Multilingual Legal Terminology on the Jibiki Platform:
The LexALP Project

Gilles Sérasset, Francis Brunet-Manquat
Université Joseph Fourier,
Laboratoire CLIPS-IMAG, BP 53
38041 Grenoble Cedex 9 - France,
`Gilles.Serasset@imag.fr`
`Francis.Brunet-Manquat@imag.fr`

Elena Chiocchetti
EURAC Research
Viale Druso 1
39100 Bozen/Bolzano - Italy
`Elena.Chiocchetti@eurac.edu`

Abstract

This paper presents the particular use of "Jibiki" (Papillon's web server development platform) for the LexALP[1] project. LexALP's goal is to harmonise the terminology on spatial planning and sustainable development used within the Alpine Convention[2], so that the member states are able to cooperate and communicate efficiently in the four official languages (French, German, Italian and Slovene). To this purpose, LexALP uses the Jibiki platform to build a term bank for the contrastive analysis of the specialised terminology used in six different national legal systems and four different languages. In this paper we present how a generic platform like Jibiki can cope with a new kind of dictionary.

1 Introduction

One of the most time-consuming hindrances to supranational law drafting and convention negotiation is the lack of understanding among negotiators and technical writers. This is not only due to the fact that different languages are involved, but mainly to the inherent differences in the legal systems. Countries that speak the same language (like France and part of Switzerland) may use the same word to represent different legal concepts[3], as defined in their respective legal traditions. The same concept may be referred to in different ways according to the legal system[4]. Also, terms that may superficially seem to be translations of each other can represent different legal notions[5].

In order to concretely address these problems, several institutions representing translators, terminologists, legal experts and computational linguists joined in the LexALP project, co-funded by EU's INTERREG IIIb Alpine Space programme. The objective of the project is to compare the specialised terminology of six different national legal systems (Austria, France, Germany, Italy, Switzerland and Slovenia) and three supranational systems (EU law, international law and the particular framework of the Alpine Convention) in the four official languages of the Al-pine Convention, which is an international framework agreement signed by all countries of the Alpine arc and the EU. This contrastive analysis serves as a basis for the work of a group of experts (the Harmonising Group) who will determine translation equivalents in French, Italian, German and Slovene (one-to-one correspondence) in the fields of spatial planning and sustainable development for use within the Convention, thus optimising the understanding between the Alpine states at supranational level.

The tools that are to be developed for these objectives comprise a corpus bank and a term bank. The corpus bank is developed by adapting the bistro system (Streiter et al., 2006; Streiter et al., 2004). The term bank is based on the Jibiki plat-

---

[1] Legal Language Harmonisation System for Environment and Spatial Planning within the Multilingual Alps

[2] `http://www.convenzionedellealpi.org`

[3] E.g.: In the German-speaking province of Bolzano Italy the *Landeshauptmann* is the president of the provincial council, with much more limited competence that the Austrian *Landeshauptmann*, who is head of one of the states (*Bundesland*) that are part of the Austrian federation.

[4] See for instance the European Union use of *chien drogue* while French legislation calls them *chien renifleur*.

[5] For example, in Italy an *elezione suppletiva* is commonly held whenever an elected deputy or senator either resigns or dies. In Germany in such cases the first non-elected candidate is called to parliament. *Ersatzwahlen* are a rare phenomenon, foreseen in some very specific cases.

form (Mangeot et al., 2003; Sérasset, 2004).

This paper details the way the Jibiki platform is used in order to cope with a new dictionary structure. The platform provides dictionary access and edition services without any new and specific development.

After a brief overview of the Jibiki platform, we describe the choices made by the LexALP team for the structure and organisation of their term bank. Then, we show how this structure is described using Jibiki metadata description languages. Finally, we give some details on the resulting LexALP Information System.

2 Jibiki, The Papillon Dictionary Development Platform

2.1 Overview

The Jibiki platform has been designed to support the collaborative development of multilingual dictionaries. This platform is used as the basis of the Papillon project web site[6].

This platform offers several services to its users:

- access to many different dictionaries from a single easy to use query form,

- advance search for particular dictionary entries through an advanced search form,

- creation and edition of dictionary entries.

What makes the Jibiki platform quite unique is the fact that it provides these services *regardless of the dictionary structure*. In other words it may be used by any dictionary builder to give access *and* collaboratively edit *any* dictionary, provided that the resulting dictionary will be freely accessible online.

2.2 Jibiki Platform Architecture

The Jibiki platform is a framework used to set up a web server dedicated to the collaborative development of multilingual dictionaries. All services provided by the platform are organised as classical 3-tier architectures with a presentation layer (in charge of the interface with users), a business layer (which provides the services per se) and a data layer (in charge of the storage of persistent data).

In order to adapt the Jibiki platform to a new dictionary, the dictionary manager does not have

---

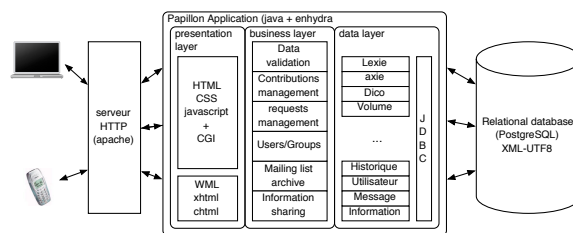[6]http://www.papillon-dictionary.org/

Figure 1: The Jibiki platform general architecture

to write specific java code nor specific dynamic web pages. The only necessary information used by the platform consists in:

- a description of the dictionary volumes and their relations,

- a mapping between the envisaged dictionary structure and a simple hypothetical dictionary structure (called CDM)[7],

- the definition of the XML structure of each envisaged dictionary volume by way of XML schemas,

- the development of a specific edition interface as a standard xhtml form (that can be adapted from an automatically generated draft).

3 The LexALP Terminology Structure

3.1 Overview

The objective of the LexALP project is to compare the specialised terminology of six different national legal systems and three supranational systems in four different languages, and to harmonise it, thus optimising communication between the Alpine states at supranational level. To achieve this objective, the terminology of the Alpine Convention is described and compared to the equivalent terms used in national legislation. The resulting terminology entries feed a specific term bank that will support the harmonisation work.

As the project deals with legal terms, which refer to concepts that are proper of the considered national law or international convention, equivalence problems are the norm, given that concepts are not "stable" between the different national legislations. Standard terminology techniques for other fields can not be applied to the field of law, where the standardisation approach (Felber, 1987;

---

[7]This mapping is sufficient for simple dictionary access

Felber, 1994) is not applicable. For this, we chose to use "acceptions" as they are defined in the Papillon dictionary (Sérasset, 1994) to represent the equivalence links between concepts of the different legal systems (Arntz, 1993).
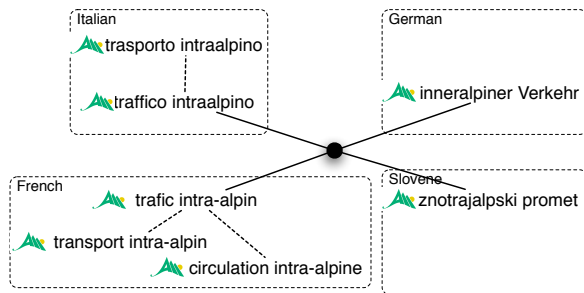

Figure 2: An Alpine Convention concept in four languages

The example given in figure 2 shows a concept defined in the Alpine Convention. This concept has the same definition in the four languages of the Alpine Convention but is expressed by different *denominations*. The Alpine Convention also uses the terms "circulation intra-alpine" or "transport intra-alpin" which are identified as synonyms by the terminologist.

This illustrates the first goal of the LexALP project. In different texts, the same concept may be realised by different terms in the same language. This may lead to inefficient communication. Hence, a single term has to be determined as part of a *harmonised quadruplet of translation equivalents*. The other denominations will be represented in the term bank as non-harmonised synonyms in order to direct drafting and translating within the Alpine Convention towards a more clear and consistent terminology use for interlingual and supranational communication.

In this example, the lexicographers and jurists did not identify any existing concept in the different national laws that could be considered close enough to the concept analysed. This is coherent with the minutes from the French National Assembly which clearly states that the term "trafic intra-alpin" (among others) should be clarified by a declaration to be added to the Alpine Convention.

Figure 3 shows an analogous quadrilingual example where the Alpine Convention concept may be related to a legal term defined in the French laws. In this example the French term is distinguished from the Alpine Convention terms, because these concepts belong to different legal sys-
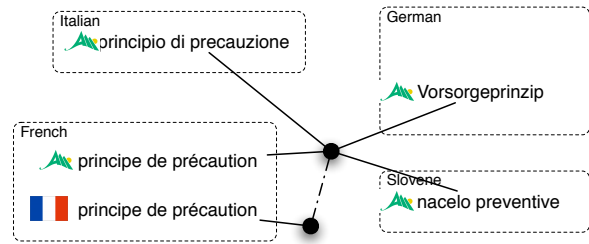

Figure 3: A quadrilingual term extracted from the Alpine Convention with reference to its equivalent at French national level

tems (and are not identically defined in them). Hence, the terminologists created distinct acceptions, one for each concept. These acceptions are related by a *translation link*.

This illustrates the second goal of the project, which is to help with the fine comprehension of the Alpine Convention and with the detailed knowledge necessary to evaluate the implementation and implementability of the convention in the different legal systems.

As a by-product of the project, one can see that there is an indirect relation between concepts from different national legal systems (by way of their respective relation to the concepts of the Alpine Convention). However, establishing these indirect relations is not one of the main objectives of the LexALP project and would require more direct contrastive analysis.

3.2 Macro- and Micro- Structures

The LexALP term bank consists in 5 volumes (for French, German, Italian, Slovene and English) containing all term descriptions (grammatical information, definition, contexts etc.). The translation links are established through a central acception volume. Figure 2 and 3 show examples of terms extracted from the Alpine Convention, synonymy links in the French and Italian volumes, as well as inter-lingual relations by way of acceptions.

All language volumes share the same microstructure. This structure is stored in XML.

Figure 4 shows the xml structure of the French term "trafic intra-alpin", as defined in the Alpine Convention. The term entry is associated to a unique identifier used to establish relations between volume entries. Each term entry belongs to one (and only one) legal system. The example term belongs to the Alpine Convention legal

```
<entry id="fra.trafic_intra-alpin.1010743.e"
    lang="fra"
    legalSystem="AC"
    process_status="FINALISED"
    status="HARMONISED">
 <term>trafic intra-alpin</term>
 <grammar>n.m.</grammar>
 <domain>Transport</domain>
 <usage frequency="common"
        geographical-code="INT"
        technical="false"/>
 <relatedTerm isHarmonised="false"
            relationToTerm="Synonym"
            termref="">
  transport intra-alpin
 </relatedTerm>
 <relatedTerm isHarmonised="false"
            relationToTerm="Synonym"
            termref="">
  circulation intra-alpine
 </relatedTerm>
 <definition>
  [T]rafic constitué de trajets ayant leur
  point de départ et/ou d'arrivée à l'inté-
  rieur de l'espace alpin.
 </definition>
 <source url="">Prot. Transp., art. 2</source>
 <context url="http://www...">
  Des projets routiers à grand débit pour
  le trafic intra-alpin peuvent être réalisés,
  si [...].
 </context>
</entry>
```

Figure 4: XML form of the term "trafic intra-alpin".

system[8] (code `AC`). The set of known legal systems includes of course countries belonging to the Alpine Space (Austria, France, Germany, Italy, Slovenia and Switzerland[9]) but also international treaties or conventions. The entry also bears the information on its status (harmonised or rejected) and its process status (to be processed, provisionally processed or finalised).

The term itself and its part of speech is also given, with the general domain to which the term belongs, along with some usage notes. In these usage notes, the attribute `geographical-code` allows for discrimination between terms defined in national (or federal) laws and terms defined in regional laws as in some of the countries involved legislative power is distributed at different levels.

Then the term may be related to other terms. These relations may lead to simple strings of texts (as in the given example) or to autonomous term entries in the dictionary by the use of the `termref` attribute. The relation itself is specified in the `relationToTerm` attribute. The current schema allows for the representation of relations

between concepts (synonymy, hyponymy and hyperonymy), as well as relations between graphies (variant, abbreviation, acronym, etc.).

Then, a definition and a context may be given. Both should be extracted from legal texts, which must be identified in the source field.

An interlingual acception (or axie) is a place holder for relations. Each interlingual acception may be linked to several term entries in the language volumes through `termref` elements and to other interlingual acceptions through `axieref` elements, as illustrated in figure 5.

```
<axie id="axi..1011424.e">
  <termref
    idref="ita.traffico_intraalpino.1010654.e"
    lang="ita"/>
  <termref
    idref="fra.trafic_intra-alpin.1010743.e"
    lang="fra"/>
  <termref
    idref="deu.inneralpiner_Verkehr.1011065.e"
    lang="deu"/>
  <termref
    idref="slo.znotrajalpski_promet.1011132.e"
    lang="slo"/>
  <axieref idref=""/>
  <misc></misc>
</axie>
```

Figure 5: XML form of the interlingual acception illustated in figure 2.

4 LexALP Information System

4.1 Overview

Building such a term bank can only be envisaged as a collaborative work involving terminologists, translators and legal experts from all the involved countries. Hence, the LexALP consortium has set up a centralised information system that is used to gather all textual and terminological data.

This information system is organized in two main parts. The first one is dedicated to corpus management. It allows the users to upload legal texts that will serve to bootstrap the terminology work (by way of candidate term extraction) and to let terminologists find occurrences of the term they are working on, in order for them to provide definitions or contexts.

The second part is dedicated to terminology work per se. It has been developed with the Jibiki platform described in section 2. In this section, we show the LexALP Information System functionality, along with the metadata required to implement it with Jibiki.

---

[8]Strictly speaking, the Alpine Convention does not constitute a legal system per se.

[9]Also Liechtenstein and Monaco are parties to the Alpine Convention, however, their legal systems are not terminologically processed within LexALP.

4.2 Dictionary Browsing

The first main service consists in browsing the currently developed dictionary. It consists in two different query interfaces (see figures 6 and 7) and a unique result presentation interface (see figure 10).
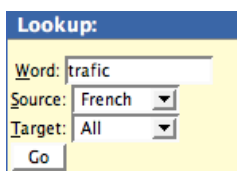

Figure 6: Simple search interface present on all pages of the LexALP Information System

```
<dictionary-metadata
 [...]
 d:category="multilingual"
 d:fullname="LexALP multilingual Term Base"
 d:name="LexALP"
 d:owner="LexALP consortium"
 d:type="pivot">
 <languages>
   <source-language d:lang="deu"/>
   <source-language d:lang="fra"/>
   <target-language d:lang="deu"/>
   <target-language d:lang="fra"/>
   [...]
 </languages>
 [...]
 <volumes>
   <volume-metadata-ref name="LexALP_fra"
     source-language="fra"
     xlink:href="LexALP_fra-metadata.xml"/>
   <volume-metadata-ref name="LexALP_deu"
     source-language="deu"
     xlink:href="LexALP_deu-metadata.xml"/>
     [...]
   <volume-metadata-ref name="LexALP_axi"
     source-language="axi"
     xlink:href="LexALP_axi-metadata.xml"/>
 </volumes>
 <xsl-stylesheet name="LexALP" default="true"
   xlink:href="LexALP-view.xsl"/>
 <xsl-stylesheet name="short-list"
   xlink:href="short-list-view.xsl"/>
</dictionary-metadata>
```

Figure 8: Excerpt of the dictionary descriptor

In the provided examples, the user of the system specifies an entry (a term), or part of it, and a language in which the search is to be done. The expected behaviour may only be achieved if :

- the system knows in which volume the search is to be performed,

- the system knows where, in the volume entry, the headword is to be found,

- the system is able to produce a presentation for the retrieved XML structures.

However, as the Jibiki platform is entirely independent of the underlying dictionary structure

```
<volume-metadata
 [...]
 dbname="lexalpfra"
 dictname="LexALP"
 name="LexALP_fra"
 source-language="fra">
 <cdm-elements>
   <cdm-entry-id index="true"
     xpath="/volume/entry/@id"/>
   <cdm-headword d:lang="fra" index="true"
     xpath="/volume/entry/term/text()"/>
   <cdm-pos d:lang="fra" index="true"
     xpath="/volume/entry/grammar/text()"/>
   [...]
 </cdm-elements>
 <xmlschema-ref xlink:href="lexalp.xsd"/>
 <template-entry-ref
     xlink:href="lexalp_fra-template.xml"/>
 <template-interface-ref
     xlink:href="lexalp-interface.xhtml"/>
</volume-metadata>
```

Figure 9: Excerpt of a volume descriptor

(which makes it highly adaptable), the expected result may only be achieved if additional metadata is added to the system.

These pieces of information are to be found in the mandatory dictionary descriptor. It consists in a structure defined in the Dictionary Metadata Language (DML), as set of metadata structures and a specific XML namespace defined in (Mangeot, 2001).

Figure 8 gives an excerpt of this descriptor. The metadata first identify the dictionary by giving it a name and a type. In this example the dictionary is a *pivot* dictionary (DML also defines *monolingual* and *bilingual* dictionary types). The descriptor also defines the set of source and target languages. Finally, the dictionary is defined as a set of volumes, each volume being described in another file. As the LexALP dictionary is a pivot dictionary, there should be a volume for the artificial language axi, which is the pivot volume.

Figure 9 shows an excerpt of the description of the French volume of the LexALP dictionary. After specifying the name of the dictionary, the descriptor provides a set of *cdm-elements*. These elements are used to identify standard dictionary elements (that can be found in several dictionaries) in the specific dictionary structure. For instance, the descriptor tells the system that the headword of the dictionary (cdm-headword) is to be found by applying the specified xpath[10] to the dictionary structure.

With this set of metadata, the system knows that:

---

[10]an xpath is a standard way to extract a sub-part of any XML structure
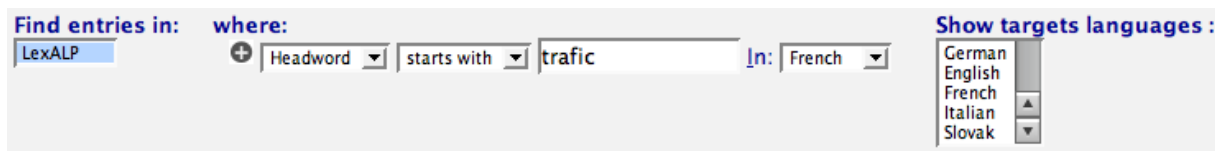
Figure 7: Advanced search interface

- requests on French should be directed to the `LexALP_fra` volume,

- the requested headword will be found in the text of the `term` element of the volume `entry` element,

Hence, the system can easily perform a request and retrieve the desired XML entries. The only remaining step is to produce a presentation for the user, based on the retrieved entries. This is achieved by way of a `xsl`[11] stylesheet. This stylesheet is specified either on the dictionary level (for common presentations) or on the volume level (for volume specific presentation).

In the given example, the dictionary administrator provided two presentations called `LexALP` (the default one, as shown in figure 10) and `short-list`, both of them defined in the dictionary descriptor.

This mechanism allows for the definition of presentation outputs in `xhtml` (for online browsing) or for presentation output in `pdf` (for dictionary export and print).

4.3   Dictionary Edition

The second main service provided by the Jibiki platform is to allow terminologists to collaboratively develop the envisaged dictionary. In this sense, Jibiki is quite unique as it federates, on the very same platform the construction and diffusion of a structured dictionary.

As before, Jibiki may be used to edit *any* dictionary. Hence, it needs some metadata information in order to work:

- the *complete* definition of the dictionary entry structures by way of an XML schema,

- a template describing an empty entry structure,

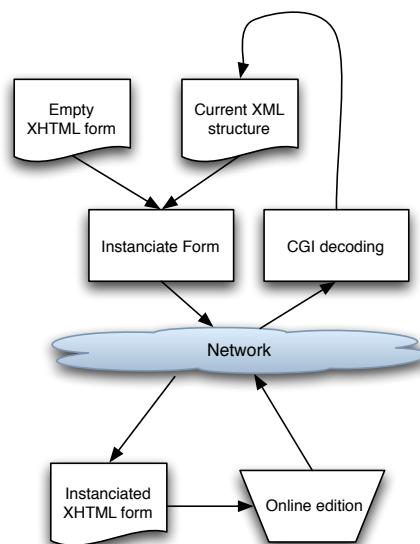[11] XSL is a standard way to transform an XML structure into another structure (XML or not).

Figure 11: Basic flow chart of the editing service

- a xhtml form used to edit a dictionary entry structure (which can be adapted from an automatically generated one).

When this information is known, the Jibiki platform provides a specific web page to edit a dictionary entry structure. As shown in figure 11, the XML structure is projected into the given empty XHTML form. This form is served as a standard web page on the client browser. After manual editing, the resulting form is sent back to the Jibiki platform as CGI[12] data. The Jibiki platform decodes this data and modifies the edited XML structure accordingly. Then the process iterates as long as necessary. Figure 12 shows an example of such a dynamically created web page.

After each update, the resulting XML structure is stored in the dictionary database. However, it is not available to other users until it is marked as `finished` by the contributor (by clicking on the `save` button). If the contributor leaves the web page without saving the entry, he will be able to retrieve it and finish his contribution later.

---

[12] Common Gateway Interface

942

TRAFIC INTRA-ALPIN, n.m. [HARMONISED/FINALISED]

TRANSPORT INTRA-ALPIN [kwic] (Synonym)

CIRCULATION INTRA-ALPINE [kwic] (Synonym)

[T]rafic constitué de trajets ayant leur point de départ et/ou d'arrivée à l'intérieur de l'espace alpin.

Source: *Prot. Transp., art. 2*

Context: Des projets routiers à grand débit pour le trafic intra-alpin peuvent être réalisés, si [...] les besoins en matière de transports ne peuvent être satisfaits, ni par une meilleure utilisation des capacités routières et ferroviaires existantes, ni par l'extension ou la construction d'infrastructures ferroviaires ou fluvio-maritimes, ni par l'amélioration d'un transport combiné, ni par d'autres mesures relatives à l'organisation des transports [...]. [goto]

INNERALPINER VERKEHR, n.m. [HARMONISED/FINALISED]

Verkehr mit Ziel und Quelle im Alpenraum inklusive Verkehr mit Ziel oder Quelle im Alpenraum.

Source: *Prot. Verk., Art. 2/Cieri*

Context: Die Vertragsparteien verpflichten sich zu einer nachhaltigen Verkehrspolitik, die [...] den inneralpinen und alpenquerenden Verkehr durch Steigerung der Effektivität und Effizienz der Verkehrssysteme und durch Förderung umwelt- und ressourcenschonenderer Verkehrsträger unter wirtschaftlich tragbaren Kosten gewährleistet [...]. [goto]

TRAFFICO INTRAALPINO, n.m. [HARMONISED/FINALISED]

TRASPORTO INTRAALPINO [kwic] (Synonym)

Traffico o trasporto con origine e destinazione all'interno del territorio alpino incluso il traffico o trasporto con origine o destinazione nel territorio alpino.

Source: *Prot. Trasp., art. 2/Cieri*

Context: Le Parti contraenti si impegnano ad attuare una politica sostenibile dei trasporti tesa a [...] ridurre gli effetti negativi e i rischi derivanti dal traffico intraalpino e transalpino ad un livello che sia tollerabile per l'uomo, la fauna e la flora e il loro habitat [...]. [goto]

ZNOTRAJALPSKI PROMET, n.m. [HARMONISED/UNPROCESSED]

Promet s ciljem in izvorom v alpskem prostoru (notranji promet), vključno s prometom s ciljem ali izvorom v alpskem prostoru.

Source: *Prot. Promet, čl. 2*

Context: Cestni projekt višjega reda za znotrajalpski promet se lahko udejanji le takrat, če [...] potreb po prevoznih zmogljivostih ni moč zadovoljiti z boljšo izkoriščenostjo obstoječih cestnih in železniških zmogljivosti, z rekonstrukcijo ali novogradnjo železniških in plovnih infrastruktur ter z izboljšanjem kombiniranega prometa kot tudi z nadaljnjimi prometno-organizacijskimi ukrepi [...]. [goto]

Figure 10: Query result presentation interface

Figure 12: Edition interface of a LexALP French entry

943

At each step of the contribution (after each update) and at each step of dictionary editing (after each save), the previous state is saved and the contributor (or the dictionary administrator) is able to browse the history of changes and to revert the entry to a previous version.

5 Conclusion

In this article we give some details on the way the Jibiki platform allows the diffusion and the online editing of a dictionary, regardless of his structure (monolingual, bilingual (directed or not) or multilingual (multi-bilingual or pivot based)).

Initially developed to support the editing of the Papillon multilingual dictionary[13], the Jibiki platform proved useful for the development of other very different dictionaries. It is currently used for the development of the GDEF (Grand Dictionnaire Estonien-Français) project[14] an Estonian French bilingual dictionary. This article also shows the use of the platform for the development of a European term bank for legal terms on spatial planning and sustainable development in the LexALP project.

Adapting the Jibiki platform to a new dictionary requires the definition of several metadata information, taking the form of several XML files. While not trivial, this metadata definition does not require any competence in computer development. This adaptation may therefore also be done by experimented linguists. Moreover, when the dictionary microstructure needs to evolve, this evolution does not require any programming. Hence the Jibiki platform gives linguists great liberty in their decisions.

Another positive aspect of Jibiki is that it integrates diffusion and editing services on the same platform. This allows for a tighter collaboration between linguists and users and also allows for the involvement of motivated users to the editing process.

The Jibiki platform is freely available for use by any willing team of lexicographer/terminologists, provided that the resulting dictionary data will be freely available for online browsing.

In this article, we also presented the choices made by the LexALP consortium to structure a term bank used for the description and harmonisation of legal terms in the domain of spacial planning and sustainable development of the Alpine Space. In such a domain, classical techniques used in multilingual terminology cannot be used as the term cannot be defined by reference to a stable/shared semantic level (each country having its own set of non-equivalent legal concepts).

References

Reiner Arntz. 1993. Terminological equivalence and translation. In H. Sonneveld and K. Loening, editors, *Terminology. Applications in Interdisciplinary Communication*, pages 5–19. Amsterdam et Philadelphia, John Benjamins Publishing Company.

Helmut Felber, 1987. *Manuel de terminologie*. UNESCO, Paris.

Helmut Felber. 1994. Terminology research: Its relation to the theory of science. *ALFA*, 8(7):163–172.

Mathieu Mangeot, Gilles Sérasset, and Mathieu Lafourcade. 2003. Construction collaborative d'une base lexicale multilingue, le projet Papillon. *TAL*, 44(2):151–176.

Mathieu Mangeot. 2001. *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*. Thèse de nouveau doctorat, spécialité informatique, Université Joseph Fourier Grenoble I, Septembre.

Gilles Sérasset. 1994. Interlingual lexical organisation for multilingual lexical databases in nadia. In Makoto Nagao, editor, *COLING-94*, volume 1, pages 278–282, August.

Gilles Sérasset. 2004. A generic collaborative platform for multilingual lexical database development. In Gilles Sérasset, editor, *COLING 2004 Multilingual Linguistic Resources*, pages 73–79, Geneva, Switzerland, August 28. COLING.

Oliver Streiter, Leonhard Voltmer, Isabella Ties, and Natascia Ralli. 2004. BISTRO, the online platform for terminology management: structuring terminology without entry structures. In *The translation of domain specific languages and multilingual terminology*, number 3 in Linguistica Antverpiensia New Series. Hoger Instituut voor Vertalers en Tolken, Hogeschool Antwerpen.

Oliver Streiter, Leonhard Voltmer, Isabella Ties, Natascia Ralli, and Verena Lyding. 2006. BISTRO: Data structure, term tools and interface. *Terminology Science and Research*, 16.

---

[13]http://www.papillon-dictionary.org/
[14]http://estfra.ee/