

II. Maintenance and Optimisation of Multilingual Digital Environment

Universal Dictionary of Concepts¹

Igor Boguslavsky, Vyacheslav Dikonov
UPM/IITP RAS, IITP RAS
Moscow-Madrid
bogus@iitp.ru, dikonov@iitp.ru

Abstract

A universal dictionary of concepts, developed as a part of the ongoing effort to create a semantic intermediary language for global information exchange, is presented. The article describes basic principles and contents of the dictionary and outlines the current state of the project. The dictionary can evolve into an open and freely available language-independent resource with many potential applications. For example, the extensible dictionary of concepts can serve as a pivot to uniformly record and link meanings of words of different languages and facilitate creation of bi- and multilingual dictionaries. Another possible use is word sense markup of corpora. It could bring rich extra benefits due to the fact that the same set of concepts is going to be linked with major world languages including Russian, English, Spanish etc. and supported by multiple text analysis tools. There is a possibility of cooperation and exchange between this dictionary project and other projects, which could enhance the output and eventually spare a lot of parallel effort.

Keywords: Universal Networking Language, universal dictionary of concepts, universal word, ontological structure, argument structure, semantic web, conceptual network

1. Introduction

This article is dedicated to the creation of a new linguistic resource – the Universal Dictionary of Concepts (UDC), also known as the UNL Dictionary. It is a part of a broader international effort to develop a semantic intermediary language named the Universal Networking Language (UNL) (Boguslavsky et al., 2005; Iraola, 2003). Although the dictionary is closely associated with the UNL language, it has considerable value of its own and can be used as a standalone resource for different scientific and practical tasks not related with UNL.

1.1. What is UNL?

UNL is an artificial language for global information exchange in computer networks (<http://www.undl.org>). Unlike Esperanto, it is not a language for direct oral communication, but a formal way to record the meaning of a natural language text. The goal of the UNL project is to produce a worldwide standard for language-neutral storage and exchange of textual information in multilingual environment. A document written in UNL can be automatically deconverted into a text in any language. Traditional automatic translation systems often fail to produce correct translation because of inherent ambiguity of the source natural language. UNL offers a possibility to edit the intermediate representation of text and/or interactively guide an enconversion system to achieve practically unambiguous representation of the source text. When used as a pivot, it ensures that the meaning of the document is always adequately expressed. UNL is a powerful tool to capture the meaning of a text and preserve it through translation and linguistic processing. It is also well suited for precise search, knowledge extraction, and AI applications.

¹ The study and preparation of these results have received funding from the EC's Seventh Framework Programme [FP7/2007-2013] under grant agreement 211938 MONDILEX. The authors are also grateful to the Russian Foundation of Basic Research for partially supporting this research (grant No. 08-06-00367).

The UNL project offers much more than the dictionary. Other linguistic resources include specifications of the language and multiple software tools, which provide translation to UNL (conversion) and from UNL (deconversion) into different languages of the world. There are several groups of linguists and computer scientists participating in the UNL project and supporting different natural languages. Such groups work in Russia (English, Russian), Spain (Spanish), France (French), Egypt (Arabic), India (Hindi, Marathi, Urdu), Brazil (Portuguese) and several other countries.

1.2. UNL Representation of Text

The UNL representation of a text is a semantic hypergraph. It consists of nodes linked with semantic role relations and embellished with attributes, which convey various grammatical meanings and attitudes of the author. A node can contain either a single lexical unit of UNL or another graph, as shown in Figure 1. The latter type is known as hypernodes.

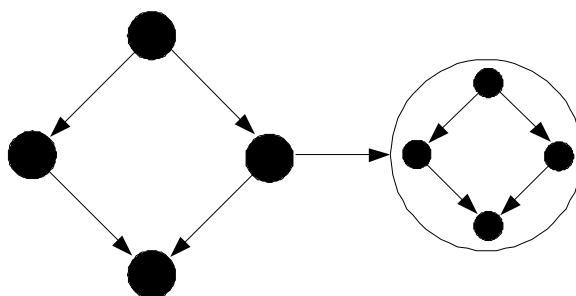


Fig.1. A possible structure of a UNL graph

The basic lexical units of UNL are called universal words (UW). Each UW stands for one single concept.

Although, the principal elements of UNL graphs (UWs, relations and attributes) are technically different in form and function, all of them are just different ways to represent semantic concepts. In some cases it is even possible to choose between using an UW or an attribute, e.g. to express a modal meaning, or prepositional UW and a relation, e.g. for space and time circumstantials. For example, the UW *to(icl>how,plt<uw,obj>thing)* is equivalent to the relation *plt* (target place) and *allow(icl>do,equ>permit,agt>volitional_thing,obj>uw,ben>volitional_thing)* can be an equivalent of the modal attribute of permission. Thus, a UNL graph can be viewed as a pure set of interconnected concepts.

2. Concepts

The concepts of UNL represented in the Universal Dictionary of Concepts are equivalent to the word senses commonly distinguished by explanatory dictionaries. For example, according to the Merriam-Webster, Collins Cobuild, Oxford and other dictionaries of the English language the word *baby* can be used to express the following five concepts:

- a human child,*
- a cub of a mammal animal,*
- an attractive girl,*
- a childish person,*
- a favorite thing, idea or project.*

Each of them is a separate lexical unit in UNL and has a unique identifier (UW). This may seem simple enough, but in fact it is not.

If we take several explanatory dictionaries of the same language, it becomes obvious that there is no unity between the authors in how many senses each word really has and how to define them. As of today, there is simply no exact scientific method to draw borders between different concepts pertaining to the same word of a natural language. The only guide here is common lexicographic practice and practical need to distinguish between different ideas, objects and phenomena of the real world. Therefore, a concept is a word sense ascribed to a natural language word in a set of typical contexts.

It is possible to argue that the concepts from the example above are not elementary and should be viewed as compositional constructs containing simpler elements, e.g. "*baby of a human*", "*baby of an animal*", "*woman whom I treat as gently as a baby*", etc. UNL does not follow this approach and refrains from any attempts to decompose the word senses into smaller semantic units. There are both practical and theoretical reasons for this decision. An essential goal of UNL is to provide a simple and easy to understand and edit representation of the text meaning. Disassembling of every word into a plethora of primitives does not help to achieve it. From a theoretical point of view UNL is a shallow semantic language, which presupposes the possibility of deeper (more detailed) semantic analysis in accordance with the principles of stratification and compositionality. The notion of concepts adopted by UNL and UDC fits well with the lexicographic tradition and facilitates the reuse of data already collected in explanatory dictionaries, thesauri and wordnets.

3. Universal Dictionary of Concepts

UDC describes the inventory of concepts used by UNL and serves as the authoritative and exhaustive lexicon of that language. A UW which is not present in UDC should not be used. Any new UW must be submitted to the dictionary. This is an important point for maintaining the lexical compatibility of UNL documents and software tools for automatic translation into natural languages.

3.1. Highlighted Features

- The Universal Dictionary of concepts strives to include and integrate conceptual lexicons of all natural languages.
- The dictionary is characterized by total absence of polysemy.
- Each concept is represented by a universal word (UW). Normally, there should be only one UW per concept.
- The dictionary does not tolerate homonymy, i.e. when one UW is used to express several different concepts.
- The dictionary does not provide any kind of grammatical or morphological information for the simple reason that there is no use for it in UNL.
- All concepts are derived from natural languages. None of them may be invented artificially and the existence of each concept must be justified by some practical need or supported by lexicographic evidence in some natural language. A small number of special abstract concepts, such as *uw*, *thing(icl>uw)*, *abstract_thing(icl>thing)*, etc. have to be privileged because of internal needs.
- If the dictionary lacks a concept, a new UW can be created on demand.
- The dictionary is more than a simple list. It organizes the concepts into a complex semantic network. The structure of this network is outlined in section 5.2.

3.2. Bringing All Tongues Together

It is a common linguistic fact that each natural language has its own unique set of concepts and there are concepts which are specific to certain languages. In fact, we should not expect that concepts which are truly identical for several languages will constitute the majority. Even very common facts and notions can be treated differently by other languages. For example, the English general concept of "*grandmother*" (*the mother of one of the parents*) does not exist in Swedish. Instead, two different words and concepts are used: "*mormor*" (*the mother's mother*) and "*farmor*" (*the father's mother*). UDC will include all three concepts and many more.

In order to be able to record all natural languages accurately the Universal Dictionary of Concepts should grow into the "Summa Lexicographica" of the human kind. This is an immense challenge, which no single group of linguists can meet. The Universal Dictionary can never be considered complete and can grow forever, because the scientific and cultural progress always adds new concepts. However, a dictionary does not have to be complete in order to be usable. There is a practical threshold where the number of registered concepts becomes sufficient for adequate recording of most texts.

4. Universal Words

This section provides only a brief overview of the UW format. More information and rules for UW construction can be found in (Boguslavsky, manuscript).

Universal Words (UW) are used in the dictionary in order to represent the concepts unambiguously. The inventor of the UW format H. Uchida made a lot of effort to achieve intuitive understanding of the concepts on the basis of the UWs alone, without any additional explanation. Nevertheless, most UWs are supplied with a short definition and an example (currently only in English).

A UW consists of a headword and a list of constraints used to differentiate between different concepts associated with the headword and provide additional information. A constraint consists of a UNL relation and another UW, usually reduced to its headword. The general UW format is:

headword(relation>uw>uw,relation>uw,...)

The headword is usually an English word.

cut(icl>wound>thing)

If the new concept is expressed by a phrase, the phrase becomes the headword. Spaces are replaced with underscores.

morse_code(icl>code>thing, equ>morse)

If there is no corresponding word in English and the concept is a hyponym of some already existing one, we should only change or add constraints. The first of the following three UWs stands for a general concept of entering into a marriage. The other two are its hyponyms describing two aspects of the action differentiated by some languages.

marry(icl>do, agt>person, obj>person)
marry(icl>do, agt>man, obj>woman) marry(icl>do, agt>woman, obj>man)

If the new concept is culture-specific and has no hypernym in English, we can use the native word transliterated into Latin and supplement it with constraints that would link it with the nearest commonly known class of objects.

tarator(icl>soup(icl>food)>matter)
lapot(icl>footwear>..., equ>bast_sandal, com>russian_peasantry)

UW constraints convey only a minimal amount of information required for identification of concepts. There are three types of constraints: ontological, semantic and argument.

Ontological constraints reflect the most important links between concepts: hypernymy (icl), meronymy (pof), instantiation (iof).

tongue(icl>concrete_thing, pof>body)
madrid(iof>city)

Semantic constraints are used to show the difference between several concepts associated with one headword: synonymy (equ), antonymy (ant), association (com).

ably(icl>how, equ>competently, ant>incompetently, com>able)

Argument constraints reflect the semantic frame of the concept: agent (agt), object (obj), second object (cob), source (src).

buy(icl>get>do, agt>person, obj>thing, cob>thing, src>thing)

More detailed information about the relations between UWs is going to be stored in the semantic network of the Universal Dictionary of Concepts.

5. Structure of the Dictionary

The Universal Dictionary of Concepts must include three principal components:

1. the repository of concepts, commonly referred to as the dictionary of UNL;
2. the network of relations between concepts, which is known as the UNL Knowledge Base (UNLKB)²;
3. the local dictionaries, which link concepts with words of various natural languages.

5.1. Inventory of Concepts

The inventory of concepts is a collection of all concepts available in the dictionary and the UNL language in the form of a flat list of UWs. There is no distinction between UWs for concepts coming from different languages. **All concepts are equal as separate lexical units** of UNL and listed together.

In principle one concept should be represented by only one UW. However, it is hardly possible to avoid a situation when several different UWs for the same concept appear. It may happen due to technical and organizational reasons in a decentralized community and the dictionary must provide adequate means to handle this situation.

The first and easiest case is when an already existing UW is modified in order to correct an error, achieve better disambiguation or supply missing information. The old version of the UW cannot be deleted immediately, because it can be used by existing UNL documents (or linked to by other resources). Simple deletion would render such documents incompatible with the dictionary. Although all UNL-related software tools must be able to process documents with unknown UWs, the percentage of such UWs should not exceed the level when it starts to affect the quality of translation. The dictionary has to support per-UW history of changes, allowing to trace any registered version of the UW and prevent reintroduction of deprecated UWs in the same version of the dictionary.

The second source of different UWs for the same concept is the very nature of human language and categorization processes. Each natural language contains a certain amount of exact synonyms which may or may not drift apart with time, e.g. *everyone* and *everybody* in English. It is extremely difficult to build a definitive list of them. Therefore, people will keep adding multiple UWs based on such words even if the corresponding concept already has an UW.

Both processes effectively create groups of UWs resembling synsets used by the Wordnet family of dictionaries. Such groups could be distinguished among all synonyms, viewed as close yet different concepts.

5.2. Network of Concepts

The concepts create a semantic network linked by the relations of hypernymy, meronymy, instantiation, synonymy, antonymy, association and various other relations describing argument frames. The goal of the semantic network is to provide description of the links between concepts, that exist in the human languages and minds, and make it as objective as possible.

The network of concepts consists of three separate structures formed by a) the ontological relations, which link the concepts with different semantic classes, b) semantic relations, which reflect similarity or contrast between concepts, and c) argument relations, which specify what classes of concepts can fill argument slots of each concept.

5.2.1. Ontological Structure

The ontological structure consists of the **icl** (hypernymy), **pof** (meronymy) and **iof** (instantiation) relations. They can be supplemented with some other types of relations, such as **val** (value of) and **scn** (domain of).

² In older UNL publications UNLKB can be referred to as the Master Entries dictionary. This name is related with the idea of Master Definitions of UWs – an extended form of UWs, which contains full set of relations with any other concepts. Currently the master definitions are not used, but they can easily be derived from UNLKB.

The *icl* and *iof* relations have a privileged status because it is obligatory for every UW to specify at least one more general ontological class through these relations. A concept should be linked to all classes, an immediate member of which the concept is. The result is a hierarchy of ontological relations embedded into a network of other relations. Hypernymic classes are hierarchical by nature and with certain approximation can be arranged in the form of a tree, although the real relations between them can be more complex (see Figure 6). UDC offers a more robust and realistic way to represent the relations between classes of concepts than a regular tree. The resulting base structure is a hybrid one. It combines features of a tree and a network. The branches may split and later join, as shown in Figure 2, yet there is a common root.

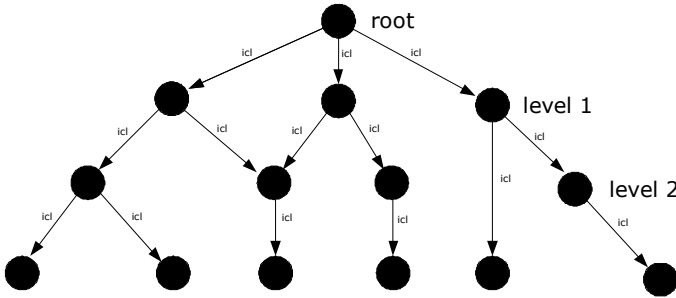


Fig.2. Ontological structure

The abstract root class is named “uw” (any universal word) and divided into further abstract classes of objects, attributes, actions, states, etc. It is possible to talk about different levels of the ontological structure, but a concept in UDC may belong to more than one level or branch.

Ontological relations make it possible to trace the relative semantic volumes of concepts and find more general terms if no direct translation is possible into the target language. For example: while translating the Russian word *жениться*, which means literally “to acquire a wife” and has no exact equivalent in English, we should replace it with the more general concept “to become married”, which has a straightforward translation.

5.2.2. Semantic Structure

The semantic structure has a different layout. It consists of the semantic relations **equ** (synonymy), **ant** (antonymy) and **com** (association). The equ relation does not distinguish between real and quasi-synonyms and can be supplemented with other technical means to mark sets of UWs denoting the same concept. The semantic relations unite groups of concepts and do not form any hierarchy. Therefore, the resulting structure is a pure decentralized network, as shown in Figure 3.

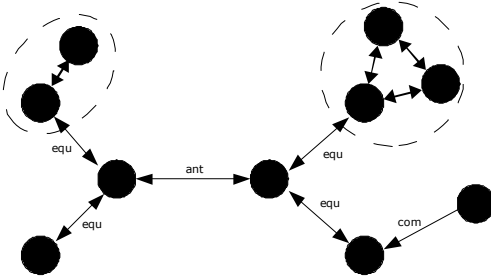


Fig.3. A fragment of semantic structure

There is no requirement for the semantic structure to be connected, unlike the ontological one. It may consist of multiple isolated fragments.

5.2.3. Argument Structure

The argument structure is a collection of argument relations, e.g. **agt** (agent), **obj** (object), **ptn** (partner), **ben** (beneficiary), **plt** (target place), **src** (source), **gol** (resulting state), etc., connecting each concept with an argument frame and general class concepts, which unite all specific concepts that normally fill respective argument slots. In most cases the argument relations point to concepts which belong to a relatively compact group of the most general ontological classes, which occupy the topmost levels of the ontological structure (Figure 4).

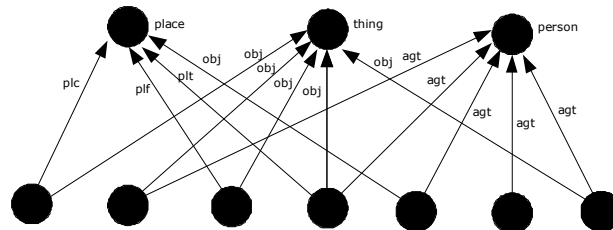


Fig.4. Argument structure

All three structures link the same concepts and are superimposed on each other, forming the network of concepts of UDC.

5.3. Local Dictionaries

Local dictionaries are optional parts of the Universal Dictionary. They are used to connect concepts with the vocabularies of different natural languages. Each language should have a local dictionary in order to be supported. The local dictionaries can be just flat lists enumerating pairs of concepts and their translations into the target language. The natural language words may be supplied with grammatic information.

A translation does not have to be one word. Some concepts represented by a single word in one language may be translated into another by multiword phrases and abbreviations, e.g. *senior pupil* or *VIP*.

However, not all concepts can be translated into all languages even descriptively. If there is a need to translate such a concept, a nearest general term or a more specific one can be found via the network of concepts. Figure 5 provides an example. It outlines relations between Russian (left) and Bulgarian (right) words for *pen*, *handle*, *knob*, *stem* and *tiller* with UWs as a pivot. There is no direct equivalent in Russian for the Bulgarian word *дръжка* in the sense of *stem of a plant*. The translation must be chosen by tracing the ontological (icl) links between *stem of a fruit* and *stem of a flower*. Additionally, there are two alternative Bulgarian translations for the concept *pen*.

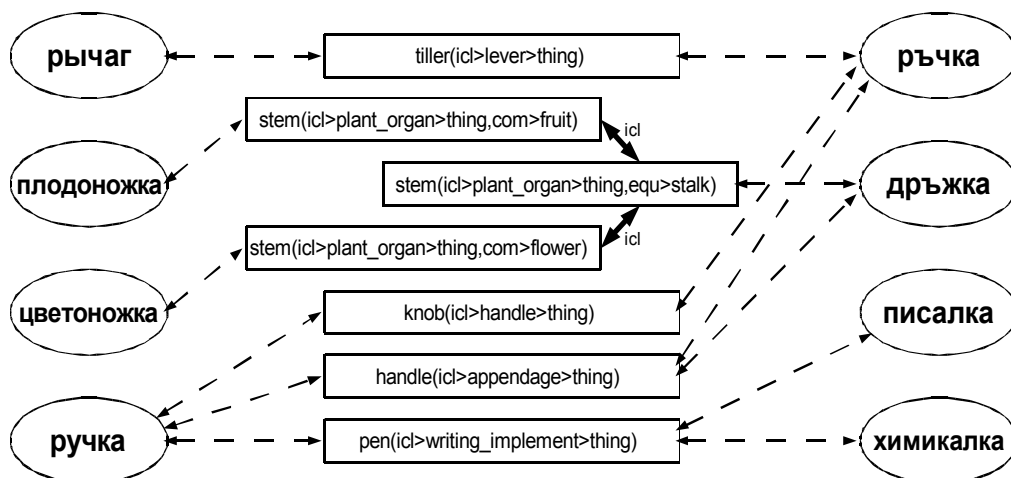


Fig.5. Concepts and possible links between some Russian and Bulgarian words

6. Universal Dictionary of Concepts and Wordnet

The Universal Dictionary of Concepts is quite similar to the well-known Wordnet family of dictionaries in many important aspects. Both have concepts as their basic units and define similar relations between them. A lot of data have been imported from Princeton Wordnet (Bekios et al., 2007). Even more information, including concepts and relations (Iraola 2003), can be imported from different existing Wordnets into the Universal Dictionary of Concepts. However, there are some important differences between UDC and Wordnets.

6.1. Relation to Natural Languages

Each Wordnet describes the lexical system of a particular language and each language is maintained separately. Wordnets may be interconnected by means of the Inter-Language-Indexes (ILI), which describe the relations between the concepts of certain versions of the original Princeton Wordnet (typically 1.5 or 1.6) and concepts of other national Wordnets. However ILIs play a subsidiary role. Only some non-English Wordnets are linked to the original Princeton Wordnet and such links get outdated as soon as a new version of it is released.

The Universal Dictionary of Concepts can be compared to several Wordnets linked through ILI, but it has no bias towards any particular language. The emphasis is given to the unified inventory of concepts and their relations. Links to vocabularies of natural languages are provided through optional local dictionaries and do not have to be discarded when changes are made in the repository of concepts and the semantic network.

The fact that most of the UW headwords come from English and the constraints in so many UWs are motivated by the need to describe the polysemy of English words, might suggest that the dictionary uses English as a pivot or “gold standard” to describe other languages. However it is not quite true. English headwords and constraints were chosen for mere practicality, because most linguists understand this language and it uses the most common and well supported A-Z script in the world. It is also a fact that not all UW headwords are English.

Concepts coming from any language receive identical status. Concepts originating from different languages can have direct links between each other. Non-English concepts may also be used as a base for modification and as constraints to describe other concepts. For example:

```
samovar(icl>boiler>concrete_thin,com>tea)
tula_samovar(icl>samovar>concrete_thing,com>tula(iof>city))

sauna(icl>sweating_room>place,com>finnish,com>dry)
parilka(icl>sweating_room>place,com>russian,com>steam)
venik(icl>massage_tool>...com>parilka(icl>sweating_room))
```

If the number of concepts unique to other languages increases, the statement about the special role of English in UDC will lose ground.

6.2. Hierarchical Structures

Wordnets organize the noun and verbal concepts into hypero-hyponymic hierarchies represented as trees. Such structures are easy to search and analyze, but pure tree classification does not support partially intersecting classes and works well only for the top classes of ontology. For example, Princeton Wordnet has concepts of (*tennis*) *racket*, and (*hockey*) *puck* as well as a class for “*sports implements*”. However, *racket* is a member of the class of sports implements and *puck* is not. Instead it is a member of the class of “*disk objects*”. Moving *puck* to the “*sports implements*” class in a pure tree structure would cause losing information that it is a disk.

UDC is able and strives to accommodate a different less formally hierarchical approach. The basic ontological structure is a network graph which has only some features of a tree. It is normal to have multiple parents to the same daughter node, which allows for more complex relations and more fine-grained classification. Every concept should be linked to all possible immediate hypernyms. For example, the word *sushi* in Wordnet is a direct daughter of the concept *dish* (food). Suppose that we want to introduce further ontological divisions by nationality (*sushi* is a Japanese dish) and primary

ingredient (sushi is made of fish). It is not possible to decide which of the two classes has to be placed higher in the hierarchy, because these classes specify intersecting sets of concepts (Figure 6)³.

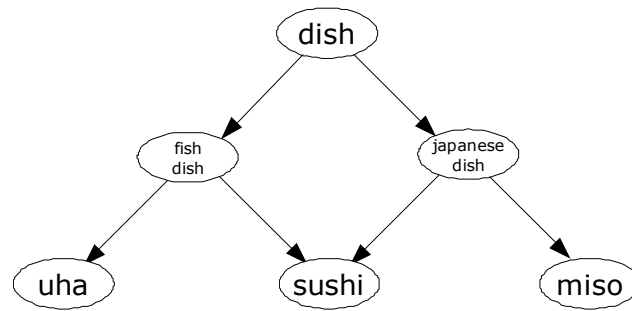


Fig.6. Multiple parent classes

Using a network instead of a tree has some implications. A tree structure allows to trace every concept to its deepest root classes with full confidence, whereas the hybrid network structure permits multiple paths, leading to different high-level classes for the same concept, even when it creates confusion. For example, the class “*functional thing*”, which includes the concept of *hammer*, is a daughter of both “*abstract thing*” and “*concrete thing*”, thus making *hammer* a possibly non physical object! This problem can be remedied in UWs by providing a secondary direct link to the relevant top class.

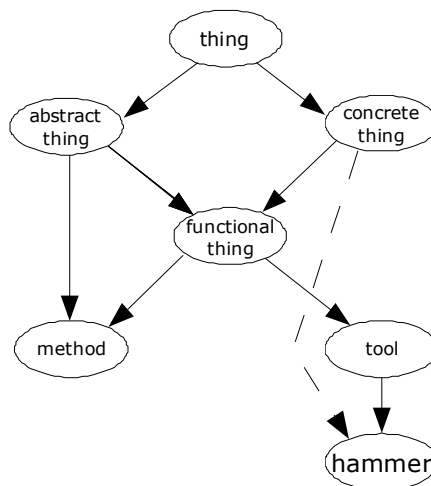


Fig.7. Additional link to the relevant top class

According to Figure 7, the UW for the concept *hammer* should be *hammer(icl>tool>concrete_thing)*. Knowing two ends allows to trace the ontological relations between any concept and the relevant top class and produce full hierarchy.

6.3. Other Features

Wordnet does not make the difference between hypernymy as a relation between classes (e.g. the class of “*living things*” includes the class of “*plants*”) and instantiation as the relation between an individual and a class to which it belongs, e.g. *Deli* is a member of the class “*cities*”. In UDC two different relations are used for such cases: **icl** for hypernymy in *plant(icl>living_thing)* and **iof** for instantiation in *Deli(iof>city)*.

³ Princeton Wordnet provides a way to include a synset into several classes at the same level of its hierarchy too, but this is not common. For example, *key* in the sense of “*a kilogram of a narcotic drug*” is described as both “*a mass unit*” and “*a metric unit*” at the same level and this split is immediately joined at the next level under the “*units of measurement*” class.

UDC does not limit itself by certain parts of speech like Princeton Wordnet and provides full set of concepts for prepositions, conjunctions and some words with special grammatical functions, e.g. modal verbs.

UDC provides more detailed semantic frame information, not limited to the verbal concepts. The roles are annotated with UNL relations and prototype semantic classes of the arguments are given where Princeton Wordnet offers only “somebody” and “something”.

Some wordnets preserve syntactic information about the words, such as part of speech, gender, animacy, etc. (Сухоногов, Яблонский, 2004), while other are coupled with morphology engines. This is not the case in the Universal Dictionary because such information is unneeded in the UNL language. Its proper place is in the local dictionaries.

7. Development of the Dictionary

The development process should follow the essential principles of **division of labor, gradual development, reuse of existing data and decentralization**. A community model, where everyone checks everyone and all significant disputes are resolved by experts, is the best option, because no single authority can have enough resources and expertise to verify everything.

Every time when a significant amount of changes is done and no formal objections received, a snapshot of the dictionary should be taken and released as a new version. From that moment all participating parties must update their tools to use the new dictionary. An automated system to propagate UW changes to local copies utilized by linguistic processors supporting UNL is required to ensure smooth transition to any new versions of the dictionary.

7.1. Current Status

At the moment of writing the Universal Dictionary is under active development. It has already passed a number of important milestones including: adoption of the common UW guidelines (Boguslavsky, manuscript) and creation of the initial set of UWs completely covering the general vocabulary of English. The current version of the dictionary includes about 200 000 UWs generated on the basis of the Princeton Wordnet (Bekios et al., 2007) and about 9 000 UWs (Диконов, 2008) created manually to fill in the gaps found in Wordnet. The manually written UWs cover English prepositions, conjunctions, and certain other words left out of Wordnet. A significant portion of them replaces the automatically generated UWs for the most frequent English verbs and nouns in order to improve the quality of the UWs.

The existing inventory of UWs was merged with the dictionaries of the linguistic processor ETAP (Диконов, 2008), developed by the members of the Russian group, and is used for text conversion from and deconversion to English and Russian. The automatically generated UWs are available online at <http://www.unl.fi.upm.es/unlweb>.

The French group develops an infrastructure for the central data repository and exchange of data between different groups. Considerable effort is made by different participants towards massive revision and correction of the generated UWs.

The next step can be enriching the semantic network beyond the links already available in the form of UW constraints.

7.2. Availability

The Universal Dictionary is going to be released to the public under a free license as soon as the first version will be ready, which presupposes merging in more UWs from other UNL groups and putting in operation the infrastructure for automated data exchange.

The essential principles to be maintained are:

- The Universal Dictionary of Concepts will be available to the public free of charge.
- The data may be used freely for any purpose, though commercial use may be a subject to special conditions.
- Everyone will be given the right to expand the resource and fix errors, provided that all modifications will be returned to the community of dictionary users and editors.

8. Possible Uses and Related Projects

The dictionary of concepts can be used as a standalone resource to match words of different languages for automatic generation of multilingual dictionaries, provided that all such languages have local dictionaries.

Universidad Politécnica de Madrid (UPM) runs a project named Patrilex (Boguslavsky et al., in print) which is aimed at experimental verification of this approach. The practical goal is to produce a multilingual dictionary of terminology in the domain of culture and national heritage for the Spanish Ministry of culture. A special custom set of UWs for the relevant terms is being built and independently translated into English, Spanish, Russian and Arabic. The translators receive flat lists of the UWs without any additional information and independently write local dictionaries for their languages. The resulting multilingual dictionary will be assembled automatically and verified to detect any problems.

Another possible use is to annotate lexical meanings after word sense disambiguation, e.g. for semantic annotation of corpora. There is a need for a reference corpus of UNL, but it is not yet created. The most relevant effort in this field is the project to translate the Encyclopedia Of Life Support Systems (EOLSS) into several languages via UNL.

The overall progress of the UNL project may seem slow, but current projects show that it is real. A quantum leap is expected as soon as the first public version of the Universal Dictionary is released and the tools for automatic conversion of text into UNL documents reach industrial quality. Every new related project and contribution make this perspective closer.

References

- Bekios, J., Boguslavsky, I., Cardeñosa, J., Gallardo, C. (2007). "An Efficient Method for Building Multilingual Lexical Resources" In: Proceedings of the *Fifth International Conference Information Research and Applications* i.TECH 2007, Varna, Bulgaria, v.1. Sofia, Ithea, pp. 39-45.
- Boguslavsky, I., Cardeñosa, J., Gallardo, C. (in print). "A Novel Approach to Creating Disambiguated Multilingual Dictionaries".
- Boguslavsky, I., Cardeñosa, J., Gallardo, C., and Iraola, L. (2005). "The UNL Initiative: An Overview" Computational Linguistics and Intelligent Text Processing.
- Boguslavsky, I.M. (*manuscript*). "Guidelines for UW construction".
- Диконов, В. (2008). Развитие системы построения семантического представления текста с использованием языка-посредника UNL на базе лингвистического процессора ЭТАП-3. *Информационные технологии и системы, ITiS'08*, Геленджик.
- Iraola, L. (2003). "Using WordNet for linking UWs to the UNL UW" *International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies*, December 2 - 6, 2003, Alexandria, EGYPT.
- Сухоногов, А.М., Яблонский, С.А. (2004). Разработка русского WordNet. RCDL2004, Пущино, Россия.
- Web site of the UNL project, <http://www.undl.org>