Facilitating Communication Between Languages and Cultures: a Computerized Interface and Knowledge Base

Claire-Lise Mottaz Jiang,[1] Gabriela Tissiani,[2]
Gilles Falquet,[1] Rodolfo Pinto da Luz [3]

[1] CUI, University of Geneva, Switzerland,
(Claire-Lise.Mottaz, Gilles.Falquet)@cui.unige.ch
[2] CNPq Researcher at CUI, University of Geneva, Switzerland
Gabriela.Tissiani@cui.unige.ch
[3] Instituto UNDL Brasil, Florianópolis, Brazil
Luz@undl.org.br

The Universal Networking Language (UNL) deals with communication, information, knowledge, language, epistemology, computer sciences, and related disciplines. This interdisciplinary endeavor calls for theoretical and applied research, which can result in a number of practical applications in most domains of human activities. Specially, it can help solving some of the most critical problems emerging from current globalization trends of markets and geopolitical interdependence among nations. This paper presents a project that aims to contribute with UNL KB (UNL Knowledge Base) theoretical and practical. The goal is to make possible people from various linguistic and cultural backgrounds to participate at UNL KB construction in a distributed environment.

1 Introduction

This paper presents a project that will be developed by the following partners: Information System Interfaces (ISI) Research Group at University of Geneva, the UNDL Foundation, and the United Nations Institute for Training and Research (UNITAR). This project is part of the Geneva International Academic Network Programme (GIAN). It involves creation of ontologies, for the Universal Networking Language (UNL) Knowledge Base (KB).

The project argues that the construction of these KB ontologies will contribute to the United Nations initiative of creating the multilingual infrastructure on UNL. Its infrastructure is meant to facilitate communication among natural languages on the Internet and includes development of a broad knowledge base from diverse linguistic sources and cultural backgrounds [10].

The UNL multilingual infrastructure is an interdisciplinary undertaking that involves both linguistic and engineering aspects. Its main components are (1) a formal, language-independent, non-ambiguous artificial language (UNL) and (2) a system that manages the interfaces between natural languages and the UNL over computer networks. The UNL itself comprises a vocabulary - a list of concepts, called "univer-

© J. Cardeñosa, A. Gelbukh, E. Tovar (Eds.)
Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 359–369.

sal words" represented in a language-independent way as character strings—and a knowledge base that explicit the relations between universal words (UNL KB).

The UNL KB construction raises several challenging problems, because of its particularities (size, high number of contributors, distributed environment, linguistic and cultural issues). Examples of problems to be tackled include:

- How to coordinate multiple, distributed (remote) contributors?
- How to deal with multilingual and multi-cultural issues in order to create a "global" knowledge base?
- Which infrastructure is needed to enable a distributed, asynchronous work and still end up with a coherent knowledge base?
- How to maintain the knowledge base to ensure its validity over time?

The overall goal of this project is to create a framework and tools to support the development and the evolution of the UNL knowledge base. This project includes both applied and theoretical research. As there is no known straightforward engineering solution to this set of problems, theoretical studies will be carried out to support a practical realization.

2 Related Work

The proposed project will use several research fields such as studies on knowledge bases, knowledge representations, and ontologies.

2.1 Knowledge Bases and Knowledge Representation Models

For the present project, we are particularly interested in works that propose models and languages for the representation of concept definitions. In the artificial intelligence field, bases containing knowledge about concepts are usually called "ontologies" or "terminological knowledge bases". Many languages and models have been devised to describe terminological knowledge. Formal terminological knowledge representation systems include for example KL-ONE, CLASSIC, LOOM, OIL, OCML, and OWL. These systems are based on first order logic, description logic, or on frame systems. Research in this area is growing fast to respond to the "semantic web" initiative of the W3C consortium. It focuses on engineering of formal ontologies (how to create and maintain them) and on ontologies use to integrate heterogeneous resources.

2.2 Ontology Engineering Tools

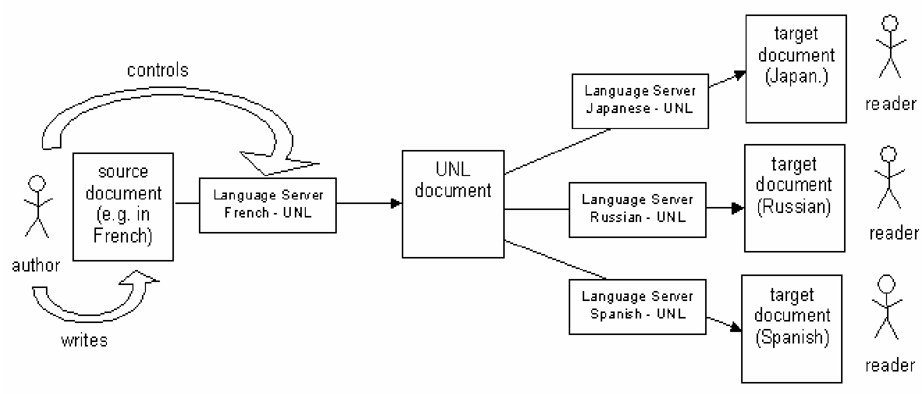
Since the mid 1990s, many ontology engineering systems have been developed (Protégé, OntoEdit, WebODE, etc.). These systems can be classified in two categories, depending on type of knowledge representation language they rely on:

- description logics based (for example Ontosaurus)
- frame based (Ontolingua, Protégé, OntoEdit, etc.)

Different presentation styles are used to display the content of the ontology. The first form is indented text, which is used in many HTML interfaces. In order to hide the formalism and provide user-friendly tools, graphical interfaces have been developed, such as node-link diagrams (graphs), tabular views, or hyperbolic trees. Graphs have been extensively used to visualize knowledge structures in artificial intelligence (semantic networks, conceptual graphs and database design (from entity-relationship diagram to object and class diagrams).

3 UNL System and the UNL KB

Today, a vast proportion of information available on the Internet is written in only a few languages, among which English ranks first. Computer translation systems have been used to provide users with a means to read information written in languages they do not understand. However, two important problems remain unsolved: 1) these systems usually function only with "dominant" languages, such as English, French, Spanish, Russian or Chinese; 2) moreover, they perform well only in specific conditions.



Fg. 1. UNL System.

The purpose of introducing the Universal Networking Language (UNL) in communication networks is to achieve accurate exchange of information among different languages. The UNL is an artificial language for computers, unlike Esperanto, which is also artificial, but for humans. In UNL, concepts and sentences are represented formally and non-ambiguously, through logical expressions. The UNL is language-independent; it provides the possibility to work at the semantic level, and enables the construction of a comprehensive "library" from various types of knowledge expressed in diverse indigenous sources.

The UNL multilingual infrastructure emerges from the convergence of linguistic and epistemic research with electronic and digital media, computers, and communication networks. The result comprises the following elements:

– the UNL specifications
– the Universal Word (UW) Dictionary

  Universal Words represent in logical expressions the meaning of the words in natural languages, normally represented in alphabet characters or ideograms. An UW denotes a concept.

– the UNL Knowledge Base (binary semantic relations between Universal Words).

The goal of this project is to theoretically study the distributed construction of very large knowledge bases and to provide a framework and tools to build the UNL KB.

3.1 How Does The UNL System Work?

In order to produce an equivalent UNL for a natural language document, one can use the UNL editor of his/her corresponding Language Server. This process is called “enconversion” and it can be either completely automatic, or interactive or completely manual. Finally, the UNL viewer is used by the reader of the document to "deconvert" the UNL text into his or her natural language, by using the UNL viewer of hi/her appropriate Language Server [10, 11].

3.2 The Structure of the UNL KB

The UNL KB structure and mechanism is based on a hierarchy formed of binary relations between UWs. Every UW must be defined in the KB and linked to the other related existing UWs [9].

The UNL Knowledge Base stores UWs that are inter-linked to each other by one of the relations present at UNL Specification. The UNL KB defines an UW by its relations with other existing UWs. Therefore, this requires also the designation of which level it should be situated as well as under which subordinate UW it should be set (on icl case). To create a new UW it is necessary to label the desired concept and also define its relation list (UW) [9] [10]. This list comprises of a relations set with other UWs required to define a concept such as a label. Subsequently, to build up the new UW in the UNL KB, it is indispensable to position it in the existing hierarchy. Finally, all new UW input in KB gate (web application that allows the access to the KB) must be homologated by the UNDL Foundation before its final inclusion into the KB.

As this whole process is text based, it requires a great human effort to manage it. Moreover, the specific view of the UNL expertise is required to update the KB. Even thus, it is not easy to manage these processes. It requires a great human effort to coordinate the multiple remote contributions to the KB, as well as an easy interface to do so.

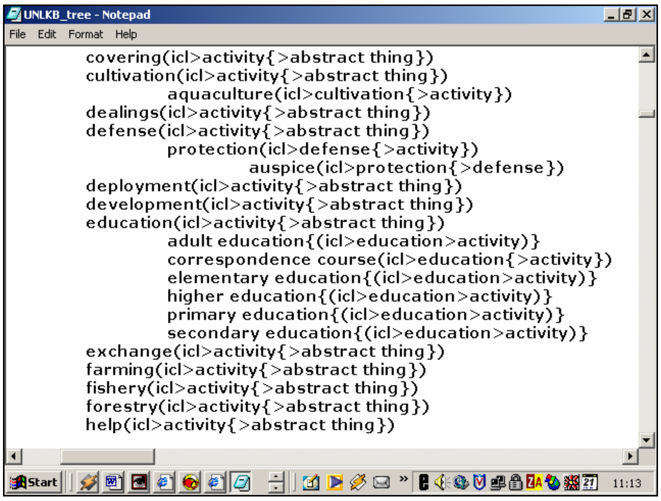


Fig. 2. The present UNL KB

4 The Complexity of the UNL KB

The main reasons for considering dealing with UNL KB a complex task are: 1) the inclusion of new UWs and 2) the management of the remote contributions into the database.

The first one is related to the creation complexity of new UWs, which includes the difficulty to relate them to existing ones. Thus, this process comprises the new UW position at the KB as well as the description of its binary relations. To avoid ambiguities, new concepts must be described according to different contexts to assure the various meanings among the languages.

The second related problem is associated to management of new inputs into the KB. It is difficult to supervise all contributions that come into the “KB gate”, to be farther transferred to the KB. Even if the process of homologating new UWs is necessary, the control of shared work can become less complicated with an efficient interface.

Thus we believe that a graphic user interface can facilitate these procedures through the visually representation of the KB textual schema. Indeed, it depends on the use of proficient presentation techniques to allocate the UWs.

It is also important to consider that this interface must allow the representation of the same UW described by different points of view. For instance, it can help the organization of the different UWs’ definitions according to the diverse domain’s dictionaries.

In this case, the search for existing UWs can be enhanced using a focused investigation for concept’s definitions instead of a traditional search for “headwords”. Therefore, it will make easier comparisons between UWs and it will add the classification of similarly concepts represented by them. By investigating the similarities between concepts, one can avoid the repetition of similar UWs. Under these settings, the UNL KB can include also a database documentation to support the managing of new UWs.

Finally, it is believed that those problems could be overcome by adoption of new KB formalisms. The proposed research builds on research works in several fields such knowledge representation models and languages, knowledge engineering, hypertexts and human-computer interaction.

5 The Proposal

The project addresses the problem of developing and maintaining a voluminous knowledge base (the UNL KB) in a distributed environment. The problem will be attacked along different axis, like the design and prototype of infrastructure, user interfaces for the UNL KB [10].

The first part of the project will focus on designing and prototyping a storage infrastructure for the UNL KB. The first step will design a model to store UNL definitions in a relational database. Then, this model will be extended to include "management information", such as versions, validation, point of views, access rights, roles, that are needed to support the knowledge base construction process.

This part of the project comprises not only enlargement of the UNL KB but also the design and prototype of various user interfaces for it. These will support different tasks and users, for instance authors, reviewers, or checkers. The first step will be to provide navigational and graphical interfaces to explore the knowledge base and to be familiarized with it.

The second step will be to create interfaces to add, edit, and delete objects in the knowledge base. These include auxiliary interfaces to support its maintenance. In this phase, we will use a wide range of paradigms, such as forms, hypertext, high-density graphical objects (hyperbolic trees, fisheye views, etc), or 3D objects [6] [7] [10]. We will also study which kind of inferences are possible using the UNL KB, not only to present richer information for users, but also to help contributors to expand further the UNL KB and to check consistency of their contributions. It will help to improve the knowledge representation structure and to provide a collaborative environment; hence, the data structure could be set up to represent comments, issues, arguments, decision, etc. For instance, in a multi viewpoint approach, every definition can be related to a viewpoint. Thus through a hypertext interface, the KB can be more comprehensible to the user since s/he will be able to view corresponding annotations when browsing or navigating the ontology. The possibility of navigation when building the knowledge base can help the whole process of UW development. For instance, it can be applied to create an interface that allows the user to investigate the UWs syntax by a visual approach.

In the last phase, we will establish a methodology for collaborative building of the UNL KB. This methodology will include the manual addition of new concepts from diverse natural languages as well as the importation of concepts from existing ontologies. In particular, it will describe the concept review and validation process, including resolution of definition conflicts.

Each part of the project will include a theoretical study, as the issues raised by a distributed construction of a large knowledge base are not yet fully understood and applicable solutions are still missing.

This work aims at progressing towards a complete environment that truly supports the process of ontology development (in contrast to ontology editors whose only goal is to enable the "encoding" of an already well-specified ontology).

6 Interfaces Specification

In another project, we applied the Lazy interface specification language to create visualization and manipulation interfaces for ontologies expressed in description logics [4]. As the UNL formalism shares some similarities with description logics, we intend to use the same kind of techniques to develop tools for the UNL KB. In this section we present a few examples of such ontology interface.

Since there are many tools to develop database applications and interfaces, it seems natural simply to store ontology in a database and to build ontology-engineering tools with database development tools. However, these tools are intended to develop "typical" database applications and are not specifically targeted at knowledge or ontology management applications. In particular, they usually provide form-based or table-based views of the data.

A hypertext view is a derived (computed) hypertext that represents the contents of some underlying information source. The idea is to provide the user with an easy to use hypertext interface that enables him/her to navigate within the information source. Thus it replaces database querying or other complex access mechanisms with just hypertext link following.

The Lazy language was designed to specify and implement hypertext views on relational and object databases. The language has been applied to generate different Web applications. Since the language is purely declarative, hypertext views can be specified without any programming.

The hypertext specification language is based on the concept of node schema. A node is comprised of - a set of parameters - a content specification (made of element specifications) - link specifications - a selection condition (what database objects to select). The instantiation of a node consists in interpreting a node schema for a given set of arguments and on at current state of database. An instance of a node schema is obtained by first selecting the objects (e.g. relation tuples) of the data collection(s) that satisfy the selection expression. Then the content and link specifications are evaluated on each selected object to generate node contents and links to other nodes.

In the Lazy model there are three types of links: reference links (the well known web links), inclusion links (to include the content of another node at this location), and expand in place links (clicking on such a link will open the target node at the link location, i.e. within the source node). These last two types of links are essential to build usable interfaces. We will see in the next sections, examples of the first prototype of this project, illustrating the usefulness of these types of links. For instance, by adding an expand-in-place link to the same node schema, we immediately obtain a typical "class browser" view.

```
node uwWithIcl[u]
  { <b>(headw), " " ,
    expand href iclOf[u]("&lt;icl")
  }
from UW selected by UW.id = u
```

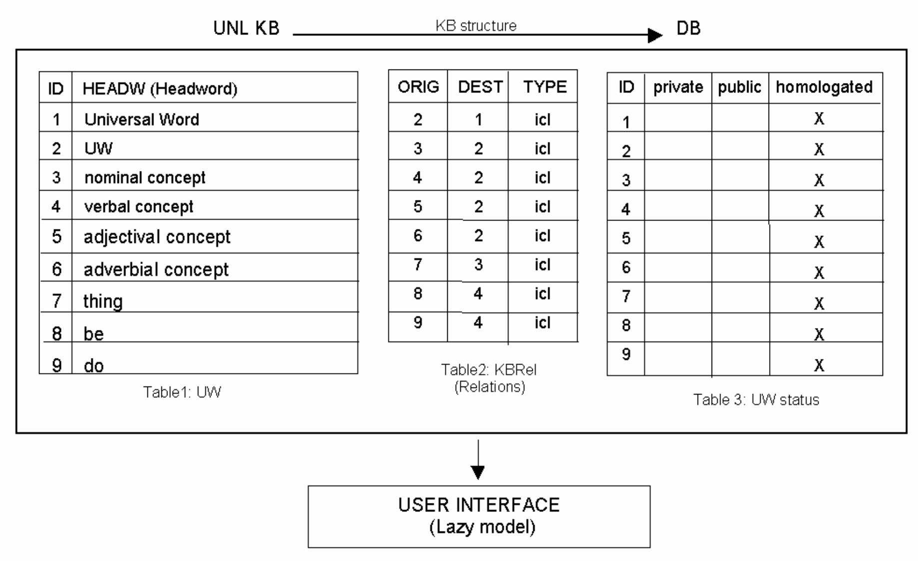


Fig. 3. Database schema for the UNL KB

7 UW Validation Process

Since the UWs homologation has been doing manually, the new UWs cannot be validated everyday, as the original UNL project proposes. In order to make it easier, we will add some workflow information to the KB. Each UNL concept can be stored into a relational database described as the proposed DB structure (figure 3). By adopting three different statuses for each UW- private, public and homologated – we can allow contributors to insert new UWs, even if they are not yet validated.

The figure 4 shows an hypertext view from part of the first prototype for the proposed UNL KB. It is made of nodes instances connected through hyperlinks. The figure 5 shows the navigational evolution of the node “uwWithIcl”, where all UNL concepts appear linked by reference and expand links.

8 Dealing with Collaborative Issues

Ontology or knowledge base building is usually a collaborative task that includes both domain experts and ontology experts. In the case of UWs, there will be always many different point of views of the same concept according to the various domains

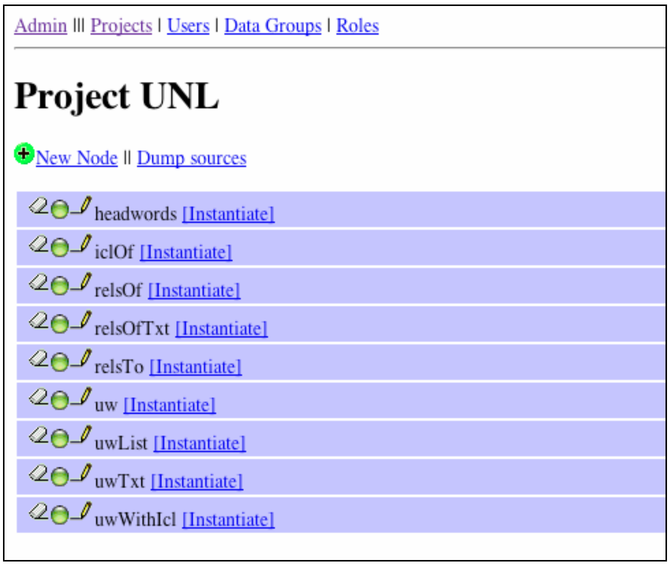


Fig. 4. First nodes of the proposed UNL KB interface.

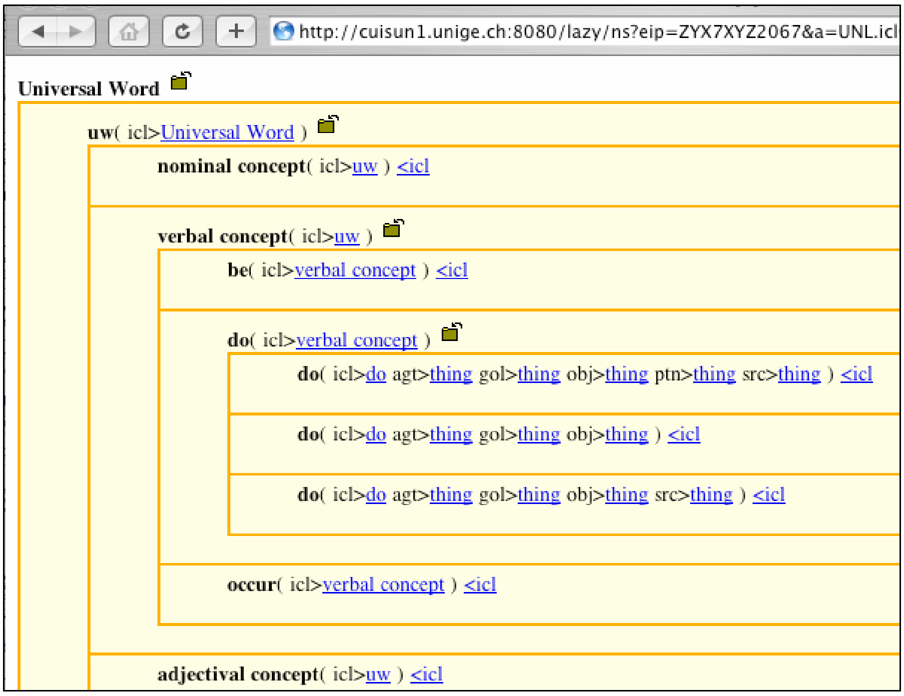


Fig. 5. Evaluation of the link "Universal Word"

and culture backgrounds. Thus, debates and conflicts on concept definitions are inevitable. Conflicts can occur either on meaning or on form of a definition, which includes the difficult of defining the headwords to write its definition [1] [4]. As happens with description logics ontologies, we believe that the development environment should not only include the UNL KB itself but also various kinds of "management" information. In other words, all sort of information that helps to build the KB (such as a viewpoint/conflict management mechanism, documents, notes, external ontologies, and databases). In fact, the use of a database as a storage infrastructure and use of Lazy to specify the interfaces enables to build an extensible environment easily, simply by adding relevant tables and interface nodes (figure 6).

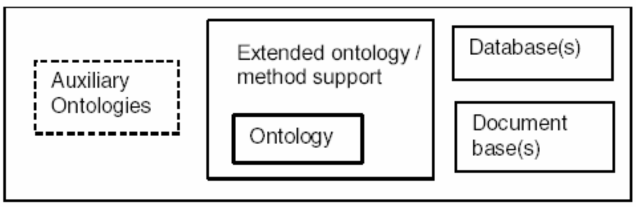


Fig. 6. Ontology environment

9 Final Considerations

The first outcome of this project will be a theoretical study of the different aspects of the distributed construction of large-scale knowledge bases. This study will be materialized in the form of several publications about specific issues and their resolution.

The second, and most important, contribution will be software applications, technical specifications, user manuals, etc. to help developers and contributors to work in a distributed environment. The applications will be packaged as "open software" and will be deployed by the UNDL Foundation to run the UNL System.

In parallel, educational and training activities will prepare a larger number of people to get involved in the collaborative development of the UNL knowledge base.

Finally, a website will be created, firstly to facilitate the wide distribution of the various software applications and documentation, but also to provide an online demonstration, for educational purposes. This website will act as a showcase and it is expected that it promotes the use of UNL and encourage new participants to take part in its development.

References

1. Falquet, G., Mottaz Jiang, C.-L., "A Framework to Construct Hypertext Interfaces for Ontology Engineering", submitted to the IJCAI03 workshop on "Knowledge Management and Organizational Memories", 2003.

2. Falquet, G., Mottaz Jiang, C.-L. "A Model for the Collaborative Design of Multi Point of View Terminological Knowledge Bases" in R. Dieng and N. Matta (Eds) Knowledge Management and Organizational Memories, Kluwer, 2002.
3. Falquet, G., Mottaz Jiang, C.-L. "Navigation hypertexte dans une ontologie multi-points de vue", in Proc. NimesTIC'01 conf., Nîmes, France, December 2001.
4. Falquet G., Mottaz C.-L., "Conflict Resolution in the Collaborative Design of Terminological Knowledge Bases". in Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, R. Dieng and O. Corby (Eds.),Lecture Notes in Artificial Intelligence 1937, Springer-Verlag, 2000
5. NG, G. K. C. Interactive Visualisation Techniques for Ontology Development, PhD thesis, University of Manchester, UK, 2000.
6. OntoWeb Consortium, A Survey of Ontology Tools, 2002, www.ontoweb.org/download/deliverables/D13_v1-0.zip.
7. OntoWeb Consortium, A Survey on Methodologies for Developing, Maintaining, Evaluating and Reengineering Ontologies, 2002, www.ontoweb.org/download/deliverables/D1.4-v1.0.pdf
8. Schneiderman, B. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 3rd edition, Addison-Wesley, Reading, Massachusetts, 1998.
9. Papers presented at the International Conference on Universal Knowledge, Goa, 2002. http://www.cfilt.iitb.ac.in/icukl2002/ papers/index_of_papers.html
10. Uchida H., Zhu M. and Della Senta T., The UNL, a Gift for a Millennium, Institute of Advanced Studies, United Nations University, 1999
11. Uchida H., Zhu M., The Universal Networking Language beyond Machine Translation, presented in the International Symposium on Language in Cyberspace, Seoul, 2001