

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

Mutsuko Tomokiyo, CLIPS (Grenoble) & ENST (Paris),

Peter Weyer-Brown, ENST (Paris)

Mathieu Mangeot, Université de Savoie (Chambéry)

Mathieu.Mangeot@univ-savoie.fr

Plan

- **Base de données lexicale « Papillon » et une théorie lexicographique « Sens-Texte »**
- **Application du dictionnaire à la traduction automatique dans le cadre de « Universal Networking Language (UNL) »**
- **Perspectives d'avenir**

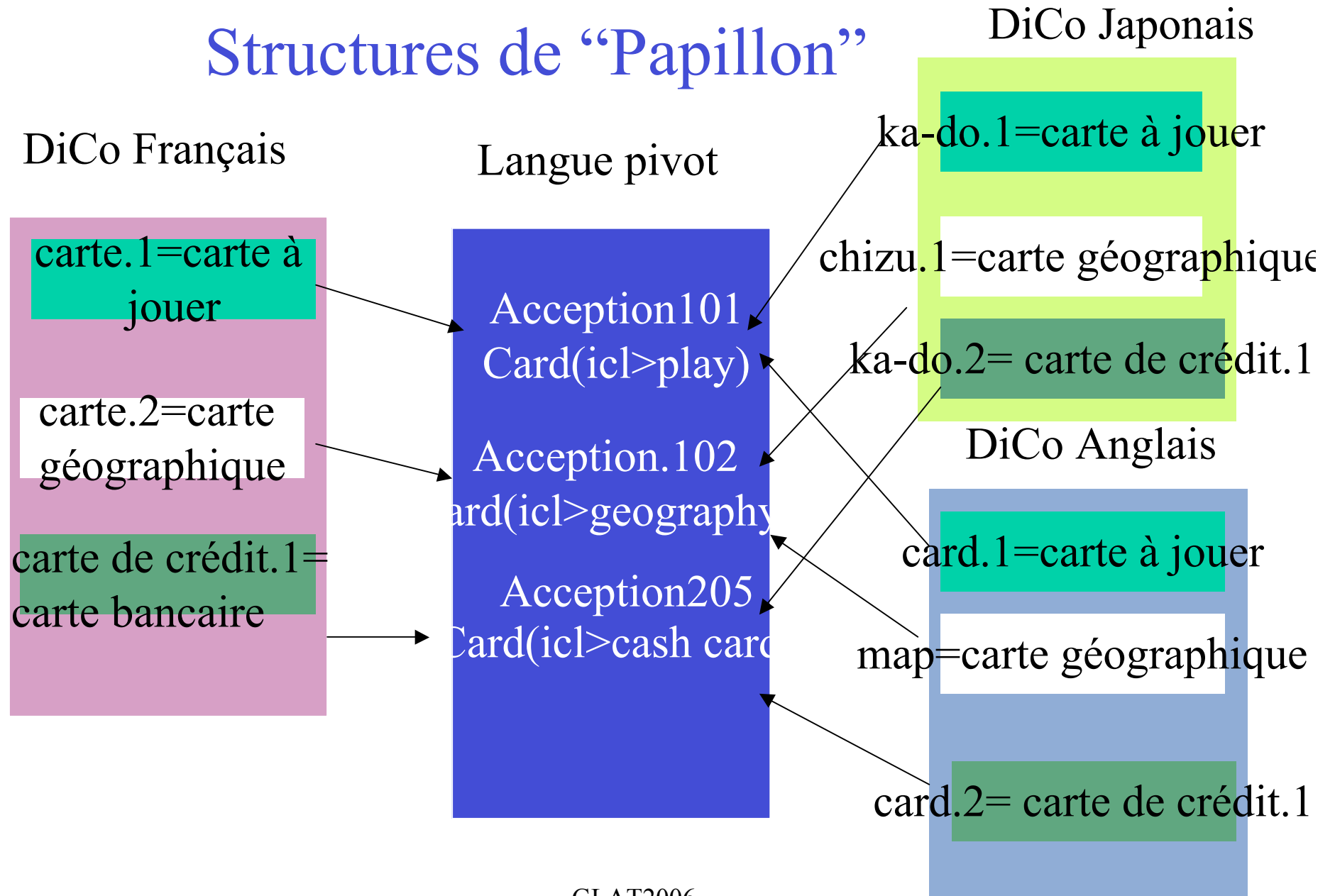
Base de données lexicales “Papillon”

Lancement du projet: 1999-

Partenaires : Consortium de GETA-CLIPS-
imag (Grenoble), NII (Tokyo) et Ambassade
de France au Japon (Tokyo)

Réalisation: Base de données lexicales
multilingue

Structures de “Papillon”



Théorie lexicographique “Sens-Texte” (1)

Description de lexies multidimensionnelle :

- l’aspect morpho-syntaxique, sémantique et pragmatique
- une analyse combinatoire et explicative

(DEC, 92)

Théorie lexicographique “Sens-Texte” (2)

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, ...qui définisse les mêmes correspondances entre sens et textes que celles qu'établissent les locuteurs
(Mel'cuk, 97)

Zones principales dans le dictionnaire DEC

- 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, 05)

Fichier de patron pour le dictionnaire du français (1)

Format XML au moyen d'une DTD:

- de décrire les lexies de manière homogène,
- de les réviser,
- de les importer dans la base de données, à fin de les rendre disponibles en ligne au format HTML

Fichier de patron pour le dictionnaire du français (2)

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- $Author: --> <!-- $Date: --> <!-- $Revision : 1.0 -->
<!-- ?xml-stylesheet type="text/css" href="../stylesheets/papillon-fr.css --
>
<lexied:id="Undefined" xmlns=http://www.clips.imag.fr/geta/
services/dml xmlns:d="http://www-clips.imag.fr/geta/services/dml"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www-clips.imag.fr/geta/services/dml
http://www-clips.imag.fr/geta/services/dml/papillon_deu.xsd">
<history><modification><author/><date/><comment/></modification>
</history>
```

Fichier de patron : lexies

<headword >"évaluer.1" </headword >

<prononciation= "evalµe" </prononciation>

<pos>v.tr.</pos>

Fichier de patron : usage

<language-levels>

<politeness grade="neutral"/>

<usage grade="NA"/>

<domain="NA"/>

</language-levels>

Fichier de patron : Régime.1

e.g. lexie.1 évaluer

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

Fichier de patron : Régime.2

<government-pattern>

<mod nb="1">

<actor>

<sem-actant>X</sem-actant> = †

<synt-actant>I</synt-actant> =

<surface-group><surface>N</surface></surface-group>

</actor>

<actor>

<sem-actant>Y</sem-actant> = †

<synt-actant>II</synt-actant> =

<surface-group> <surface>N</surface></surface-group>

</actor>

</mod>

</government-pattern>

Fichier de patron : Fonctions lexicales.1

- d'un point de vue de la synonymie :

```
<lexical-functions>
```

```
<function><name>Qsyn</name>
```

```
<valgroup><value>fixer</value> </valgroup>
```

```
</function>
```

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

Fichier de patron : Fonctions lexicales.2

d'un point de vue de rapports paradigmatiques :

```
<lexical-functions>
```

```
<function><name>S0 (évaluer)</name>
```

```
<valgroup><value>évaluation</value> </valgroup>
```

```
</function>
```

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

Fichier de patron : Fonctions lexicales.3

d'un point de vue de la transformation paraphrastique :
e.g. évaluation

```
<lexical-functions>
```

```
<function><name>Oper1 (évaluation)</name>
```

```
<valgroup><value>faire</value> </valgroup>
```

```
</function>
```

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


Fichier de Patron : Fonctions lexicales.4

d'un point de vue de modificateurs :

e.g. évaluation

```
<lexical-functions>
```

```
<function><name>A3(évaluation) </name>
```

```
<valgroup><value>rapide</value> </valgroup>
```

```
</function>
```

e.g. mesure

```
<lexical-functions>
```

```
<function><name>Magn(mesure) </name>
```

```
<valgroup><value>drastique</value> </valgroup>
```

```
</function>
```

(DEC, 92)
GLAT2006

Fichier de patron : exemples et expressions figées

<examples>

<example dml:id="fra.évaluer.1.e1"> On y évalue la faisabilité et réalisme économique des nouvelles technologies radio.</example> <!--P.10 -->

<full-idioms>

<idiom dml:id="évaluation.1.i1">évaluation budgétaire</idiom>

<idiom dml:id="évaluation.1.i2">évaluation fiscale</idiom>

<!--Logos -->

</full-idioms>

Application du dictionnaire à la traduction automatique

Universal Networking Language (UNL) :

- Langue de codage**
- Formalisme**
- Environnement de la traduction automatique avec un analyseur, un générateur et un dictionnaire « UWs (Universal words) »**

Dictionnaire de UNL : UWs

langue pivot: mots anglais désambiguïsés avec des restrictions

e.g.

state(icl>do>express) -> énoncer en français

state(icl>country) -> nation en français

state(icl>abstract thing) -> condition en français

State(icl>government) -> Etat en français

(UNL center, 03)

Formalisme de UNL : Hyper-Graphes sémantiques

e.g. Paul a acheté deux livres en ville.

<unlsem>

agt(buy.@entry.@past,Paul(icl>name>first name))

obj(buy.@entry.@past, book(icl>thing).@pl)

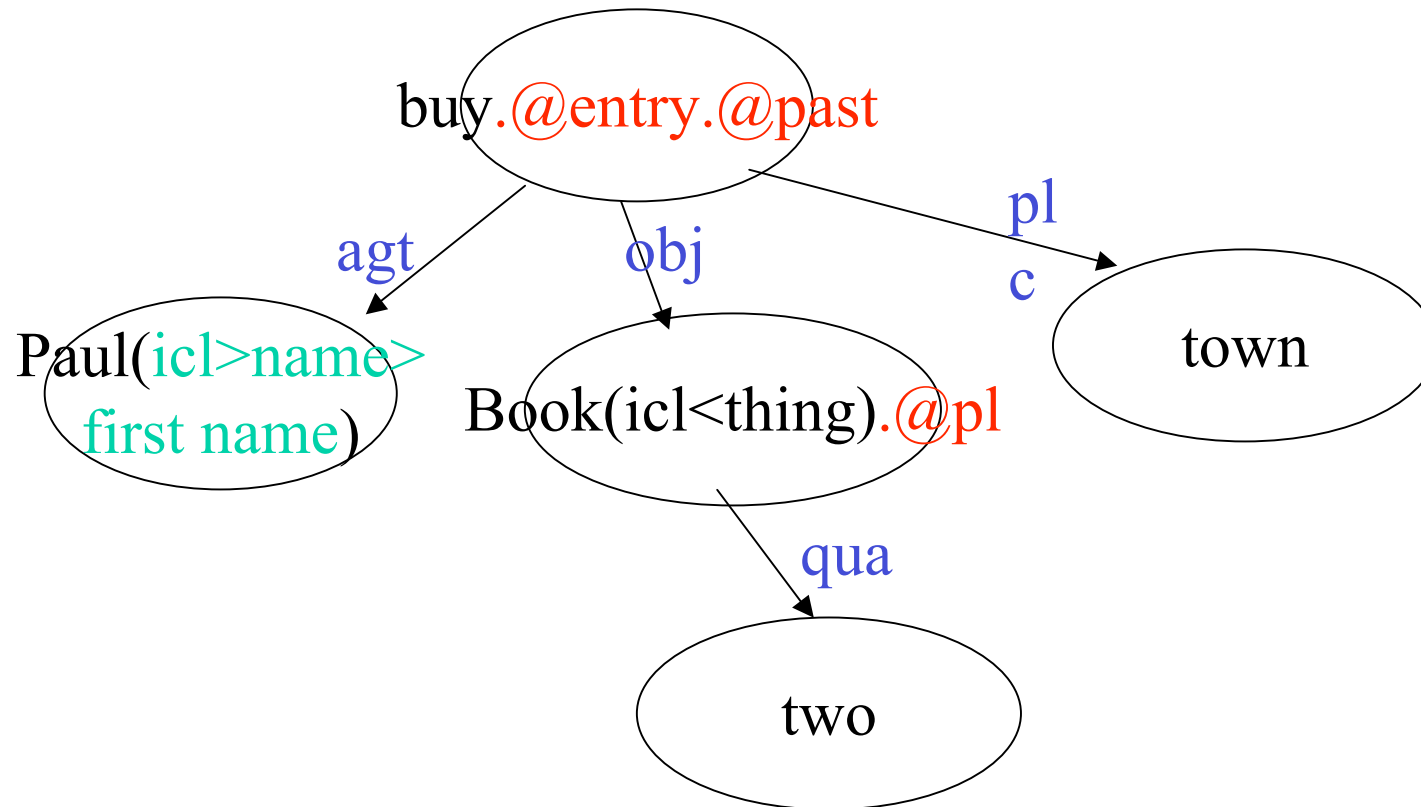
plc(buy.@entry.@past, town)

qua(book(icl>thing).@pl, two)

</unlsem>

(tomokiyo, 05)

UNL graphes -> déconvertisseur de GETA



(tomokiyo, 05)

Correspondance de l'information dans le dictionnaire UWs et le dictionnaire de Plaque

Dans le dictionnaire d'UWs bilingue :

a c h e t e r - b u y ,
AUX(AVOIR),CAT(CATV),GP2(A),VAL1(GN),VAL2(GN)

a c h e t e r - b r i b e (o b j > h u m a n) ,
AUX(AVOIR),CAT(CATV),GP2(A),VAL1(GN),VAL2(GN)

Correspondance de l'information dans le dictionnaire UWs et le dictionnaire de Plaquette

Dans le dictionnaire de Plaquette pour **acheter-bribe** :

<semantic-formula> action : action intellectuelle

<sem-label>quelqu'un</sem-label>

<actor>X</actor>

<sem-label>quelqu'un</sem-label>

<actor>Y</actor>

</semantic-formula>

Correspondance de l'information entre le dictionnaire UWs et le dictionnaire de Plaquette

« évaluer » : 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..."(page 10),

Dans le dictionnaire UWs "assess" a une restriction sur son objet direct :

évaluer-assess(obj>abstract thing>feasibility, realism),
AUX(AVOIR), CAT(CATV), GP2(A), VAL1(GN),
VAL2(GN)

Dans le dictionnaire de la plaquette :

```
<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>
```

Réflexions et perspectives (1)

- Pour la terminologie : sous-catégorisation des mots (Dickman, 1999)

e.g.

tri-classifying(icl>la poste)

tri- odd trick(icl>games>bridge)

tri-carrier tricycle(icl>vehicle)

trier-sort(obj>courier)

tournée-delivery(icl>distribution)

Réflexions et perspectives (2)

- Pour établir une liste de lexies un logiciel KWIC bilangue « Koyori » est utilisé.
- Pour intégrer les données de la Plaque sur Papillon au dictionnaire de UNL, il faut développer un outil.
- Pour la génération de l'anglais, on utilisera un décodeur de UNL.