

## Preface

This volume<sup>1</sup> is an attempt to compile and illustrate all the open lines of research within the UNL initiative. The included papers constitute a selection of the most significant papers presented in several international conferences and workshops during the last four years that served as a meeting point for the UNL consortium. In general, papers are not restricted to UNL although they are clearly predominant; they clearly illustrate the wideness and flexibility of this UNL initiative, launched by the United Nations aiming at the elimination of linguistic barriers.

Since the starting of the UNL project in 1996, the participants in the project from initially 15 languages have made substantial progress in technical matters and the organizational aspects involved as well. This book attempts to provide a survey on the approaches and theoretical studies around UNL, since research on UNL is not only devoted to studies on interlinguas, MT or any NLP related issues, the intrinsic properties of UNL make it a firm candidate to support a wide variety of applications ranging from e-learning platforms to management of multilingual document bases. Such a variety of applications, their theoretical basis and subsequent methodological inquiries are at core of this volume.

### What is UNL? Its motivation and purpose

The emerging needs and use of Internet for cultural and educational dissemination and commercial expansion of the peoples collide with linguistic diversity, which in principle diminishes the potential of Internet as a vehicle of knowledge for everybody. Aware of this problem, the Institute of Advanced Studies of the University of the United Nations University (UNU/IAS) launched the UNL project in 1996 with the initial participation of 15 languages (German, Arab, Chinese, Spanish, French, Hindi, Indonesian, English, Italian, Japanese, Latvian, Mongol, Portuguese, Russian, Thai). In short, the UNL Programme was initially conceived to support multilingual services in Internet being an alternative to classical machine translation systems.

The UNL system revolves around a unique artificial language (Universal Networking Language) that pretends to capture the meaning of written documents. This language is based on the representation of concepts and its relations. The definition of this language has been possible thanks to the collaboration of more than one hundred people, prestigious researchers, and scientists of all around the world, that worked during the first three years of the project to produce a final version of the UNL specifications<sup>2</sup>.

---

<sup>1</sup> Earlier versions of the papers at pages 10, 109, 117, 125, 145, 215, 230, 254, 268, 276, 309, 347, 359, 370, 380 have been published in the Proceedings of Convergences'03, Alexandria, Egypt. Earlier versions of the papers at pages 3, 10, 27, 38, 101, 261, 326 have been published in the Proceedings of LREC-2002.

<sup>2</sup> UNL Specifications, v.3.1 available at <http://www.unl.org/unlsys/unl/UNL%20Specifications.htm>

## **The UNL organization**

The UNL initiative has often been regarded as “hidden organization”. The first years of the project (1996-2000) were devoted to the definition of the interlingua and to the development of the essential components required to undertake the basic process in UNL (mainly dictionaries and language generators). During this period, the organization was closed and limited to a number of participants, because of the need to define the specifications of the language.

By the end of this period, the UNL project reached a significant degree of quality in the development of components, linguistic resources and technical specifications; and the specifications were finally produced. Once the specifications were finished, they were made public and accessible to all the international community, so that collaboration and *participation in this initiative is completely open*.

As a consequence of this degree of development, the Board of the United Nations University, in its fifth meeting in 2000, agreed on the creation of a new institution responsible for the organization and promotion of the UNL in the future under the umbrellas of the United Nations. This new entity was the UNDL Foundation, with headquarters in Geneva<sup>3</sup>. The development of the components of different languages was assigned to the so-called Language Centres, constituted by the initial teams in each country in charge of the development of the essential components of UNL.

The year 2004 represents a turning point in the evolution of UNL for two main reasons. First, it is the year where a new period coordinated and fostered by the Language Centres starts for the debugging, updating and expansion of linguistic resources and developed components of their representative languages, in order to respond to the institutional and marketable challenges at a pre-competitive level in the support of multilingual services. Second, it is the year where the UNL patent has been approved in USA for the UN (US Patent No. 6704700 B1, March 2004). It has been the first software patent of the United Nations.

## **Open nature and scientific dissemination of UNL**

Since 2002, an open annual conference around the convergence of language, culture and knowledge is being held as a meeting point for researchers, politicians, linguists and engineers. The most recent edition of this open conference was *Convergences'03*, held in Alexandria, Egypt. The most significant papers from this conference have selected and included in this volume. Additionally, an international workshop on UNL and Interlinguas was organized in 2002 (International Workshop on UNL, other Interlinguas and their Applications, held at Las Palmas de Gran Canaria, May, 2002), papers from this workshop are also compiled in this volume. Finally, we include the papers of the current edition of the UNL Workshop, held in Mexico D.F, February, 2005.

These conferences and workshops try to be a forum where all the interested people in this initiative find a vehicle for communication and exchange of knowledge. The

---

<sup>3</sup> [www.undl.org](http://www.undl.org)

UNL is a great initiative that could never succeed and advance if the number of participants is limited to the initial ones. The heterogeneity of the authors and languages involved in this selection of papers shows the open nature of UNL.

## Research on UNL: Current Trends

Apart from the mere applied studies of UNL, there is a current important trend on theoretical studies of UNL, even though there is a final version of the specifications of the language, dating to July 2003.

The rationale for such theoretical research is the need for standardization and homogenization on the use of the Interlingua both at the applied level and at the theoretical level. The UNL Specifications turned out to be subject to different personal interpretations, thus *creating own UNL dialects*. This is not desirable for an interlingua, that claims to be language independent and that, in fact, turned out to be “person-dependent”. For this reason, it is important and desirable to foment theoretical studies on UNL, both from the linguistic point of view and the knowledge point of view.

From a scientific point of view, UNL follows the approach of the concept of Interlingua, as an “artificial” language aiming at the neutral representation of linguistic meaning. In this sense its roots can be sought in the tradition of MT interlinguas and in the tradition of Knowledge Representation formalisms.

When viewed as an interlingua, UNL differs from some of its predecessors and current Interlinguas in the generality of appliance, that is, UNL is not restricted to a number of languages or to a given domain. Thus, its design pretended to show the highest degree of language independence while retaining natural language expressiveness in order to support multilingual generation tasks.

Of course, the staging of UNL is such a general enterprise that requires research and efforts. This process can be divided into several periods:

- Creation of deconversion and enconversion modules, (see Part 3) that is, development of the basic tools to undertake the basic architecture of the UNL system (enconversion and generation), along with dictionaries. Although basic, it is *conditio sine qua non* to have powerful generation systems. This a fruitful trend in the UNL consortium, with three different approaches:
  1. The official one: those using a common engine provided by the UNL Center.
  2. The integrative ones: those that have integrated UNL into pre-existing MT systems, following the transfer-based architecture, showing the flexibility of UNL with good results.
  3. The new ones: those that have noticed the drawbacks of the *official* components, and have decided to create new architectures for generation

It should be noticed that emphasis is put on the deconversion process, quantitatively proven by the number of papers devoted to generation. Teams usually develop generation systems, not so much enconversion systems, although the integrative usually includes both processes in UNL.

- Application of UNL in other contexts (see Part 3). Should UNL be considered as an interlingua, it can be applied in fields and tasks other than multilingual generation, being the main one Knowledge representation and Knowledge Management.
- Use of external lexical and ontological resources. It is important as well, and following the spirit of the integrative approaches, the use of external lexical resources such as Wordnet to enhance some of the processes of UNL, especially in the lexicographic part (see Part 3, also). This is also a trend and the philosophy of UNL: integration and complementation of resources is encouraged, rather than confrontation. And this is the spirit of the consortium and of every work in UNL.

From an engineering point of view, research is taken on:

- Creation of methodologies in the workflow.
- Standardization of UNL, integration of UNL into current standards.

Why such studies methodologies and standards? Because of the heterogeneity and diversity of the current consortium, it is needed such a process of standardization and methodologies, since the short and medium term objective of UNL is its staging in the market, where standards and methodologies are required in order to pursue higher productivity and quality. The areas of linguistic engineering together with knowledge engineering are claiming for such methodologies and processes of standardization.

## **The Future**

After some time developing components and systems to support the multilingual services, UNL researchers and new teams have discovered that the UNL could be support of other applications as crosslingual information retrieval, knowledge repositories, automatic building of ontologies from texts once represented in UNL and much more. UNL could be useful in new possible applications in areas where a common conceptual representation is needed, independent of any particular language. For doing it, new necessities emerge; particularly when putting together semantics and multilingualism. More theoretical studies are needed, along with the tuning up of resources and tools, the proper standardization of the interlingua and processes for enconverting and deconverting, and of course the integration and definition of the lexical component of UNL.

## The Structure of the Book

The volume is divided into four parts.

### Part 1. Introduction

This first part is an introduction to the language itself, and its purpose is to set up the reader in the UNL context. These introductory papers posit the general philosophy of the language (paper at page 3) and provide a general introduction to the language itself and to the context of multilingual generation, one of the main and most basic “applications” supported by UNL (paper at page 10).

### Part 2. Fundamentals

This part is dedicated to theoretical studies on UNL. As already said, UNL is mainly an interlingua. There are many aspects that have to be taken into account when designing an interlingua, such as its expressiveness, degree of language-independency, accuracy and formality of the language, etc. Most of these issues are covered in this part. Thus, the part opens up with an experiment on the common understandability of UNL by different humans and the admissible degree of indeterminacy and ambiguity in an Interlingua (paper at page 27). Pure theoretical studies on the universality of UNL and its adequacy from a representational and linguistic point of view follow (papers at pages 51 to 101). It has to be pointed out that this part is not exclusively devoted to UNL, but to the field of interlinguas in general (paper at page 38; paper at page 109).

All these papers point at the proper designs of the Interlingua. However, there is another important aspect worth of consideration in any artificial language, namely, the syntactic formalism of the formal language and its adequacy to the declared purpose. These topics are addressed in papers at page 117 and at page 125, where the emphasis is put on the syntactic properties of UNL expressions and its consequences to other issues such as analysis or proper deconversion. Finally, there is a (recurrent) thematic shift; UNL is not viewed as an interlingua to support linguistic tasks, but as a language for knowledge representation (papers at page 138 and at page 145).

These two sides of UNL (an interlingua to support linguistic tasks and a as knowledge representation language) determine the nature of the applications dealt with in Part 3.

### Part 3. Applications

The core applications of UNL are those that support the tasks of NL analysis and generation (*enconversion* and *deconversion* in the UNL jargon). When dealing with NLP tasks, the scene is quite heterogeneous: from the use of common generation tools provided by the UNL Center (as shown in papers at pages 215 and 241), to the integration of existing MT translation systems based on the transfer architecture to support

an Interlingua architecture (papers at pages 157 and 230). Other languages are supported with new tools, but differs in their configuration and architecture (maybe reflecting language variety, maybe reflecting different ways to support generation and of course, as an advanced over common tools, like Deco). Chinese, Brazilian Portuguese, Arabic or Armenia are example of this, where very different paradigms are illustrated in order to undertake the generation task (papers at pages 167, 175, 195, and 210, respectively).

Papers at pages 254 to 276 illustrate the development of workbenches to support the processes of edition, generation and training and with the creation of multilingual platforms within the UNL framework.

In parallel with the theoretical studies of Part 2, UNL also presents and applied dimension when conceived as a language for knowledge representation (papers at pages 337 and 359). These papers present the use of UNL as an extension (or complementation) to the expressiveness of standard languages such as XML (illustrated in papers at pages 300 and 309), as the communication language among agents, developed in paper at page 326, or as the support of case-based reasoning systems (paper at page 347). It is also remarkable the possibility of complementation and integration with other lexical and ontological resources such as WordNet (papers at pages 370 and 380) to the enhancement of the processes of knowledge acquisition and representation within the UNL context. Finally, paper at page 286 shows how to extend the expressivity of UNL in order to represent and formalize meaning coming for oral sources.

#### **Part 4. Methodologies**

Finally, the volume ends up with the methodological work. Methodologies target at the creation of methodologies to support multilingual services (papers at pages 395 and 413) and for the optimization of knowledge intensive tasks (paper at page 430). Needless to say, methodologies conforms an integral part of the UNL R+D activities, as long as productivity, quality and a real consolidation of UNL are pursued both at the scientific and commercial levels.

Mexico D.F, 16th February 2005

Jesús Cardeñosa  
Alexander Gelbukh  
Edmundo Tovar  
Editors