

LAUDATIO

for Professor PhD **Nancy IDE**, Vassar College, United States of America,
on the occasion of awarding the title of *Honorary Professor* of Alexandru Ioan
Cuza University of Iași

25 March 2010

www.uaic.ro

Laudatio

In honour of
Professor PhD **Nancy IDE**,
Vassar College, United States of America

Nancy Ide holds degrees in Neuropsychology, English Literature, Linguistics and Computer Science. Her dissertation, defended in 1982, used statistical methods to determine patterns of textual lexical-semantic cohesion almost a decade before statistical approaches to text analysis were embraced by the computational linguistics community. From 1985 to 1995, Nancy Ide served as a president of the Association for Computers and the Humanities; in 1996 she took over the editorship of the *Association's journal, Computers and the Humanities*. A few years later, Professor Ide revamped the journal to reflect changes in the field and changed its title to *Language Resources and Evaluation*. She continued as a co-editor-in-chief of the journal to this time. In 1995, she co-founded the Springer book series *Text, Speech and Language Technology*, which contains over 30 volumes on state-of-the art methods in computational linguistic research.

In 1987 Professor Ide founded the Text Encoding Initiative (TEI) and served as the chair of its Steering Committee periodically until 1996. She also worked extensively within several TEI working groups, most notably the Dictionary working group, within which she led the development of a standard for representing lexical information, which

has served as the basis for more extensive standards for lexical and terminological data developed later in the International Standards Organization (ISO). The TEI itself, which is still actively maintained and developed and remains the *de facto* standard for encoding humanities texts after 22 years, significantly influenced the development of the web standard Extensible Mark-up Language, XML.

Professor Ide has published extensively in the areas of word sense disambiguation and lexical semantics. In the early 1990's she worked with extracting semantic relations from machine-readable dictionaries, and developed a neural network representation for use in WSD. She was also among the first to explore and evaluate the use of cross-lingual information for WSD. She has co-authored seminal papers in the field, including the 1998 introduction to a special issue of *Computational Linguistics* on WSD, which remains one of the most cited surveys in the field to date, and a more recent book chapter co-authored with Yorick Wilks assessing the state and future of WSD research. More recently, she developed a method and metric for determining semantic similarity to be used for bootstrapping lists of semantically-related words as well as determining inter-sense similarities.

Professor Ide spent several years doing research in Europe, and from 1993 to 1997 served as a co-project leader for the MULTEXT Project, which was at the time the largest EU project ever funded in the field of computational linguistics. The MULTEXT Project was the first to develop large-scale lexical and corpus resources and tools for multiple languages in a common format. Several cornerstones of current work

on lexicon and corpus content and design as well as tool architectures were laid in the MULTEXT Project, including the pipeline architecture for NLP tools and the concept of standoff annotation for representing linguistic information added to text. Within the MULTEXT and EAGLES projects, Professor Ide was responsible for developing the Corpus Encoding Standard (later, the XML Corpus Encoding Standard or XCES), which is still in widespread use for representing linguistic corpora and their annotations. In 1994 Professor Ide spearheaded and led the follow-on European project MULTEXT-EAST, which applied the methods and tools developed in MULTEXT to six Eastern European languages. The famous “Orwell 1984 Corpus”, which now includes parallel, linked annotated versions in over 20 languages, was one of the most important products of that project. It is still heavily used as one of the few massively parallel corpora in existence. For instance, the Romanian edition of the corpus includes XML markings for part-of-speech and morphology, sentence segmentation, noun phrases, co-reference and, partly, discourse structure and *veins theory* (which Nancy Ide co-authored).

Professor Ide has been active in the development of standards for language resource representation and management for over 20 years, starting with her founding the TEI and development of the Corpus Encoding Standard. She also worked on several standards within the EU EAGLES Project in the early and mid-1990’s. In 2000 she joined the team of experts in the International Standards Organization (ISO) TC 37 SC4 Language Resource Management and serves as the Convener for the Working Group on basic descriptors and mechanisms for language resources. In this context she has recently developed the

Linguistic Annotation Framework (LAF) and its adherent Graph Annotation Format (GrAF), a powerful common representation formalism for linguistic data and associated linguistic information. Professor Ide is currently co-PI on the project “Sustainable Interoperability for Language Technology” (SILT), funded by the US National Science Foundation to work toward interoperability of language resources, which is a parallel “sister” project to the EU-funded FLaReNet project. In this context Professor Ide is working on the development of a software infrastructure to enable resource interoperability, and working to provide open access to linguistic data of all kinds.

In 1998 Professor Ide initiated a project to develop an American National Corpus, comparable to the British National Corpus, widely used in computational linguistics research. This on-going project has been funded for several years by the US National Science Foundation and aims at the development and enhancement of automatic annotation methods via machine learning, and the development of representation and access mechanisms for large-scale linguistic data and annotations. The project involves both the FrameNet and WordNet teams concerned with the harmonization of these two resources, and the improvement of inter-annotator agreement studies for sense annotation.

Between 2002 and 2004, Professor Ide worked on another NSF-funded project to exploit semantic web technologies for the study of historical documents. The project classified texts from the Franklin D. Roosevelt Library regarding Japanese-American relations in the six months prior

to the bombing of Pearl Harbour in 1941. The texts were heavily annotated and supported by an extensive ontology of entities and events, which enabled unprecedented access to the information in these texts and enabled the discovery of previously unrecognized relations and characteristics. Within this project, Professor Ide developed a model for “historical ontology” that adjusts its representation of relations among ontological elements according to the time frame referenced by a user query.

In 2007, Professor Ide became the president of the Association for Computational Linguistic Special Interest Group on Annotation (SIGANN). SIGANN sponsors the Linguistic Annotation Workshop, which Professor Ide has organized for the past four years.

Professor Ide has also organized and/or chaired numerous workshops over the past decade on a wide variety of topics ranging from architectures for language processing tools, frameworks and methods for linguistic annotation to word sense disambiguation and psycholinguistic approaches to sense distinction. She also initiated the Conference on Global Interoperability for Language Resources, held in Hong Kong every two years.

Professor Ide has been a member of the Computer Science Faculty at Vassar College since 1982, where she teaches advanced level courses in Theory of Computation and Language Theory and Automata, Compiler Design and Computational Linguistics. In the late 1980’s, as a member of a small consortium of representatives of selective liberal arts colleges, she contributed to the development of a Computer

Science curriculum suited for such institutions that has since served as the foundation for Computer Science curricula in liberal arts colleges throughout the US. She became the chair of the Computer Science Program, which at the time, in 1989, offered only a degree in Mathematics and Computer Science. The Department of Computer Science was founded at Vassar College under her leadership in 1991, offering a stand-alone Computer Science degree for the first time. She served as the chair of the Department for 20 years until 2009, supporting the development of the program into what is widely recognized as one of the best Computer Science programs in a liberal arts college in the US.

The year 1997... We are in Tuşnad waiting for the invited professors to arrive from Otopeni to the EUROLAN Summer School on Natural Language Processing - the third one in a series of international schools set up in 1993 by the Faculty of Computer Science of our University and held biennially since then. For the first time the organizers have decided to move EUROLAN outside Iaşi, in a place where the local configuration would make professors and students interact closely. Apart from few Romanian teachers (Dan Tufiş, Liviu Ciortuz and Dan Cristea) there also participated Paula Monachesi – from Utrecht, Massimo Poesio – from Essex, Nicoletta Canzolari – from Pisa, John Sinclair from the Tuscan Word Centre in Montecatini – the director of the Collins series of dictionaries, inventor of corpus linguistics, Professor Aravind Joshi from Philadelphia – inventor of tree adjoining grammars and centering and Nancy Ide from Vassar, Poughkeepsie – the already famous inventor of TEI. As always, the school went on for two intensive weeks, during which lectures and practical sessions were

given to PhD and gifted undergraduate students from Romania and other six countries, in front of computers (connected to the Internet, in the small Carpathian resort, after an extraordinary effort made by the organizers). Nancy Ide was touched by the friendly atmosphere; she took the time to discuss with students and even initiate collaboration projects with the organizers. Professor Ide was so impressed with the idea of this series of schools that she has become a member of the Program Committee ever since, contributing to making this event increasingly important in Europe and worldwide. Since the Tuşnad EUROLAN School in 1997, she came back to Romania six more times (four times in Iaşi), getting involved in the organization of these schools, sharing her experience with us and our students, teaching and doing research.

Nancy Ide has greatly contributed to the development of related fields to computational linguistics and natural language processing, so that her name will always be one of the most prominent. She has been developing close scientific relations with Alexandru Ioan Cuza University of Iaşi. This year in June, the eighth generation of master students will be graduating our Faculty of Computer Science in the field of Computational Linguistics. Many of them took their PhD at our Faculty or abroad (I have counted 17 PhD graduates and 12 PhD candidates, but I am sure there are more). Among many other native Romanian researchers, these people, now spreads all over the world perform remarkable scientific work in our field. If you ask any of them who Nancy Ide is, you will hear a different story. They met her during their studies at our Faculty, but afterwards they met her again in large conferences, in important scientific gatherings, in amphitheatres, in

international projects, or on the Web, by reading her papers. One way or another, Professor Ide is continuously present in this scientific field.

It is a privilege to offer Professor **Nancy Ide** the distinction of *Honorary Professor* of Alexandru Ioan Cuza University of Iași in recognition of both her scientific merits and the long-term collaboration with our University and other Romanian higher-education institutions.

The members of the Committee wish her a long and active life!

Members of the Awarding Commission

President,

Professor PhD **Vasile IȘAN**

Rector of Alexandru Ioan Cuza University of Iași

Members:

Professor PhD **Dan CRISTEA**

Vice-dean of the Faculty of Computer Science, Alexandru Ioan Cuza University of Iași

Professor PhD **Horia-Nicolai TEODORESCU**

Vice-rector of Gheorghe Asachi Technical University of Iași

Director of the Institute for Computer Science of the Romanian Academy – Iași branch

Correspondent member of the Romanian Academy

Professor PhD **Eugen MUNTEANU**

Faculty of Letters of Alexandru Ioan Cuza University of Iași

Director of Alexandru Philippide Institute for Romanian Philology of the Romanian Academy – Iași branch

Director of the Center for Biblical-Philological Studies “Monumenta Linguae Dacoromanorum” of Alexandru Ioan Cuza University of Iași

Iași, 25 March 2010