Some Lexical Issues of UNL

Igor Boguslavsky

Institute for Information Transmission Problems, Russian Academy of Sciences
19, Bolshoj Karetnyj, 101447, Moscow, Russia
bogus@iitp.ru

Abstract. The Universal Networking Language (UNL) developed by Dr. H. Uchida at the Institute for Advanced Studies of the United Nations University is a meaning representation language designed for multi-lingual communication in electronic networks, information retrieval, summarization and other applications. We discuss several features of this language relevant for correct meaning representation and multi-lingual generation and make some proposals aiming at increasing its efficiency.

1 UNL Approach to the Lexicon

The Universal Networking Language (UNL) developed by Dr. H. Uchida at the Institute for Advanced Studies of the United Nations University is a meaning representation language designed for multi-lingual communication in electronic networks, information retrieval, summarization and other applications.

Formally, a UNL expression is an oriented hypergraph that corresponds to a natural language sentence in the amount of information conveyed. The arcs of the graph are interpreted as semantic relations of the types agent, object, time, reason, etc. The nodes of the graph can be simple or compound. Simple nodes are special units, the so-called Universal Words (UWs) which denote a concept or a set of concepts. A compound node (hypernode) consists of several simple or compound nodes connected by semantic relations.

In addition to propositional content ("who did what to whom"), UNL expressions are intended to capture pragmatic information such as focus, reference, speaker's attitudes and intentions, speech acts, and other types of information. This information is rendered by means of attributes attached to the nodes.

After 6 years of the UNL project development, it is possible to take stock of what has been achieved and what remains to be done. In this presentation, I am going to concentrate on one of the central problems with which any artificial language is faced if it is designed to represent meaning across different natural languages. It is a problem of the language vocabulary.

I would like to single out three distinctive features of the UNL dictionary organization.

1. **Flexibility.** There is no fixed set of semantic units. There is only a basic semantic vocabulary that serves as a building material for free construction of derivative