

On the Aboutness of UNL

Ronaldo Teixeira Martins,^{1,2} Maria das Graças Volpe Nunes.^{2,3}

¹Faculdade de Filosofia, Letras e Educação – Universidade Presbiteriana Mackenzie
Rua da Consolação, 930 – 01302-907 – São Paulo – SP – Brazil

²Núcleo Interinstitucional de Linguística Computacional (NILC)
Av. Trabalhador São-Carlense, 400 – 13560-970 – São Carlos – SP – Brazil

³Instituto de Ciências Matemáticas e da Computação – Universidade de São Paulo
Av. Trabalhador São-Carlense, 400 – 13560-970 – São Carlos – SP – Brazil

ronaldomartins@mackenzie.com.br, gracan@icmc.usp.br

Abstract. This paper addresses the current status, the structure and role of the UNL Knowledge Base (UNLKB) in the UNL System. It is claimed that the UNLKB, understood as the repository where Universal Words (UWs) are named and defined, demands a thorough revision, in order to accomplish the self-consistency requirement of the Universal Networking Language (UNL). In order to emulate human cognition and constitute the “aboutness” of the UNL, the UNLKB should be decentralized, distributed and reorganized as a network of networks, allowing for multicultural information and dynamic data.

1 Introduction

The Universal Networking Language (UNL) is an “electronic language for computers to express and exchange every kind of information” [Uchida, Zhu & Della Senta, 1999]. It can be defined as a knowledge-representation formalism expected to figure either as a pivot language in multilingual machine translation (MT) systems or as a representation scheme in information retrieval (IR) applications. It has been developed since 1996, first by the Institute of Advanced Studies of the United Nations University, in Tokyo, Japan, and more recently by the UNDL Foundation, in Geneva, Switzerland, along with a large community of researchers—the so-called UNL Society—representing more than 15 different languages all over the world.

Formally, the UNL is a semantic network believed to be logically precise, humanly readable and computationally tractable. In the UNL approach, information conveyed by natural language utterances is represented, sentence by sentence, as a hyper-graph composed of a set of directed binary labeled links (referred to as “relations”) between nodes or hyper-nodes (the “Universal Words”, or simply “UW”), which stand for concepts. UWs can also be annotated with attributes representing context-dependent information.

As a matter of example, the English sentence ‘Peter kissed Mary?!’ could be represented in UNL as (1) below:

```
(1) [S]
    {unl}
    agt(kiss(agt>person,obj>person).@entry.@past.@interrogative.@exclamative,
    Peter(iof>person))
    obj(kiss(agt>person,obj>person).@entry.@past.@interrogative.@exclamative,
    Mary(iof>person))
    {/unl}
    [/S]
```

In (1), ‘agt’ (agent) and ‘obj’ (object) are relations; ‘Peter(iof>person)’, ‘Mary(iof>person)’ and ‘kiss(agt>person,obj>person)’ are UWs; and ‘@entry’, ‘@past’, ‘@interrogative’ and ‘@exclamative’ are attributes.

Differently from other semantic networks (such as conceptual graphs [Sowa, 1984, 2000] and the RDF [Lassila & Swick, 1999]), UNL relations and attributes are predefined in the formalism. As of the 3.2 version of the UNL Specification (UNL Center, July, 2003), the set of relations, which is supposed to be closed and fixed, consists of 44 elements that conveys information on ontological relations (such as hyponym and synonym), on logical relations (such as conjunction and condition), and on semantic case or thematic role (such as agent, object, instrument, etc.) between UWs. The set of attributes, which is subject to increase, currently consists of 72 elements, and cope with speaker’s focus (topic, emphasis, etc.), attitudes (interrogative, imperative, polite, etc.) and points-of-view (need, will, expectation, etc.) towards the event. This feature brings UNL to represent not only denotative but also connotative, non-literal, information. The set of UWs, which is open, can be extended by the user, but any UW should be registered and defined in the UNL Knowledge-Base in order to be used in UNL documents.

Under the UNL Program, natural language analysis and understanding is referred to as a process of “enconverting” from natural language (NL) into UNL. This enconverting process, which has been carried out in a somewhat computer-aided human basis, is said to be not only a mere encoding (i.e., to rephrase the original sentence using different symbols), but truly a translation from the source sentence into a new target language - the UNL -, which is thought to be as autonomous and self-consistent as any NL, and whose graphs are expected to be language-independent and semantically self-governing. As it targets the information conveyed by the source text rather than its syntactic or even its semantic structure, the UNL is assumed to be different from other interlingua-based approaches, and to be more akin to the knowledge representation paradigm than to the machine translation techniques. As a matter of fact, and at least for the time being, UNL has been mainly used for multilingual document generation, through a process referred to as “deconverting”, which consists in automatically providing NL outputs that can be said to be functionally (yet not formally) equivalent to the information conveyed by UNL graphs.

In this paper, we address the current structure of the UNL Knowledge Base and some of the problems we have been facing during the process of creation and definition of UWs inside EPT-WEB, an English-to-Portuguese UNL-based MT project. The paper is organized as follows: Section 2 brings some additional information on UWs and their internal structure; Section 3 analyzes the concept of Master Definition (MD) and the structure of the UNL KB; in Section 4, we explore some problems and short-

comings of the current version of the UNL KB; finally, in Section 5, we suggest some changes and enhancements in the UNL KB structure.

The authors should acknowledge that some of the opinions and definitions indicated below do not necessarily represent the official perspective on the UNL and may not coincide with those supported by the UNL Center.

2 Universal Words (UWs)

Universal Words, the words of UNL, are composed of a root (usually referred to, in UNL Specifications, as “headword”) and a suffix (“the constraint list”). The latter comes between parentheses and is used mainly when the root is believed to be ambiguous. Examples of UWs are presented below:

- (2a) ‘Universal Word’
- (2b) ‘UW(equ>Universal Word)’
- (2c) ‘Peter(iof>person)’
- (2d) ‘apple(icl>fruit)’
- (2e) ‘kiss(agt>person,obj>person)’
- (2f) ‘explain(icl>express(agt>thing,gol>person,obj>thing))’
- (2g) ‘Manyoshu(icl>Japanese poem)’

Both root and suffix are normally made out of English words, but it is also possible to find “extra UWs” (2g above), when no simple (single) English word can be suited to convey the intended meaning, as in the case of culture-dependent concepts, such as special sorts of clothing, dish, furniture, etc. In those cases, foreign (transliterated, if necessary) words may be used to label the root of a UW.

The fact that the vocabulary of UNL is mainly derived from English may introduce an undesired natural language bias which can be said to be not only ethnocentric (in the sense all foreign concepts would be reduced to the ones carved up by the English language) but mainly counter-effective, as it would lead UNL to be a mere sort of controlled English. However, it is claimed that UWs, as labels, do not have meaning themselves. They would be just unique strings of characters that are used to refer to concepts. In this sense, the root (the headword), as well as the suffix (the constraint list), do not play any role other than disambiguating and ensuring uniqueness to the UW. The obvious resemblance between UWs and English words would be rather accidental, in order to cope with the commitment that UNL, as a semantic network, should be, to some extent, humanly readable. The use of English words would make UWs to be mnemonic and would facilitate the use of UNL by humans, but it would be completely useless and ineffective from the machine-tractability point-of-view. Yet it may seem to convey some meaning, the machine would consider ‘apple(icl>fruit)’ as meaningful as any arbitrarily assigned memory address.

In order to UNL to be really self-consistent and language-independent, the meaning of a UW, i.e., its value, should be entirely derived from a set of relations assigned in the UNL itself. The meaning of ‘apple(icl>fruit)’ should not come from a human comprehension or an external language that would never be replicable by the machine, but, instead, should be stated in a purely intensional (non-mental) dimension, a

sort of electronic (possible) world, which would represent the sense and the reference for UNL words and expressions. This digital (and artificial) world, and not the human analogical one(s), would be the “aboutness” of UNL, as it would comprise the truth-condition requirements for UNL expressions to be “meaningful”. Inside the UNL System, such synthetic world has been referred to as the UNL Knowledge-Base (or simply the UNLKB), a huge network where nodes (concepts) would be interconnected as to emulate the structure of human cognition.

As a matter of example, the meaning of ‘apple(icl>fruit)’ should be defined by a set of binary relations such as those indicated by (3) below:

- (3a) icl(apple(icl>fruit), fruit(pof>plant))=1;
- (3b) obj(eat(agt>thing, obj>thing), apple(icl>fruit))=1;
- (3c) aoj(round(aoj>thing), apple(icl>fruit))=1;
- (3d) pof(apple(icl>fruit), apple tree(icl>tree))=1;

This means that “apple(icl>fruit)” would be the concept that concomitantly a) assigns an ‘icl’ (a-kind-of) relation to the concept labeled by the UW ‘fruit(pof>plant)’; b) receives an ‘obj’ (object) relation from the concept labeled by the UW ‘eat(agt>thing, obj>thing)’; c) receives an ‘aoj’ (attribute of thing) relation from the concept labeled by the UW ‘round(aoj>thing)’; and finally d) assigns a ‘pof’ (part-of) relation to the concept labeled by the UW ‘apple tree(icl>tree)’. The value of ‘apple(icl>fruit)’ would be the sum (and nothing but the sum) of all relations in which it takes part in the UNLKB. This is a rather negative definition, given that it does not state positively the meaning of “apple(icl>fruit)”, but only the relations that it may take. The set of UWs would be therefore a sort of sign system where the value of a given sign would solely derive from its position in the network. This is to say that, at least at the lexical level, UNL would consist of “un système où tout se tient” (Meillet, 1901; Saussure, 1916), following hence the structuralist approach that “every language is a system, all parts of which organically cohere and interact [... where] no component can be absent or even different, without transforming the whole” (Gabelentz, 1901). This would be a *sine qua non* condition for the autonomy and self-consistency of the UNL.

It should be stressed that this negative (relational) definition does not necessary coincide with the positive, contentful one, normally ascribed by a human. In the example above, for instance, nothing has been said about the relation that the UW ‘apple(icl>fruit)’ takes with other UWs such as ‘red(icl>color)’ or ‘apple pie(icl>pie)’. This means that, in UNL, at least in the given situation, such features do not participate in the definition of ‘apple(icl>fruit)’, which would be therefore incomplete from a human point-of-view. But given that complete definitions are not to be easily achieved, because they can be self-contradictory (as apples can be red or green, for instance) and dynamic (different users, or even the same user at different times, may have different experiences or reactions towards apples), the UNLKB is not expected to define, in an exhaustively way, all the meaning intended by any concept.

Actually, in the UNL Program, there seems to be at least two different representational levels for defining UWs. The first would be related to the UNLKB itself and would target the (alleged) systematic part of the meaning, in a sense very close to the one intended by the concept of “semantic markers” (Katz & Fodor, 1963). On the

other hand, the unsystematic part of meaning (the “distinguishers”) would be treated in the UNL Encyclopaedia, which is a huge UNL document base, also organized as a network, where idiosyncrasies and additional information on UWs are expected to be stored. Here we will focus only on the UNLKB structure and on the boundaries between relations that should be necessarily included in a UW definition.

3 The UNL Knowledge-Base (UNLKB)

The UNLKB is a semantic network whose entries have the structure exemplified in (3) above. They comprise a binary directed relation (extracted from the UNL relation-set) between two UWs, along with a degree of certainty, which can range from 0 (completely false) to 255 (completely true). Any UNL-relation can hold between UWs in the UNLKB, and a single UW may receive and assign many different relations from and to other UWs. However, in order to assure replaceability, inference and cross-reference (inheritance) inside the network, any UW should be linked by at least one of the three ontological relations, namely “icl” (a-kind-of), “iof” (an-instance-of) or “equ” (equal-to). The relation “pof” (part-of), formerly used, has no longer been adopted, as it does not allow for direct inheritance.

One can say that linking a UW to any other by means of “icl”, “iof” or “equ” is to compose a UW Thesaurus, or the UNL Ontology, but it should be stressed that such network is only a part of the UNLKB. Inside the UNL System, this subnetwork has been referred to as the “UW System”, and it constitutes a lattice structure, given that a child-node may have many different parent-nodes. This hierarchical network also comprises an inheritance mechanism, so that all information assigned to a given parent-node could be directly inherited by its children-nodes. In this sense, if (4) below had been stated in the UNLKB, there would be no need for (3b), provided that it could be easily inferred from (3a):

- (4) obj(eat>thing,obj>thing), fruit(pof>plant))=1;
 (3a) icl(apple>fruit),fruit(pof>plant))=1;
 (3b) obj(eat>thing,obj>thing), apple(icl>fruit))=1;

The need for the UNLKB has been subject to criticism inside the UNL Project, but it should be observed that knowledge-based MT systems have proved to provide better results than those that are only language-based (Nirenburg, Raskin et al., 1986). Inside the UNL System, the UNLKB is intended to assure robustness and precision both to the NL-UNL enconverting and to the UNL-NL deconverting. In the former case, the UNLKB would be used as a sort of word sense disambiguation device; in the latter, the UNLKB, through the replaceability operations, would allow for the deconversion of UWs not predicted by the target language dictionaries. Additionally, the power of the UNLKB for intelligent searching and semantic inference and reasoning should not be underestimated.

In order to discipline and organize the creation of UWs, the UNL Center has proposed a particular technique for both naming (labeling) and defining a UW at a single movement: this is the Master Definition (MD), introduced in 2000. The MD for nam-

ing the UW “apple(icl>fruit)” and defining it in the UNLKB (through an “icl” relation to the UW “fruit(pof>plant)”) is presented in (5) below:

(5) apple(icl>fruit{pof>plant})

The MD is said to facilitate (and regulate) the labeling of a UW, which would derive its suffix (the constraint list) from its definition in the UNLKB. The name of the UW would be the same as the MD without the strings included inside the curly braces. This would motivate the UW name and keep it as mnemonic as possible for human use (although, as said before, the name of a UW has nothing to do with its “mechanical” meaning). However, it should be noticed that the concept of MD brings itself many shortcomings, mainly the facts that: 1) due to the simplification of syntax, the MD is not capable of conveying any degree of certainty other than 1; and 2) MDs can only be used to define the UW by means of ‘icl’, ‘equ’ or ‘iof’; any richer definition would require longer strings and more expensive strategies that may fall out the MD scope. Nevertheless, and at least for the time being, the UNLKB has been entirely defined as a hierarchy of MDs.

4 Problems and Limitations of the Current Version of the UNL KB

The current version of the UNLKB has been developed and constantly updated by a single person, the director of the UNL Center, and can be downloaded from the UNDL web site [<http://www.undl.org>]. The comments below refer to the version as of November 28, 2004. Among the problems, three have been specially selected and are going to be addressed in this paper, all of them related to our experience in creating UWs in an ongoing UNL-based MT Project carried out by NILC. The problems are the syntactic bias of the UNLKB; the synonymy of UWs; and the lack of criteria for categorization.

The syntactic bias of the UNLKB can be mainly ascribed to the third and the fourth uppermost levels of the UW System. According to the current version of the UNLKB, the topmost level of the UW System is the UW ‘Universal Word’ itself, followed by ‘UW(equ>Universal Word)’ and, next, by four different UWs, each of which has their own children, leading to the basic structure depicted in Figure 1 below.

From Figure 1, it is possible to see that the UNLKB structures the lexical semantics of UNL in a rather syntactically-biased way. Every UW should be linked – either directly or indirectly – to the four grammatical nodes: “nominal concept”, “verbal concept”, “adjective concept” and “adverbial concept”, taken as semantic primitives. Even though these concepts may lead indeed to semantic notions, they normally refer to syntactic and morphological markers, as indicated in the case of the English word “reading” in the examples below, adapted from the WordNet [<http://www.cogsci.princeton.edu/>]:

- (6a) She is **reading** a book.
- (6b) She enjoys **reading** books.
- (6c) She disapproved his **reading** of Shakespeare.
- (6d) She bought some **reading** material.

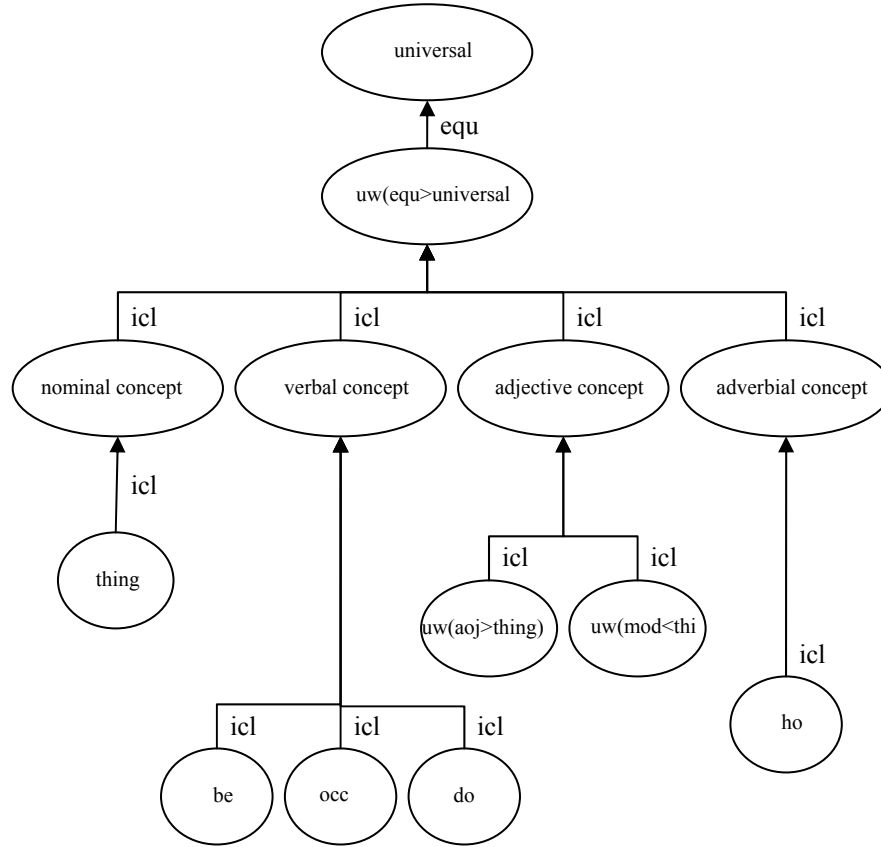


Figure 1: the topmost levels of the UW System

In the UNL representation for those sentences, it is likely to find the following:

- (6a) read(agt>person,obj>information)
- (6b) reading(icl>action)
- (6c) reading(icl>information)
- (6d) reading(aoj>thing)

Although UNL is fine-grained enough and capable of disambiguating between different uses of “reading”, it should be primarily attached to the semantic content conveyed by each occurrence. It should be observed, however, that those semantic values may have very little to do with the part-of-speech information, which is actually only a feature of English, rather than a universal character. In Portuguese, for instance, “reading” in (6b) would be normally translated by a verb, instead of a noun, as indicated in (6b’) below:

- (6b’) Ela gosta de **ler** livros.

A similar problem occurs in (6c) and (6d), which would be indirectly linked to “nominal concept” and “adjective concept”, respectively, because in English they would play the roles of a noun and an adjective. In Portuguese, however, “reading” would be normally translated as a noun in (6c) and as prepositional phrase in (6d), according to (6c’) and (6d’) below:

(6c’) Ela desaprovou sua **leitura** de Shakespeare.

(6d’) Ele comprou algum material **para ler (de leitura)**.

Both cases illustrate that, at least from the perspective of Portuguese, the definition of “reading” in (6a) to (6d) would be different from the one achieved if English is to be taken as the source language. This proves a language-dependent feature that may interfere in the definition of UWs, assigning a semantic value to operations that are rather syntactic (such as nominalization, deverbalization, adjectivation, etc). If UNL is really expected to represent the ideational content of utterances, rather than its syntactic or semantic structure, it should consider such sort of cross-language mismatches, in order to be as language-independent as possible.

Two additional illustrations of the syntactic bias of the UNLKB can be reached if we consider the representation of UWs conveying information on places and adjectives. In both cases, we have to consider the realm of ‘adjective concepts’ and ‘adverbial concepts’, which, differently from ‘nominal concepts’ and ‘verbal concepts’, do not correspond to a real taxonomy, but to a flat list where there is no internal hierarchy among the elements.

As to places, for instance, the UNLKB comprises two different UWs corresponding to the English word “here”: a) simply ‘here’, without any suffix, under ‘adverbial concept’; and b) ‘here(icl>place)’, under ‘nominal concept’. It is possible to say that they cover the meanings intended respectively by (7a) and (7b) below (extracted once again from the WordNet):

(7a) in or at this place; where the speaker or writer is; "I work here"

(7b) the present location; this place; "where do we go from here?"

However, it should be stressed that, in both cases, ‘here’ is essentially a place, regardless of its grammatical role (i.e., its part-of-speech) in the sentence. Although this information may be kept in ‘here(icl>place)’, it is definitely lost in the basic UW ‘here’, which is directly located under ‘how’ (and hence under ‘adverbial concept’). This is especially unproductive because the “adverbial role” of “here”, if any, could be alternatively represented by means of the relation “plc” (place) or even “man” (manner).

A third illustration may come from the classification of adjectives. In UNL, adjectives - which also correspond to a plain list instead of a hierarchy - are said to be either predicative, or attributive, or both. This is the case for “good”, that can be found either as ‘good(mod<thing)’ or ‘good(aoj>thing)’, in order to cope with (8a) and (8b), respectively:

(8a) A good boy

(8b) The boy is good.

It is under dispute, however, if the opposition between predicative and attributive, that maybe is relevant for English, really holds in every language. Is this a real gen-

eral semantic phenomenon or simply a language-dependent syntactic event? If semantic, should not it be represented by relations or attributes instead of UWs? Is it really useful to register, in the UW dictionary, both ‘good(aoj>thing)’ and ‘good(mod<thing)’, given that they mean the same?

As a matter of fact, the repertoire of UWs seems to indicate that the lexicalization of UWs is exaggeratedly based on the surface structure of English sentences and on the lexical items of English. This can be further attested because of the presence of variants, antonyms and synonyms in the UW dictionary.

For instance, one will find, in the UNL KB current version, both ‘behavior(icl>action)’ and ‘behaviour(icl>action)’. The difference between them is not semantic, but strictly orthographic. There is no reason for cataloging such kind of spelling difference in a semantic database.

The same should apply for pairs of antonyms such as give/receive, borrow/lend, etc. These verbs are supposed to convey the same meaning in a reversed subcategorization frame: $give(x,y) = receive(y,x)$. Once “give” and “borrow” are there, would there be any reason for including “receive” and “lend” as well?

- (9a) give(agt>thing,gol>person,obj>thing)
- (9b) receive(agt>thing,obj>thing,src>thing)
- (10a) borrow(agt>thing,obj>thing)
- (10b) lend(agt>thing,gol>person,obj>thing)

This sort of overlapping among UWs does not affect only antonyms and can be found all over the UNLKB. Let us consider two last examples: is there any real need for registering, in the same knowledge base, all the words appearing in (11) and (12) below? Are the semantic differences between them really relevant? Are they going to be preserved in languages other than English?

- (11a) begin(agt>thing,obj>thing)
- (11b) commence(icl>begin(agt>thing,obj>thing))
- (11c) start(icl>begin(agt>thing,obj>thing))
- (12a) nurse(icl>medical assistant)
- (12b) nurse({icl>person>human,} icl>occupation{>work})

The examples referred to above prove that economy has not been an asset of the UNLKB. Obviously, one can claim that synonyms and variants are to be represented, because there is no perfect synonymy and UNL is supposed to be as comprehensive and fine-grained as any natural language. But again, and provided that there is no perfect lexical matching between languages, would UNL be wide enough to comprehend the vocabulary of every existing natural language? How to prevent combinatorial explosion inside the UNLKB? How to prevent that the proliferation of UWs will not affect the maintenance of UNL resources, will not introduce different dialects to UNL, and will not cause the entropy of the whole system?

In addition to the syntactic bias and the synonymy, there are many other problems that could be pointed out inside the UNLKB, but most of them are far much easier to handle. Due to the discrepancy on the use of braces, for instance, there are many duplicated entries inside the network, such as (13) below:

(13a) eau de cologne{(icl>perfume>functional thing)}

(13b) eau de cologne(icl>perfume{>functional thing})

And there are also some class inconsistencies. Tigers and panthers, for instance, are normally defined as belonging to the species of felines, but, in the UNLKB, they have been categorized directly under ‘mammal(icl>animal)’, differently from ‘cat(icl>feline)’:

(14a) tiger(icl>mammal{>animal})

(14b) panther(icl>mammal{>animal})

(14c) cat(icl>feline{>mammal})

In the same way, specific languages and types of languages have been categorized at the same level, as indicated in (15) below:

(15a) spoken language{(icl>language>system)}

(15b) Russian(icl>language{>system})

(15c) inflectional language{(icl>language>system)}

Circularity may also be found, as in (16) and (17):

(16) thing{(icl>nominal concept)}

abstract thing{(icl>thing)}

event(icl>abstract thing{>thing})

thing(icl>event{>abstract thing})

(17) figure(icl>figure{>attribute})

The main problem, however, concerns the lack of (uniform) criteria for categorizing concepts. In (18) below, for instance, the concept conveyed by the English words “film” and “movie” is said to be linked to the concept of “abstract thing”. Why that? Why not “concrete thing”? Or why not “functional thing” instead? What about instances of films, such as “Gone with the wind”? Would they also be considered a sort of “abstract thing”?

(18) abstract thing{(icl>thing)}

art(icl>abstract thing)

cinema(icl>art{>abstract thing})

film(icl>cinema{>art})

movie(icl>cinema{>art})

Such categorization turns to be even more astonishing if we consider the case for “book”, which is also located under the “abstract thing” branch of the UNLKB, as indicated in (19):

(19) abstract thing{(icl>thing)}

information{(icl>abstract thing)}

document(icl>information)

book(icl>document{>information})

book of general works{(icl>book>document)}

manuscript{(icl>book of general works)}

rare book{(icl>book of general works)}

book of geography{(icl>book>document)}

On the other hand, both “landscape” and “scenery”, and even “beauty spot”, are categorized under “concrete thing”, as seen in (20):

```
(20) concrete thing {(icl>thing,icl>place>thing)}
      natural world {(icl>concrete thing,icl>place>thing)}
      landscape(icl>natural world)
      scenery(icl>landscape{>natural world})
      beauty spot(icl>scenery{>landscape})
      scene(icl>scenery{>landscape})
```

The absence of categorization guidelines and protocols cause the UNLKB to be excessively impressionist, in the sense it contains, to a considerable extent, subjective and personal ideas towards the world and the structure of events. Although some of those decisions may sound quite reasonable from a given perspective, it is clear that they cannot be taken for granted. They are rather culture- and even individual-dependent and will be subject to an everlasting dispute. As a matter of fact, this is said to be the main reason why knowledge-based approaches have been discarded as a feasible strategy for language processing and, inside the UNL Program, this is probably the reason why there is so much resistance on adopting a more fine-grained level of lexical description.

In fact, outside the UNL Center, it has been observed a relatively flat use of the suffixes of UWs, as if their only role was to assign some part-of-speech information to the roots and to disambiguate between nouns, verbs, adjectives and adverbs. Even though UWs as simple as ‘book(icl>thing)’ or ‘book(icl>do)’ can be really uncomplicated and effortless, they are not trouble-free, as they may not totally disambiguate English words and assure precision and robustness to both enconverting and deconverting. In the WordNet, for instance, the English word ‘book’ (presented in Table 1, below), as a noun, may take ten different senses, some of which may not be translated, in Portuguese, by the same single word. In those circumstances, a low-level use of suffixes would not only be insufficient, but mostly misleading. To reduce all senses of “book” to ‘book(icl>thing)’ would be no better than declaring that “book” is a sort of “abstract thing”.

Consequently, the best solution for the limitations pointed out above ought not to be to extinguish the UNLKB, and cause the UNL System to be a strictly language-based representation formalism (which would turn UNL into a mere metalingua), or to deprive the UNLKB, restricting its power and the granularity of its representation. Actually, the answer is to keep improving the UNLKB, but in a rather different perspective, as suggested in the next section.

5 On the ideal structure of the UNLKB

The UNLKB urges to be decentralized. The development of the UNLKB cannot be a single-man activity, regardless of how good this man can be. In order to avoid one-sided decisions on categorization and in order to push for classification standards and protocols, the UNLKB has to be conceived as a really multilateral and multicultural endeavor. In this sense, the UNLKB has to be plural rather than singular. This is to say that there should be allowed many different UNLKBs, or that the UNLKB should

Table 1. English-to-Portuguese correspondence for the noun “book”

English	Definition	Portuguese
1. book	a written work or composition that has been published printed on pages bound together; "I am reading a good book on economics"	livro
2. book , volume	physical objects consisting of a number of pages bound together; "he used a large book as a doorstep"	brochura
3. ledger, leger, account book, book of account, book	a record in which commercial accounts are recorded; "they got a subpoena to examine our books"	registro
4. book	a number of sheets ticket or stamps etc. bound together on one edge; "he bought a book of stamps"	álbum
5. record, record book, book	a compilation of the known facts regarding something or someone; "Al Smith used to say, 'Let's look at the record'; "his name is in all the recordbooks"	registro
6. book	a major division of a long written composition; "the book of Isaiah"	livro
7. script, book , playscript	a written version of a play or other dramatic composition; used in preparing for a performance	livro
8. book , rule book	a collection of rules or prescribed standards on the basis of which decisions are made; "they run things by the book around here"	livro
9. Koran, Quran, al-Qur'an, Book	the sacred writings of Islam revealed by God to the prophet Muhammad during his life at Mecca and Medina	Livro
10. Bible, Christian Bible, Book, Good Book, Holy Scripture, Holy Writ, Scripture, Word of God, Word	the sacred writings of the Christian religions; "he went to carry the Word to the heathen"	Livro

behave as a distributed network, where many different (even contradictory) repositories of information are somehow interrelated, according to some specific topology that should be addressed by the UNL Center. Ultimately, the UNLKB has to be turned into a network of networks, or the UNL World Wide Web itself.

The fact is that the UNL Center's KB can no longer be considered the sole parameter for verifying the correctness and adequacy of UNL expressions, but should be taken as a reference to be pursued by any Language Center involved in the UNL Program. The UNL Center's KB should play the role of a Core KB, which, at least for

time being, would be responsible for regulating and indexing any other satellite KBs to be available in the UNL Web. However, in order to be taken as such a reference, the UNL Center's KB should undergo some general categorization protocols that are to be discussed and obeyed. These protocols must take into consideration the fact that the UNLKB is a semantic network where UWs are interconnected to emulate human perception and categorization, and that human cognition may vary a lot, between different cultures and even among different subjects. Such considerations ought to govern the whole process in order to avoid excessively naive approaches on ontology and should benefit from the extensive use and study of knowledge representation strategies that have been carried out inside the Artificial Intelligence.

Whatever the case may be, it should be stressed that the answers to the questions presented in the last section are not exactly as simple as they may seem. Actually, they involve the whole philosophy behind UNL and what UNL is supposed to represent. Banishing synonyms, antonyms, variants and other alleged excesses from the UNLKB may obviously impoverish and weaken the representation power of the UNL and will bring consequences that should be considered in the UNL environment. From the human speaker's perspective, "John gave a book to Mary" and "Mary received a book from John", although quite related, may convey different meanings. The same would hold for the difference between "behavior" and "behaviour", which may be used to attest a dialect. In both cases, however, UNL would be far much closer to the semantic and syntactic surface structure of natural language sentences than it would be advisable. Maybe UNL should focus, at least in its very beginning, on a sort of deeper information structure that could be more easily extracted from natural language utterances, so that it would be possible to represent a part, yet infinitesimal, of its alleged meaning. This would be not only more straightforward and faster, but it would also allow for extending the knowledge on natural language syntax and semantics so to provide better results somewhere in the future.

References

- Katz, J. and Fodor, J. The structure of a semantic theory, *Language*, 39, pp. 170-210, 1963.
- Lassila, O. and Swick, R. R. Resource Description Framework (RDF): model and syntax specification. W3C Recommendation, 1999.
- Nirenburg, S, Raskin, V et al. "On knowledge-based machine translation", In *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 1986,
- Sowa, J. F., *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA, 1984.
- Sowa, J. F., *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.
- Uchida, H. and Zhu, M. UNL annotation. Version 1.0. UNL Centre/UNDL Foundation, Geneva, 2003.
- Uchida, H., Zhu, M. and Della Senta, T. *A gift for a millennium*, IAS/UNU, Tokyo, 1999.
- UNL Centre. *Enconverter specifications*. Version 3.3. UNL Centre/UNDL Foundation, Geneva, 2002.
- UNL Centre. *UNL Specification*. Version 3.2. UNL Centre/UNDL Foundation, Geneva, 2003