

Some controversial issues of UNL: linguistic aspects.

Igor Boguslavsky

Universidad Politécnica de Madrid (Spain)/IITP RAS (Russia)
igor@opera.dia.fi.upm.es

Abstract. We discuss several linguistic aspects of the Universal Networking Language (UNL); in particular, those connected with Universal Words (UWs), UNL relations, and hypernodes. On the one hand, the language should be rich enough and provide sufficient means to express the knowledge that might be required in the applications it is intended for. On the other hand, it should be simple enough to allow uniform and consistent use across languages and by all encoders. The major expressive device of UNL used for overcoming lexical divergence between languages is so-called restrictions. They have three functions, which are relatively independent of each other: the ontological function, the semantic function, and the argument frame function. We discuss various types of restrictions and propose new expressive means for describing UWs. Sample dictionary entries are given which incorporate our proposals. We propose several new UNL relations and discuss when and how hypernodes should be introduced.

1 Background

Among many problems that developers and users of a meaning representation language are facing, two somewhat conflicting requirements are standing out. On the one hand, the language should be rich enough and provide sufficient means to express the knowledge that might be required in the applications it is intended for. The more complex and knowledge-demanding the application, the more complex the design of the meaning representation language becomes. On the other hand, it should be simple enough to allow uniform and consistent use across languages and by all encoders. In the case of UNL, the latter problem is particularly serious, since the encoders work in different countries, belong to different linguistic schools, and have different linguistic traditions. Therefore, uniform understanding and use of UNL by all partners is difficult to achieve.

Since the start of the project in 1996, a large number of UNL-encoded documents have been accumulated that were produced by the project participants from 16 language groups each working on its native language. The analysis of these documents clearly shows two things: *UNL is still lacking means to express meaning adequately*, and *there is not enough uniformity in the UNL use among the partners*. To some extent, UNL has developed its own dialects. Despite the existence of the UNL Specifications, divergences between the dialects tend to grow. This tendency clearly manifests itself in the fact that all deconverters (=generators) are doing much better when dealing with the UNL documents produced by the authors of the deconverter than with those provided by other teams. If it goes on this way, the