Building lexical resources: towards programmable contributive platforms

Mathieu Mangeot*†
*Laboratoire LLS
Campus de Jacob Bellecombette, BP 1104
F-73011 Chambry CEDEX, France
Email: mathieu.mangeot@imag.fr

Hong Thai Nguyen†
†Laboratoire GETALP-LIG
385 rue de la bibliothque, BP 53
F-38041 Grenoble CEDEX 9, France
Email: hong-thai.nguyen@imag.fr

***Abstract*—Lexical resources are very important in nowadays society, with the globalization and the increase of world communication and exchanges. There are clearly identified needs, both for humans and machines. Nevertheless, very few efforts are actually done in this domain. Consequently, there is an important lack of freely available good quality resources, especially for under-resourced languages. Furthermore, the majority of existing bilingual dictionaries is built with one language as English. Therefore, if one wants to translate from one language (that is not English) to another, it uses English as a pivot. And even for English native speakers, it creates a lot of misunderstandings that can be critical in many situations. In order to create and extend freely available good quality rich lexical resources for under-resourced languages online with a community of voluntary contributors, Jibiki, an online generic platform for managing (lookup, editing, import, export) any kind of lexical resources encoded in XML, has been developed. This platform is successfully used in several dictionary construction projects. Concerning the data, a serious game has been launched in order to collect precious lexical information such as collocations that will be integrated later into dictionary entries. Work is now done on extending our platform in order to reuse the resulting resources and enriching them by synchronization with the other systems (language learners and translators environments, machine translation systems, etc.).**

I. Introduction

There should be no need to underline the importance of lexical resources in nowadays society, with the globalization and the increase of world communication and exchanges. There are clearly identified needs, both for humans (tourism, communication, translation) and machines (analysis, machine translation, Natural Language Processing applications in general).

Nevertheless, very few efforts are actually done in this domain. Most of the time, the expensive construction costs prohibit companies to launch new projects. Consequently, there is an important lack of freely available good quality resources, especially for under-resourced languages (Pi-languages [1]).

Furthermore, the majority of existing bilingual dictionaries is built with English as one language and another language. Therefore, if one wants to translate from one language (that is not English) to another, it uses English as a pivot. And even for English native speakers, it creates a lot of misunderstandings that can be critical in many situations.

The main issue we are trying to solve for some time already is to find a way to develop freely available good quality rich lexical resources for under-resourced languages online with a community of voluntary contributors. Another issue is to reuse the resulting resources in NLP systems and enriching them by synchronization between the lexical database and the different systems (language learners environments, translators environments, machine translation systems, etc.)

We will first explain what should be a perfect lexical resource, considered as a kind of ”graal” in our domain. The second part will present the efforts done concerning the tools for developing such resources. The third part will focus on new ways to collect lexical data via serious games. In the last part, we will discuss the ways to extend the existing lexical database system in order to exchange data with other systems via API and synchronization tools. The conclusion will try to resume what has been done and what is left in order to reach our ultimate goal.

II. The graal of Lexical Resources

In this section, we will present what we consider as an ultimate goal for lexical resources, starting from the analysis of what is existing for Vietnamese.

A. Current situation for Vietnamese

The table I briefly shows the situation of online lexical resources for Vietnamese. We can see that most of the languages pairs available are Vietnamese-English or Vietnamese-French. There is no resource available for language pairs with neighbouring countries (Thai, Lao, Khmer, etc.). Concerning the coverage, only Bamboo indicates the number of entries, and it is far from a broad coverage of the languages, except for English (¿ 400,000 entries). Furthermore, this is the only resource that allows contributors to edit and add entries. Concerning the copyrights, only two resources are open-source while the other are protected. Concerning the information available for each entry, most of the resources are very limited. They should be considered as lexicons rather that real dictionaries.

B. Lack of lexical resources

The situation for Vietnamese is roughly the same for every other language except English. There is a lack of

Table I
ONLINE LEXICAL RESOURCES FOR VIETNAMESE

Name	URL	Language	Notes
VDict	http://vdict.com	en-vn, vn-en, fr-vn, vn-fr, en-en(WordNet), FOLDOC (Computing Dictionary)	Fuzzy search: yes Edition: no Data copyright: GNU
Bamboo	http://baamboo.com/?tab=Vietdic	en-vn(402.688), fr-vn(63.369), jp-vn(114.282), vn-vn(36.862), vn-en(379.749), vn-fr(43.339), vn-jp(84.847), abbreviation dict. (145.517)	Fuzzy search: yes Edition: yes (Wiki mode) Data copyright: protected
VietFun	http://dict.vietfun.com/	like VDict with: de-vn, vn-de, ru-vn, no-vn	Fuzzy search: yes Edition: no Data copyright: GNU
E-Lexicon	http://www.edusoft.com.vn/e-lexicon/	en-vn, vn-en, fr-vn, vn-fr, specially domain dictionaries: architecture and building, stock.	Installation on machine, plug-in to MS Office Data copyright: EDUSOFT com.
LacViet MTD	http://www.vietgle.vn/	en-vn, vn-en, vn-fr, fr-vn, abbreviation dict.	Installation on machine with media Most use by public from 1998 New version on mobile Data copyright: LacViet com.

freely available online lexical resources of good quality and broad coverage, especially for under resourced languages (Pi-languages [1]). It is very frequent that if one wants to translate from one language to another, it uses English as a pivot, therefore increasing misunderstandings and errors.

Nowadays, the costs of building a new bilingual dictionary from scratch is too high. This is why, most of the time, the dictionary publishing houses prefer to edit new versions of existing dictionaries instead of creating new language pairs. There is also a problem of IPR[1]. A company that invested so much in building a dictionary cannot give it for free. Consequently, we should look into the LINUX community paradigm: building dictionaries with a community of voluntary contributors.

C. Ideal situation

These problems where identified and detailed in M. Mangeot's Ph.D. thesis [2]. A perfect solution, the graal of lexical resources would be a broad coverage multilingual pivot database with rich detailed monolingual entries and interlingual links usable by humans and machines, editable online and freely available.

Starting in 2000, Papillon[3] project was launched, a new multilingual database construction project that began to address these problems.

The macrostructure consists in one monolingual volume for every language of the dictionary and one pivot volume in the middle (see Figure 1).

When a new entry in a language A is added, it must be then linked to the interlingual volume. These links are created either by reusing existing bilingual dictionaries lang A - lang B, or by entering manually the link from an existing translation. The link lang A -¿ lang B becomes lang A -¿ pivot -¿ lang B. If the entry of lang B was already linked to other languages, automatically, the entry of lang A will also benefit from these links: lang A -¿ pivot -¿ lang B, lang C, lang D, etc. This idea

[1]Intellectual Property Rights

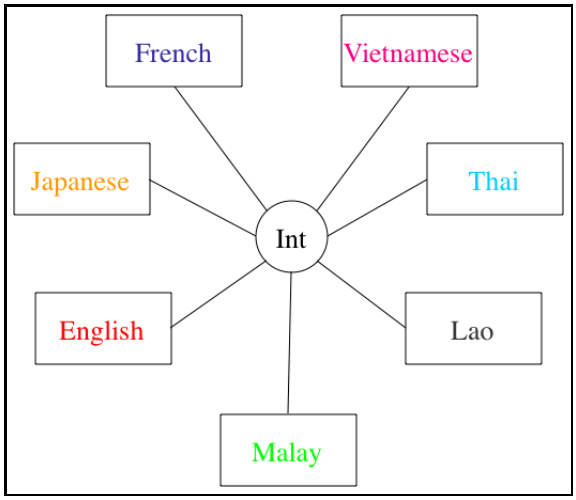

Figure 1. Multilingual Pivot Macrostructure

of a pivot volume that plays the role of a lexical center is very useful for under-resourced languages.

The microstructure of the monolingual entries is very rich and detailed. It is is based on the structure used for the formal lexical database DiCo [4] of the OLST laboratory in Universit de Montral. The encoding methodology is directly borrowed from the Explanatory and Combinatorial Lexicology (ECL)[5], which is part of the Meaning-Text Theory. This theory gives the necessary information to go from a meaning to its realizations in a given language. The resulting dictionary microstructure is independent from the languages and the information is theoretically usable by humans and machines. Therefore, it is very suitable for our purpose.

Each entry or lexical unit is made of a name, grammatical properties (mainly a part of speech), a semantic formula which can be seen as a formal definition. In the case of a predicative lexie, it describes the entire predicate and its arguments, a government pattern which describes the syntactic realization

of the arguments of the predicate, a list of lexico-semantic functions. There is a fixed number of 56 basic functions that can be applied in any language. These functions can be combined to create more elaborated ones; a list of examples; a list of full idioms.

The specifications of Papillon project refer to this graal of lexical resources. But just like high quality, broad coverage fully automatic machine translation of any text, it is not reachable in one shot. With the time, Papillon project has become a kind of meta-project[6] with many sub-projects, each one corresponding to one particular aspect of the initial goal. As we will detail later, the tool aspects are covered by the Jibiki project, the data collection by the JeuxDeMots project and finally, the construction of new dictionaries by the Motamot project.

III. How to build lexical resources, especially for under-resourced languages

A. Specifications and requirements of the system

As we already mentioned, nowadays, the construction costs of a new resource are too high. There is a need of a community-driven resource building process like Wikipedia. The platform needs to be online in order for many people to access it.

Unfortunately, it is not possible to use an existing wiki platform because the entry structure of a dictionary is not free. It has to be the same for all the entries. Some commercial environments exist like Tshwanelex [2] or the IDM Dictionary Production System [3] but our goal is to develop dictionaries for under-resourced languages. Hence, there is a need of a freely available and customisable platform.

Another key problem is how to deal with heterogeneous entry structures if we want to query, edit and reuse them all at once.

Therefore, there is a need of an online platform with users/groups management, heterogeneous dictionary lookup interface and generic entry structure edition interface.

B. Jibiki: an online generic platform for managing lexical resources

The Jibiki platform [7], [8], [9] was developed to answer these needs. It is a community web site primarily developed for the Papillon project. This platform is entirely written in Java using the ”Enhydra” web development framework. All XML data is stored in a standard relational database (Postgres). This community web site proposes mainly two services: a unified interface to simultaneously access many lexical resources at once (monolingual, bilingual dictionaries, multilingual databases, etc.) and a specific edition interface to contribute to the dictionaries stored on the platform.

1) The unified lookup interface: This service[4] currently gives access to thirteen (13) multilingual, bilingual and monolingual dictionaries, representing more than one million entries. Every available dictionary will be queried according to its own structure from a multi-criteria search interface (see 2). Moreover, all results will be displayed in a form that fits the structure. Any monolingual, bilingual or multilingual dictionary may be added in this collection, provided that it is available in XML format. With the Jibiki platform, giving access to a new, unknown, dictionary is a matter of writing two XML files: a dictionary description and an XSL stylesheet. For currently available dictionaries, this took an average of about one hour per dictionary.

The description file gathers dictionary meta-information and a minimum set of information in the dictionarys XML structure. The Jibiki platform defines a standard structure of an abstract dictionary containing the most frequent subset of information found in most dictionaries. This abstract structure is called the Common Dictionary Markup [10]. To describe a new dictionary, one has to write an XML file that associate CDM elements to pointers in the original dictionary structure.

Along with this description, one has to define an XSL style sheet that will be applied on requested dictionary elements to produce the HTML code that defines the final form of the result. If such a style sheet is not provided, the Jibiki platform will itself transform the dictionary structure into a CDM structure and apply a generic style sheet on this structure.

The entire process of writing all the meta data files for one dictionary and importing the data takes around one hour. It is then immediately possible to lookup and edit the newly imported dictionary.

2) The key feature: an online generic editor: The main purpose of the Jibiki platform is to gather a community around the development of one or several dictionaries. Thus, the crucial challenge that has been faced was to provide a way to edit the dictionary entries directly on the platform. It was specifically difficult because we wanted to be able to edit any kind of dictionary entry (the editor had to adapt itself to the structure of the entries) and to edit them online with a simple browser (it had to be bult only with a combination of HTML forms and simple javascripts). Java applets could not be used because of compatibility problems.

The editor works with a template XHTML interface that is instanciated with the entry that the user wants to edit. This template can be generated automatically from a description of the entry structure in XML schema. It can be modified afterwards for improving the rendering on the screen. Thus, the only data needed to edit a dictionary entry on the jibiki platform (apart from the dictionary metadata described previously) is the XML schema of the structure of the entry and furthermore, any type of dictionary entry as long as it is encoded in XML.

HTML forms are very limited. The available interactors are text fields, radio buttons, check boxes and pop up menus.

[2] http://tshwanedje.com/tshwanelex/

[3] http://www.idm.fr/

[4] http://papillon-dictionary.org

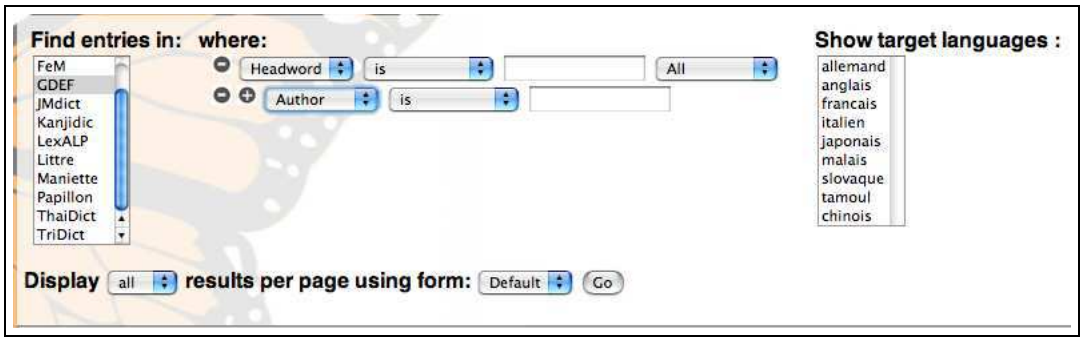

Figure 2. Advanced search interface

It was not enough to be able to edit complex entries. Thus, there was a necessity to build more complex interactors from the combination of the previous ones in order to handle lists (adding,deleting, moving an item on a list) and links (links to entries in the same volume or other ones). These elements can be themselves complex objects containing lists of other objects, etc.

Any user, who is registered and logged in to the jibiki platform web site, may contribute to the stored dictionaries by creating or editing an entry. Moreover, when a user asks for an unknown word, s/he is encouraged to contribute it to the dictionary. Contribution is made through a standard HTML interface (see Figure 3).

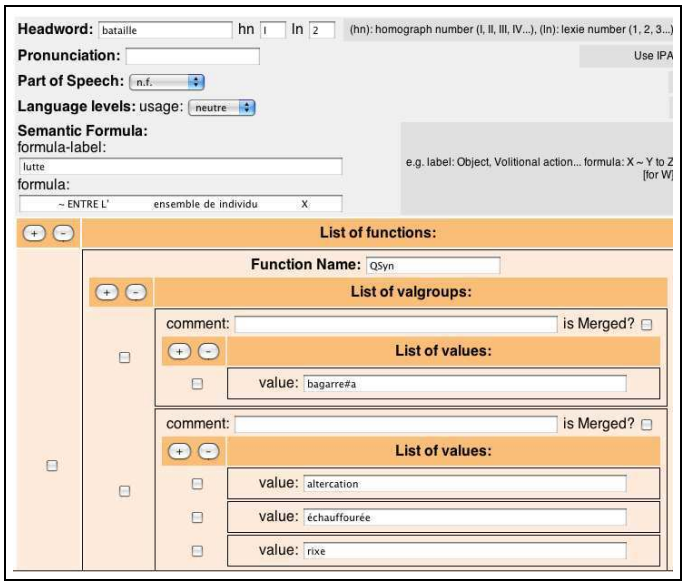

Figure 3. Editing interface for Papillon entry

The source files of the jibiki software platform are available freely online via subversion on the laboratory sourceforge[5].

C. Dictionary building projects with Jibiki Platform

In this section, we present some dictionaries building projects that use the jibiki platform and detail each one specificity.

1) GDEF Estonian-French bilingual dictionary: The GDEF project[11] aims at building a high quality bilingual Estonian-French dictionary for professional translators. The figure 4 shows an example entry and lookup interface.

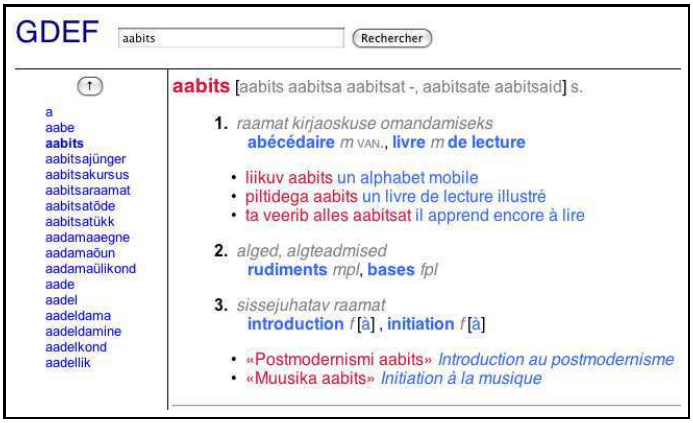

Figure 4. GDEF entry

In order to obtain high quality, the building process is very strict. First, the Estonian monolingual volume has been bootstrapped with the Estonian part of an Estonian-Russian bilingual dictionary. The French monolingual volume has been filled with the freely available Morphalou[6] lexicon.

The entry writing and revision process must go into the following steps:

- First a payed contributor completes an existing Estonian entry of creates a new one if it does not exist;
- The entry is next revised by a reviewer;
- It is then validated by a validator.

These process is implemented via the entries status. First, every entry has a draft status. Then, when a contributor opens it in the edition interface, the status becomes ”not finished”. It stays in this status until the contributors decides that s/he has finished the work and clics on the ”finish entry” button. Then, the status of the entry becomes ”finished” and it appears in the reviewer interface. It follow the same kind of steps until it

[5]http://jibiki.ligforge.imag.fr

[6]http://www.cnrtl.fr/lexiques/morphalou

is marled as ”validated” by the validator. It is them available to the public. The dictionary is available on the web[7].

2) LexALP multilingual terminological Database: The aim of the LexALP[8] project [12] was to build a multilingual (English, French, German, Italian and Slovene) terminological database on the legal terms of the alpine convention.

The particularity of this project is that each term of the domain can have several translations in the same language. For example, a specific term will not be translated in Swiss, Austria and Germany in the case of German and France and Swiss in the case of French. Furthermore, it can happen that a term has several equivalents in one language in one country. Therefore, in order to take this into account, the chosen macrostructure is a pivot one as seen in figure 5.

Figure 5. LexALP entry

The database is available on the web[9].

3) MOTAMOT multi-bilingual dictionary for under-resourced languages: The MOTAMOT project is the newest one. It consists in building a multilingual database via bilingual dictionaries. The languages considered are the followings: French, English, Khmer and Vietnamese. It it planned in the future to add other languages. It is targeting communities of voluntary contributors. Anybody will contribute. In order to have an idea of the quality of the data, every contributor and every entry will have a number of stars that will represent his/her level. The figure 6 shows that the French entry ”abaisser” has received 3 stars.

Every people will contribute to a given bilingual dictionary, e.g. an entry with langA-langB. In the backgroud, an interlingual link will be built and added to the pivot volume. If the entry langB is linked to another entry lang C, a draft link with only one star will be added to the langA-langC dictionary and will wait for validation. A preliminary version of the platform is available on the web[10]. This is ongoing work and some results should be available after one year.

Figure 6. Motamot draft entry

IV. How to collect important specialized lexical data

A. Remaining issues for collecting data

The success of the Wikipedia project could make anybody think that it should not be very difficult to build a dictionary online. But this is not the case. Even if their size is regularly increasing, the dictionary Wiki projects (wiktionary, wiktionaryZ), did not meet the same success as their elder brother Wikipedia. Early experiments showed some problems when building a dictionary online with a community of voluntary contributors.

The first problem is the motivation. when people contribute to an encyclopedia entry, they are recognized as experts in their field. Their work is rewarded by the fame. But, for a dictionary entry, it is not the case because the entries are rather small compared to encyclopedias and every entry has the same structure. The contributions are much more anonymous. Therefore, it is more difficult to motivate people to contribute to writing dictionary entries compared to encyclopedia ones.

Another problem is that only specialized lexicographers can contribute to complex and detailed entries. Non specialists people that are willing to contribute cannot help very much. For example, the DiCo database [4] contains around 1,000 entries. Nevertheless, any native speaker has an in-depth knowledge of the language albeit implicit. Anybody is able for example to distinguish correct or wrong utterances of their language.

Lets confess it, writing dictionaries is not fun! But wouldn’t there be a more funny way to exploit the implicit knowledge that each one has about her/his mother tongue? All these observations led us to think about collecting important specialized lexical data through online word games.

B. JeuxDeMots: collecting data through serious word games

JeuxDeMots game tries to address the latter issues. It aims at building a rich and evolving lexical network, that could be compared to a certain extend to the famous WordNet [13] database.

The principle is the following: a game needs two players. When a player A initiates a game, an instruction is displayed concerning a type of competency corresponding to a lexical relation (synonym, antonym, domain, intensifier, etc.) and a word W is chosen randomly in the database. The player A has then a limited amount of time for giving propositions that answer the instruction applied to the word W.

The same word W with the same instruction is proposed to another player B and the process is the same. The two

[7]http://www.estfra.ee

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

[9]http://217.199.4.152:8080/termbank/LexALP.po

[10]http://papillon.imag.fr

half-games the one of the player A and the one of the player B are not simultaneous but asynchronous. For each common answer in A and B propositions, the two players earn a certain amount of points and credits. For the word W, the common answer of A and B players are entered into the database. It participates to the construction of a lexical network linking terms with typed and weighted relations, validated by pairs of players. The relations are typed by the instructions given to the players and weighted with the number of pair players that proposed them.

The structure of the lexical network that should be obtained is built upon the notions of nodes and relations between nodes like in [14]. Every node of the network is a lexical unit gathering all its lexies or word senses. The relations between nodes come from lexical functions as the ones presented by [5]. The figure 7 shows the relations gathered for the French word ”dictionnaire”.

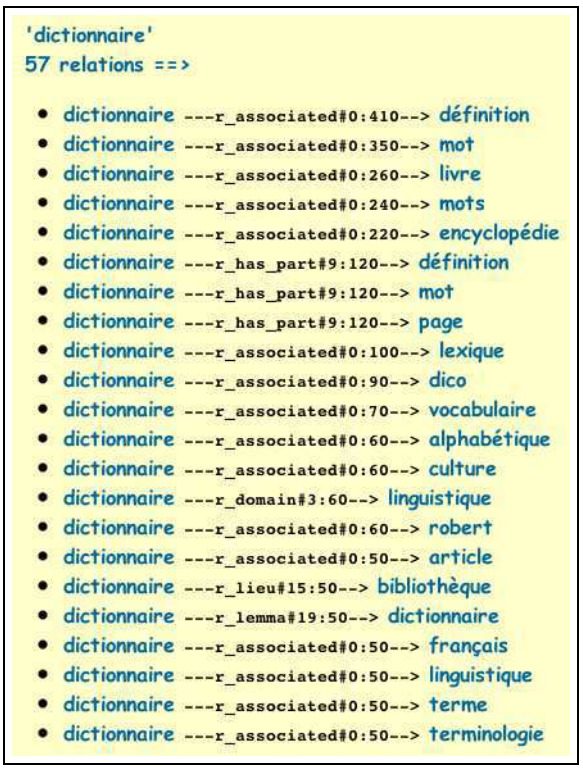

Figure 7. JeuxDeMots network for ”dictionnaire”

The first version of the game for French was launched in July 2007. There are also existing versions in English, Japanese and Thai. They are available on the web [11]. The game is easily translatable into another language. People interested can contact us if they want to launch a game in their language.

C. Intermediate results of the jeuxdemots projects

After 18 months, there are more than 1,000 active players that created around 150,000 games with 20,000 of them are pending (waiting for another player). There are now more than 170,000 relation occurrences created. The players entered 10,000 new terms or forms mainly linked to the news. There are now more than 160,000 terms in the database.

The relation results obtained with JeuxDeMots (JDM) were compared to the French Euro WordNet (EWF). JDM has around 6 times more terms than EWF that has 23,000 terms. Concerning the relations, JDM has now more than twice more relations than EWF. Furthermore, these numbers are constantly increasing in the case of JDM. A sample of 100 terms the most frequently used by the JDM users was analyzed. It appears that in 97% of the cases, the associations are correct. The remaining 3% of the cases correspond to errors (most of the time typos and spelling errors) or misunderstandings. Data collected with JDM bring a lot of originality, but the precision rate is less important that the data obtained with EWF. This lack of precision is rapidly decreasing when the relations have more weight.

Another study tried also to compare the results of lexical functions obtained with JDM with the manually entered lexical functions found in the DiCo database [4]. The Magn function (intensifier) was selected and compared the two corpora. There are 14 words in common with Magn function available. On 81 function results, there are 11 results in common. Then, the results were gathered following their weight: w=50: 0/50; w=60: 3/12; w=70 : 2/9; w=80 : 0/2; w=90 : 2/2; w=100: 1/1; w=110: 1/2; w=130: 0/1; w=340: 1/1; w=350: 1/1. If we consider the results with a weight bigger than 80, there are 6 results in common out of 8.

While the corpus is still to small to give solid conclusions, we can see that when there is a significant weight, the results of JDM are very similar to the ones of DiCo database. This intermediate result is very promising for the future. There are plans to tweak the JDM game in order to propose more often words that are also in the DiCo database for a certain lexical function. It will then be possible to obtain a bigger corpus for comparison.

V. How to integrate these resources with other systems

A. Description of the needs

We showed previously that the jibiki platform is ready to be used in many resources building projects. Nevertheless, there lack of the possibility to interact with other systems (like Machine Translation) in order to mutualize the entry building efforts. There is also a need to give possibility to use semi-automatic techniques to detect inconsistencies and enhance the quality of the data. The following part will describe an extension of the Jibiki platform in order to answer to these needs.

B. PIVAX: an extension for synchronization and data handling

PIVAX is the first online[12] contributive lexical database system allowing to create, maintain and manage the lexical resources of Machine Translation systems using a ”lexical pivot”. These resources can be heterogeneous because

[11] urlhttp://www.jeuxdemots.org

[12] http://javalig3.imag.fr/pivax/

their language-specific components are developed at different places, with different linguistic approaches and computational tools. Only the most basic language-specific lexical information must be stored in PIVAX, so that developers can use their own tools and protect proprietary information. In order to use the resources between different systems, a new macrostructure with three layers is proposed (figure 8). For each natural

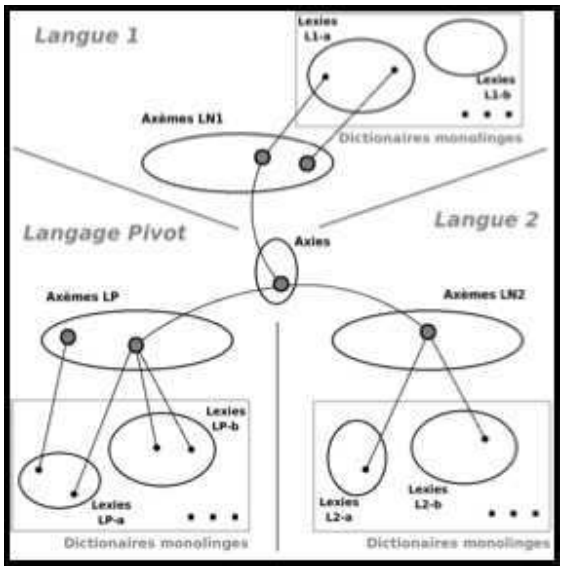

Figure 8. PIVAX Macrostructure

language (NL) supported, and each formal interlingua, there are:

- One or more volumes for lexies (and associated information). Lexies correspond to word senses in dictionaries.
- A unique volume for axemes (”monolingual acceptions”). An axeme links synonymous lexies of the same language.
- A unique volume for axies (interlingual acceptions), linking synonymous axemes

With this structure, the solution to collect all lexical Vietnamese resources becomes relative simple:

- for monolingual Vietnamese dictionaries, a volume is created for each dictionary, e.g. the dictionary of abbreviations or vietnamese dictionaries from Bamboo;
- for bilingual dictionaries, they are separated into two monolingual volumes and resulting relations are put into axeme and axie volumes.

This generic structure and platform gives bases to mutualize (and normalize) all resources available. For each entry, like Jibiki platform, it gives possibilities to define any kind of microstructure. In most cases, some common information are separated: lemma, POS, example, etc... All other private, non-sharing or encoding information specific for each system are commented in order to hide them from user interface. For example, it could be possible to define a microstructure for VDict dictionary, and then, the editing interface is generated automatically.

Navigation and research interface are inspired from PARAX, an environment created and used by a computational linguist [15]. In the following screen capture 9, we can see the French word tester from Ariane MT system volume. It is connected to other words tester from Systran MT system.

C. Usages of PIVAX for integration with other systems

We observed some problems from the first application of PIVAX for U++C[13] during the creation of Universal Words for UNL[14] system. It is not really used efficiently for some reasons:

- Subjective: each partner has its own private tools, and don’t want to re-begin a new environment;
- Objective: lack of synchronization tools between the center database and the local systems

The solutions are using and integrating PIVAX in our different projects, giving accessibilities via constraints and contexts. One of the projects is EOLSS[15]/UNL-FR, translation of 1/300 of the EOLSS on-line encyclopedia (25 documents, about 220,000 words) from English into the 5 other Unesco languages in 6 months. To support this project, a post-edition environment called SECTra_w [16] has been developed. In SECTra_ws interface, 3 panels are added for terminology consultation and contribution. When the user post-edits a segment, all translations of the occurrences in the segment will be displayed in the first panel. A robot collects this draft resource automatically. When the user submits a translation, by default, used translations will be taken from the first panel to the second panel. If the user changes words directly on the dictionary panel or adds some new words, theses contributions are considered stable and are copied to the third panel. In another project called OMNIA, image research & indexing via ontology, PIVAX is used like an environment for presenting a multilingual ontology[17].

Towards a more programmable platform, our own lexical resource would be considered as a lexical graph, and the refinement work in lexical resource becomes graph manipulation. Then, a type of generic SLLP (Specialized Language for Linguistic Programming) allows users to define themselves some kinds of constraints on the lexical graphs and associated actions for lexical graph manipulation.

VI. CONCLUSION

The road to the perfect lexical resource is long and difficult. However, some progress has been done step by step. Concerning the tools, the Jibiki online lexical database managing platform is a success. It is now used in very different projects, from bilingual dictionaries to multilingual termbanks. Concerning the data, the intermediate results of the JeuxDeMots projects for French are very promising. It would be very interesting now to launch other languages versions in order to experiment

[13]open and free association of researchers, bussiness entities and people with a common interest in the development of useful applications to society based in the UNL language, http://www.unl.fi.upm.es/consorcio/

[14]Universal Networking Language: http://www.undl.org/

[15]Encyclopedia Of Life Support Systems), http://www.eolss.net/

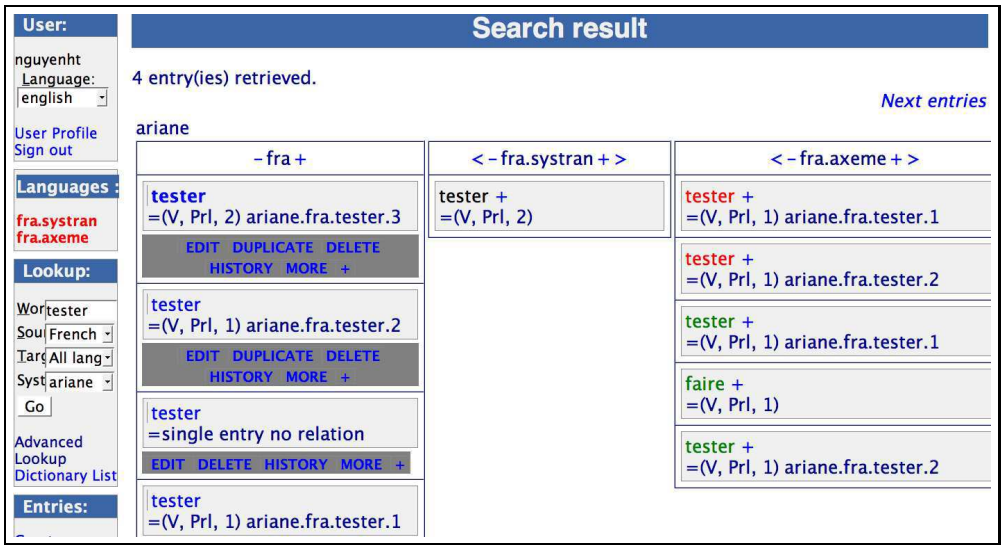

Figure 9. Displaying search results in columns

with multilingual data. The Motamot project seems also very promising. There is a plan to start the contribution phase before two years from now. The project will also be a way to experiment the validity of our theory developed around the monolingual word senses and interlingual links. It will then be possible to say if the next step has been achieved successfully. Anybody who is motivated and wants to join the project is welcomed. It is mainly based on voluntary work and aims to build a reference lexical resource. Concerning the development APIs and the synchronization between databases, the internal collaborations in our team has showed interesting possibilities. Lets meet again in two years for a new conclusion.

ACKNOWLEDGMENT

The MOTAMOT project has been partly funded by the Action de Recherche en Rseau program of the Agence Universitaire de La Francophonie[16].

REFERENCES

[1] V. Berment, "Méthodes pour informatiser des langues et des groupes de langues "peu dotées"," Ph.D. dissertation, Université Joseph Fourier Grenoble I, Grenoble, France, 18 mai, 277 p. 2004.

[2] M. Mangeot, "Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue," Thèse de nouveau doctorat, Spécialité Informatique, Université Joseph Fourier Grenoble I, Septembre 2001.

[3] C. Boitet, M. Mangeot, and G. Sérasset, "The papillon project: cooperatively building a multilingual lexical data-base to derive open source dictionaries and lexicons," in *Proc. of the 2nd Workshop NLPXML 2002, Post COLING 2002 Workshop*, G. Wilcock, N. Ide, and L. Romary, Eds., Taipei, Taiwan, 1 September 2002, pp. 93–96.

[4] A. Polguère, "Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for french," in *Proceeding of EURALEX'2000, Stuttgart*, 2000, pp. 517–527.

[5] I. Mel'čuk, A. Clas, and A. Polguère, *Introduction à la lexicologie explicative et combinatoire*, ser. Universites francophones et champs linguistiques. Louvain-la Neuve: AUPELF-UREF et Duculot, 1995.

[6] M. Mangeot, "Papillon project: Retrospective and perspectives." in *Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine, LREC workshop*, P. Zweigenbaum, Ed., Genoa, Italy, 22 May 2006, p. 6.

[7] G. Sérasset and M. Mangeot, "Papillon lexical database project: Monolingual dictionaries and interlingual links," in *NLPRS-2001*, Tokyo, 27-30 November 2001, pp. 119–125.

[8] M. Mangeot and G. Sérasset, "Frameworks, implementation and open problems for the collaborative building of a multilingual lexical database," in *Proc. of SEMANET Workshop, Post COLING 2002 Workshop*, G. Ngai, P. Fung, and K. W. Church, Eds., Taipei, Taiwan, 31 August 2002, pp. 9–15.

[9] M. Mangeot, G. Sérasset, and M. Lafourcade, "Construction collaborative d'une base lexicale multilingue," *Traitement Automatique des Langues*, vol. 44, no. 2, pp. 151–176, February 2004.

[10] M. Mangeot, "An xml markup language framework for lexical databases environments: the dictionary markup language," in *LREC Workshop on International Standards of Terminology and Language Resources Management*, Las Palmas, Spain, 28 May 2002, pp. 37–44.

[11] M. Mangeot and A. Chalvin, "Dictionary building with the jibiki platform: the gdef case," in *LREC 2006*, Genova, Italy, 21-26 May 2006, pp. 1666–1669.

[12] G. Sérasset, "Multilingual legal terminology on the jibiki platform: The lexalp project," in *Proc. of Papillon 2005 Workshop*, M. Lafourcade, Ed., Chiang Rai, Thailand, 11-13 December 2005, pp. 64–73.

[13] G. A. Miller, R. Beckwith, C. . Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: an on-line lexical database," *International Journal of Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.

[14] A. Polguère, "Structural properties of lexical systems: Monolingual and multilingual perspectives," in *Workshop on Multilingual Language Resources and Interoperability (COLING/ACL 2006)*, Sydney, 17-21 July 2006, pp. 50–59.

[15] É. Blanc, "Parax-unl: a large scale hypertextual multilingual lexical database," in *NLPRS'99: the 5th Natural Language Processing Pacific Rim Symposium*, Beijing, China, 1999, p. 4.

[16] C.-P. Huynh, C. Boitet, and H. Blanchon, "Sectra_w.1: an online collaborative system for evaluating, post-editing and presenting mt translation corpora," in *LREC 2008: 6th Language Resources and Evaluation Conference*, Marrakech, Morocco, 26-30 May 2008, p. 6.

[17] D. Rouquet and H.-T. Nguyen, "Multilinguïsation d'une ontologie par des correspondances avec un lexique pivot," in *Toth 2009*, Annecy, France, juin 2009, p. 19.

[16] http://www.ltt.auf.org