HAL
open science

Construction of an open-source multilingual lexical system targeted on French and Japanese through contributive and automatic methods

Mathieu Mangeot

▶ To cite this version:

Mathieu Mangeot. Construction of an open-source multilingual lexical system targeted on French and Japanese through contributive and automatic methods. [Research Report] Laboratoire d'Informatique de Grenoble. 2015. hal-01294562

HAL Id: hal-01294562
https://hal.science/hal-01294562v1

Submitted on 29 Mar 2016

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.

Construction of an open-source multilingual lexical system targeted on French and Japanese through contributive and automatic methods

Research project in Japan

Mathieu MANGEOT-NAGATA
Department of Digital Media
Hosei University
3-7-2 Kajino-cho, Koganei,
Tokyo 184-8584 Japan
Mathieu.Mangeot@imag.fr

Introduction

This research is located in the natural language processing (NLP) domain, at the intersection of computer science and linguistics, more specifically on multilingual lexicography and lexicology.

In a first long stay in Japan from November 2001 to March 2004, we made the observation that the French-Japanese lexical resources available on the Web were almost nonexistent. Which gave birth to the Papillon project of building a multilingual lexical database with a pivot structure (Sérasset et al., 2001). Since then, progress has been made in several areas (technical, academic, social) (Mangeot, 2006), but the production of real data has made very little progress. On the other hand, reuse of lexical resources is trendy (WSD, use of open-source resources like Wiktionary or Dbpedia, fusion with ontologies, etc.). Even if they can consolidate and expand the coverage of existing resources, these experiences always start from data created manually by lexicographers.

Based on this observation, we defined the following project which involves building a multilingual lexical system with focus on the French-Japanese language pair. Construction will be based on the reuse of existing resources (Franco-Japanese lexicons, Wiktionary) and automatic operations (reification of translation links, word senses disambiguation) and also on a community of volunteer contributors working on the Web. They will be asked to contribute either via serious lexical games, or directly on dictionary entries according to their level of expertise and knowledge in the field of lexicography or bilingual translation.

Resources generated will be royalty free and designed to be used both by humans via bilingual dictionaries and tools for automatic language processing (analysis, machine translation, etc.).

We will begin with a brief inventory of bilingual lexicography in general and focus on French-Japanese in particular. We then present recent advances in the field of construction of lexical resources online. Then, we describe in more detail the lexical system that we plan to build. We conclude with a description of the steps involved in this construction.

1 State of the art of bilingual lexicography

Lexicography is now at a turning point that began in the end of the twentieth century with the rise of the computer. Paper dictionaries are increasingly difficult to find an audience and their electronic versions are far from being commercial successes.

The main difficulty of the current bilingual lexicography lies in the prohibitive costs of building large amounts of data. For example, the Electronic Dictionary Research project (EDR, 1993) whose

aim was to build a Japanese-English dictionary took more than 1,200 person/years of work. The selling price of about € 84,000, is much lower than the actual costs of construction, costs that will probably never be recouped.

Anyway, these costs are too high for an individual. Therefore, only institutions can acquireit. On the other hand, the data provided at this price can only be used by some machine translation systems based on specific techniques.

In front of these unmanageable costs, publishers end up living on their laurels and offer mainly new editions of existing dictionaries. Few publishers have the courage to embark on the construction of a new high quality bilingual dictionary from scratch.

On the other hand, even in the most comprehensive dictionaries, there is almost always a lack of information especially on collocations. Scarce resources that take them into account do not do it systematically.

Despite the advent of the Internet, there are now few good quality lexical resources freely available online. Most are in fact bilingual lexicons made by volunteers non-specialist in lexicography.

Multilingual lexicography as such is in fact its infancy. Indeed, there is no real way to print a true "multilingual dictionary." On the contrary, it is quite possible to find multilingual terminology databases (as IATE) or small lexicons and multilingual phrasebooks.

2 State of the art of Japanese bilingual lexical resources

Although French and Japanese languages are considered well resourced in terms of tools and linguistic resources, the French-Japanese couple is considered an under-resourced language pair. There are indeed no royalty free high quality bilingual lexicons. French-Japanese aligned bilingual corpora and machine translation systems are logically also rare.

For historical as well as practical reasons, the Japanese society quickly put emphasis on English. The English-Japanese couple is one of the best equipped at present with very substantial resources like the EDR dictionary (1993) and machine translation systems are among the most efficient.

2.1 French-Japanese published dictionaries

Existing high quality Japanese-French dictionaries are published dictionaries that exist only on paper or in compiled electronic dictionaries (denshi-Jishou). There is no online interface for consultation.

2.1.1 French→Japanese dictionaries

Le Dico (Hakusuisha, 1993) contains 34,000 entries.

The Crown (Sanseido) contains 47,000 entries.

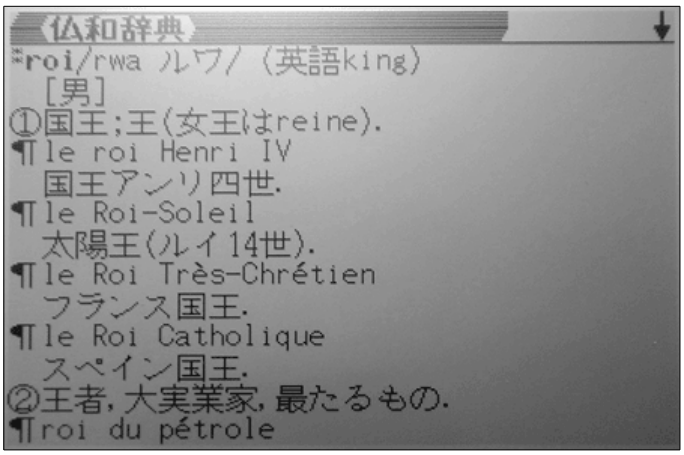

Figure 1: Screen shot of the Crown dictionary in electronic version

2.1.2 Japanese→French dictionaries

The Royal (Obunsha, 1992) contains 42,000 entries.

The Concise (Sanseido) contains 38,000 entries.

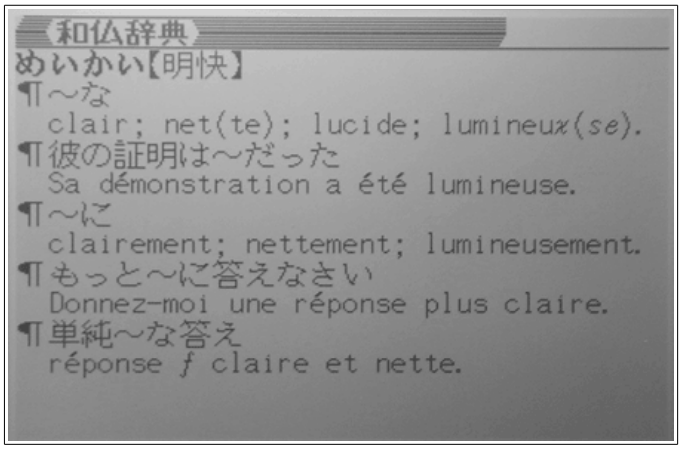

Figure 2: Screen shot of the Concise dictionary in electronic version

Conclusion

Advantages: These dictionaries are very complete with definitions and usage examples.

Disadvantages: not free, these dictionaries are protected by copyright. They are not transferable and non-reusable in other projects. The latest edited versions date of twenty years ago. Furthermore, many of these dictionaries have been designed for Japanese speakers (Le Dico, The Royal, etc). Therefore, there is no latin transcription of Japanese words. Their reading requires mastery of kanji, which greatly reduces their usability.

2.2 Wiktionary projects

The French Wiktionary has currently 2.2 Mo of entries of which 1.2 Mo are French and a little less than 7,000 are Japanese translations of which half of them are a proper names (this is often a simple transcription of the headword in Japanese syllabary). There are also some translations included from the website dictionnaire-japonais.com (see below). Translations are specified in the entry level and not word sense level. There is no description of the translation context (gloss, examples), or information on the Japanese translations (part-of-speech, etc).

The Japanese Wiktionary has 83,000 entries of which 26,000 are Japanese and 2,800 are French entries translated into Japanese (there are inflected forms or verbal forms, eg 32 entries for the verb "aimer"). The coverage is very poor.

Wiktionary projects are interesting and trendy but they have several limitations:

- The entry structure is free. It is not possible to use the same microstructure for all entries.
- Although it is possible to describe in a Wiktionary of language A, a word sense of an entry in language B, the interface is not designed to prepare bilingual dictionaries. For example, the description of the reverse link language A → language B must be done by hand in the Wiktionary of language B.

- It is also not possible to automatically add existing data from other sources to construct a draft for later refinement.
- Contributions are anonymous. It is not possible to use a high quality data or reviewing / validation system.

2.2 Online Japanese bilingual resources

2.2.1 Japanese→English dictionary: JMdict

The JMdict[1] (Japanese-Multilingual Dictionary) (Breen, 2004) is a project led by Jim Breen. It contains 165,000 entries Japanese-English entries with some translations in other languages: German (from WaDokuJiten), 31,000 French equivalents (from dico FJ), Russian, etc.

Search Key: 食べる Current Dictionary: Jpn-Eng General (EDICT)
Options:[G]oogle search, [GI] Google images, [S]anseido dictionary, [A]LC dictionary (Eijiro), [Ex]ample sentences, [V]erb conjugations, [F] Feedback, [L]esson from JapanesePod101.com, [JW] Japanese WordNet, [W] Japanese Wikipedia,[Edit] Edit this entry,[Promote] Move to JMdict/EDICT.

食べる(P); 喰べる(iK) 【たべる】 (v1,vt) (1) to eat; (2) to live on (e.g. a salary); to live off; to subsist on; (P) [Edit] [V][Ex][L][G][GI][S][A][W] [JW] [L][G][GI][S][A]
エジプトでは何を食べて生活していますか。 What do they live on in Egypt?[Amend]
ガンガン食べる; がんがん食べる 【ガンガンたべる(ガンガン食べる); がんがんたべる(がんがん食べる)】 (exp,v1) (sl) to pig out; to chow down [Edit] [V][G][GI][S][A] [G][GI][S][A]
生で食べる 【なまでたべる】 (exp,v1) to eat raw (fresh) [Edit] [V][L][G][GI][S][A]
一口食べる 【ひとくちたべる】 (v1) to eat a mouthful [Edit] [V][L][G][GI][S][A]
ぼりぼり食べる 【ぼりぼりたべる】 (exp,v1) to eat with a munching or crunching sound [Edit] [V][L][G][GI][S][A]

Figure 3: entry 食べる of the JMdict dictionary

Advantages: resource coverage, royalty-free and available for free download. It is regularly revised and completed.

Disadvantages: unidirectional Japanese→English. There is no reverse dictionary English→ Japanese. The microstructure is limited: translation contexts are not described. It also lacks of a definition and examples.

2.2.2 Japanese→German dictionary: WaDokuJiten

The WaDokuJiTen[2] from Ulrich Apel (Apel, 2002) consists of more than 280,000 entries. Its coverage and its microstructure are more developed than the JMdict.

Nr.	Japanisch	Lesung	Deutsch	Worttyp	
1	食べる	たべる	[1] essen; speisen; zu sich nehmen; fressen; probieren. [2] leben von.	下一他	
2	食べるのを遠慮する	たべるのをえんりょする	nicht essen.	サ変自	
3	食べるとしゃきしゃきする	たべるとしゃきしゃきする	beim essen knusprig sein.	サ変自	

Figure 4: entry 食べる of the WaDokuJiTen dictionary

Advantages: JMdict the most complete in terms of coverage and information free of charge

1 http://www.csse.monash.edu.au/~jwb/jmdict.html
2 http://www.wadoku.de

and available for free download.

Disadvantages: like JMdict, the dictionary is unidirectional. It does not use examples to illustrate the translation contexts.

This dictionary is to date the most comprehensive Japanese-other language lexical resource available for free download. It is a goal for our resource in terms of coverage.

2.3 French-Japanese online resources

2.3.1 Projet Dico FJ

The dico FJ project, pioneer in the field, was launched in early 2000 by Jean-Marc Desperrier (Desperrier, 2002). It contains a little more than 10,000 entries from translation of Japanese-English JMdict dictionary. There has been no change since 2003.

Advantages: free of charge and available for free download.

Disadvantages: more thant the same disadvantages of JMdict, there are translation errors due to the fact that some contributors unfamiliar with the Japanese language directly translates English translations instead of Japanese headwords, which increases the number of misinterpretations.

2.3.2 dictionnaire-japonais.com

The dictionnaire-japonais.com[3] project contains a little over 28,000 entries. It is a significant improvement compared to other French-Japanese dictionary online projects. Each user can directly contribute by adding entries. The community of contributors seems quite active as evidenced by the project forum. Information available for each entry is relatively limited to a "grammatical type" a "category" (domain), a language level, and sometimes an "origin of the word" (etymology).

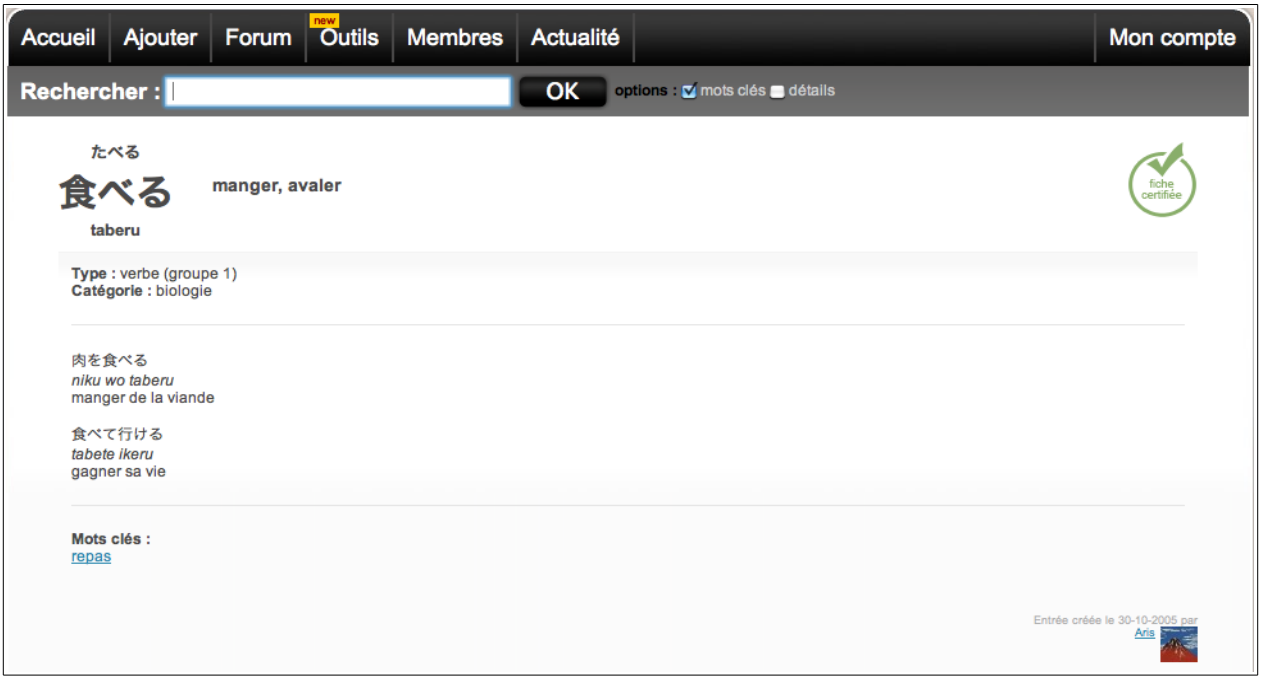

Figure 5: entry 食べる of the dictionnaire-japonais.com

Advantages: available online, covering a little more vocabulary than the dico FJ dictionary,

3 http://www.dictionnaire-japonais.com

active community of volunteer contributors.

Disadvantages: more than the FJ disadvantages , it seems that the a atesource files are not available for download.

2.4 Lexicons from machine translation systems

The UNL[4] project defines a semantic pivot language called UNL. For each natural language N, it is then possible to program language N→ UNL enconverters and UNL → language N deconverters. French and Japanese partners of the project have each produced a lexicon for their conversion systems. These lexicons are available online[5].

The Japanese-UNL lexicon contains 1,266,694 entries. Here is the Japanese entry 食べる (to eat):

[食べる]{} "eat(icl>food)";

The French-UNL lexicon contains 520,305 entries. Here are the entries corresponding to the verb "manger":

[manger]{AUX(AVOIR),CAT(CATV),GP1(DE),VAL1(GN)}"consume(icl>event)";

[manger]{AUX(AVOIR),CAT(CATV),VAL1(GN)}"eat";

Advantages: these lexicons have a large coverage large and can be used to build a skeleton dictionary.

Disadvantages: the microstructure is very small: a lemma in French or Japanese and its equivalent in UNL, sometimes with information from systems such as part-of-speech (CAT). These lexicons are not available for free download.

2.5 Conclusion

The Japanese-French lexicons available online have a small size and are all oriented from Japanese to French.

Most French-Japanese dictionaries also lacks specific information concerning Japanese. There is no example to our knowledge of a dictionary with both kanji (ideograms), kana (syllabaries) and romaji (roman alphabet transcription). Dictionaries without romaji are designed for Japanese speakers. Dictionaries without kanji are designed for French beginners. It is also difficult to find some important information such as counters (objects are not counted the same way: one car = ichi dai =, one dog = ippiki, etc.) or language levels.

In conclusion, for personal use, it is possible to find good quality printed (or electronic) dictionaries printed if one knows how to read kanji, but when there is a need for a free online dictionary, there is often no other choice but to use an English-Japanese dictionary, which, as we know, can only increase errors of understanding and translation.

3 Advances in the construction of online resources

Our Ph.D. thesis (Mangeot, 2001) defines a number of theoretical foundations in the area. The first concrete steps have been made with the Papillon project: requirements definition, structures and types of desired information, implementation of the Jibiki platform for handling lexical resources online. Followed a reflection on how to contribute online indirectly through serious games with the JeuxDeMots project. Several resources construction projects were also conducted using the Jibiki

4 http://www.undl.org/
5 http://www.undl.org/uwgate/

platform but none was based on the French-Japanese language pair. The ongoing GDEF project (Mangeot et al., 2006) aims to build a high quality Estonian-French dictionary for translators. The LexALP project (Sérasset et al, 2006) has produced a multilingual terminology (German, French, Italian, Slovenian) with a pivot structure for the vocabulary of the Alpine Convention. The MotÀMot project (Mangeot et al., 2010) focused on the recovery of an existing French-Khmer dictionary.

3.1 On contributing aspects: Wikipedia and Wiktionary projects

The Wikipedia online contributive encyclopedia met a big resounding success. Although the Wiktionary projects are developing, the success is not yet there for some languages (26,000 entries for Japanese) and they are rarely used to build bilingual dictionaries. The WiktionaryZ project, who claimed to ward off Wiktionary defects does not have the desired effect.

One hypothesis to explain this problem is the motivation. Indeed, when a person contributes to a Wikipedia article, it is rewarded with fame. It will then be recognized as an expert in her field. This is not possible with a dictionary. The contributions focus on small parts of highly targeted information and are therefore anonymous. On the other hand, there is a technical aspect related to the structure. An encyclopedia article is more or less free while a dictionary entry must follow a very precise microstructure (headword, grammatical information, semantic blocks, translation blocks, examples, etc.). It is therefore not possible to reuse a wiki platform to build a dictionary with a well defined structure.

Once accepted the idea that writing dictionary entries is not as much fun as working on a Wikipedia article, one must find ways to motivate a community of volunteers to contribute to a dictionary. L Serious lexical games are a first track. It is also possible to highlight the contributors, for example through a table of top contributors of the month. And finally, the social networks such as Facebook should also bring grist to the mill.

3.2 On technical aspects: the Jibiki platform

Jibiki (Mangeot et al., 2004) is a generic platform for managing online lexical resources with users and groups handling, heterogeneous resources lookup and generic dictionaries entries editing. This is a community website originally developed for the Papillon project. The platform is programmed entirely in Java, based on the "Enhydra" framework. All data is stored in XML format in a relational database (Postgres). This website offers two main services: a unified interface for simultaneous access to many heterogeneous resources (monolingual, bilingual dictionaries; multilingual databases, etc.) and an editing interface in order to contribute directly to specific dictionaries available on the platform.

The editor is based on a HTML interface model instantiated with the entry to edit. The model can be automatically generated from a description of the structure of the entry with an XML schema. It can then be modified to improve the screen rendering. The only information required is the XML schema representing the structure of this entry. In addition, it is possible to edit any type of dictionary entry as soon as it is encoded in XML.

Several lexical resources construction projects used or still use this platform successfully. This is the case for example of the GDEF Estonian-French bilingual dictionary project[6]. The code for this platform is open source and downloadable from the LIG Laboratory forge[7].

6 http://www.estfra.ee/
7 http://jibiki.ligforge-imag.fr/

3.3 On the data collecting through serious games: the JeuxDeMots projects

JeuxDeMots[8] (Lafourcade & Joubert, 2008) is an attempt to answer to the anonymous contributing problem. This project aims to build a lexical network rich and scalable, which can be compared to some degree to the famous WordNet base (Miller et al., 1990). The principle is as follows: one game requires 2 players. Player A starts a game. The type of relation to build (synonyms, opposites, domain, etc.) is displayed, and a word M randomly selected in a database of words. Player A then has a limited time to respond with proposals corresponding to, according to him, the word M and the type of game. This same word with the is then asynchronously offered to another player B, the process is identical.

The two half-games, the one of player A and of player B are not simultaneous but asynchronous. For each common response to proposals A and B, both players earn a number of points. The structure of the lexical network is based on the concepts of nodes and relations between nodes to build a network type of (Polguère, 2006). Each node consists of a lexical unit (word or phrase) involving all its lexical items and the relationships between nodes reflect the lexical functions. The first version of the game for French was launched in July 2007. There are also English, Arabic, Japanese, Vietnamese, Thai and more recently Portuguese versions (Mangeot et al., 2012). They are available on the Web.

4 Description of the lexical system to build

4.1 Microstructure of entries based on the Meaning-Text Theory

The microstructure of the entries gathered into monolingual volumes is a simplification of the Papillon project one. Each entry is based on the vocable. A word is either a group of lexies (a word sense), or an idiom.

The lexie consist of a name, grammatical properties, a semantic formula that can be seen as a formal definition - in the case of a predicative lexie, the formula describes the predicate and its arguments and the syntactic realization of the arguments - and a list of lexical semantic features - there are 56 basic lexical functions applicable to any language that can be combined together - a list of examples and finally to a list of idioms.

To cope with different contributor skill levels, the editing interface can adapt itself and display appropriate information. For example, a beginner contributor will be prompted for a simple gloss to characterize a lexie, while an expert linguist will describe a complete semantic formula. Similarly, some contributors only have access to the list of lexical functions to fulfill.

4.2 Pivot macrostructure pivot via bilingual interfaces

The macrostructure is derived from the Papillon project with a monolingual volume for each language and a center pivot volume. This macrostructure has been tested and validated in the LexAlp[9] project (Sérasset et al., 2006) that aimed to build a multilingual terminology for the vocabulary of the Alpine Convention. This project also uses the Jibiki platform (Mangeot et al., 2004) as well as for its development and online consultation.

When a new entry in a language A is added, it must be connected to interlingual volume. These links are created either by reusing existing bilingual dictionaries language A → language B, or by adding them manually from a translation. The link language A → language B then becomes language A → pivot → language B. If the entry of language B is already connected to another entry

8 http://jeuxdemots.org/
9 http://217.199.4.152:8080/termbank/LexALP.po

of C language, then the entry of language A also benefit from these links.

However, in order not to confuse users, they contribute through an interface with a classic bilingual dictionary view. Each bilingual link language A → B language added via this interface will actually be translated in the background by the creation of two interlingual links as well as a pivot link representing the original translation link in order to obtain finally language A → pivot axie → language B. This idea has been used for the MotÀMot[10] project (Mangeot et al., 2010) which led to the construction of a bilingual French-Khmer dictionary.

4.3 Establishing bilingual and interlingual links

If a contributor wants to add a translation link between a vocable in language A Va and a vocable of language B Vb, s/he can establish that link at different levels.

The ideal solution is to connect a word sense Sa vocable Va to another word sense Sb of vocable Vb. In this case, the link is bijective and Sb is also connected to Sa.

If the vocable Vb has not yet any specific word sense or if the contributor can not choose a word sense, s/he can connect directly Sa to Vb. In this case, a new word sense Sb' of the vocable Vb is created and tagged with a draft quality and the link as well as the word sense Sb' are marked for further refinement.

In the case of existing data recovery, it is often impossible to attach information to a specific word sense. In this case, at the end of the vocable Va is added an information to indicate that one of the word senses of vocable Va is connected to one of the word senses of vocable Vb, but this information will not be added to Vb. It will of course be marked for emergency refinement!

Thanks to the pivot macrostructure, if two links language A → language B and language B → language C exist, then a link language A → language C will be automatically created and its level will be marked as draft and to be revised.

4.4 Data and contributors quality levels

Each piece of information for each entry is assigned a level of quality. The levels range from 1 star for a draft (recovered data whose quality is not known) to 5 stars, for an entry certified by an expert (eg, a translation link validated by a sworn translator).

Likewise, the contributors will be assigned a skill level (1 to 5 stars as well). 1 star is a beginner level unknown in the community and 5 stars is the level of a recognized expert.

Then, when a level 3 contributor revises a level 2 entry, the entry automatically rises to level 3. Similarly, if the work of a contributor is systematically validated without corrections by other contributors of a better level, s/he can automatically be switched to the next level after a certain threshold (eg 10 contributions).

To go further, we plan to analyze the work of contributors. If a person contributes massively eg on a particular domain, the system will automatically send her regular contribution proposals in its domain.

5 Working plan

5.1 Gathering existing resources

The first stage of the project is to collect existing resources that we are allowed to reuse (primarily

10 http://jibiki.univ-savoie.fr/motamot/

royalty free). These resources are lexical (dictionaries, lexicons) and corporal (aligned bilingual corpus or similar). The latter will be used for extracting bilingual examples.

5.1.1 Dictionaries available on the Internet

- JMdict dictionary with its 31,000 French equivalents;
- French and Japanese Wiktionary translations.

5.1.2 Resources of NLP projects

- Data of the ongoing sakura-survitra project with Kyoto university;
- French-UNL and Japanese-UNL lexicons from the UNL project with United Nations Universoty and Tokyo Soft company;
- Pre-terminological data from Mohammad Daoud's co-supervised with Pr. Kyo Kageura from University of Tokyo (Daoud et al. 2010), (Daoud, 2010);
- Any data from other Japanese partners (Hosei University, National Institute of Informatics, Laboratory of Prof. Fuji Ren at University of Tokushima, Laboratory of Prof. Yoshinori Sagisaka at Hokkaido University, Laboratory of Prof. Yves Lepage at Waseda University, etc.).

5.1.3 Personal lexicons from translators and interpreters

In the absence of a reference dictionary with broad coverage or specialized terminology databases, it is common for translators and interpreters to build their own lexicons most often using Excel files. These resources could be reused in the project.

5.2 Production of a skeleton dictionary to be revised

Once collected, the resources must be converted into a single format and then automatically merged. Two entries with the same headword and the same part-of-speech are merged into a single entry with several word senses. The translation links must be reified: from a French article with a Japanese translation Fj, a Japanese article J and a pivot entry A connecting the two entries A are generated (Fj ⇒ F← A→ J).

The second step consists in merging the different word senses of an entry resulting from a previous fusion. Word sense disambiguation techniques (Navigli, 2009) such as ant colony algorithms (Schwab et al. 2011) can then be applied.

When the first version of the resource is ready, it is then put online to be corrected and enriched semi-automatically by other programs and also manually by a community of volunteer contributors.

5.3 Semi-automatic enrichment of the resource

We plan to enrich the resource with the JeuxDeMots data via two main methods:

1. Lexical elicitation: recent research has proved that "serious lexical games", such as JeuxDeMots, are quite useful to gather monolingual lexical resources. Promising research is currently done about ways to develop bilingual or multilingual games on top of JeuxDeMots games. While JeuxDeMots has already produced a large lexical network for French, an important objective of the project is to considerably expand the audience of the (existing)

Japanese JeuxDeMots[11].

2. Collocation discovery and translation: an important part of the needed resources is made of collocations, and in particular of multiword terms. Statistical processing and analogous processing have been successfully used to find such terms in specialized English corpora (e.g. in the Genia corpus), and to propose translations of them in other languages.

5.4 Animating a community of contributors around the project

A first conclusion of our experiments is that JeuxDeMots projects based on the work of voluntary contributors must always be animated constantly otherwise, they can be abandoned rapidly past the early stages of discovery. One of our main tasks will be to lead a community of contributors through project advertising, forums moderating, social networks, etc. It will also be necessary to recruit experts to supervise contributors and lead also the community. Tools can also be set up to reward contributors (best contributor of the month, etc.).

Finally, we plan to start again the task of scientific communication around the project by organizing a new Papillon seminar in Japan, giving presentations in partner laboratories and participating in conferences in the domain.

References

Apel U. (2002) *WaDokuJT - A Japanese-German Dictionary Database.* Papillon 2002 Seminar, 16-18 July 2002, NII, Tokyo, Japan, 13 p.

Berment V. (2004) *Méthodes pour informatiser des langues et des groupes de langues "peu dotées".* Thèse de nouveau doctorat, Université Joseph Fourier Grenoble I, Grenoble, France, 277 p.

Breen JW. (2004) *JMDict: a Japanese-multilingual dictionary.* In: Coling 2004 workshop on multilingual linguistic resources, Geneva, Switzerland, pp. 71-78.

Daoud M., Kageura K., Boitet C., Kitamoto A. & Mangeot M. (2010) *Multilingual Lexical Network from the Archives of the Digital Silk Road.* Proc. of OntoLex 2010, 6thWorkshop on Ontologies and Lexical Resources, hosted by COLING 2010, Beijing, 22 August 2010, 9 p.

Daoud M. (2010) *Utilisation de ressources non conventionnelles et de méthodes contributives pour combler le fossé terminologique entre les langues en développant des "préterminologies" multilingues.* Thèse de nouveau doctorat, spécialité informatique, Université Joseph Fourier Grenoble I, 192 p.

Desperrier J-M. (2002) Analyze [sic] of the results of a collaborative project for the creation of a Japanese-French dictionary. In: Proceedings of Papillon 2002 Seminar, Tokyo, Japan.

EDR (1993) *EDR Electronic Dictionary Technical Guide.* Project Report, n°-042, Japan Electronic Dictionary Research Institute Ltd., 16 August 1993, 144 p.

Lafourcade M. & Joubert A. (2008) *JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes.* In JADT 2008 : 9es Journées internationales d'Analyse statistique des Données Textuelles, Lyon, France, pp. 657–666.

Mangeot M. (2001) *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue.* Thèse de nouveau doctorat, spécialité informatique, Université Joseph Fourier Grenoble I, 280 p.

Mangeot M. (2006) *Papillon project : Retrospective and perspectives.* In P. Zweigenbaum, Ed., Acquiring and Representing Multilingual, Specialized Lexicons : the Case of Biomedicine, LREC workshop, Genoa, Italy, 6 p.

Mangeot M., Sérasset G. & Lafourcade M. (2004) *Construction collaborative d'une base lexicale multilingue.* Traitement Automatique des Langues, vol. 44(2), pp. 151–176.

Mangeot M. & Chalvin A. (2006) *Dictionary building with the Jibiki platform : the GDEF case.* In LREC 2006, Genova, Italy, pp. 1666–1669.

Mangeot M. & Sereysethy Touch S. (2010) *MotÀMot project: building a multilingual lexical system via bilingual dictionaries,* SLTU 2010: Second International Workshop on Spoken Languages Technologies

11 http://jeuxdemots.liglab.fr/jpn/

for Under-Resourced Languages, Penang, Malaysia, 3-5 May 2010, 6p.

Mangeot M. & Ramisch C. (2012) *A Serious Lexical Game for Building a Portuguese Lexical-Semantic Network.* Proceedings of the ACL 2012 3rd Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP, Jeju, Republic of Korea, jul 2012.

Mel'čuk I., Clas A. & Polguère A. (1995) *Introduction à la lexicologie explicative et combinatoire.* Universités francophones et champs linguistiques. Louvain-la Neuve : AUPELF-UREF et Duculot, 256 p.

Miller G. A. et al. (1990) *Introduction to WordNet : an on-line lexical database.* International Journal of Lexicography, 3(4), pp. 235–244.

Navigli R. (2009) *Word Sense Disambiguation: a Survey.* ACM Computing Surveys, 41(2), ACM Press, 2009, pp. 1-69.

Polguère A. (2000) *Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French.* In Proceeding of EURALEX'2000, Stuttgart, pp. 517–527.

Polguère A. (2006) *Structural properties of lexical systems : Monolingual and multilingual perspectives.* In Workshop on Multilingual Language Resources and Interoperability (COLING/ACL 2006), Sydney, Australia, pp. 50–59.

Schwab D., Goulian J. et Guillaume, N. (2011) *Désambiguïsation lexicale par propagation de mesures sémantiques locales par algorithmes à colonies de fourmis.* 18ème conférence sur le Traitement Automatique des Langues Naturelles (TALN 2011), 27 juin - 1er juillet 2011, Montpellier.

Sérasset G., Brunet-Manquat F. & Chiocchetti E. (2006) *Multilingual Legal Terminology on the Jibiki Platform: The LexALP Project.* COLING/ACL2006 Conference, 17-21 july 2006, Sydney, Autralia.

Sérasset G. & Mathieu Mangeot M. (2001) *Papillon Lexical Database Project: Monolingual Dictionaries & Interlingual Links.* Proc. of NLPRS 2001, Hitotsubashi Memorial Hall, National Center of Sciences, Tokyo, Japan, 27-30 November 2001, pp. 119-125.