HAL
open science

Multilingual Legal Terminology on the Jibiki Platform: The LexALP Project

Gilles Sérasset

To cite this version:

Gilles Sérasset. Multilingual Legal Terminology on the Jibiki Platform: The LexALP Project. Proc. of Papillon 2005 Workshop, Dec 2005, Chiang Rai, Thailand. pp.64–73. hal-00965707

HAL Id: hal-00965707

https://hal.science/hal-00965707v1

Submitted on 25 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multilingual Legal Terminology on the Jibiki Platform: The LexALP Project

Gilles Sérasset
Université Joseph Fourier, Laboratoire CLIPS-IMAG, BP 53
38041 Grenoble Cedex 9, France,
Gilles.Serasset@imag.fr,
http://clips.imag.fr/geta/gilles.serasset/

Abstract

This paper presents the particular use of "Jibiki" (Papillon's web server développement platform) for the Lexalp[1] project. LexALP's goal is to harmonize the terminology of the Alpine Convention's[2] four languages (French, German, Italian and Slovenian) so that member states are able to cooperate effectively. For this, the project uses the Papillon platform in order to build a term bank used to compare the specialized terminology of six different national legal systems in four different language, and to harmonize it, optimizing the understanding between various alpine states in environmental matters at a supranational level. In this paper we present how a generic platform like the Papillon one is to be used in order to cope with a new kind of dictionary.

1 Introduction

One of the most blocking problem in supranational law or convention negotiation comes from the lack of understanding among negociators and writters. This lack of understanding is not only due to the fact that different languages are involved, but also, and mainly, to the fact that different legal systems are involved. Countries using the same language (like France and part of Switzerland) may use the very same word to represent very different legal concepts defined in their respective legal texts. Terms, that are usually considered as translations of each others may also represent very different legal notions.

In order to concretely address these problems, several institutions representing translators, lexicographs, legal experts and computational linguists joined in the LexALP project, funded by EU's INTERREG IIIb Alpine Space program. The objective of the project is to compare the specialised terminology of six different national legal systems in four different languages, and to harmonise it, optimising the understanding between the various Alpine states in environmental matters at the 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 reusing the bistro system (Streiter et al., 2006; Streiter et al., 2004). The term bank is based on the Jibiki platform (Mangeot-Lerebours 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 organization 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[3].

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

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

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

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-tiers 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).

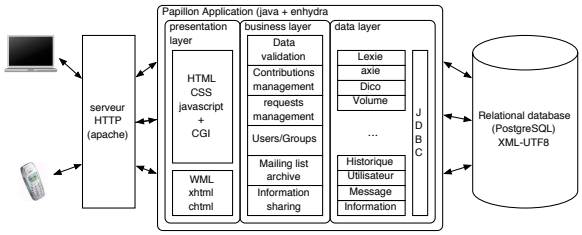

Figure 1: The Jibiki platform general architecture

In order to adapt the Jibiki platform to a new dictionary, the dictionary manager does not have to write specific java code nor specific dynamic web pages. The only necessary information used by the Jibiki plateform 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)[4],
- the definition of the XML structure of each envisaged dictionary volumes 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)

[4]This mapping is sufficient for simple dictionary access

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 in four different languages, and to harmonise it, optimising the understanding between the various Alpine states in environmental matters at the supranational level. To achieve this objective, partners are collecting terms used in the Alpine convention (a trans-national convention on spatial planning and durable development gathering countries from the alpine region) and their counterparts in national laws. The collected terms are structured in a specific term bank that will support the future harmonisation work.

As the project deals with law terms, each term refers to a concept that is proper to the considered national law or international convention, we have to deal with contrastive problems as concepts are not "stable" between the different national laws. Here, standard terminology techniques are not applicable, as the "normalisaion" approach(Felber, 1987; Felber, 1994) is not applicable. For this, we chose to use "acceptions" as they are defined in the Papillon dictionary (Sérasset, 1994a; Sérasset, 1994b) to represent the translation links between concepts of the different legal systems (thus adopting a method closer to (Arntz, 1993)).

The example given in figure 2 shows a concept defined in the Alpine Convention. This concept is realised by terms in the languages of the convention. The Alpine Convention also uses the terms "circulation intra-alpine" or "transport intra-alpin" which are identified as synonyms by the lexicographer.

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 is considered as a mistake. Hence, a single term has to be chosen as an

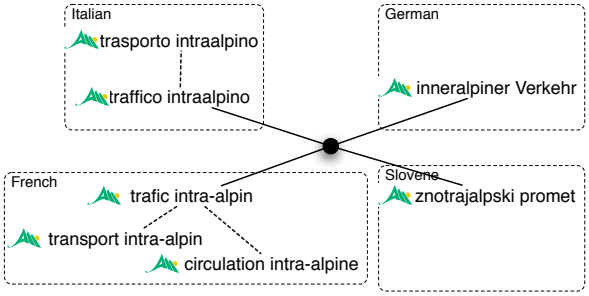

Figure 2: A quadri-lingual term extracted from the Alpine Convention

harmonised term. The others will be represented as non harmonised synonyms in order to help Alpine Convention writers and translators to use the correct term.

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

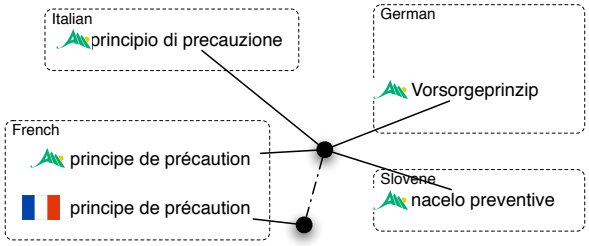

Figure 3: A quadri-lingual term extracted from the Alpine Convention with reference to a French translation

Figure 3 shows an analogous quadri-lingual 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 their concepts are not identically defined in their respective legal systems. Hence, the lexicographer 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 jurists with the fine comprehension of the Alpine Convention and with the detailed knowledge necessary to evaluate the implementability of the convention in their legal system.

As a by-product of the project, one can see that there is a 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.

3.2 Macro- and Micro- Structures

The LexALP term bank consists in 5 volumes (for English, French, German, Italian and Slovene) containing the term descriptions. 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 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.

```
<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".

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 system[5] (code `AC`). The set of known legal systems includes of course countries belonging to the alpine space (Austria, France, Germany, Italy, Liechtenstein, Slovenia, Switzerland) 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 some of the countries involved allows several level of legislative authorities.

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 (along with its source) and a context may be given. The definition and context should be extracted from a legal text that is 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 languages volumes through `termref` elements and to other interlingual acceptions through `axieref` elements, as illustrated in figure 5.

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

```
<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 lexicographers, translators and jurists from all the involved countries. Hence, the LexALP consortium has set up a centralised information system that is used to gather all textual and lexicographical data.

This information system is organized in 2 main parts. The first one is dedicated to corpus management. It allows the users to upload legal texts that will serve to bootstrap the lexicographical work (by way of candidate term extraction) and to let lexicographers find occurrences of the term they are working on, in order for them to provide definition or contexts.

The second part is dedicated to lexicographical 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.

4.2 Dictionary Browsing

The first main service consists in browsing the currently developed dictionary. It consists in 2 different query interfaces (see Figures 6 and 7) and an unique result presentation interface (see Figure 10).

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

Figure 7: Advanced search interface

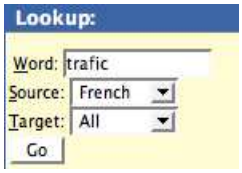

Figure 6: Simple search interface present on all pages of the Lexalp Information System

```
<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

```
<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

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 (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-Lerebours, 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

Figure 10: Query result presentation interface

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[6] to the dictionary structure.

With this set of metadata, the system knows that:

- 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`[7] 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) 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 lexicographers 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 dictionary.

As before, Jibiki may be used to edit *any* dictionary. Hence, it needs some metadata in-

[6] an xpath is a standard way to extract a sub-part of any XML structure

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

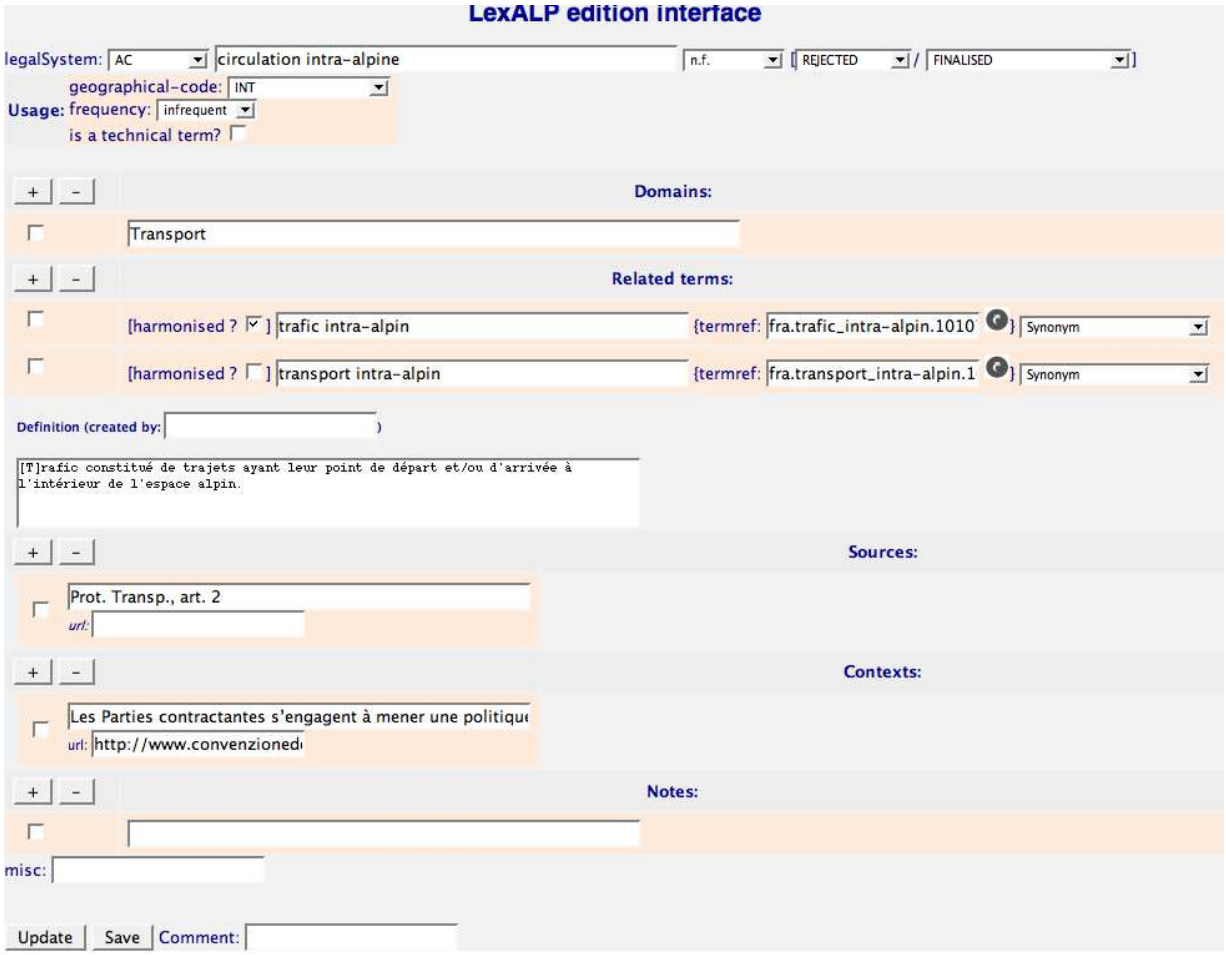

Figure 11: Edition interface of a Lexalp French entry

formation in order to work:

- the *complete* definition of the dictionary entry structures by way of a XML schema,
- a template describing an empty entry structure,
- 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 12, 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 edition, the resulting form is sent back to the Jibiki platform as CGI[8] data. The Jibiki plateform decodes this data and modifies the edited XML structure accordingly. Then the process iterates as long as necessary. Figure 11 shows an example of such a dynamically created web page.

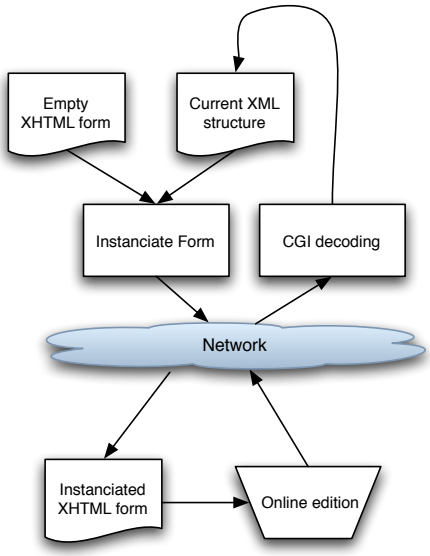

Figure 12: Basic flow chart of the edition service

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.

At each step of the contribution (after each update) and at each step of the dictionary edition (after each save), the previous state is saved and the contributor (or the dictionary administrator) is able 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 edition of a dictionary, regardless of his structure (monolingual, bilingual (directed or not) or multilingual (multi-bilingual or pivot based)).

Initially developed to support the edition of the Papillon multilingual dictionary[9], 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[10] an Estonian French bilingual dictionary. This article also shows the use of the platform for the development of an European term bank for legal terms on spatial planning and 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. Hence this adaptation may be done by experimented lexicographers. Moreover, when the dictionary microstructure needs to evolve, this evolution does not require any programming. Hence the Jibiki platform gives lexicographers great liberty in their decisions.

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

The Jibiki platform is freey available for use

[8] Common Gateway Interface

[9] `http://www.papillon-dictionary.org/`

[10] `http://estfra.ee/`

by any willing lexicographer team, 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 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 his 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-Lerebours, Gilles Sérasset, and Mathieu Lafourcade. 2003. Construction collaborative d'une base lexicale multilingue, le projet Papillon. *TAL*, 44(2):151–176.

Mathieu Mangeot-Lerebours. 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. 1994a. An interlingual lexical organisation based on acceptions, from the parax mock-up to the nadia system. In *ICLA-94*, pages 21–33, July.

Gilles Sérasset. 1994b. 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.