Prologue

UNL is an ongoing worldwide initiative starting in 1996. Almost 10 years have passed a big span of time for a project. We could say that UNL didn't meet its expectations. But let's have a closer look to UNL, the project, its basics and objectives. A closer look at its objective will reveal that this affirmation is gratuitous and unmotivated.

The Problem: Linguistic Diversity

UNL was launched by IAS/UNU to erase linguistic barriers. Linguistic barriers collide with the enhancement of linguistic diversity and the value that native languages as one of the main vehicles to express one's cultural identity. Apart from socio-cultural issues, linguistic diversity also knows an economic and political dimension. Institutions like the United Nations or the European Union have to face everyday with the barriers that linguistic diversity imposes. It is well known the enormous amount of documentation that these institutions produce everyday, which have to be produced in all their official languages: 6 for the UN, 25 for the European Union. It is simply unfeasible to rely on human translators for the production of all these amount of documentation.

Aware of this, the IAS/UNU launched the UNL project, aiming at the real access of information in the own native language and not recurrent to dominant languages. UNL is basically an artificial language where contents expressed in natural languages can be converted to and subsequently, contents written in UNL can be generated into any natural language, provided that the adequate tools are built.

MT and Multilinguality

From the technological point of view, multilinguality has been tackled by Machine Translation. In the evolution of the area of MT, there is variety of architectures to undertake the task of translating the *contents* of one text written in a given language into another language. Transfer-based systems could be regarded as the most productive and of better quality. But they are hindered by the exponential growth in the modules to be developed when the number of involved languages increases. A transfer-based system involving N languages need to develop N*(N-1) modules. An astronomic number to create real multilingual platforms.

Further, although there are some very good systems, the quality of these systems seem to be limited, since after years of refinement, the MT system does not surpass a given degree of quality. Besides, the development of transfer based MT systems is usually reduced to the so-called majority languages (English, French, German and even Spanish or Italian), but it is fairly rare to find a good quality and wide coverage MT system covering English and Polish, let's say.

Transfer based MT is not the only option, Interlingua-based systems represents an alternative to transfer systems. Interlingua-based MT does not work on pair of lan-

guages, but translation is carries out to and from an artificial language that serves as a pivot for all the natural languages involved in the system. This architecture tries to overcome the exponential growth of transfer-based systems, since the number of modules to develop for N languages is 2*N and the inclusion of new languages into the system does not affect the other language modules. In this way, UNL follows the architecture of Interlingua-based MT systems.

Usually, Interlinguas are abstract formal (or semi formal) languages that captures the meaning of texts in a language independent way. Ideally, the Interlingua should not be close to a given particular language and should not include linguistic devices proper of natural languages. In this way, Interlingua-based systems seem the most plausible (and even the unique) option to tackle massive multilinguality.

But Interlinguas has been often rejected within the scientific community and since their boom in the 80ies, there have no commercial application of Interlinguas and the systems developed under this trend were laboratory products. Why is this so? Let's have a look at the properties of interlinguas.

Problems with Interlinguas

Interlinguas are *semantic languages* designed to represent the meaning of any given text, ideally satisfying the following conditions:

  (a) They are language neutral.
  (b) They are precise, unambiguous, formal languages

Being so, they usually show the following characteristics:

− Interlinguas are intimately tied up with ideas about the representation of meaning, being meaning the most abstract and deepest level of linguistic analysis (that should be common to all languages, far enough from surface representation of languages).
− An Interlingua is "another language" in the sense that it has autonomy and thus its components need to be defined: vocabulary and "relations" mainly. Besides, and Interlingua is an artificial language that should be as expressive as natural languages.

Here we find the main bottleneck of interlinguas: its proper design and definition. Defining an Interlingua involves the following parameters:

  (a)   A language whose "atoms" are not dependent on any given natural language so that the ambiguity of natural languages is eliminated.
  (b)   A language whose "atoms" are not dependent on a given natural language so that the concepts and ideas expressed in different natural languages can be easily and naturally expressed in the Interlingua.
  (c)   A language that is as expressive as a natural language so that what can be expressed in natural languages can be transposed to the Interlingua, and from the interlingua to other natural languages.

These three conditions make interlinguas hard to design. It is quite difficult to find the equilibrium between language independency, degree of abstraction and expres-

siveness in a formal device such an Interlingua. Maybe this difficulty in the design of interlinguas is the reason why they have not been successful at least in open domains within massively multilingual environments. The examples of interlingua-based systems are domain dependent and quite limited in the number of languages.

Is UNL a Viable Solution?

The panorama appears quite despairing. While Interlinguas are theoretically biased and difficult to put into practice, transfer based systems have proved to be unattainable when dealing with massive multilinguality. Maybe the concept of Interlingua should be revisited, and re-adapted to real necessities and to real scenarios. This is the spirit of UNL. UNL, by its definition and by its most basic architecture is definitely an Interlingua-based system. Its targets are the support of multilinguality, not restricted to a given domain or to a given family of languages. Thus, the design of a interlingua like UNL encounters all the possible barriers that an Interlingua may encounter (especially to find a real language independent representation).

So why we could considered UNL as different, as a new viable technology if interlinguas were rejected a long time ago? First, let's remember the main objective of UNL:

− to generate and produce contents in any natural language in any domain.
− to support multilingual services.

That is, there is a primacy of generation and coverage of languages and domains, which means that a **very expressive formalism** has to be designed in order to represent such a variety of contents coming from any natural language.

Let's illustrate this fact by have a closer look at the vocabulary of the Interlingua, one of the most difficult and polemic issues of UNL and of any Interlingua. UNL utilizes the so-called Universal Words as the semantic atoms of the Interlingua (no decomposable). They exhibit the following main characteristic:

They are based on English headwords.

From this very simple definition, we can conclude that UNL is language biased (English) and thus:

1. UNL is based on a natural language:
2. It hinders logical relations and inferences (facilitated by primitive based solutions)
3. Its vocabulary is a potential source of ambiguity
4. Its vocabulary fosters lexical and conceptual mismatches among languages.

So is there any advantage in the UW system and in the overall essence of UNL? Well, if theoretical reasons do not support the design of open-domain interlinguas, let's look at the practical or pragmatic ones.

   (a) UNL is based on a natural language. At first sight could be a drawback, however, the expressiveness of a natural language is inherited by the Interlingua, thus allowing for the representation of a variety of domains and concepts.

(b) UNL shows an English oriented vocabulary. At this moment, English is the lingua franca, the most accessible to work with for Indo-Europeans, Semitic, Japanese, Chinese, etc. Bilingual dictionaries usually have English as one of their target/source languages, thus the development of lexicographic resources is facilitated by choosing English as the most basic atoms of the language.

Of course, this approach (although supported by pragmatism) is far from perfect. Even at first sight, it can be considered as naïve, since it merely "suggest" well known problems in lexical semantics (like support verbs, compounds expressions, connotational meaning, etc). For this reason, theoretical research on the UNL as a language itself should be fostered within the Consortium, while respecting the basic nature of the language.

That is, UNL should be viewed rather than a perfect Interlingua as **the pillars to support multilingual services**. Its natural language orientation (apparently, its weakest points as an Interlingua) turns the language as a candidate to the support of multilinguality and facilitates converting contents to and from UNL. There are several aspects that support it. First, the creation of generators of medium quality (where post-edition is possible) is rather straightforward. Second, its flexibility and language orientation makes it possible to integrate UNL into other pre-existent MT systems (be it transfer-based be it another architecture) which extends the range of application of UNL and makes possible to alleviate the problem of exponential growth in transfer-based systems. And last, but not least, the processes of enconverting and deconverting are independent so that if generation is taken as a priority, generators are constructed first; the process of enconversion can be done manually, due to the human readability of the language.

At this point in the evolution of UNL, there appears a contradiction, UNL is still not theoretically mature, but from an applied perspective, it is. In the short term there is priority for the UNL Consortium to get feedback from previous experiences in Interlinguas, from Linguistic Theory (semantics, logic, and lexical semantics) in order for UNL to grow and find a place in the scientific community and, why not, in the market as a real approach to support multilinguality, once the applications and utilities are clear and defined within the UNL Programme.

Prospective

So is it worth another attempt? Definitely yes, the real need to overcome linguistic barriers (be it at the institutional level, be it at the social level) claims for a solution to the problem of multilinguality. Transfer based systems simply are out of question *if isolated*. This doesn't mean that they are useless: they are not. An interlingua like UNL is conceived as another autonomous languages, close enough to the superficial form of natural languages, thus integration of the Interlingua into the transfer system is possible and not a *contradiction in terminis*.

After several years of experience, we know that knowledge and language generation do not go *on a par*. Thus the final design have to be done bearing the ultimate

purpose of the interlingua (the closer to language semantics is, the better to generate languages) and probably will lead to the success of the interlingua.

A Final Word

I would like to thank the editors of this book for their invitation to write a prologue to this work and to collaborate with them in the selection and revision of the selected papers presented in this volume. Hopefully it will provide a thorough understanding of the UNL Programme, its meaning, its evolution, its shortages and its strengths.

Carolina Gallardo