HAL
open science

Représentation Sémantique de Lexique pour un Dictionnaire de Traduction Manuelle et Automatique

Mutsuko Tomokiyo, Peter Weyer-Brown, Mathieu Mangeot

To cite this version:

Mutsuko Tomokiyo, Peter Weyer-Brown, Mathieu Mangeot. Représentation Sémantique de Lexique pour un Dictionnaire de Traduction Manuelle et Automatique. Actes des Aspects méthodologiques pour l'élaboration de lexiques unilingues et multilingues, May 2006, Bertinoro, Forli, Italie. pp.12. hal-00968625

HAL Id: hal-00968625

https://hal.science/hal-00968625

Submitted on 1 Apr 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.

Représentation Sémantique de Lexique pour un Dictionnaire de Traduction Manuelle et Automatique

Mutsuko Tomokiyo, GET/ENST (Paris) et GETA-CLIPS (Grenoble), mutsuko.tomokiyo@imag.fr
Peter Weyer-Brown, GET/ENST (Paris),
weyer@enst.fr
Mathieu Mangeot, Condillac-LISTIC-Université de Savoie (Chambéry)
Mathieu.Mangeot@univ-savoie.fr

Résumé

Cet article a pour but de proposer une description de lexique, inspirée de la lexicographie explicative et combinatoire issue de la théorie «Sens-Texte», dont la structure est basée sur le dictionnaire multilingue Papillon, et d'établir un protocole de rédaction d'un dictionnaire monolingue du français et de l'anglais contribuant à la traduction manuelle et automatique de la plaquette « Recherche » de l'ENST (École Nationale Supérieure de Télécommunication).

Mots-clefs

base de données lexicale, lexicographie explicative et combinatoire (LEC), théorie "Sens-Texte"(TST,), fonction lexicale, traduction manuelle et automatique, dictionnaire multilingue Papillon, dictionnaire de terminologie, plaquette « recherche » de l'ENST

Abstract

This paper aims to propose a description of a lexicon, inspired by the Explanatory Combinatorial Lexicography, part of the "Meaning-Text" theory. The structure will be based on the Papillon multilingual dictionary. The lexicon will be built using the Jibiki online dictionary management platform. A protocol will be established in order to write this monolingual dictionary of French and English contributing to machine and human translation for the ENST (École Nationale Supérieure de Télécommunication) booklet.

Keywords

lexical database, "Meaning-Text" theory (MTT), Explanatory and combinatory lexicography (ECL), lexical function, human and machine translation, Papillon multilingual dictionary, dictionary of terminology, ENST "Research" booklet

1 Introduction

Malgré un but commun, une partie conséquente des informations nécessaires à la traduction manuelle et celles nécessaires à la traduction automatique est très différente et cela donne lieu à la construction de ressources lexicales distinctes. Notre expérience acquise dans le domaine de la construction de ressources lexicales pour dictionnaires à usage humain dans le cadre du projet Papillon d'une part et pour systèmes de traduction automatique dans le cadre du projet UNL d'autre part nous a naturellement amenés à tenter de fusionner ces deux

catégories de ressources.

C'est pourquoi nous proposons une description de lexique inspirée de la lexicographie explicative et combinatoire (LEC) issue de la théorie «Sens-Texte» (TST), dont la structure est basée sur le dictionnaire multilingue Papillon et nous établissons un protocole de rédaction d'un petit dictionnaire monolingue du français et de l'anglais contribuant à la traduction manuelle et automatique. Nous prévoyons d'expérimenter ce dictionnaire pour traduire manuellement puis automatiquement la plaquette bilingue « recherche » de l'ENST (École Nationale Supérieure de Télécommunication).

Dans la suite de cet article, nous présentons d'abord le projet Papillon et son dictionnaire multilingue à structure pivot en détaillant sa macrostructure. Nous expliquons ensuite la LEC issue de la TST et son application à la microstructure du dictionnaire Papillon. La section suivante présente le corpus d'expérimentation : la plaquette « Recherche » de l'ENST. Ensuite, nous discutons d'une application du dictionnaire à la traduction automatique dans le cadre du projet "Universal Networking Language (UNL)". Enfin, nous menons une réflexion en guise de conclusion dans la dernière partie. Dans l'Annexe 1, la description de la lexie "évaluer" est donnée.

2 Présentation du dictionnaire Papillon

2.1 Historique du projet Papillon

La base de données lexicale Papillon est construite par des personnes ayant des compétences diverses. Elle est ouverte au public au niveau de son utilisation ainsi qu'au niveau de son développement. Nous nous proposons donc de faire un dictionnaire pour la plaquette "Recherche" sur la plateforme de Papillon, et d'aborder sa traduction automatique en utilisant ce dictionnaire dans un futur proche[1].

2.2 Macrostructure du dictionnaire

La base de données Papillon s'appuie sur plusieurs volumes monolingues au niveau structurel, et établit une mise en relation automatique entre le sens d'un mot dans une langue et le sens correspondant dans l'autre langue grâce à un lien interlingue reliant les acceptions[2] des deux mots (Mangeot, 2001). La figure 1 montre la connexion de l'acception sur Papillon.

Le mot français "carte" a deux sens, et est donc décrit comme deux lexies: carte.1= carte à jouer et carte.2=carte géographique. La lexie "carte.1" est traduit en "カード(ka-do)" en japonais et la "carte.2" en "地図(chizu)". Par conséquent, les deux lexies sont enrégistrées comme l'acception différente dans la liste.

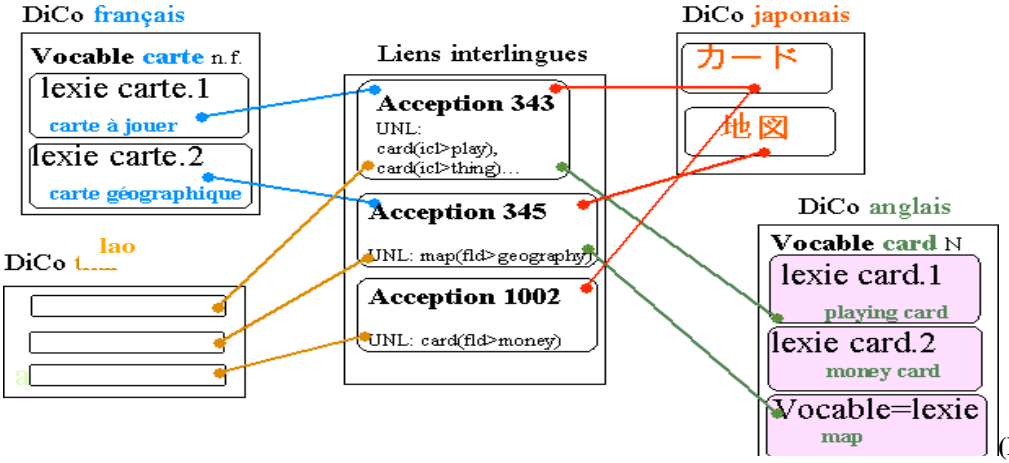
Par rapport à l'anglais, les lexies "carte.1" et "carte.2" sont traduites en "card" en anglais et "carte.1" est enregistrée comme "card.1" et "carte.2" comme "card.2" dans la liste. Mais le

---

[1] Papillon est en train d'être construit au niveau de données, et le nombre de ses entrées change tous les jours: le dictionnaire de l'anglais contient plus de 600 000 mots, et le dictionnaire du français contient 600 mots. La granularité de l'information des mots dans un dictionnaire n'est pas homogène pour l'instant, et dépend aussi des langues.

[2] L'acception est définie comme une interlangue conceptuelle commune aux langues traitées.

mot "card" a troisième sens, et il a donc troixième lexie: card.1=carte à jouer, card.2=carte géographique, et card.3=carte bancaire. Par conséquent, le "card.3" est enregistré comme l'acception différente dans la liste.



(Figure 1)

(Sérasset, 1994)

3 La Lexicographie Explicative et Combinatoire, partie de la théorie "Sens-Texte"

La description de lexique est basée sur une théorie lexicographique "Sens-Texte", qui a été proposée par Mel'cuk à Université de Montréal vers 1994 [Mel'cuk, 94].

3.1 La LEC et son application au dictionnaire pour "Recherche"

La représentation des lexies[3] dans le dictionnaire explicatif et combinatoire (DEC) est basée sur la théorie "Sens-texte" (Mel'cuk,1997)(DEC,1984)(Polguère,1998, 2000). L'approche consiste à observer des mots d'un point de vue multidimensionnel, c'est-à-dire à effectuer une analyse morphosyntaxique et sémantique. En fait, chaque mot est décrit par sa définition et de sa formation morphologique, indiqué par ses connotations et spécifié par la cooccurrence syntaxique et lexicale dans le contexte précisément donné.

Les entrées du dictionnaire ont les dix zones suivantes:

- zone vedette (lexie)
- zone phonologique (prononciation)
- zone morphologique (classe)
- zone stylistique (marques d'usage)
- zone sémantique (définition et connotations)
- zone de combinatoire syntaxique (régime, restrictions sur la cooccurrence des différents actants)
- zone de combinatoire lexicale restreinte (la cooccurrence contextuelle et ontologique)
- zone d'exemples
- zone de Nota Bene (OLST, 2005)

---

[3] La lexie est définie ici comme l'unité fonctionnelle significative du discours dans le contexte déterminé. La lexie simple peut être un mot, et la lexie complexe est une séquence figée.

D'après Mel'cuk (Mel'cuk97), *la langue est considérée comme un mécanisme, ou un système de règles, qui permet au locuteur de faire deux choses :* parler et comprendre. La tâche de linguistes est *donc de construire, pour la langue étudiée L , un système de règles — quelque chose comme un programme informatique — qui définisse les mêmes correspondances entre sens et textes que celles qu'établissent les locuteurs.*

C'est la raison symbolique pour laquelle nous avons pris cette théorie pour éditer un dictionnaire d'une part. D'autre part, le cœur conceptuel de la théorie consiste à analyser la combinatoire syntaxique de lexies et ses connotations ", et cela nous facilite de traiter l'information sur les mots dans les documents finalisés comme "Recherche". C'est la raison pratique pour laquelle nous avons appliqué cette théorie à notre dictionnaire.

3.2 Description des lexies

Dans le projet Papillon, nous avons repris la structure de description de lexies du projet DiCo qui est lui-même une simplification du DEC du point de vue de la structure. Nous l'avons ensuite encodée en XML au moyen d'une Définition de Type de Document (DTD)"[4]. Cette structure nous permet:

- de décrire les lexies de manière homogène,
- de les réviser,
- de les importer dans la base de donnée, ils seront disponibles ensuite en ligne au format HTML.

Pour assurer la traçabilité des données, nous stockons l'historique des opérations dans chaque entrée.

Voici la structure des lexies du volume du français: il y a la zone pour noter l'auteur de la lexie, la date de description et révision et la zone de l'historique.

La description de lexies commence par <headword hn="" /> et finit par </lexie

```
<?xml version="1.0"
encoding="UTF-8" ?>
  - <lexie d:id="Undefined" >
  - <history>
  - <modification>
    <author />
    <date />
    <comment />
    </modification>
    </history>
  <headword hn="" />
  <pronunciation />
  <pos />
  - <language-levels>
    <usage></usage>
    </language-levels>
  - <semantic-formula>
    <formula-label />
    <formula />
    </semantic-formula>
<government-pattern>
  - <mod nb="">
  - <actor>
    <sem-variable></sem-
variable>
    <synt-variable></synt-
variable>
  - <surface-group>
    <surface />
    </surface-group>
    </actor>
    </mod>
  </government-pattern>
- <lexical-functions>
  - <function>
    <name />
  - <valgroup>
    <value />
    </valgroup>
    </function>
</lexical-functions>
  <examples>
    <example />
    </examples>
  - <full-idioms>
    <idiom />
    </full-idioms>
    <more-info />
    </lexie>
```

N c l'exemple de lexie "évaluer" et "

Dans la Figure 1, une lexie à décrire se situe dans <headword hn=""/>, suivie par

[4] La "DTD" contient l'information du document (l'identification de version du document, langue de codage, l'auteur, la date d'écriture, la date de révision), le fichier historique, le contenu du document.

<pronunciation /> pour la prononciation de la lexie et par <pos/> pour sa classe grammaticale.

Exemple :

<headword >"évaluer"</headword >
<prononciation= "evaljue" </prononciation>
<pos>v.tr.</pos>

Dans <language-levels>, le niveau de politesse de la lexie est indiqué.

Exemple :

<language-levels>
<politeness>neutre</politeness>
</language-levels>

Dans <semantic-formula>, la définition de la lexie et une étiquette sémantique[5] sont données.

Exemple :

<semantic-formula>
<formula-label />
<formula>action : action intellectuelle</formula>
</semantic-formula>

Dans <government-pattern>, l'information syntaxique et sémantique d'une lexie est donnée par le nombre des actants de la lexie et ses classes grammaticales. Le verbe "évaluer" a 2 actants, qui sont marqués de "X" et "Y" à l'ordre de "I" et "II", quand il forme une phrase. Le premier actant est un nom comme classe grammaticale, et le deuxième actant l'est aussi.

Exemple :

<government-pattern>
- <mod nb="1">
- <actor>
<sem-variable>X</sem-variable>
=
<synt-variable>I</synt-variable>
=
- <surface-group>
<surface>N</surface>
,
</surface-group>
</actor>
- <actor>
<sem-variable>Y</sem-variable>
<surface-group>
<surface>N</surface>
</surface-group>
</actor>
</mod>
</government-pattern>

Dans <lexical-functions>, les fonctions lexicales de la lexie est analysée selon la liste des fonctions lexicales que le DEC donne (DEC92, pages 127-131). La liste contient 60 étiquettes qui indiquent diverses fonctions de lexies du point de vue multidimensionnel.

Ce sont par exemple:

- d'un point de vue de la synonymie

Exemple :

<lexical-functions>
<function><name>Qsyn[6]</name>
<valgroup><value>fixer</value> </valgroup>

---

[5] Un ensemble des étiquettes va être établi comme information ontologique.

[6] Qsyn = quasi-synonyme

```
</function>
</lexical-functions>
```

- d'un point de vue de rapports paradigmatiques

Exemple :

```
<lexical-functions>
<function><name>S0 (évaluer)[7]</name>
<valgroup><value>évaluation</value> </valgroup>
</function>
</lexical-functions>
```

- d'un point de vue de la transformation paraphrastique

Exemple :

```
<lexical-functions>
<function><name>Oper1 (évaluation)[8]</name>
<valgroup><value>faire</value> </valgroup>
</function>
</lexical-functions>
```

- d'un point de vue de modificateurs

Exemple :

```
<lexical-functions>
<function><name>A3(évaluation)[9] </name>
<valgroup><value>rapide</value> </valgroup>
</function>
```

Exemple :

```
<full-idioms>
<idiom dml:id="évaluation.1.i1">évaluation budgétaire</idiom>   <!--Logos -->
</full-idioms>
```

4 Plaquette de l'ENST, "Recherche"

"Recherche" est une plaquette[10] de l'ENST, qui explique des recherches menées dans l'école, en suivant la description faite par le directeur de la recherche à l'ENST. Elle mentionne la communication électronique, la communication satellitaires, le traitement du signal et des images, l'économie, la gestion, les sciences humaines et sociales, le traitement automatique des langues naturelles, les réseaux, etc. La plaquette est publiée à peu près tous les 5 ans en version française et en version anglaise.

Voici un extrait de "Recherche"

L'explosion de l'internet a démultiplié les sources et les volumes d'information textuelle disponibles en ligne, soulignant le besoin d'outils de traitement automatique de cette information (version française, page 11).

---

[7] S0 (évaluer) = forme nominale

[8] Oper1 (évaluation) = verbe (ici faire) sémantiquement vide qui prend le nom du deuxième actant

[9] A3 (évaluation) = modificateur typique pour 3ème actant du mot clé.

[10] Chaque plaquette est composée de 16 pages en format A4. Elle contient 7 771 mots au total et 40 619 caractères au total (espace non compris) pour la version française. Pour la version anglaise, elle contient 6 212 mots au total (espace non compris) et 37 211 caractères au total. Le ratio pour les mots différents par rapport aux mots totaux est 61%. Ces chiffres sont établis par un calcul de 1/16 page d'échantillon au hasard, et nous attendons qu'un logiciel nommé "Koyori" soit mis à jour, afin d'analyser plus précisément les plaquettes.

The Internet explosion has led to a considerable increase in the sources and volumes of textual information available online, underlining the need for automatic processing tools for this information (version anglaise, page 11).

5 Application du dictionnaire à la traduction automatique

Un de nos objectifs pour publier le dictionnaire est de l'utiliser dans les systèmes de traduction automatique dans le cadre du formalisme UNL.

5.1 "UNL" et le dictionnaire de UWs (Universal Words)

UNL est non seulement une langue de codage, mais aussi un formalisme ainsi qu'un environnement de traduction automatique, basé sur le système de pivot, fourni d'un dictionnaire nommé "UW dictionary", d'un "déconverter (générateur)" et d'un "enconverter (analyseur)". Le dictionnaire des UWs est un dictionnaire bilingue entre une sorte de langue pivot et la langue cible. La langue pivot est constituée de mots anglais désambiguïsés par des annotations diverses. Dans l'exemple suivant, l'expression comprenant le symbole "icl>" permet de désambiguïser les sens du mot "state".

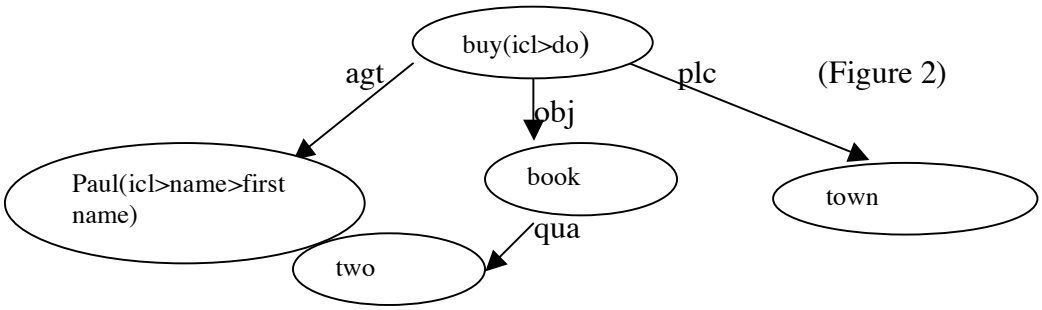
Exemple :
state(icl>express) -> énoncer en français
state(icl>country) -> nation en français
state(icl>abstract thing) ->condition en français
State(icl>government) -> Etat en français

Une expression UNL est constituée de graphes sémantiques basés sur une relation binaire entre deux mots dans un énoncé, qui est spécifiée par les étiquettes "relation" et "attribut". La "relation" dénote une relation syntaxico-sémantique entre deux mots et le nom de cette relation se situe sur l'arc reliant les deux graphes, et les deux mots se situent sur le nœud comme on le voit dans la figure 2. Un ensemble des "relation" se compose de 41 étiquettes. L'"attribut" dénote l'information morpho-syntaxique et grammaticale d'un énoncé et se compose de 30 étiquettes.

Voici un exemple d'expression UNL :

Exemple : Paul a acheté deux livres en ville.

agt(buy(icl>do).@entry.@past, Paul(icl>name>first name))
obj(buy(icl>do).@entry.@past, book.@pl.)
qua(book.@pl., two)
plc(buy(icl>do).@entry.@past, town)



(Figure 2)

Dans l'exemple, "agt", "obj", "plc" et "qua" sont l'étiquette de la "relation" entre les deux mots, et ".@entry", ".@past" et "@pl." sont l'étiquette d'"attribut". ".@entry" indique que le mot pivot "buy" est un mot dominant dans l'énoncé, ".@past" le temps de l'énoncé, "qua" la quantité, et "@pl." indique que le mot "book" est un nom en forme de pluriel.

Le transfert lexical du français au mot pivot est indiqué dans le dictionnaire UW pour le français: acheter->buy, deux->two, livre->book, et ville->town

5.2 Correspondance des informations entre le dictionnaire UWs et notre dictionnaire

La base lexicale "Papillon" est construit dans le but de fournir des outils linguistiques en modalité électronique ainsi que pour éditer des dictionnaires destinés aux traducteurs humains et aux systèmes de la traduction automatique.

Nous utilisons notre dictionnaire pour effectuer une traduction automatique dans le cadre du système UNL. Il s'agit donc la correspondance de l'information dans le dictionnaire UW à celle qui est donnée dans notre dictionnaire.

Ce qui est important pour un dictionnaire de traduction automatique est de trouver une façon de désambiguïser les significations de mots (Boitet, 2002), parce que la plupart de mots sont polysémiques et que le choix automatique du mot pour la langue cible correspondant à la langue source est une tâche difficile.

Nous citons deux exemples où deux dictionnaires proposent leur façon de désambiguïser.

Dans le dictionnaire UWs du français[11], le verbe "acheter" a deux descriptions:

- acheter-buy, AUX(AVOIR),CAT(CATV),GP2(A),VAL1(GN),VAL2(GN)

Cette description indique que le verbe "acheter" va être réécrit en "buy", qu'il prend "avoir" comme auxiliaire en forme composée, que la classe grammaticale du verbe "acheter" est le verbe transitif, que le verbe "acheter" forme un énoncé avec deux constituants comme valence, et que ses actants sont le syntagme nominal.

- acheter-bribe(obj>human), AUX(AVOIR),CAT(CATV),GP2(A),VAL1(GN),VAL2(GN)

Cette description indique que le verbe "acheter" signifie "se procurer, à prix d'argent, sa complicité ou sa complaisance" (e.g. "Cet homme d'affaires a acheté plusieurs députés.") (Logos, 1978), cet "acheter" peut être réécrit en "bribe" en mot pivot.

Pour désambiguïser les significations du verbe "acheter", l'information sur le syntagme nominal du deuxième actant est donnée dans le dictionnaire.

En effet, dans le dictionnaire que nous sommes en train d'éditer, ce genre d'information est donné dans les zones de <semantic-formula> et <lexical-functions> :

```
<semantic-formula> action : action intellectuelle
        <sem-label>quelqu'un</sem-label>
        <actor>X</actor>
        <sem-label>quelqu'un</sem-label>
```

[11] Le dictionnaire du français qui est actuellement utilisé contient 39 223 mots. On peut le visiter sur le site http://gohan.imag.fr/unldeco/

```
        <actor>Y</actor>
  </semantic-formula>

  <lexical-functions>
  <function name="Qsyn">
  <valgroup><value><reflexie xlink:href="corrompre.1">corrompre </reflexie> </value>
  </valgroup></function>
  </lexical-functions>

  <lexical-functions>
  <function name="Sinstr(acheter)[12]">
  <valgroup><value><reflexie xlink:href="argent.1">argent</reflexie> </value>
  <value><reflexie xlink:href="don.1">don</reflexie> </value>
  <value><reflexie xlink:href="promesse.1">promesse</reflexie> </value>
  </valgroup></function>
  </lexical-functions>
  <lexical-functions>
  <function name="S2C(acheter)">
  <valgroup><value><reflexie xlink:href="dirigeant.1">dirigeant</reflexie> </value>
  <value><reflexie xlink:href="juge.1">juge</reflexie> </value>
  <value><reflexie xlink:href="témoin.1">témoin</reflexie> </value>
  <value><reflexie xlink:href="député.1">député</reflexie> </value>
  </valgroup></function>
  </lexical-functions>
```

"S2c" indique la sémantique du deuxième actant du verbe. Elle peut être des mots utilisés dans le document ainsi que des mots plus abstraits tels que les mots dans un thésaurus.

En ce qui concerne le verbe "évaluer", le dictionnaire français-anglais donne 4 traductions : to value, to appraise, to assess et to estimate (Harrap's, 1989).

Dans l'énoncé "On y évalue la faisabilité et le réalisme économique des nouvelles technologies..."(version française, page 10), il est préférable que le verbe "évaluer" soit réécrit en "to assess" comme le mot pivot.

Pour cela, "assess" a une restriction sur son objet direct dans le dictionnaire UWs:
évaluer-assess(obj>abstract thing>feasibility, realism), AUX(AVOIR), CAT(CATV), GP2(A), VAL1(GN), VAL2(GN)

Dans notre dictionnaire de la plaquette, la même information est ainsi donnée dans la zone de <lexical-functions>:

```
  <lexical-functions>
  <function name="S2c(évaluer)"><valgroup>
  <value><reflexie xlink:href="faisabilité.1">faisabilité</reflexie></value>
  <value><reflexie xlink:href="réalisme.1">réalisme</reflexie></value>  </valgroup></function>
 </lexical-functions>
```

5 Réflexions

Nous avons fait, dans cet article, un rapport d'un projet pour éditer un dictionnaire

---

[12] Sinstr=nom typique pour le circonstant de l'instrument.

monolingue en français et en anglais en suivant la structure du dictionnaire Papillon, afin de l'utiliser pour la traduction automatique de la plaquette du français vers l'anglais. Les entrées du dictionnaire sont donc choisies dans la plaquette "Recherche" et "Research". Elles contiennent beaucoup de terminologie spécifique aux domaines de recherche de l'ENST.

Le traducteur recherche habituellement des informations sur la syntaxe de mots, le sens contextuel, l'expression figée, la terminologie, le nom propre, le synonyme, l'ontologie, l'usage pragmatique de mots, etc. La description par la théorie choisie ne peut pas entièrement répondre aux demandes de traducteurs. Par exemple, dans le dictionnaire thématique que nous avons proposé (Dickman, 1999), la terminologie de La Poste a été présentée sous une taxonomie « d'affaires postales »[13]. Nous pensons que l'information sur les domaines, l'expression figée, la sous-classification de terminologie, l'usage quotidien dans un domaine, etc. sont indispensables pour un dictionnaire de terminologie, alors que, pour l'instant, la sous-classification de terminologie et les expressions figées spécifiques au domaine ne sont pas encore abordées dans le dictionnaire monolingue sur Papillon.

Exemple :

clavier –keybord(icl>musical instrument, computer),
clavier- compass(icl>musical instrument>clarinet)

En ce qui concerne l'aspect pragmatique des mots, aucun dictionnaire n'en donne l'information, alors qu'elle est pourtant nécessaire pour certaines langues (Tomokiyo, 2003).

Enfin, il est question de la méthode pour établir une liste de lexies dans le dictionnaire. Nous utilisons un logiciel nommé "Koyori", qui a été développé dans le laboratoire de l'université de Nagoya au Japon (Ogawa, 2005). Il est une sorte de KWIC bilingue (keywords in context[14]), qui permet d'aligner et d'extraire une paire de mots correspondant entre deux langues différentes dans un document parallèle (par exemple, une version française et une autre version anglaise).

Références

(Boitet, 2002) Christian Boitet (2002), The translation of examples, citations and grosses in the Papillon Project, 2002 Seminar, 16-18 July 2002, NII, Tokyo, Japan

(DEC, 1992) Igo Mel'cuk (1992), Dictionnaire explicatif et combinatoire du français contemporain. les presses de l'Université de Montréal.

(Dickman, 1999) Vera Mihailovich-Dickman (1999). DICTIONNAIRE de la POSTE et des AFFAIRES POSTALES, La maison du dictionnaire. Paris.

(Harrap's,1989) Harrap's Shorter French and English Dictionary (1989) Harrap, London and Paris.

(Logos, 1978) Jean Gorodet (1978). *Logos, Grand dictionnaire de la langue française*. Bordas, Paris.

---

[13] Dans le dictionnaire français-anglais de la Poste, toutes les activités de la Poste principale et régionale sont schématisées, et toutes les étapes de traitement de courrier sont définies. Ce genre de l'information est utile pour la traduction manuelle et automatique .

[14] Le KWIC est une liste où tous les mots dans un document sont triés en forme canonique et que tous les énoncés comprenant le mot clé sont donnés avec des renseignement statistiques.

(Mangeot, 2001) Mathieu Mangeot & Gilles Sérasset (2001) Papillon Lexical Database Project : Monolingual Dictionaries & Interlingual Links, NLPRS01

(Mangeot, 2003) Mathieu Mangeot (2003) Papillon DTD. http://www-clips.imag.fr/geta/services/dml.

(Mangeot, 2003b) Mathieu Mangeot, Gilles Sérasset & Mathieu Lafourcade (2003) Construction collaborative de données lexicales multilingues, le projet Papillon. Revue TAL Traitement Automatique des Langues, ed. Michael Zock and John Carroll, Vol. 44:2/2003, pp. 151-176.

(Mel'cuk, 1997) Igor Mel'cuk (1997) Vers une linguistique Sens-Texte. Leçon inaugurale, Collège de France, Chaire internationale, 43 pages. http://www.fas.umontreal.ca/LING/olst/FrEng/melcukColldeFr.pdf

(Ogawa, 2005) Yasuhiro Ogawa and Katsuhiko Toyama. 2005. Bilingual KWIC- GUI Support Tool for Bilingual Dictionary Compilation. Proceeding of SNLP 2005. Volume II. Thailand.

(OLST, 2005) Dictionnaire explicatif et combinatoire, http://www.olst.umontreal.ca/decfr.html, Universite de Montréal

(Polguère, 2000) Alain Polguère (2000), Notions de base en lexicologie, Observatoire de Linguistique Sens-Texte(OLST)

(Polguère, 1998) Alain Polguère (1998), La théorie Sens-Texte. Dialangue, Vol. 8-9, Université du Québec à Chicoutimi, pp. 9-30. http://www.fas.umontreal.ca/LING/olst/FrEng/PolgIntroTST.pdf

(Sérasset, 1994) Gilles Sérasset (1994), Interlingual Lexical Organization for Multilingual Lexical Databases in NADIA, COLING-94, 5-9 August 1994, vol. 1/2: pp. 278-282.

(Tomokiyo, 2000) Mutsuko Tomokiyo (2000), Mathieu Mangeot & Emmanuel Planas. Papillon : a Project of Lexical Database for English, French and Japanese, using Interlingual Links, Journées des Sciences et Techniques de l'ambassade de France au Japon, Tokyo, Japon. 3p.

(Tomokiyo, 2003) Mutsuko Tomokiyo (2003). Description de propriétés pragmatiques pour les dictionnaires monolingues sur Papillon. International Workshop Papillon. Sapporo

(UNL, 2003) UNL Center & UNDL Foundation
The Universal Networking Language (UNL) Specifications. .Version 3 Edition 2
http://www.undl.org/unlsys/unl/UNL%20Specifications.htm

Annexe 1 Description de la lexie "évaluer"


```
<p:lexie p:id="fra.évaluer.1">
  <p:headword>évaluer</p:headword>
- <!-- Eng: asses ; jap: hyoukasuru-->
- <!-- Harrap's --> <!-- standardFJ -->
  <p:pronunciation encoding="romaji">evalje</p:pronunciation>
  <p:pos>v.tr.</p:pos>
- <p:language-levels>
  <p:politeness>neutral</p:politeness>
  </p:language-levels>
- <p:semantic-formula>
  <p:formula-label />
  <p:formula>action : action intellectuelle</p:formula>
  </p:semantic-formula>
- <p:government-pattern>
- <p:mod nb="1">
- <p:actor>
  <p:sem-variable>X</p:sem-variable>
  =
  <p:synt-variable>I</p:synt-variable>
  =
- <p:surface-group>
  <p:surface>N</p:surface>
  ,
  </p:surface-group>
  </p:actor>
- <p:actor>
  <p:sem-variable>Y</p:sem-variable>
  =
  <p:synt-variable>II</p:synt-variable>
  =
- <p:surface-group>
  <p:surface>N</p:surface>
  </p:surface-group>
  </p:actor>
  </p:mod>
  </p:government-pattern>
- <p:lexical-functions>
- <p:function>
  <p:name>Qsyn</p:name>
- <p:valgroup>
  <p:value>fixe</p:value>
  </p:valgroup>
  </p:function>
- <p:function>
  <p:name>Qsyn</p:name>
- <p:valgroup>
  <p:value>determiner</p:value>
  ,
  <p:value>arbitrer</p:value>
  ,
  <p:value>calculer</p:value>
  ,
  <p:value>chiffrer</p:value>
- <!-- Dictionnaire de Synonymes, Robert -->
  </p:valgroup>
  </p:function>
- <p:function>
  <p:name>So(évaluer)</p:name>
- <p:valgroup>
  <p:value>évaluation</p:value>
  </p:valgroup>
  </p:function>
  <p:function>
  <p:name>Ao(évaluer)</p:name>
- <p:valgroup>
  <p:value>évaluable</p:value>
  </p:valgroup>
  </p:function>
- - <p:function>
  <p:name>SQ2</p:name>
- <p:valgroup>
  <p:value>faisabilité</p:value>
  <p:value>réalisme économique</p:value>
  </p:valgroup>
  </p:function>
  <p:function>
  <p:name>Adv1</p:name>
- <p:valgroup>
  <p:value>précisément</p:value>
  <p:value>de manière précise</p:value>
  </p:valgroup>
  </p:function>
  </p:lexical-functions>
- <p:examples>
  <p:example p:id="fra.évaluer.1.e1">On y évalue la faisabilité et réalisme économique des nouvelles technologies radio.</p:example>
- <!--Recherche P.10 -->
  <p:example p:id="fra.évaluer.1.e2">il ne faut pas se lancer dans cette affaire sans avoir évalué les risques et la chances de réussite.</p:example>
- <!-- Logos -->
  </p:examples>
- <p:full-idioms>
  </p:full-idioms>
- <p:more-info>
- <p:axies>
  <p:refaxie xlink:href="papillon-axi#axi0015" />
  </p:axies>
  </p:more-info>
  </p:lexie>
```