

Le projet Papillon

*Construction collaborative d'un
dictionnaire multilingue*

Mathieu Mangeot Nagata

NII, Tokyo, Japon

mangeot@nii.ac.jp

Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

Point de départ

- *But initial*
 - *Construire un dictionnaire français-japonais*
- *Initiateurs*
 - *Emmanuel Planas (ancien postdoc chez NTT)*
 - *François Brown de C. (ancien attaché scientifique)*
- *Premiers partenaires*
 - *Institut National d'Informatique, Tokyo, Japon*
 - *Laboratoire GETA-CLIPS, Grenoble, France*

Motivations

- *Très peu de ressources existantes*
 - *français-japonais*
- *Manque d'informations*
 - *Compteurs numériques, kanji+kana+romaji*
- *Coûts de construction prohibitifs*
 - *EDR Dictionnaire anglais-japonais*
 - *1200 h-an; 300 000 entrées; 14,3 Mo ¥*
- *Projets de construction collaborative en cours*
 - *Edict jpn->eng, SAIKAM thai->(jpn, eng, fra)*

Article du Dico français-japonais

faciliter [fasilite ファスイリテ] -- (他動) 容易にする、(...の)
助けになる: *L'argent facilite beaucoup de choses.*
金があれば、たいがいのことが楽にできる。

-- *se faciliter* (代動) 自分の... を楽にする: *Elle a
acheté un lave-vaisselle pour se ~ la vie.* 彼女は家
事が楽になるように皿洗い器を買った。

Objectifs étendus du projet

- *Faire un dictionnaire complet de référence*
 - *Large couverture 500 -> 200 000 articles*
 - *Multilingue 2 -> 20 langues*
 - *Usages multiples débutants, experts, applications*
- *Développement en communauté*
 - *Paradigme de construction de LINUX*
 - *Contributeurs bénévoles*
 - *Mutualisation des ressources*
 - *Profils et préférences utilisateurs*

Chacun définit son dictionnaire

manger

Verbe

食べる

たべる

taberu

Si tu manges ton riz, tu auras un bonbon.

もし御飯[ごはん]を食[た]べたら、飴[あめ]をあげるよ。

moshi gohan wo tabetara,
ame wo ageruyo.

manger /マンジュエ/

Verbe 1^{er} groupe

食べる

Si tu manges ton riz, tu auras un bonbon.

シ チュ マンジュ トンリ、チュ オラ アン ボンボン。

もし御飯を食べたら、飴をあげるよ。

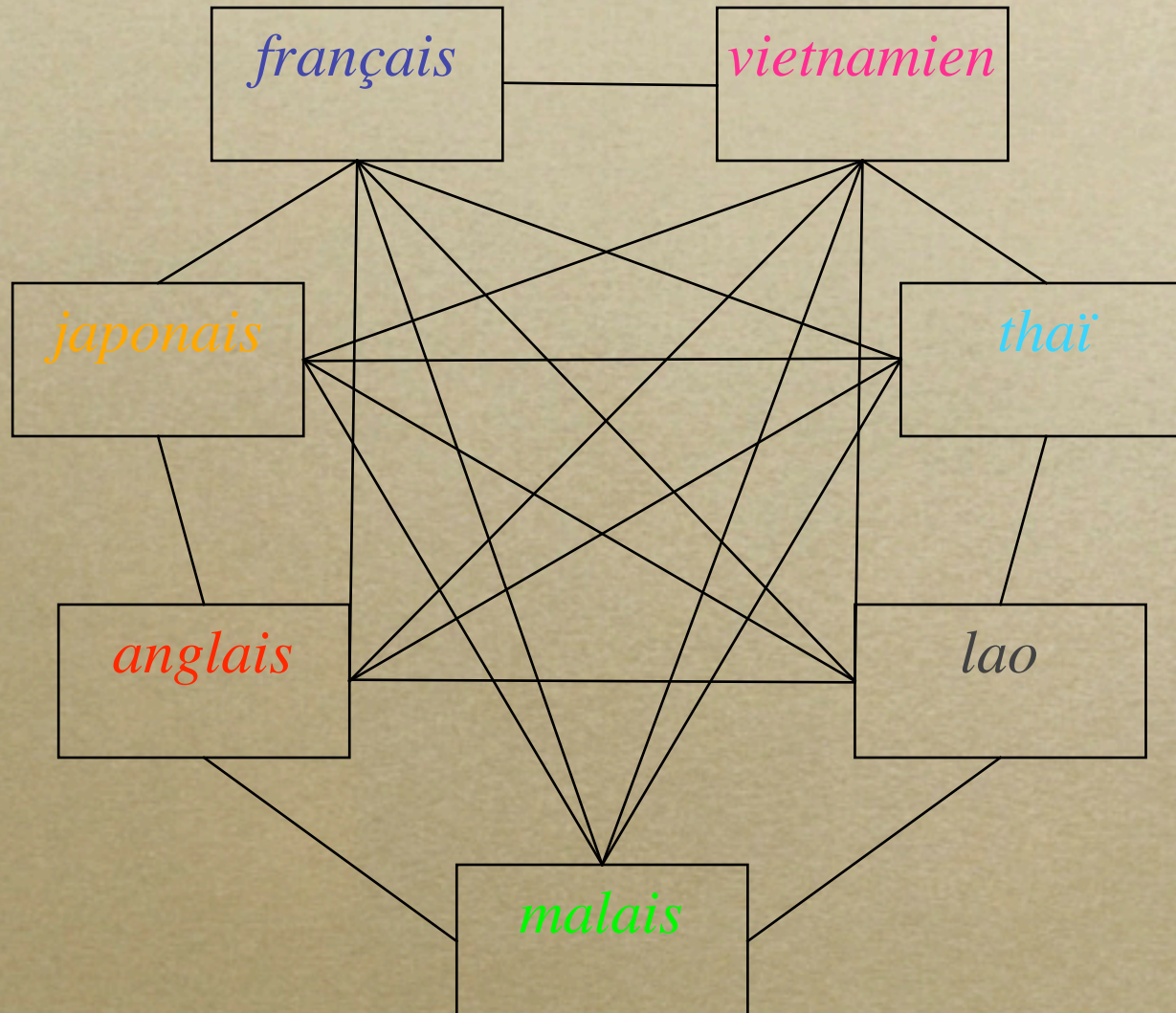
Nouveaux partenaires

Pays	Partenaires	Ressources
Japon	ATR, Yves Lepage NTT, Francis Bond	EngDict eng-kor
Australie	U. Monash, Jim Breen	Edict jpn-eng
Chine & Singapour	U. XiaMen, Mendel Shi U. Singapour, Ji Donghong	Cedict eng-zho
Allemagne	Ulrich Apel	WaDokuJT jpn-deu
Malaisie	USM Penang, T. Enya Kong	FeM fra-eng-msa
Thaïlande	U.Kasetsart, Asanee K. NECTEC, Virach S.	ThaiDict tha
Vietnam	U. Saigon, Dinh Dien U. Da Nang, Phan Khanh	VietDict vie-fra

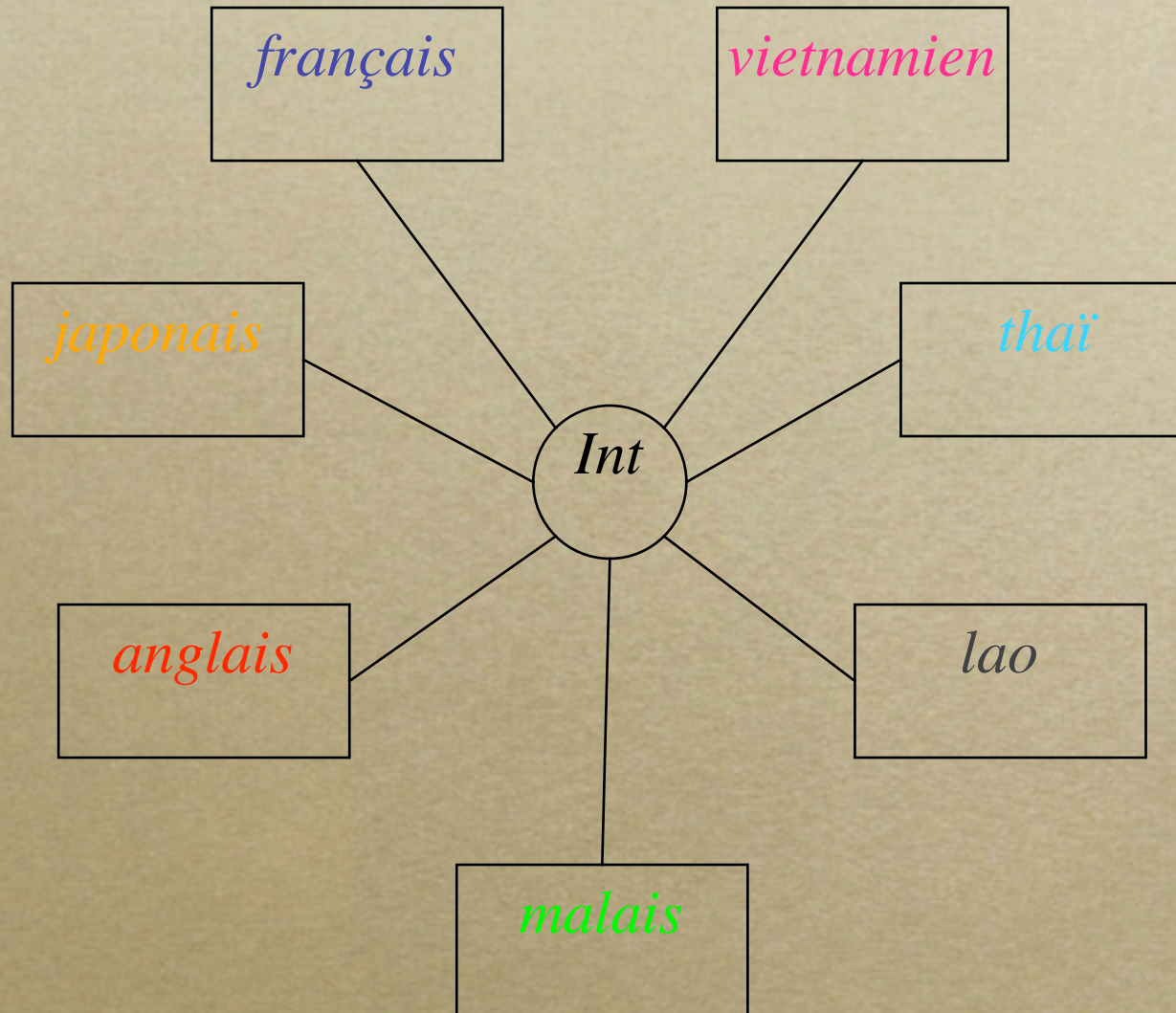
Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

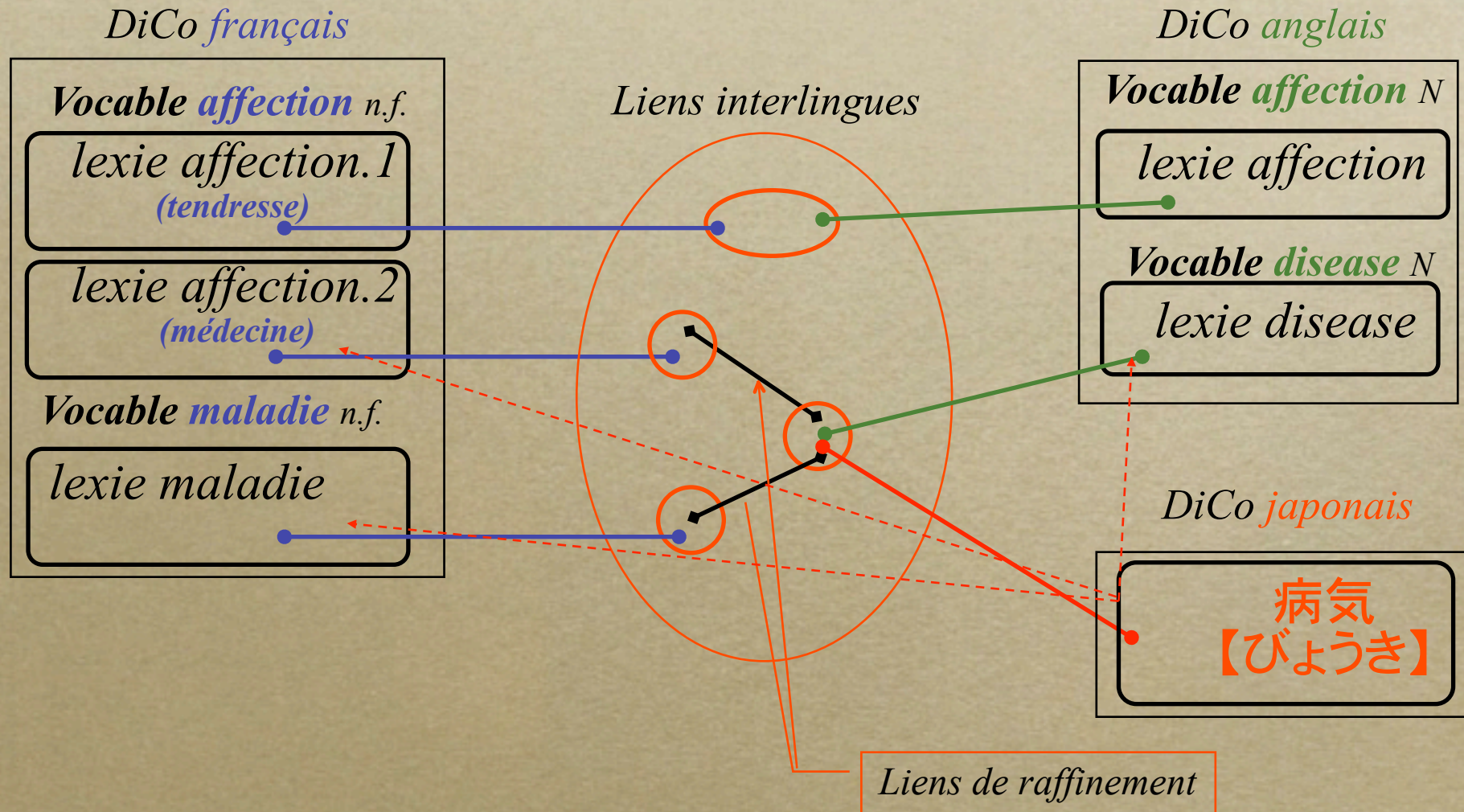
Dictionnaires bilingues



Dictionnaire pivot



Structure pivot détaillée



Dictionnaires sur mesure

français

japonais

français

japonais

japonais

anglais

français

japonais

anglais

français

thaï

anglais

français

thaï

anglais

français

Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

Microstructure des articles

- *Structure des articles du dictionnaire*
- *Un article “classique” contient un mot et ses différents sens ou traductions*

aimer /e-me-/

v.tr. ;

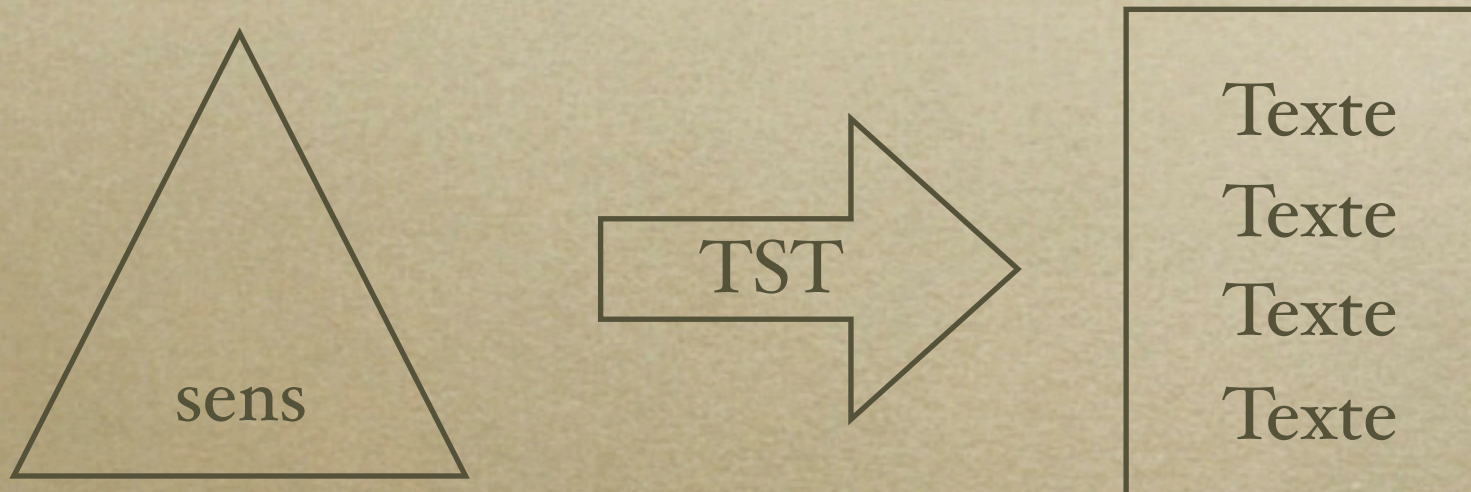
apprécier ; to like

d'amour ; to love ; to be in love with someone

- *Notre unité : le sens de mot = lexie*

Lexicographie combinatoire

- *Partie de la théorie sens-texte*
- *Alain Polguère & Igor Mel'tchuk (U. de Montréal)*



- *Dictionnaires existants : DEC, DiCo & LAF*
- *Même structure pour toutes les langues*
- *56 fonctions lexicales de base*

Exemple d'article français

- ◆ *Nom de l'unité lexicale* : MEURTRE
- ◆ *Propriétés grammaticales* : nom, masc
- ◆ *Formule sémantique* : action de tuer: ~ PAR L'individu X DE L'individu Y
- ◆ *Régime* : X = I = de N, A-poss Y = II = de N, A-poss
- ◆ *Fonctions lexicales* :
 - ◆ {QSyn} assassinat, homicide#1; crime /*Quasi synonymes*/
 - ◆ {Oper₁} accomplir, commettre, perpétrer [ART ~]; tremper [dans ART ~] /*Causer que X fasse un M.*/*
 - ◆ {S₁} auteur [de ART Ø]//meurtrier-n /*Nom pour X*/
 - ◆ {S₂} victime [de ART Ø] /*Nom pour Y*/
- ◆ *Exemple* : La mésestente pourrait être le mobile du meurtre.
- ◆ *Idiomes* :
 - ◆ _appel au meurtre_
 - ◆ _crier au meurtre_

Exemple d'article japonais

- ◆ *Nom de l'unité lexicale* : 殺人【さつじん】
- ◆ *Propriétés grammaticales* : 名詞【めいし】
- ◆ *Formule sémantique* : どうさ: 人 Y の 人 X の ~
- ◆ *Régime* : X = I = N, Y = II = N の
- ◆ *Fonctions lexicales* :
 - ◆ {QSyn} 殺戮【さつりく】, 殺害【さつがい】 /*Quasi synonymes*/
 - ◆ {Oper₁} [~を] する; [~を] 犯す /*Causer que X fasse un M.*/
 - ◆ {S₁} 殺人者【さつじんしゃ】, 殺人鬼【さつじんき】 /*Nom pour X*/
 - ◆ {S₂} 被害者【ひがいしゃ】 /*Nom pour Y*/
- ◆ *Exemple* : 喧嘩【けんか】は殺人【さつじん】の動機【どうき】になり得【え】るだろう。
- ◆ *Idiomes* :
 - ◆ _殺人剣【さつじんけん】_
 - ◆ _殺人光線【さつじんこうせん】_
 - ◆ _囑託殺人【しくたくさつじん】_

Article au format XML

```
<lexie><headword>ABOIEMENT</headword>
  <pos>nom, masc, surtout pl</pos>
  <formula>cri: ~ DE L'animal X</formula>
  <pattern>X = I = de N, A-poss</pattern>
  <lexical-functions>
    <function name="QSyn">
      <value> « Oua! Oua! »</value>
    </function>
    <function name="V0">
      <value>aboyer</value></function>
  <example>Un voisin a été réveillé par
  les aboiements de son chien.</example>
</lexie>
```

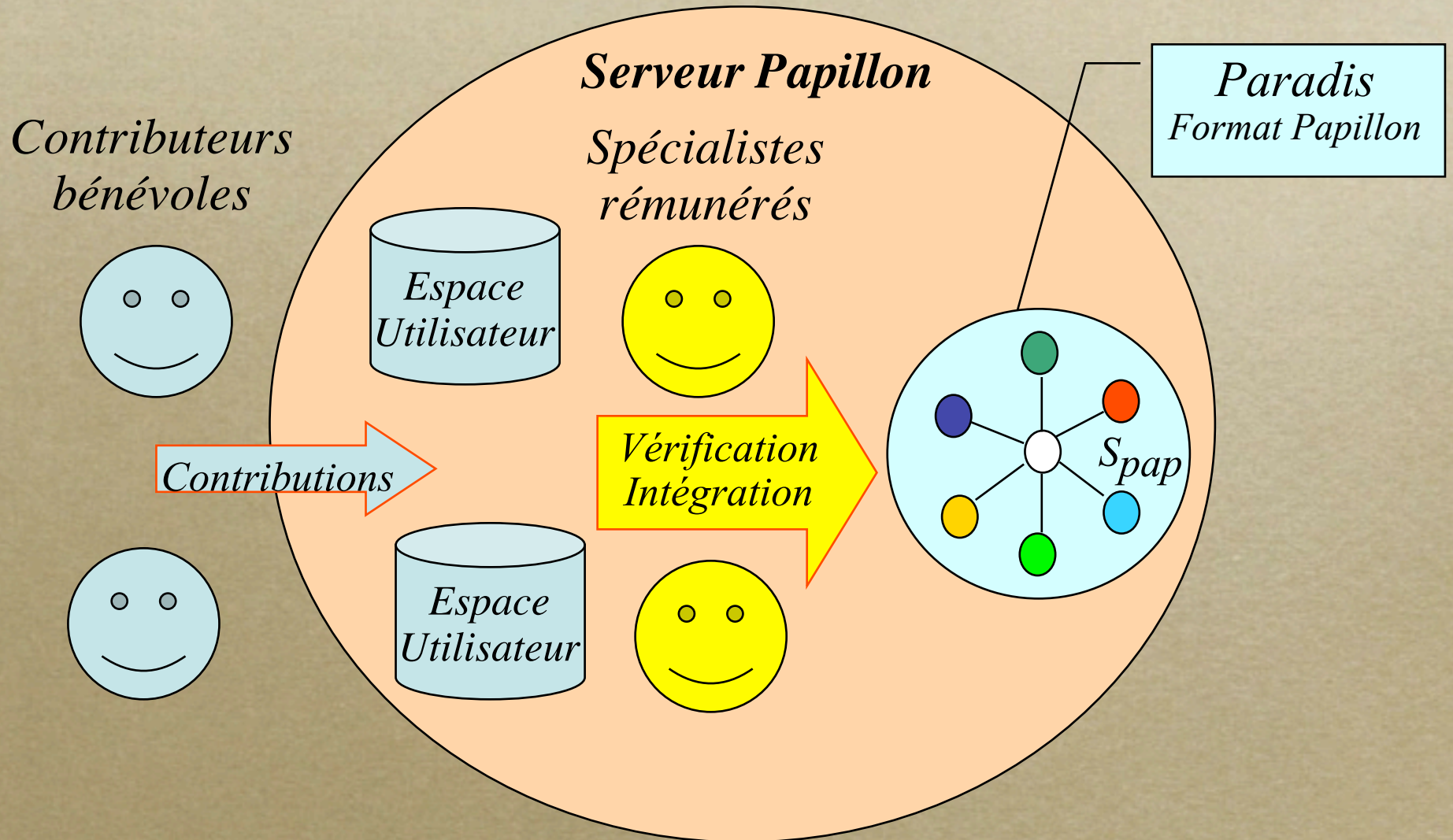
Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

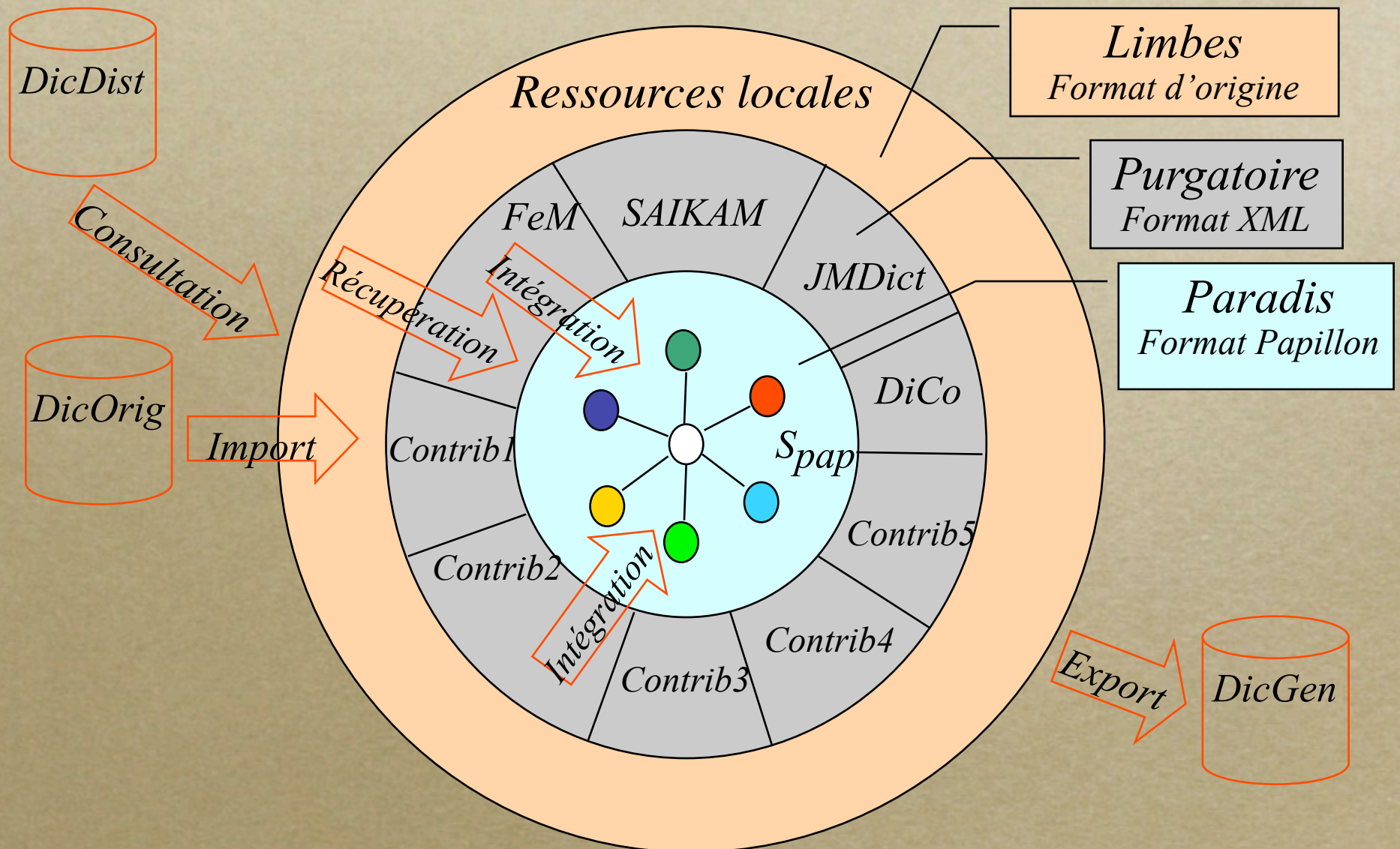
Contributions et vérifications



Plan de l'exposé

- *Présentation du projet*
- *Macrostructure du dictionnaire*
- *Microstructure des articles*
- *Consultation des données (Demo)*
- *Édition en ligne des contributions (Demo)*
- *Préparation des données existantes*

Préparation des données existantes



Découpage d'un article en sens : subjectif !

Hachette-Oxford

sentir v.t.

odorat

toucher, corps, gout

avoir conscience de

sentir v.i.

avoir une odeur

puer

révéler

Collins online

sentir v.t.

odorat

goût

toucher

répandre une odeur

avoir la saveur

dénoter, annoncer

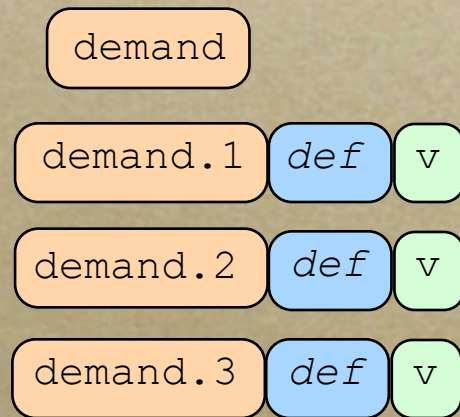
sentir v.i.

Vecteurs conceptuels

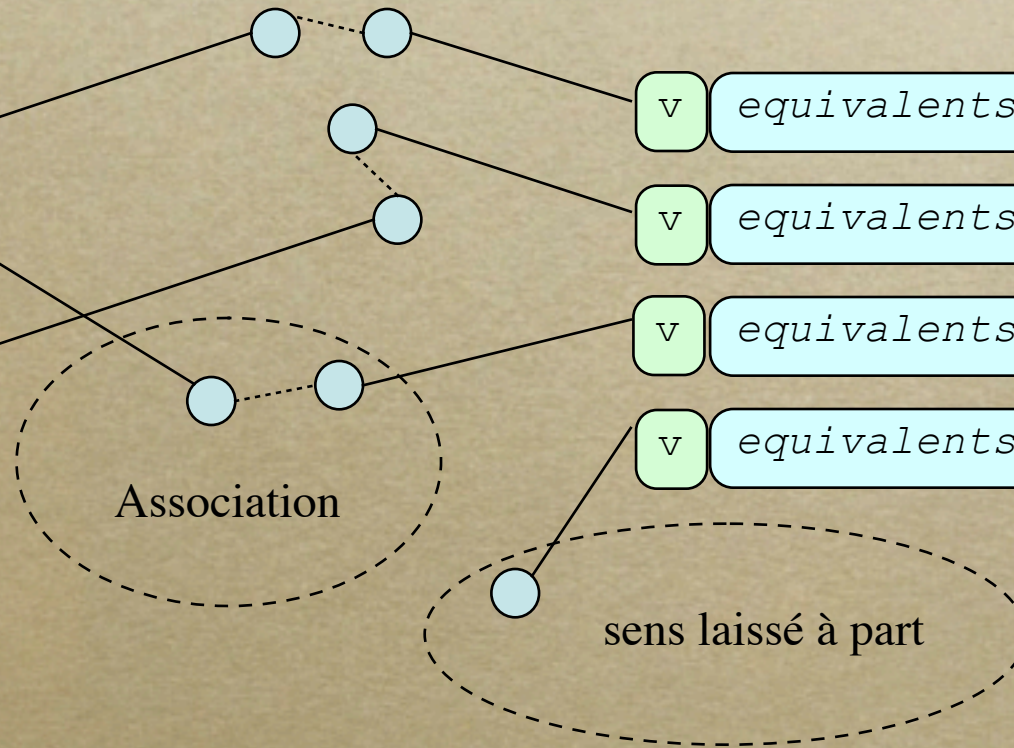
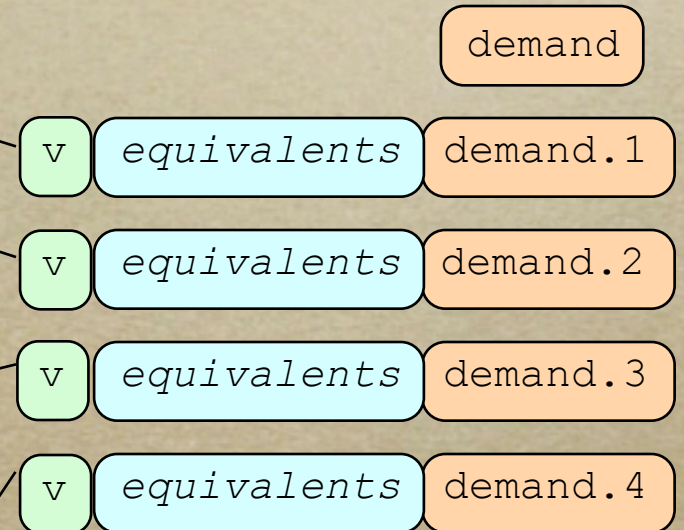
- *Une idée = 1 vecteur = 1 concept*
- *L'espace vectoriel est de dimension K*
 - *$K = \text{nb de concepts dans un thésaurus}$*
 - *Le français : thésaurus Larousse = 873 concepts*
- *Distance entre 2 vecteurs = distance angulaire*
 - *$DA(x, y) = \text{acos}(\text{sim}(x, y))$*
 - *$DA(x, y) = \text{acos}(x \cdot y / \|x\| \|y\|)$*

Relier les sens avec les vecteurs

dictionnaire monolingue
anglais



Dictionnaire bilingue
anglais-français



Conclusion

- *Important travail d'implémentation déjà réalisé*
- *En cours : récupération des données existantes*
- *Problèmes de recherche restants*
 - *Gestion des liens interlingues*
 - *Système expert pour l'édition des données*
- *Problématique sociologique :*
 - *Le projet ne peut réussir sans l'aide du public*

Nous aurons besoin de vous !

<http://www.papillon-dictionary.org/>