



# Projet Papillon



Une base lexicale pour le français et  
le japonais basée sur des liens  
interlingues

Mutsuko Tomokiyo, Mathieu Mangeot & Emmanuel Planas

GETA/CLIPS IMAG Grenoble, France

Présenté par Nicolas Dessaigne  
National Institute of Informatics, Tokyo, Japon



# Plan

- Initiateurs et partenaires du projet
- Motivations et buts du projet
- Vue générale et architecture de la base
- Structure des dictionnaires monolingues
- Méthodologie de construction
  - Intégration de ressources existantes
  - Ajout de nouvelles entrées
  - Révision des nouvelles entrées
- Consultation de la base lexicale
- Travaux en cours
- Conclusion et contacts



# Initiateurs et partenaires

- **Initiateurs :**
  - Emmanuel Planas (GETA/CLIPS)
  - François Brown de Colstoun (Ambassade de France)
- **Partenaires :**
  - **NII**: National Institute of Informatics  
(Tokyo, Japon) Responsable : Frédéric Andrès
  - **GETA/CLIPS** : Groupe d'étude sur la T.A.  
(Grenoble, France) Responsable : Mathieu Mangeot
- **Et de nombreux contributeurs bénévoles**



# Motivations du projet

- Manque de dictionnaires d 'usage (et de toute façon payants) français <-> japonais
- Manque de dictionnaires pour linguiciens
- Informations non informatisées
- Internet permet aux linguistes, traducteurs et chercheurs de coopérer facilement
- Données du projet disponibles sous licence de logiciel libre (Open Source)
- Quelques données non structurées sont disponibles



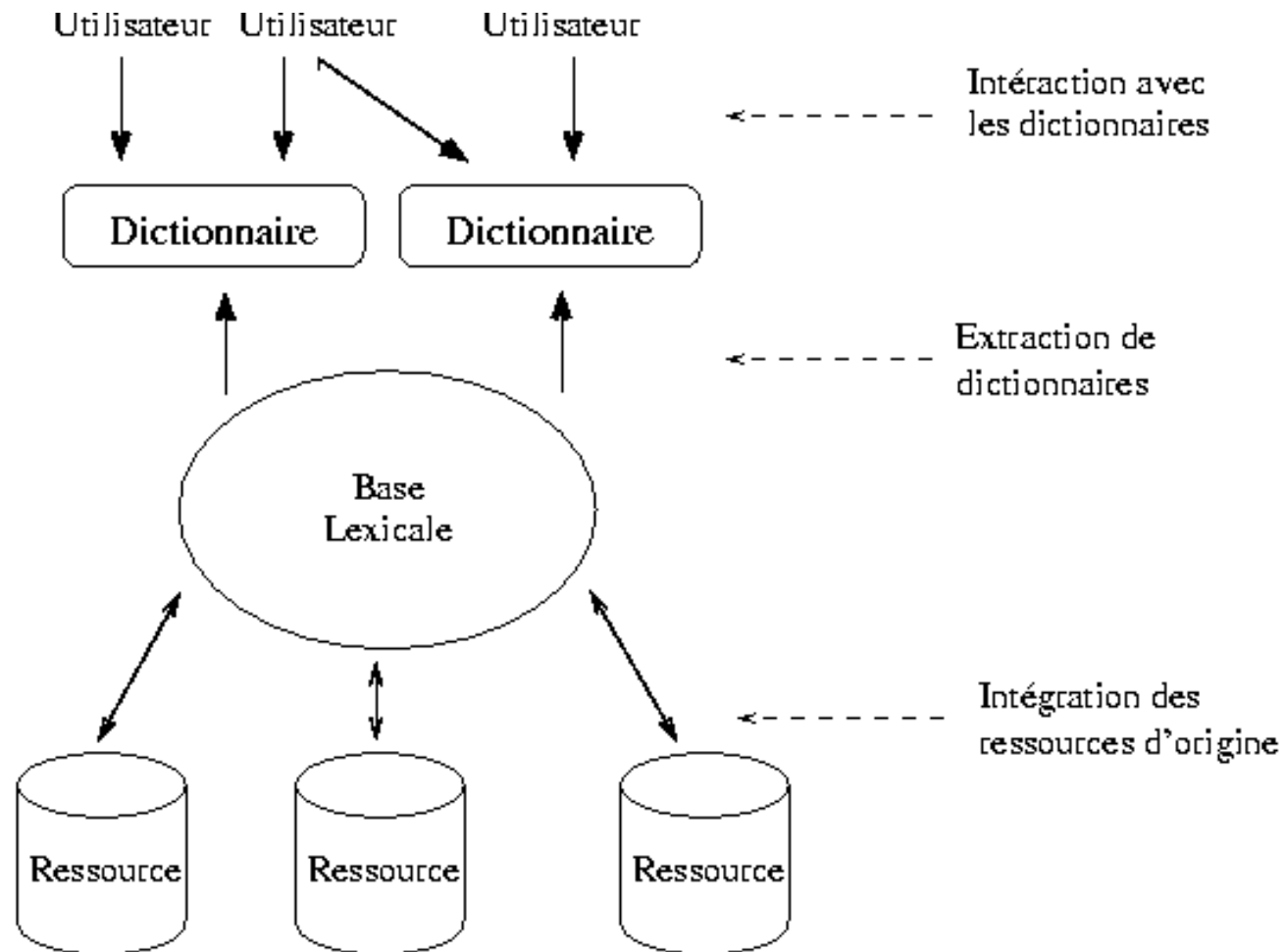
# Buts du projet :

## Production de dictionnaires

- Pour humains, aux formats habituels :
  - Consultation Internet en ligne
  - Édition papier
- Pour les humains, grâce aux BD :
  - Aide directe pour éditeurs, navigateurs ou PDAs
- Pour les machines:
  - Ressources terminologiques pour linguiciels
- Pour la science:
  - Dvpt de dict. multilingues a partir de dict. mono.



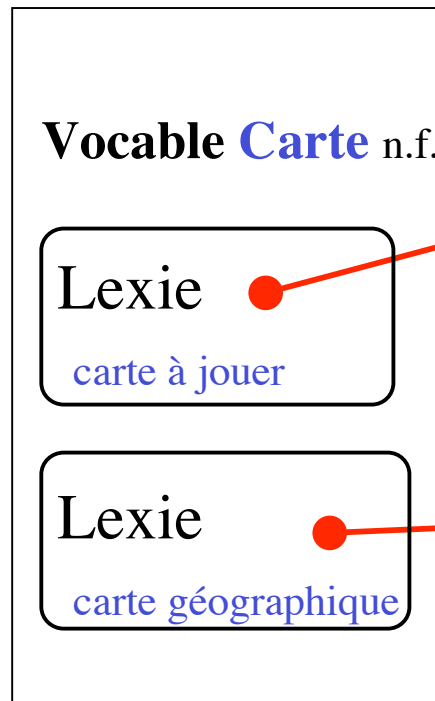
# Vue générale de la base



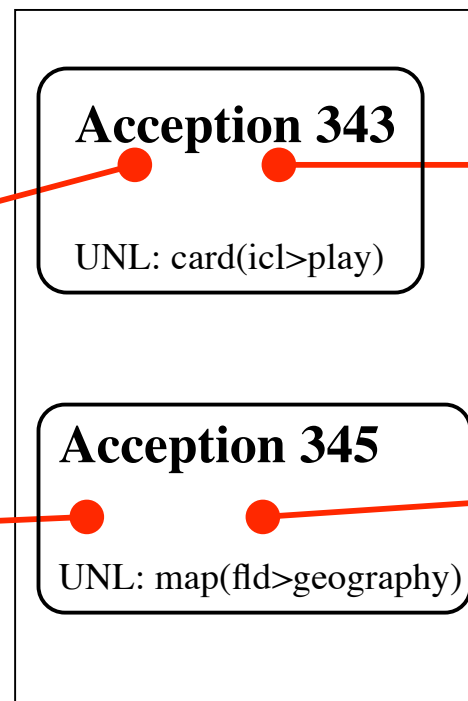


# Architecture interne de la base

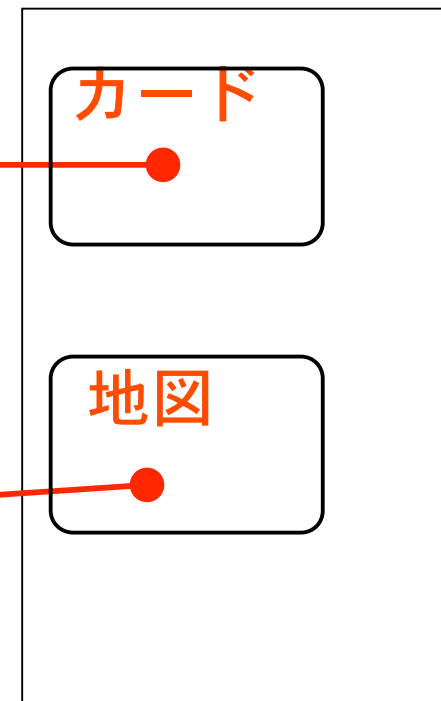
Dictionnaire français



Dictionnaire interlingue



Dictionnaire japonais





# Dictionnaires monolingues

1. Nom de l'unité lexicale : MEURTRE
2. Propriétés grammaticales : nom, masc
3. Formule sémantique : action de tuer: ~ PAR L'individu X DE L'individu Y
4. Régime : X = I = de N, A-poss Y = II = de N, A-poss
5. (Quasi-)synonymes : {QSyn} assassinat, homicide#1; crime
6. Dérivations sémantiques et collocations :
  - {V0} tuer
  - {A0} meurtrier-adj / \*Nom pour X\*/
  - {S1} auteur [de ART Ø] //meurtrier-n /\*Nom pour Y\*/
  - {S2} victime [de ART Ø] /\*Très choquant\*/
7. Exemples : La mésestante pourrait être le mobile du meurtre.
8. Idiomes :
  - appel au meurtre
  - 1. crier au meurtre





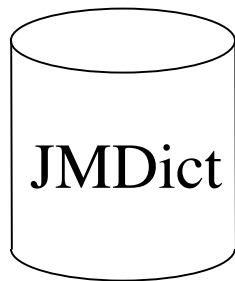
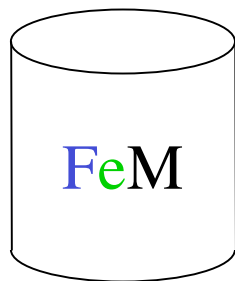
# Méthodologie de construction

- Création de la soupe lexicale
  - Intégration des données existantes
- Révision de la soupe lexicale
  - Révision des liens créés automatiquement
- Création de nouvelles données
  - Le linguiste rédige les entrées monolingues
  - Le traducteur édite les liens interlingues
- Révision de nouvelles données
  - Le lexicologue révisé les liens et les entrées

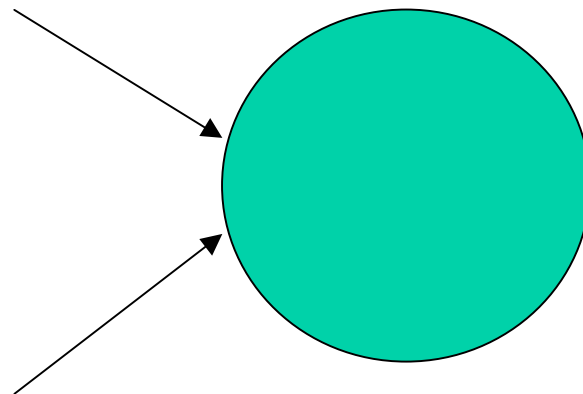


# Création de la soupe lexicale

Dictionnaires



Base lexicale





# FeM : français->anglais

FeM

[<<previous](#) [next>>](#)

**dictionnaire** /diksyone+r/  
n.m. ; **dictionary**  
les enfants qui ne connaissent pas l'ordre alphabétique ne peuvent pas  
consulter le dictionnaire ; <<

<http://clips.imag.fr/geta/services/fem/>



# JMDict : japonais->anglais

EDICT

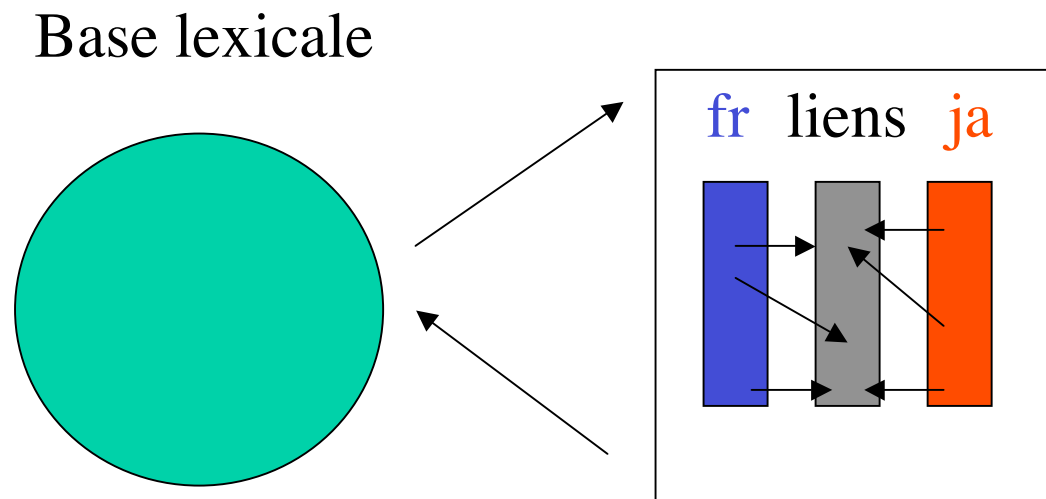
Search Key: 日本 (longest match found)

裏日本 【うらにっぽん】 Japan Sea coastal areas  
裏日本 【うらにほん】 Japan Sea coastal areas  
表日本 【おもてにほん】 Pacific side of Japan  
日本 【にっぽん】 Japan  
日本 【にほん】 Japan  
日本猿 【にほんざる】 Japanese monkey; Japanese macaque  
日本画 【にほんが】 Japanese paintings  
日本海 【にほんかい】 Sea of Japan  
日本海上自衛隊 【にほんかいじょうじえいたい】 Japan maritime self defense force  
日本学者 【にほんがくしゃ】 Japanologist; Japan scholar

<http://www.csse.monash.edu.au/~jwb/wwwjdic.html>



# Révision des liens





# Interface de révision

The screenshot shows the 'Abeille' software window with a menu bar (File, Edit, Link, Help) and three main panels: French, Interlingual links, and Japanese. Each panel has a search field and a button to view existing links. The Interlingual links panel also includes fields for 'U.W.' and 'Gloss', and a 'Confirm link' button. Below these panels are three scrollable lists of words. A blue line connects the word 'table de chevet' in the French list to 'メイトテーブル' in the Japanese list. At the bottom, a blue instruction reads: 'Click on a word in source language.'

French	Interlingual links	Japanese
<b>Table</b>	table de chevet = メイトテーブル	<b>テーブル</b>
<b>1.</b>		<b>1.</b>
table de <b>chevet</b>		メイト <b>テ</b> ーブル
table a repasser		アイロン台

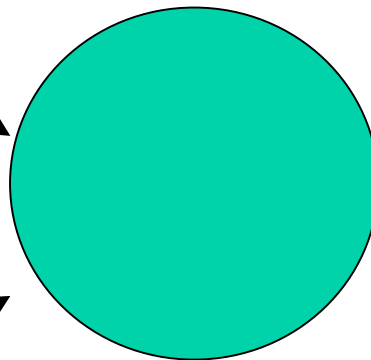


# Rédaction de nouvelles entrées

Lexicographe

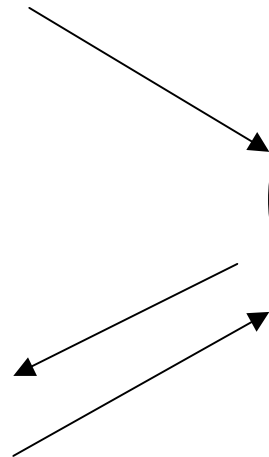
nouvelle  
entrée

Base lexicale



Lexicologue

nouvelle  
entrée  
révisée

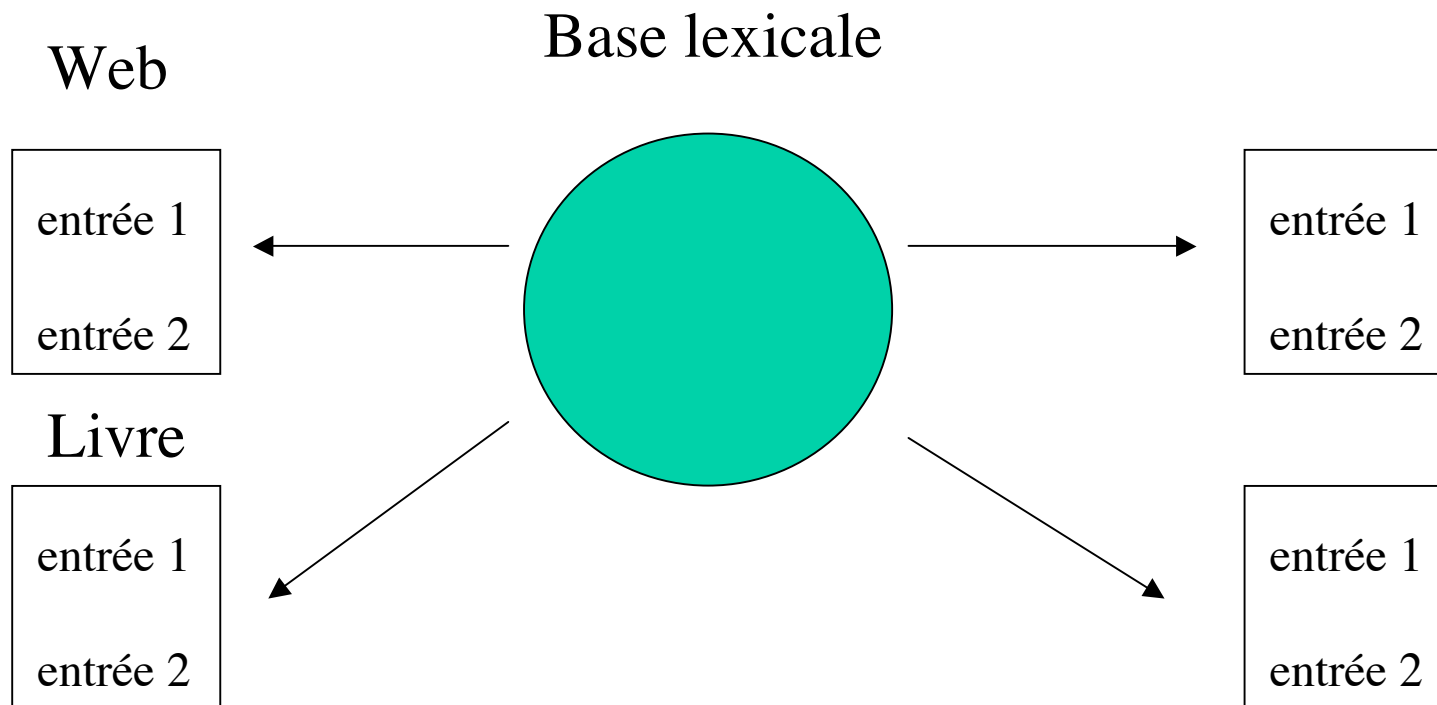




# Consultation de la base

Humains

Machines







# Travaux en cours

- Stage doctoral au **NII**, Tokyo
  - Travail sur les outils de manipulation de dictionnaires
  - Préparation de la soupe lexicale
- Séjour post-doctoral de 2 ans au **NII**, Tokyo
  - Bourse CNRS/JSPS pour le projet Papillon
  - Gestion technique du projet Papillon
  - Élaboration du serveur et des outils en collectif
- Séminaire Papillon 2001 à Grenoble, France
  - Juillet 2001 Organisé par le GETA/CLIPS



# Conclusion

- Besoins pour le développement du projet :
  - Nécessité d'un serveur central et d'une équipe d'experts
  - Développer des outils coopératifs
  - Besoin de contributeurs bénévoles
- Avantages :
  - Intégration de nouvelles langues facilitée
    - Discussions pour le Thaï en cours avec NECTEC (Bangkok)
  - Disponibilité des données grâce aux logiciels libres
  - Génération de formats multiples à partir de la base



# Contacts

- Site Web : <http://vulab.ias.unu.edu/papillon>
- Responsable : Emmanuel Planas
  - `mailto:Emmanuel.Planas@imag.fr`
- Responsable technique : Mathieu Mangeot
  - `mailto:Mathieu.Mangeot@imag.fr`
- Responsable **NII** : Frédéric Andrès
  - `mailto:andres@nii.ac.jp`



# Méthodologie de construction

<i>Commentaires</i>	<i>Français</i>	<i>Japonais</i>	<i>anglais</i>	<i>UNL</i>
Idéal: résultat de la fusion	oui	oui	oui	oui
Idéal mais pas disponible	oui	oui	non	oui
Données du GETA	oui	non	non	oui
FEM + JMdict	oui	oui	oui	non
Petites listes disponibles	oui	oui	non	non
FEM	oui	non	oui	non
JMdict	non	oui	oui	non
LADL-CNAM	oui	non	non	non
De zéro	non	non	non	non