Jibiki-LINKS: a Tool between Traditional Dictionaries and Lexical Networks for Modelling Lexical Resources

Ying Zhang, Mathieu Mangeot, Valérie Bellynck, Christian Boitet

Jibiki-LINKS: a Tool between Traditional Dictionaries and Lexical Networks for Modelling Lexical Resources

ZHANG Ying[1, 2] Mathieu MANGEOT[1] Valérie BELLYNCK[1] Christian BOITET[1]

1. GETALP-LIG, 41 rue des Mathématiques BP53, 38041 Grenoble Cedex

2. SAS Lingua et Machina, Domaine de Voluceau, Rocquencourt, 78153 Le Chesnay

{ying.zhang, mathieu.mangeot, valerie.bellynck, christian.boitet}@imag.fr

Abstract

Between simple electronic dictionaries such as the TLFi (computerized French Language Treasure)[1] and lexical networks like WordNet[2] (Diller et al., 1990; Vossen, 1998), the lexical databases are growing at high speed. Our work is about the addition of rich links to lexical databases, in the context of the parallel development of lexical networks. Current research on management tools for lexical databases is strongly influenced by the field of massive data ("big data") and by the Web of data ("linked data"). In lexical networks, one can build and use arbitrary links, but possible queries cannot model all the usual interactions with lexicographers-developers and users, that are needed, and derive from the paper world. Our work aims to find a solution that allows for the main advantages of lexical networks, while providing the equivalent of paper dictionaries by doing the lexicographic work in lexical DBs.

1 Introduction

The growing importance of IT in all human activities extends and expands the needs and usages of all key digital resources that include lexical resources. Thus, while applications valuing the linguistic processes rely on increasingly abstract representations, modelled for computer operations, it remains that models coming from the historical construction of resources foster human understanding, and therefore, the building of tools for studies centring on the humanities.

In this this section, we place the emergence of the concept of lexical database between electronic dictionaries and lexical networks. We show that this concept is still valid, that it is still necessary to enrich it, and that our work on improving tools for lexical databases helps solve real problems.

To do this, we analyse in the second section the evolution of lexical resources in 4 main steps (simple electronic dictionaries, simple lexical databases, multilevel and multiversion lexical databases, and lexical networks) and present the associated problems. In the third section, we present Jibiki-LINKS, a platform for building multilingual lexical databases that enriches the Jibiki generic platform by introducing the concept of *rich link* between the components it manages (dictionary entries and dictionary volumes). Finally, we show that it allows the construction of lexical databases such as Pivax-UNL, which support scaling up.

2 From computerized dictionaries to lexical databases with rich links

The first computerized lexical resources are electronic versions of printed dictionaries, mainly monolingual or bilingual. The use of computers has helped to overcome the constraints of the paper form. The impossibility to inverse bilingual dictionaries led to a model having a "pivot" consisting of axies[3]. Lexical pivot-based databases are invertible and transitive, but rooted on the form of the

[1] http://atilf.atilf.fr

[2] http://wordnet.princeton.edu

[3] "Axie" = "interlingual meaning," by analogy with "lexie".

Zock/Rapp/Huang (eds.): Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon, pages 87–98,
Dublin, Ireland, August 23, 2014.

symbols, while the lexical networks allow a move towards the direct manipulation of semantic tokens, regardless of their surface form, and thus of the language.

In this section, we present the evolution of approaches, distinguishing four main types of lexical resources, the limitations that motivated this evolution, and the remaining hard problems.

2.1 Simple electronic dictionaries

A simple electronic dictionary is an electronic version of a printed dictionary, or the computer representation of a new kind of the same type of dictionary, for example, the TLFi[4], the morphological and bilingual dictionaries of Apertium[5], etc. A simple electronic dictionary contains either one volume or two volumes. The electronic version of a monolingual paper dictionary is (usually implicitly) based on its *microstructure*, that is to say, on the organization of its entries in the form of a small tree organizing the information it contains. In a paper dictionary, the presentation of an entry reflects the microstructure, but the microstructure is not always directly retrievable from it (for example, parts in italics can correspond to different types of information units, such as idiom or example of use).

In absolute terms, it is always possible to represent the information specified in each entry of a dictionary according to a common structure. In reality, the structures of paper dictionaries are less rigorous than what would be required for automatic processing, so that manual editing is required.

A bilingual paper dictionary is generally based on a structure in two volumes, one for each language pair, each volume conforming to the same microstructure. There are therefore generally one volume from language A (Lg A) to language B (Lg B) and a mirror volume from Lg B to Lg A. We define the *macrostructure* of a dictionary as the organization of the volumes that make up its structure. These macrostructures constitute the bulk of the printed dictionaries.

2.2 Lexical databases

A lexical database is a tool for unifying any set of dictionaries, where each dictionary can be monolingual, bilingual or multitarget. A multilingual lexical database is composed of volumes that are monolingual, direct multilingual, or indirect multilingual, i.e. connecting the entries of different languages via a pivot structure. It has an overall macrostructure, and a microstructure for each of its volumes. A link between 2 entries is realized by the software tool as a direct link, or as 2 links going through an intermediate language, or as a semantic link, etc.

The lack of symmetry of the correspondence between the entries of bilingual dictionaries (from word senses to words, not word senses) led to the concept of *interlingual pivot*. In the pivot macrostructure developed and used for the Papillon-NADIA multilingual lexical database (Sérasset and Mangeot, 2001), there is only one monolingual volume for each language. *Lexies* are word senses (of a lexeme or an idiom) and make up the entries of these volumes. To group the lexies of different languages together, there is a pivot volume of *axies* (interlingual acceptions). An axie connects synonymous lexies. The links are established only between lexies and axies. This is the simplest macrostructure for a pivot-based multilingual lexical resource that allows for the extraction of usage dictionaries for all pairs in all directions. The concept of axie-based pivot structure has been validated by the Papillon project and then included in the Lexical Markup Framework standard (Francopoulo et al. 2009).

2.3 Multilevel and multiversion databases

In this type of lexical database, several monolingual volumes are allowed for each *lexical space*[6], A volume of *axemes* (monolingual acceptions) is introduced to link synonymous lexies of the considered lexical space. Also, various levels are introduced to tag entries according to different points of view (sublanguage, version, type of link, reliability, preference). The simple links of previous versions are replaced by *rich links* that can be established not only between lexies, axemes and axies, but also between entries and subentries, monolingually (lexicosemantic functions) or bilingually (translations).

---

[4] Trésor de la Langue Française informatisé, http://atilf.atilf.fr/

[5] http://wiki.apertium.org/wiki/User:Alessiojr/Easy_dictionary_-_Application-GSOC2010

[6] A lexical space of a natural language contains various levels (wordform, lemma, lexie, prolexeme, proaxeme); it can also contain the lexical symbols of an artificial semantic representation language (e.g., the UWs of UNL).

For example, there is a 3-level macrostructure (lexie, axeme, axie) in PIVAX (Nguyen & al., 2007) and a 4-level macrostructure (lexie, prolexeme, proaxie, axie) in ProAxie (Zhang & Mangeot, 2013), described in more detail in section 4.1. Both allow us to manage one or more monolingual volumes for each lexical space. That has been quite useful in the ANR Traouiero GBDLex-UW++ subproject, during which we stored the UNL part of many UNL-Li dictionaries (the UW interlingual lexemes, built with slightly different conventions by different UNL groups for their languages), and tried then to unify them in a new monolingual UNL dictionary (using a set of "UW++" built from WordNet and from the previous UNL dictionaries).

2.4 Lexical network

A lexical network brings together the set of words that denote ideas or realities that refer to the same theme, as well as all the words that, because of the context and certain aspects of their meaning, also evoke this theme[7]. The theme may possibly be very broad. It is possible to represent the full vocabulary of a language in a lexical network, such as, for French, the JeuxDeMots network (Lafourcade and Joubert, 2010) or RFL (Lexical Network of French (Lux-Pogodalla, Polguère 2011)).

Lexical networks are traditionally represented as graphs. Nodes represent the lexemes of one or more languages, and links represent the relationships between these lexemes (translation, synonymy, etc.). A lexical network can be monolingual or multilingual. One can create syntactic, morphological and semantic relations between lexemes.

Although lexical networks have many advantages, they are not suitable for all usages. For example, lexical networks like WordNet (Diller & al., 1990; Vossen, 1998), HowNet (Dong et al., 2010) and MindNet (Dolan and Richardson, 1996) (Richardson et al., 1998) are not browsable in alphabetical order. But we need that possibility to have an idea of the content of a lexical repository, whatever its nature, or to play word games, or to find a word one has on the tip of the tongue[8]. On the other hand, in a lexical network, the concept of volume is missing, which prevents to create a resource in a simple way when studying a new language.

For example, the lexical network DBNary (Sérasset, 2012), which is based on the Lemon model (McCrae et al., 2011), contains millions of terms, but does not allow labelling the links. To navigate in this system, one must write SPARQL queries, which is not within the reach of everyone.

2.5 Conclusion: features, limitations and hard problems

Research efforts focus today mainly on lexical networks, but much remains to be done on the preceding types (pivot, multilevel). In particular, the import of lexical databases in lexical networks causes a loss of information, especially information born by the attributes of rich links. For example, what concerns the history, the etymology or the evolution of word senses is not systematically imported into lexical networks. They therefore cannot meet the needs of the humanities, nor allow the transition to "digital humanities."

A lexical network is actually the type of structure that enables the greatest freedom of representation. Indeed, we can create entries and links arbitrarily. But the possible queries cannot model all the usual interactions with lexicographers-developers and users, which come from the world of paper, and are felt necessary. They allow us to represent all categories of lexical resources, but the analogy with the real world is lost. Thus, the practical expertise of linguists-lexicographers is lost.

We must continue to equip lexical databases, because that is the right level to transfer the techniques used by lexicographers-linguists. Also, modelling by a volume-based macrostructure allows keeping a link to the original paper world. Moreover, there are already reusable resources of these types. That is why we focus on the management of resources having multiversion and multilevel macrostructures.

3 Reuse of rich links

In this section, we present an improvement that consists in introducing into lexical databases relational

[7] http://ddata.over-blog.com/xxxyyy/3/12/82/15/GRAMMAIRE/champs-et-reseaux-lexicaux.pdf

[8] For that kind of functionality, multiple sorting on subsets of inflected forms and on arbitray types of information seems to be a necessary first level of computer aid.

information in the form of *rich links* that will bring them closer to lexical networks. An important point is that these links may bear arbitrary labels.

3.1 Presentation of the Jibiki platform

Jibiki is a generic platform that enables the construction of contributive websites dedicated to the construction of multilingual lexical databases. That platform has been developed mainly by Mathieu Mangeot (Mangeot & Chalvin, 2006) and Gilles Sérasset (Sérasset & Mangeot, 2001). It has been used in various projects (EU LexALP project, Papillon project, GDEF project, etc.). The code is available in open source, and freely downloadable by SVN from ligforge.imag.fr. With this platform, one can perform import, export, edit and search operations in lexical databases. One can also manage the contributions. Jibiki allows handling almost all lexical resources of XML type, by using different microstructures and macrostructures.

In the Jibiki approach, resources are organized in *volumes*, which makes it easier to achieve the equivalent of paper dictionaries, keeping the mental image of the representation of the dictionary, while offering new interactions allowed in the digital world. Usages of dictionaries in Jibiki are also similar to those of paper dictionaries. For example, one can consult a database in alphabetical order, indicate a source and/or target language, group lexies in vocables, navigate in a volume, etc.

3.2 Classical Common Dictionary Markup

Version 1 of Jibiki uses "CDM pointers" (Common Dictionary Markup (Mangeot, 2002)) to import, view and edit any type of microstructure without modifying it. CDM pointers are also used to index specific parts of the information, and then allow a multi-criteria search.

Each CDM pointer indicates the path (XPath) to the corresponding element in the XML microstructure of the described resource (see Figure 1). Its description is stored in a XML metadata file. When the resource is imported in the Jibiki platform, the pointers are computed, and the result is stored in a table of the (postgresql) database, for each volume. This table is considered as an indexing table.

```
<cdm-elements>
 <cdm-volume xpath="/g:volume"/>
 <cdm-entry xpath="/g:volume/g:article"/>
 <cdm-entry-id xpath="/g:volume/g:article/@g:id" />
 <cdm-headword xpath="/g:volume/g:article/g:vedette/g:mot/text()" d:lang="fra" />
 <cdm-pronunciation xpath="//g:prononciation/text()" d:lang="fra" />
 <cdm-pos xpath="//g:cat-gram/text()" d:lang="fra" />
 <cdm-definition xpath="/g:volume/g:article/g:vedette/g:mot/text()"/>
</cdm-elements>
```

Figure 1: CDM pointers for the French volume of the GDEF[9] resource (Mangeot and Chalvin, 2006)

CDM tags	FeM[10] (Gut et al., 1996)	OHD[11]	JMdict[12] (Breen, 2004)
Volume	/volume	/volume	/JMdict
Entry	/volume/entry	/volume/se	/JMdict/entry
Entry ID	/volume/entry/@id		/JMdict/entry/ent_seq/text()
Headword	/volume/entry/headword/text()	/volume/se/hw/text()	/JMdict/entry/k_ele/keb/text()
Pron	/volume/entry/prnc/text()	/volume/se/pr/ph/text()	
PoS	//sense-list/sense/pos-list/text()	/volume/se/hg/ps/text()	/JMdict/entry/sense/pos/text()
Domain		//u/text()	
Example	//sense1/expl-list/expl/fra	//le/text()	/JMdict/entry/sense/gloss/text()

Table 1: Examples of Common Dictionary Markup

[9] GDEF is a large Estonian-French dictionary that is being created by the Franco-Estonian lexicography association (see http://estfra.ee/GDEF.po).

[10] FeM is a French-English-Malay dictionary (30000 entries, 50000 lexies, 8000 idioms, 10000 examples of use).

[11] OHD is abbreviation of Oxford-Hachette Dictionary, which is a French-English dictionary.

[12] JMdict is a Japanese-multilingual dictionary.

The translation links are treated at this stage with conventional CDM pointers, as classical information elements. It is not possible to index other information carried by the links, such as weights or labels.

Hence, multilevel macrostructures cannot be modelled in a generic manner with Jibiki-v1 and traditional CDM pointers. For example, it is not possible to link the same volume to several volumes at different levels. This has forced us initially to use palliatives that did not scale up. It became necessary to modify the conceptual model. We addressed these shortcomings in a new version, Jibiki-LINKS.

Table 1 above is an example of CDM for the different resources.

3.3 New version of Jibiki with CDM LINKS

To manage multilevel macrostructures, we enriched the CDM with a richer description of the links (see Figure 2). For each link, more information can be indexed:

- the identifier of the source entry.
- the identifier of the target entry.
- the identifier of the XML element of the source entry containing the link. For example, the sense number in a polysemous entry having a translation link for each translation direction. That allows us to precisely retrieve the origin of the link.
- the link name. It is used to distinguish between different types of links in a single entry, such as a translation link and a synonymy link.
- the target language (three-letter code ISO 639-2 / T).
- the target volume.
- the type of link. Some types are predefined, because they are used by the algorithms that compute the rich links (translation, axeme, axie), but it is possible to use other types of links.
- a label whose text is arbitrary.
- a weight whose value must be a real number.

These links can be established between two entries of the same volume or between two different volumes. The same volume may group entries connected to several volumes.

To realize the implementation of rich links, we separated the module processing the links from the module processing other CDM pointers. It means we have two CDM tables in the database associated to each volume. The first stores CDM traditional pointers, and the second CDM LINKS. All information of LINKS can be found in this table.

```
<cdm-elements>
  <cdm-volume    xpath="/p:volume"/>
  <cdm-entry    xpath="/p:volume/p:vocable"/>
  <cdm-entry-id    xpath="/p:volume/p:vocable/@p:id"/>
  <cdm-headword    xpath="/p:volume/p:vocable/p:lemma/text()"/>
  <cdm-headword-variant    xpath="/p:volume/p:vocable/p:altspelling/text()"    d:lang="eng"/>
  <cdm-pos xpath="/p:volume/p:vocable/p:pos/text()" />
  <cdm-sense-id    xpath="/p:volume/p:vocable/p:lexie/@p:id"/>
  <links>
     <link name="axeme" xpath="/p:volume/p:vocable/p:lexie/p:entryref">
          <type xpath="@type" />
          <volume xpath="@volume" />
          <value xpath="@p:idref" />
          <lang xpath="@lang" />
          <label xpath="@p:relation-mono" />
     </link>
  </links>
</cdm-elements>
```

Figure 2: CDM-LINKS for the English volume of the CommonUNLDict resource

3.4 Approach by rich links in searching in a complex lexical network

To explain how we create arbitrary links, let us give an example. A free label is available for each link. For example, in a lexical resource including SMS, in French "A+" has a link to "Over" with a "SMS" label, in English "L8R" corresponds to "later" with a "SMS" label, and the label of the link between "Over" and "later" is "translation."

A ProAxie macrostructure (Zhang & Mangeot, 2013) has been implemented on the Jibiki-Links platform. We present another example of rich links for semantic search in section 4.1.

3.5 Algorithms for computing rich links

The computer implementation is based on two algorithms. The first collects the links, and the second builds the result. More precisely, the first looks for all possible links in the set of all rich links of all volumes, for a desired entry. The second recursively performs the following steps: (1) selection of the start entry; (2) search of the links to other entries; (3) treatment of labels; (4) recursive call of the algorithm on the connected entry; (5) integration of the XML code of the entry connected to the start entry; (6) display.

4 Experimentation

4.1 Examples of multilevel macrostructures

We have already installed several multilevel macrostructures on Jibiki-LINKS. Here are 3 examples.

MotÀMot: trilingual lexical database with a pivot structure (Mangeot & Touche, 2010)

This project (2009-2012) has computerized a French-Khmer classical dictionary, initially in Word, into a Jibiki database (see http://jibiki.univ-savoie.fr/motamot/).

The macrostructure is composed of a monolingual volume for each language and a central pivot volume. However, in order not to confuse users, the contributing interface shows a classic view of a bilingual dictionary. Each bilingual link language A → language B added via this interface is actually translated into the background by creating two interlingual links as well as an axie link representing the original translation, to finally get: language A → pivot axie → language B (see Figure 3).

If a contributor wants to add a translation link between a vocable Va of language A and a vocable Vb of language B, s/he can establish this link at different levels. The ideal solution is to connect a word meaning (lexie) La of the vocable Va to another word meaning Lb of the vocable Vb. In this case, the link is bijective and Lb is also connected to La.

If the contributor cannot choose between word meanings, s/he can connect directly the word meaning La to the vocable Vb and the link is tagged for refinement.

With the pivot macrostructure, if two links language A → language B and language B → language C exist, then it will automatically create a link language A → language C tagged for refinement.

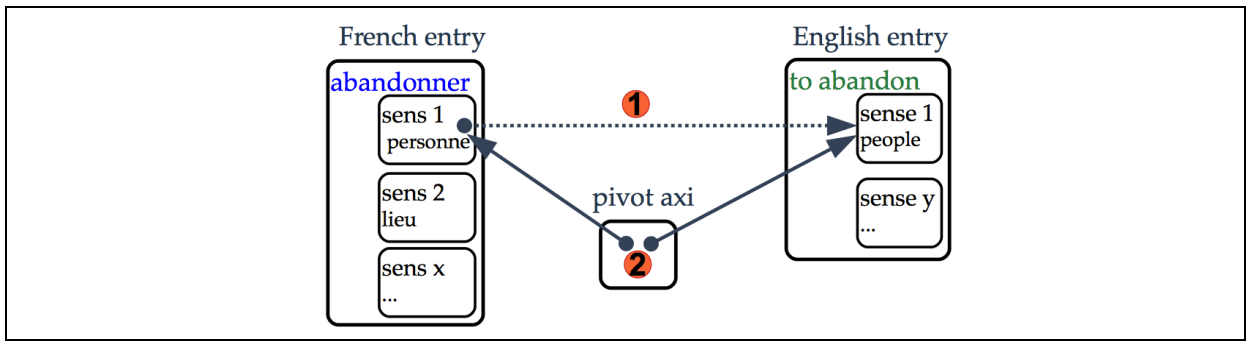

Figure 3: Example of MotÀMot

ProAxie: multilingual extension of ProxlexBase (Tran, 2006)

The ProAxie macrostructure aims at solving the problem of linking several terms that refer to one and the same referent, in particular for the management of acronyms (Zhang et Mangeot, 2013). In this macrostructure, there are two different layers. The base layer consists of two types of volume: volumes of lexies and volumes of axies. The axies are used to connect the lexies that match each other exactly. For example, one translates "ONU" by "UN" (see Figure 4) from French into English.

The "Pro" layer allows us to propose to users translations having the same referential meanings. This layer includes the volumes of *prolexemes* (Tran, 2006) and one volume of *proaxies*. A prolexeme entry links lexies having the same meaning with a label (aka, acronym, definition, etc.). A proaxie entry connects prolexemes of different languages. If one cannot find the translations directly using the lower layer, one will get the translations proposed by the "Pro" layer.

For example, for "Nations-Unies", translations by "United Nations" and "UN" will be proposed, with the "alias" label.

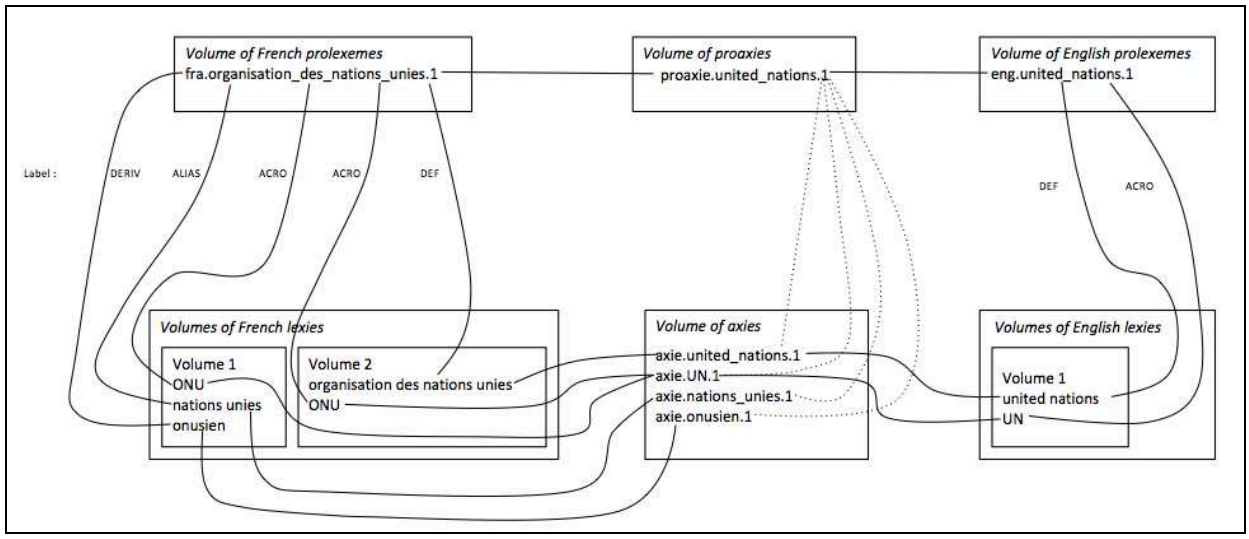

Figure 4: Example of ProAxie

For each natural language, there are one or more volumes of lexies, and a single volume of prolexemes. For each dictionary, there is a volume of axies and a volume of proaxies.

This gives three levels of translation, classified according to the precision obtained.

(1) The system finds a lexie directly, using the volume of axies. That is the first and most accurate level of translation.

(2) The system searches a link to the prolexemes volume of the source language with a certain label. When it finds the link in the proaxies volume, it follows the prolexeme link of the target language, and finally arrives at the volume of lexies in the target language, and finds a lexie that has the same label. That is the second, intermediate level.

(3) The system finds the lexies going through prolexemes and proaxies, without a corresponding label. These proposed lexies constitute the third and least accurate level.

Pivax: lexical multilingual multiversion database with 3 levels

The Pivax macrostructure has three levels: *lexie, axeme* and *axie* (Nguyen & al., 2007). Axemes are monolingual acceptions, and group monolingual lexies having the same meaning. Axies group synonymous axemes of different languages in a central "hub". In some situations, a lexical database has several volumes for a single language. For example, when there are several editions, or when the lexical resource is created for a machine translation system: one may have one volume coming from Systran, one from Ariane/Héloïse, one from IATE[13], etc. This macrostructure allows us to manage multiple volumes in the same language. Given a language, there are one or more volumes of lexies and a single volume of axemes. For any Pivax database, there is only one volume of axies. The links between the lexies and the axemes and between the axemes and the axies are rich links with attributes such as type, target volume, target language, free label, weight, etc.

4.2 CommonUNLDict: toward scaling up with a resource of Pivax type

In this section, we present the CommonUNLDict resource that uses the Pivax macrostructure. We have implemented this resource on the Pivax-UNL platform, which is an instance of Jibiki-Links. Users can easily use this resource via the link http://getalp.imag.fr/pivax/Home.po.

Resource created by linguists

Thanks to CDM-LINKS, all types of XML formats can be used in an instance of Jibiki-LINKS without modification. One needs only simple knowledge about XML to create a resource for Jibiki-LINKS. In addition, very useful available tools can be used to create an XML file, such as oXygen[14] that allows the creation of a DTD using a graphical interface.

[13] "A single database for all EU-related terminology (InterActive Terminology for Europe) in 23 languages opens to the public", 2007)

[14] http://www.oxygenxml.com

The CommonUNLDict resource has been created by the Russian lexicographer and linguist Viacheslav Dikonov (Dikonov & Boguslavsky, 2009). Figure 5 shows the graph of a monolingual volume structure using oXygen. In this example, each volume contains a large quantity of vocables, and each vocable includes one or more lexie. We will explain this structure in section 3.2.3.

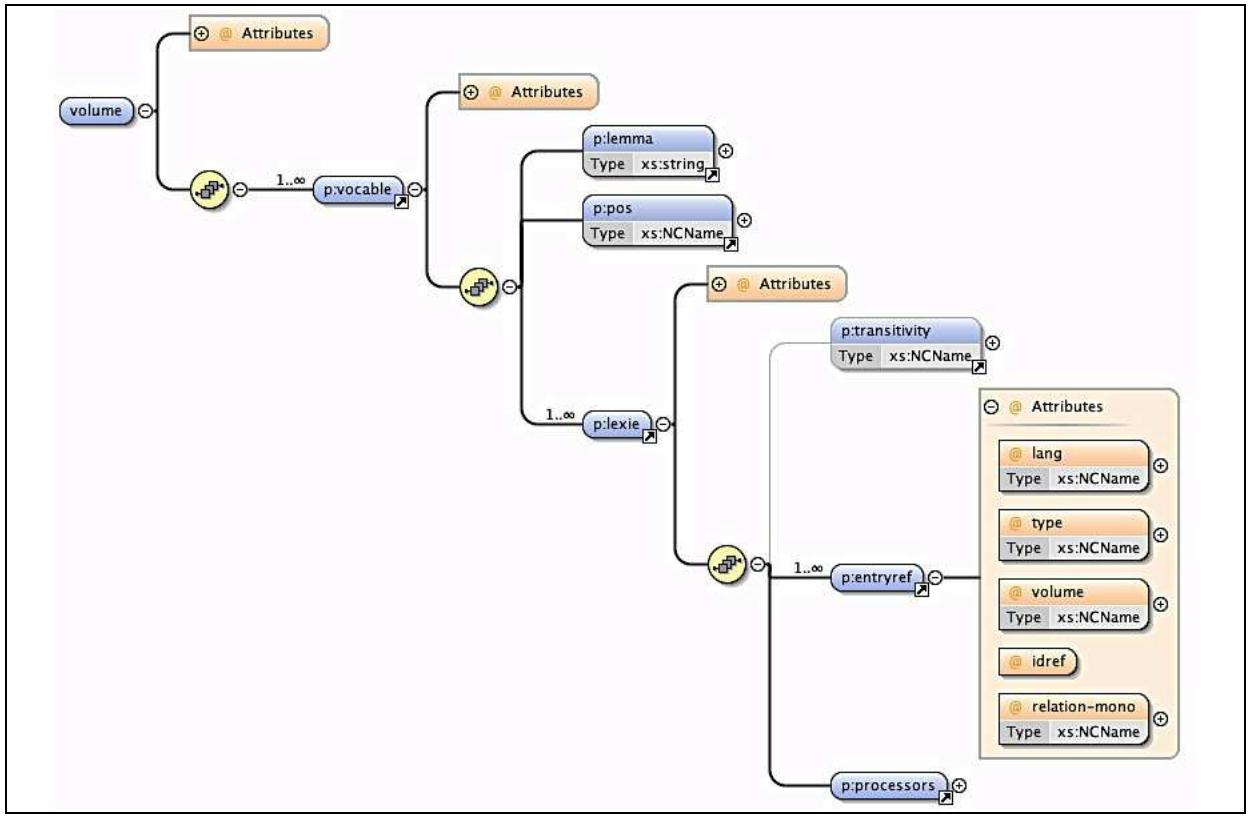

Figure 5: Structure of a monolingual volume

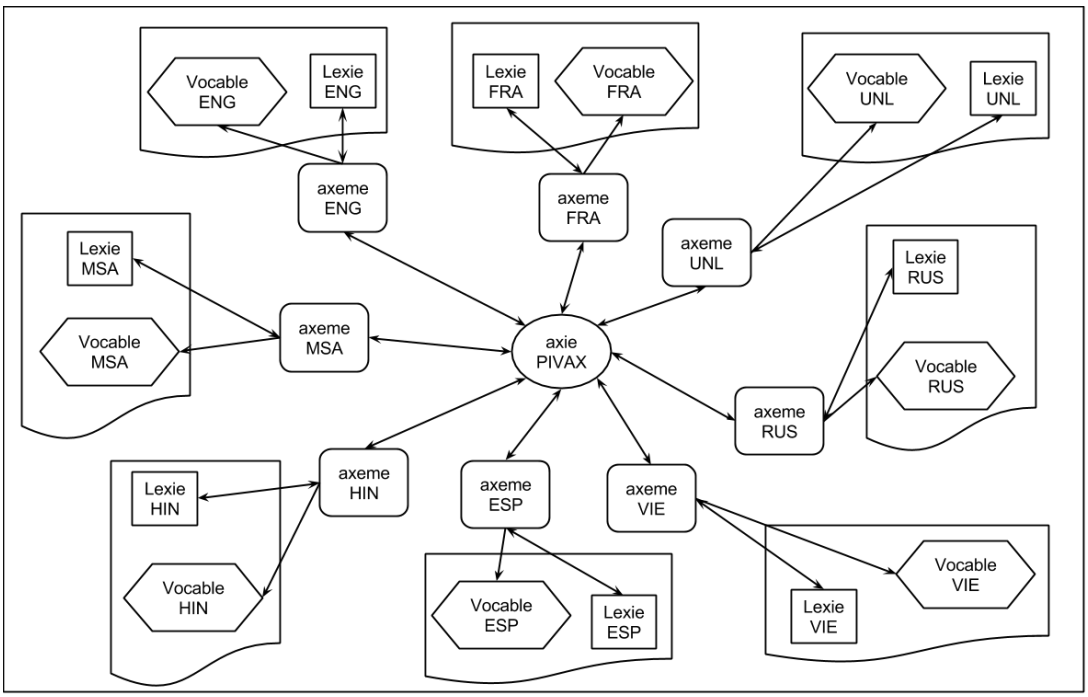

Figure 6: Macrostructure of CommonUNLDict

Macrostructure of CommonUNLDict

CommonUNLDict contains 8 languages (7 natural languages, French, English, Hindi, Malay, Russian, Spanish, Vietnamese, and the UNL language) and 17 volumes (8 volumes of monolingual data, 8 volumes of monolingual axemes, and 1 volume of axies ("interlingual meanings"). The macrostructure of CommonUNLDict is diagrammed in Figure 6. For each language, there is only one volume of monolingual data (vocables and lexical items) and a single volume of axemes. For the whole CommonUNLDict, there is only one volume of axies.

Microstructure of CommonUNLDict

The microstructure is the structure of the entries (Mangeot, 2001). In the CommonUNLDict resource, there are three types of entries (vocables, axemes and axies) and 720 K entries in total.

See Table 2.

Volume	*Language*	*Entries*
CommonUNLDict_axi	axi	82804
CommonUNLDict_eng	English	45471
CommonUNLDict_eng-axemes	English	82069
CommonUNLDict_esp	Spanish	7080
CommonUNLDict_esp-axemes	Spanish	22254
CommonUNLDict_fra	French	27537
CommonUNLDict_fra-axemes	French	48312
CommonUNLDict_hin	Hindi	31255
CommonUNLDict_hin-axemes	Hindi	50380
CommonUNLDict_msa	Malay	37342
CommonUNLDict_msa-axemes	Malay	31699
CommonUNLDict_rus	Russian	28475
CommonUNLDict_rus-axemes	Russian	45020
CommonUNLDict_unl	unl	82804
CommonUNLDict_unl-axemes	unl	82804
CommonUNLDict_vie	Vietnamese	6585
CommonUNLDict_vie-axemes	Vietnamese	8819

Table 2: Number of entries of CommonUNLDict

All volumes of the same type have the same microstructure. The example below (see Figure 7) shows the microstructure of a volume of *vocables*. Each entry of vocable type allows us to describe all detailed information, such as part of speech (POS), pronunciation, etc. Each vocable includes one or more lexies (word senses). Figure 2 shows an example. Therefore the number of axemes is greater than or equal to the number of vocables. In this microstructure, the "entryref" attribute allows us to manage the links between lexies and the entries of axeme type.

```
<p:vocable p:id="fra.ADN.n">
                <p:lemma>ADN</p:lemma>
                <p:pos>n</p:pos>
                <p:gender>m</p:gender>
                <p:lexie p:id="CommonUNLDict.lexie.fra.ADN.1">
                        <p:entryref type="axeme" volume="CommonUNLDict_fra-axemes" p:idref="CommonUNLDict
(icl&gt;polymer&gt;thing,equ&gt;deoxyribonucleic_acid)" lang="FRA" p:relation-mono="OTHER"/>
                        <p:processors>
                                <p:processor p:name="Ariane" p:access="Public">
                                <p:procref type="entry" id="ADN" var="CAT(CATN),GNR(MAS)" lang="FRA"/>
                                </p:processor>
                        </p:processors>
                </p:lexie>
</p:vocable>
```

Figure 7: Microstructure of a volume of lexies

- In this example, the value of "type" is the type of link, the value of "volume" is the target volume, the value of "idref" is the identifier of the axeme entry, the value of "lang" is the target language, and the value of "relationship-mono" is the label.

- The microstructure of the entries of axeme type allows us to describe the links with entries of lexie type and the links with entries of axie type. The microstructure of the axies allows us to describe the links with the entries of axeme type.

Response time and use case

The tests were performed with an instance of Jibiki-LINKS installed on a machine with an Intel Core i3 processor at 3.3 GHz with 8 GB of RAM.

The tool used to perform queries is wget. The command is run directly on the server to avoid the latency due to the network. We give three examples in Table 3, which show the number of links computed by the system, of entries displayed, of queries, of different languages, and the average response time. The response time, less than 1 second in these cases, is generally satisfactory. For better understanding, there is some details about the example "manger" (see Figure 8). We search "manger" in French, and find one entry with id "fra.manger.v" in the French vocable volume. The search direction is "up". This entry links to another entry of the volume of French axemes, whose id is CommonUNLDict.axeme.fra.eat(icl>consume>do, agt>living_thing, obj>concrete_thing, ins>thing)[15]. This axeme entry links with one axie entry and the vocable entry fra.manger.v. Because the search direction is "up", we just go to the axie entry. When we arrive in the volume of axies, the search direction is changed to "down". The axie entry links to 6 different axeme entries. We search each axeme entry and its links. Because the search direction is "down", we only take into account vocable entries links. For each axeme entry, we find at least one vocable entry. In other cases, one vocable entry has more than one lexie, so it links to one or several axeme entries, and there are more links.

Search argument	Links	Displayed entries	Number of requests	Different languages	Average time (ms)
French vocable "manger"	14	6	10	6	19.7
French vocable "recherche"	66	27	10	6	73.5
UNL "search(icl>action)"	51	20	10	6	56

Table 3: Response time on three examples

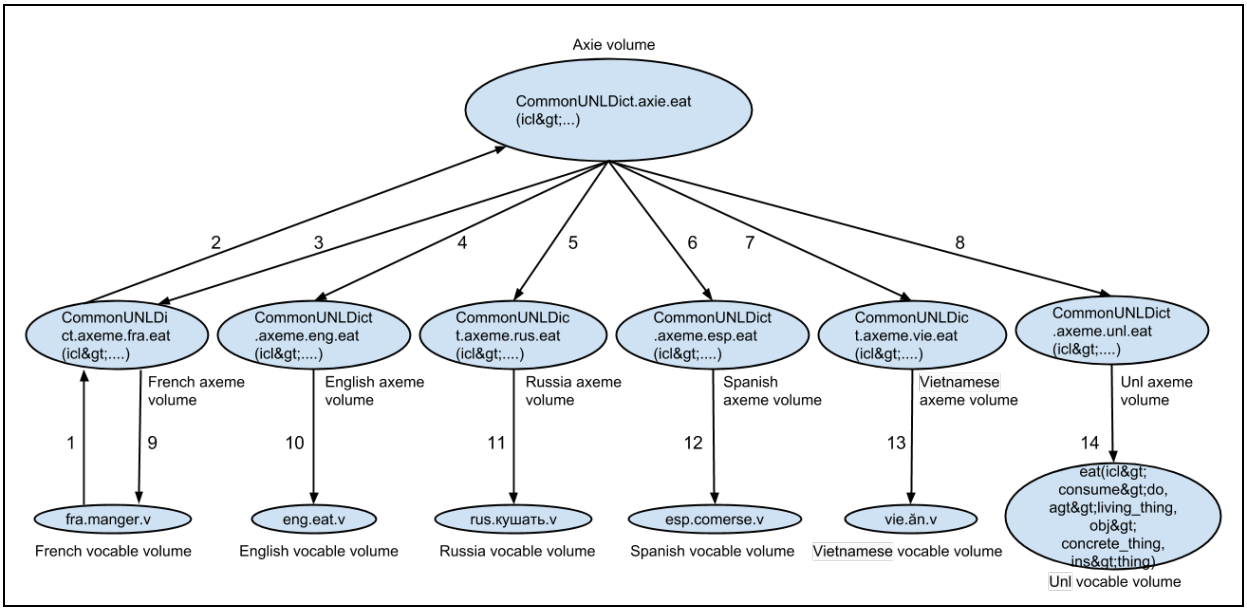

Figure 8: Links in the case of "manger"

Figure 9 shows the display of the interface for a classical search in a Web browser.

[15] In order to better display figure, we have simplified the id in figure 8.

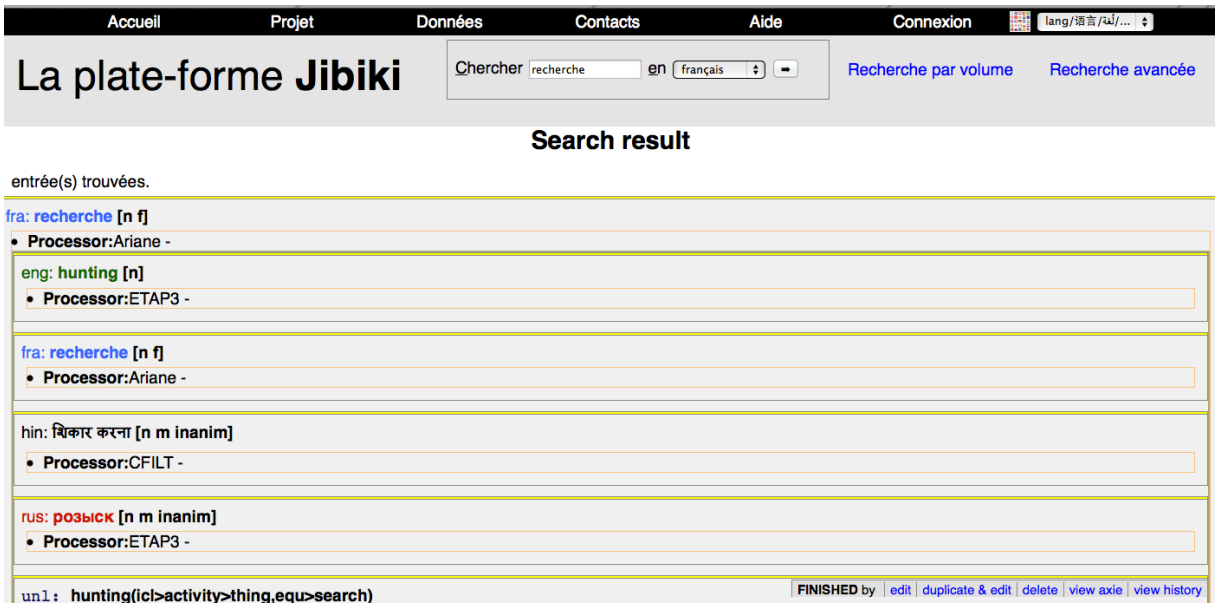

Figure 9: Display of the interface for a classical search

5 Conclusion and perspectives

In this article, we analysed the different types of lexical resource and presented a method of modelling lexical resources using *volumes*. This method allows us to manage complex resources while providing facilities for manipulation and treatment equivalent to those of a paper dictionary.

Jibiki-LINKS is a new version of the Jibiki platform, which can manage resources based on multilevel macrostructures using rich links, bearing attributes such as target volume, weight, type, language, open label, etc. To realize the implementation of rich links, we separated the module processing the links from the module processing other CDM pointers. Jibiki-LINKS has been used to implement the MotÀMot, ProAxie and Pivax macrostructures.

On the Pivax-UNL platform, another instance of the Jibiki-LINKS-based Pivax macrostructure, we have installed the volumes corresponding to the CommonUNLDict resource of V. Dikonov, and tested our platform with that resource.

There is also a UW (UNL interlingual lexemes) resource of 8G entries that was created from DBpedia by David Rouquet. In that resource, there are several volumes for the same language. As links were poorly structured, we are currently working on this resource in order to recompute them. We hope to be able to import this resource, and to make tests at that very large scale in the near future.

To sum up, lexical databases equipped with rich links allow for importing XML-based electronic dictionaries without loss of information, whether they have been elaborated from source or printable forms (such as Word, rtf, ps, pdf) or directly produced in XML from a relational database, or using a dedicated editor knowing their microstructures. They also allow us to automatically produce from them a pivot-based macrostructure organised in *volumes*, and after that to edit and improve them, using a mixed textual and graphical interface to merge or split lexies, axemes or axies, or to enrich the links with appropriate labels. The introduction of rich links to multilevel lexical databases enhances them with a very interesting aspect of the lexical networks while keeping the classical ways of using dictionaries and of performing lexicographic work.

References

EU-IATE (2007) A single database for all EU-related terminology (InterActiveTerminology for Europe) in 23 languages opens to the public. *Press release*. Brussels. 2007-06-28.

Breen, J. W., (2004) JMdict : a Japanese-Multilingual Dictionary. In Gilles Sérasset, Susan Armstrong, Christian Boitet, Andrei Pospescu-Belis, and Dan Tufis, editors, post COLING Workshop on Multilingual Linguistic Resources, Geneva, Switzerland, 28th August. International Committee on Computational Linguistics.

Dikonov V., Boguslavsky I., (2009) Semantic Network of the UNL Dictionary of Concepts. *Proceedings of the SENSE Workshop on conceptual Structures for Extracting Natural language SEmantics Moscow*, Russia, July 2009, 7 p.

Diller, G.A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.J. (1990) Introduction to WordNet: an on-line lexical database, *International Journal of Lexicography 3(4)*, pp. 235-244.

Dolan, W.B. & Richardson, S.D., (1996) Interactive Lexical Priming for Disambiguation. Proc. *MIDDIM'96, Post-COLING seminar on Interactive Disambiguation*, C. Boitet ed. Le Col de Porte, Isère, France. 12-14 août 1996. vol. 1/1 : pp. 54-56.

Dong, Z.D., Dong, Q., Hao, C.L., (2010). HowNet and Its Computation of Meaning. In Actes de *COLING-2010*, Beijing, 4 p.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M. and Soria, C. (2009). Multilingual resources for NLP in the lexical markup framework (LMF). *In journal de Language Resources and Evaluation, March 2009, Volume 43,* pp. 55-57.

Gut, Y., Ramli, P. R. M., Yusoff, Z., Kim, Ch. Ch., Samat, S. A., Boitet, Ch., Nédobejkine, N., Lafourcade, M., Gaschler, J. and Levenbach, D. (1996). *Kamus Perancis-Melayu Dewan, dictionnaire français-malais.* Dewan Bahasa Dan Pustaka, Kuala Lumpur, 667 p.

Lafourcade, M., Joubert, A. (2010). Computing trees of named word usages from a crowdsourced lexical network. *Investigationes Linguisticae*, vol. XXI, pp. 39-56

Lux-Pogodalla, V., Polguère, A. (2011) Construction of a French Lexical Network: Methodological Issues. Proceedings of *the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI-2011 Workshop*. Ljubljana, 2011, pp. 54-61.

Mangeot, M. (2002). An XML Markup Language Framework for Lexical Databases Environments: the Dictionary Markup Language. In Actes de *LREC-2002*, pp. 37-44.

Mangeot, M & Chalvin, A. (2006). Dictionary Building with the Jibiki Platform: the GDEF case. In Actes de *LREC-2006*, Genoa, pp. 1666-1669.

Mangeot, M. & Touch, S., (2010) MotÀMot project: building a multilingual lexical system via bilingual dictionaries. Proc. *SLTU 2010: Second International Workshop on Spoken Languages Technologies for Under-Resourced Languages*, Penang, Malaysia, 2010, 6 p.

McCrae, J., Spohr, D. and Cimiano, P., (2011) Linking lexical resources and ontologies on the semantic web with lemon. Proc. *ESWC'11*, Berlin, pp. 245-259.

Nguyen, H.T., Boitet, C. and Sérasset, G. (2007). PIVAX, an online contributive lexical data base for heterogeneous MT systems using a lexical pivot. In Actes de *SNLP-2007*, Bangkok, 6 p.

Richardson, S.D., Dolan, W.B. and Vanderwende, L. (1998) MindNet: acquiring and structuring semantic information from text, no. MSR-TR-98-23.

Sérasset, G. (2012) Dbnary: Wiktionary as a LMF-based Multilingual RDF network. In Actes de *LREC-2012*, Istanbul, 7 p.

Sérasset, G. & Mangeot, M. (2001). Papillon Lexical Database Project: Monolingual Dictionaries and Interlingual Links. In Proc. *NLPRS-2011*, Tokyo, pp. 119-125.

Tran, M. (2006). Prolexbase : Un dictionnaire relationnel multilingue de noms propres : conception, implémentation et gestion en ligne. *Thèse de doctorat*, Tours, pp. 54-57.

Vossen, P., (1998) EuroWordNet: A Multilingual Database with Lexical Semantic Networks, *Computers and the Humanities,* 32(2-3).

Zhang, Y. & Mangeot, M., (2013). Gestion des terminologies riches : l'exemple des acronymes. In Actes de *TALN-2013*, Les Sables d'Olonne, 8 p.