acknowledged by everybody in the field as the starting point for the creation of the field of LRs. The final recommendations coming out from the Workshop had as results several European projects (ACQUILEX, ET-7, MULTILEX, MULTEXT, GENELEX, DELIS, etc.) and many research and organisational activities.

The lack of appropriate LRs is nowadays unanimously recognised as one of the main obstacles to the success of R&D activities in Natural Language Processing. Various American governmental agencies (NSF, DARPA, NSA, etc.) and the European Commission have stated the essential need for LRs and standards: their production needs co-operation in terms of capacities, fundings, people, and for this reason Zampolli has always promoted international (EC and joint EC-NSF projects) and national (Italian Ministry of University and Research) activities by means of which various public and private subjects, both in Italy and abroad (Europe, US, Japan and other Asian countries), could co-operate. The infrastructural role of LRs in NLP implies that creation and development of basic LRs for a language are supported with public funding, as it is currently happening in several European countries where national projects are going on.

As a matter of fact, Zampolli committed himself to promoting collaboration between the different communities of ComputationalLinguistics.

Many activities for standardisation and creation of LRs (NERC, RELATOR, EAGLES, PAROLE, SIMPLE, etc.) have been 'invented' by him and finally led to his last "creature",

the European project ENABLER, which has strong strategic and political implications besides the scientific ones.

The project has built a European Network of the various national projects dealing with NLP, in order to ensure co-ordination among these national initiatives and harmonisation of the LRs produced, having in view also future common projects (e.g. for multilingual and multimodal resources).

Antonio Zampolli devised the International Conference on Language Resources and Evaluation (LREC), which was held in Granada in 1998, Athens in 2000, Las Palmas de Gran Canaria in 2002 and Lisbon in 2004, which is probably by now the most important conference in the field. Zampolli perceived the need of a conference able of gathering together LRs and evaluation experts, as no other international conference appropriately covered these fields. LREC has become the seat where the two communities of 'speech' and 'written' – which usually do not have occasions for meeting and exchanging information - get together: this is probably the best example of Zampolli's overall point of view on our field.

The more appropriate tribute to his memory would be organising LREC 2006 in Italy.

A short biography

Antonio Zampolli was full Professor of Computational Linguistics at the University of Pisa, and founder (1968) and Director of the Linguistic Division of CNUCE, transformed in 1978 into the Institute of Computational Linguistics of the National Research Council, Pisa.

His main research interests were computational lexicology and lexicography, literary and linguistic text analysis, reusability of lexical resources, text processing, quantitative linguistics, formal grammars and parsers, multimodality, multilinguality, standards for literary and linguistic data processing, machine translation, computer-assisted language teaching, modalities and strategies for international co-operation.

He has been, among others, president of ALLC, EURALEX, GILA; vice-president of ACH, ICCL, AILA, CETIL; founder of ELRA, president of the ELRA Board and chair of the International Conference on Language Resources and Evaluation (LREC); director of the Pisa International Summer School for Literary and Linguistic Computing; member of the Steering Committee of TEI; member of the Linguistic Data Consortium Advisory Board; founder and member of the ELSNET Management Board; member of several committees of experts for the EC; representative for Literary and Linguistic Computing in the Permanent Steering Committee of the European Science Foundation and in the Executive Committee of ACL; co-ordinator of several European projects, mainly for the production and standardisation of language resources and for international co-operation; co-ordinator of two Italian national projects: "National infrastructure for the linguistic resources in the field of automatic processing of written and oral natural language" and "Computational Linguistics: mono and multilingual researches"; responsible for the software development of the national project for the Italian library of digital resources for the humanities (CIBIT); member of the Editorial Board of several scientific journals such as "Computer and the Humanities", Histoire et Mesure", Journal of Applied Linguistics", etc.; director of the Italian journal "Linguistica Computazionale"; organizer and chairman of several international Conferences.