

# Interlinguistic Divergences in Papillon Dictionary

---

Mathieu Mangeot-Nagata  
萬怒永田 真忠

NII, Tokyo, Japan  
[mangeot@nii.ac.jp](mailto:mangeot@nii.ac.jp)

Kyoko Kuroda  
黒田 京子

Pref. College, Shimane  
[kyoko.kuroda@nifty.ne.jp](mailto:kyoko.kuroda@nifty.ne.jp)

# Outline

- Motivations & Goals of Papillon Project
- Macrostructure of the Dictionary
- Microstructure of the Entries
- Construction & Edition of the Data
- Argument Structure Divergence Issue
- Proposed Solution

# Motivations

- Lack of Information
  - Numerical Specifiers, kanji+kana+romaji
- Very Few Existing Resources
  - French-Japanese
- Construction Costs Too High
  - EDR English Japanese Dictionary
  - 1,200 Person-Year; 300,000 Entries; 14.3 Mo ¥
- On Going Collaborative Construction Projects
  - Edict Japanese->English (Jim Breen)
  - SAIKAM Japanese-Thai

# Entry of “Le Dico” fra-jpn

**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。

# Goals

- Build a Reference Dictionary
  - Broad Coverage 500 -> 200,000 Entries
  - Multilingual 1 -> 20 Languages
  - Multiple Usages (beginners, experts, applications)
- Community Development
  - LINUX Construction Paradigm
  - Voluntary Contributors
  - Mutualization of the Resources
  - User Prefs & Profiles

# Define one's own Dictionary

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 /マンジェ/

動詞 1<sup>er</sup> groupe

食べる

*Si tu manges ton riz, tu auras un bonbon.*

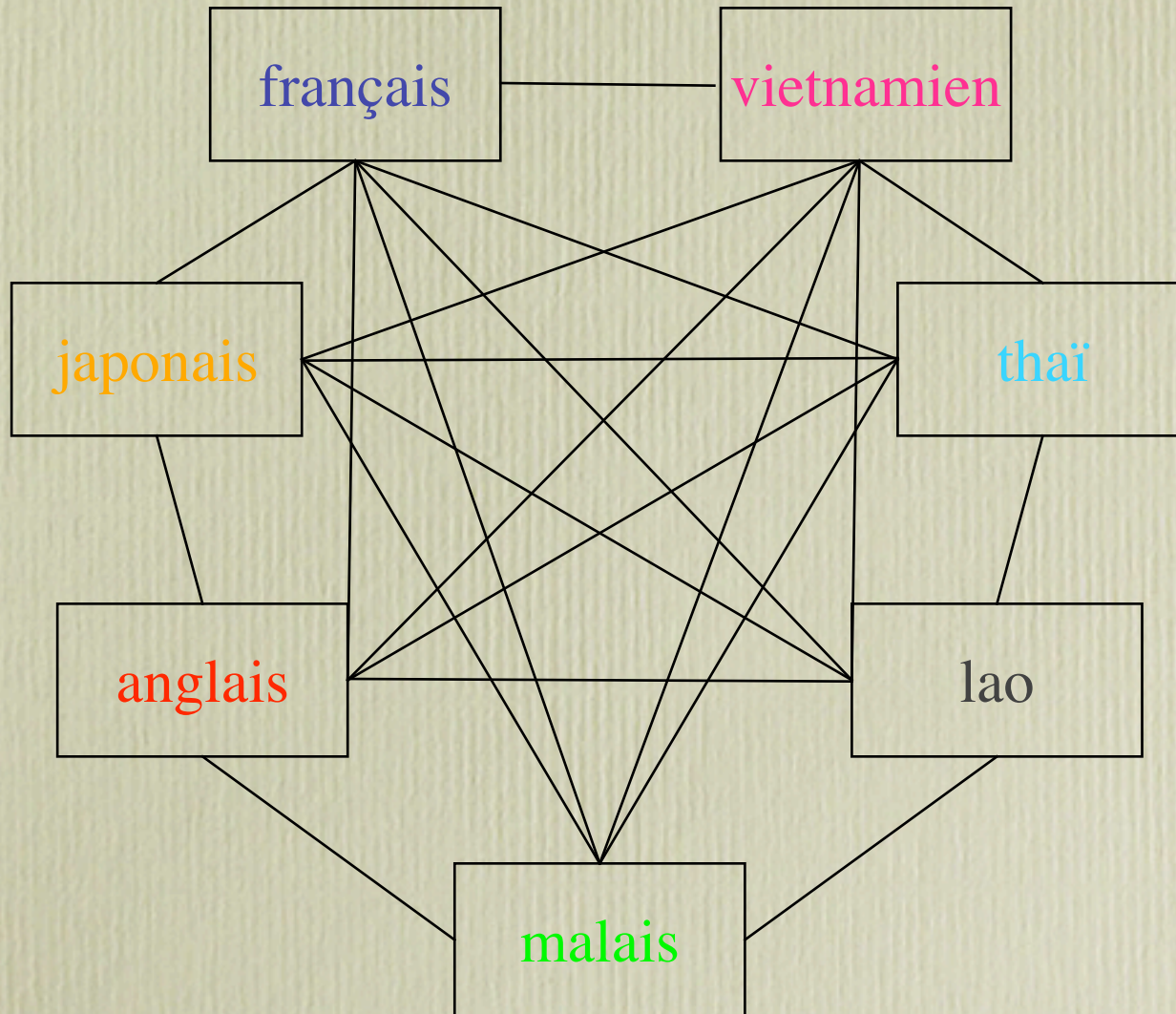
シ チュ マンジュ トン リ、チュ オラ ア  
ン ボンボン。

もし御飯を食べたら、飴をあげるよ。

# Outline

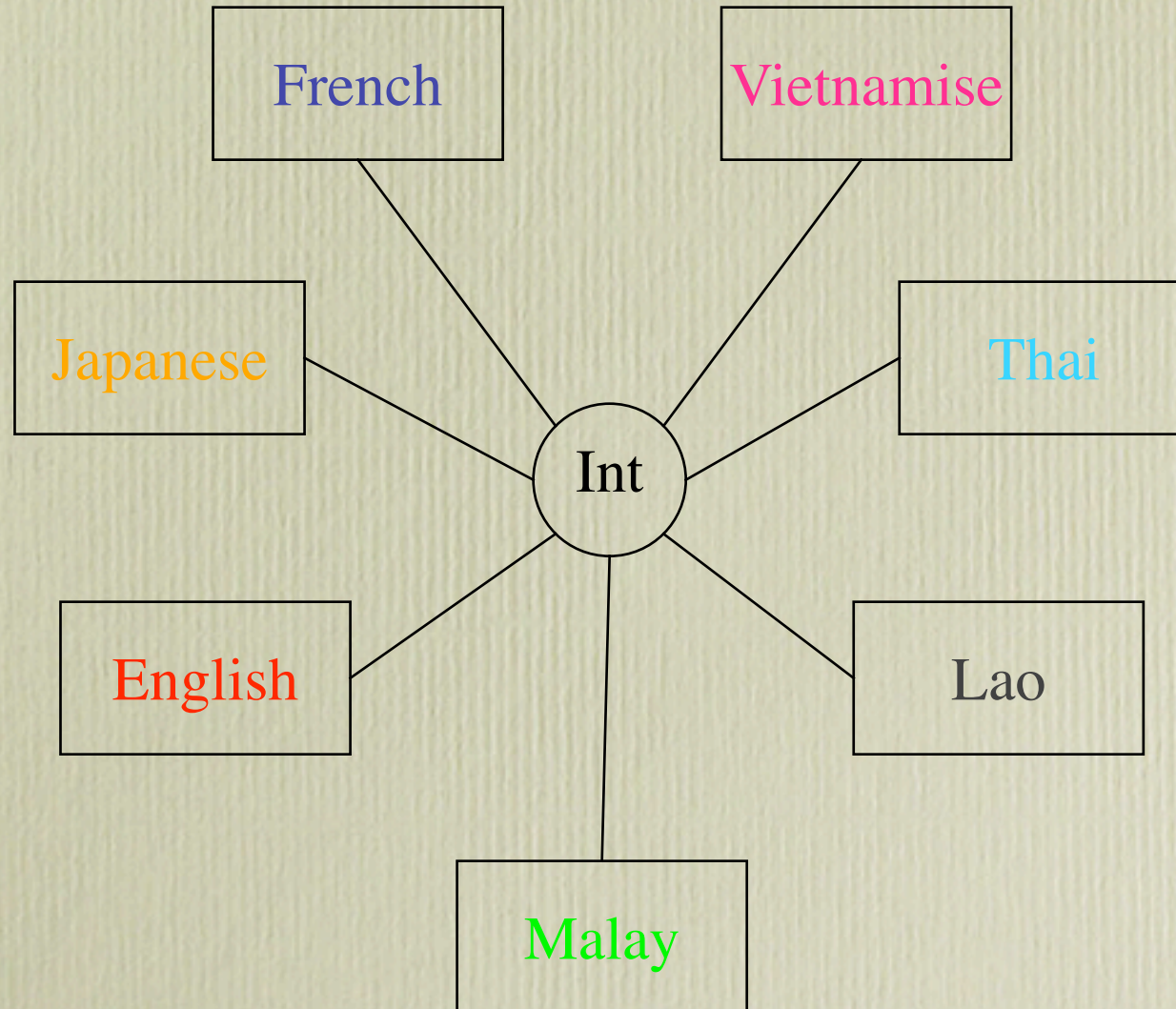
- Motivations, Goals & Status of Papillon Project
- Macrostructure of the Dictionary
- Microstructure of the Entries
- Construction & Edition of the Data
- Argument Structure Divergence Issue
- Proposed Solution

# Bilingual Dictionaries

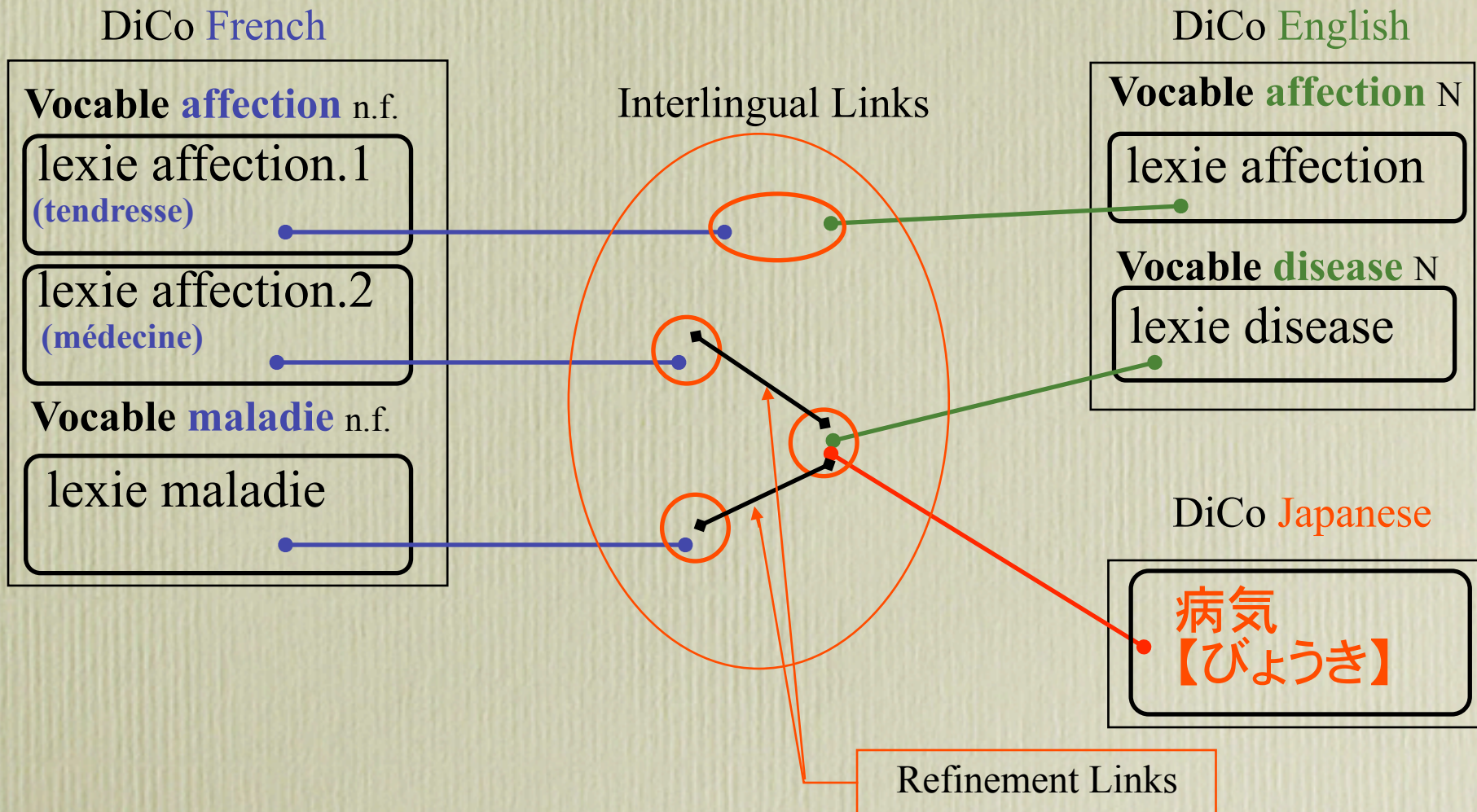




# Multilingual Pivot Dictionary



# Detailed Pivot Structure



# Outline

- Motivations, Goals & Status of Papillon Project
- Macrostructure of the Dictionary
- **Microstructure of the Entries**
- Construction & Edition of the Data
- Argument Structure Divergence Issue
- Proposed Solution

# French Entry

- Nom de l'unité lexicale : **BATAILLE**
- Propriétés grammaticales : **nom, fem.**
- Niveaux de langue : **standard**
- Formule sémantique : **lutte: ~ ENTRE L'ensemble de individu X ET L'ensemble de individu Y**
- Régime **mode 1 X+Y = I+II = entre N et N, A-poss**  
**mode 2 X = I = de N, A-poss; Y = II = contre N**
- Fonctions lexicales :
  - **{QSyn} combat, affrontement /\*Quasi synonymes\*/**
  - **{Oper<sub>1/2</sub>} adversaire**
  - **{V<sub>0</sub>} se battre**
  - **{Magn} intense, terrible**
- Exemple : **Les trois statues avaient été évacuées pendant la bataille de Caen.**
- Idiomes : **\_cheval de bataille\_ \_en bataille\_**

# Japanese Entry

- Name of the Lexical Unit: 殺人【さつじん】
- Grammatical Properties: 名詞【めいし】
- Semantic Formula: どうさ: 人 Y の 人 X の ~
- Government Pattern:  $X = I = N, Y = II = N$  の
- Lexical Functions:
  - {QSyn} 殺戮【さつりく】, 殺害【さつがい】 /\*Quasi synonyms\*/
  - {Oper<sub>1</sub>} [~ を] する; [~ を] 犯す /\*Causer que X fasse un M.\*/\*
  - {S<sub>1</sub>} 殺人者【さつじんしゃ】, 殺人鬼【さつじんき】 /\*Name for X\*/
  - {S<sub>2</sub>} 被害者【ひがいしゃ】 /\*Name for Y\*/
- Example: 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。
- Idioms:
  - 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】\_

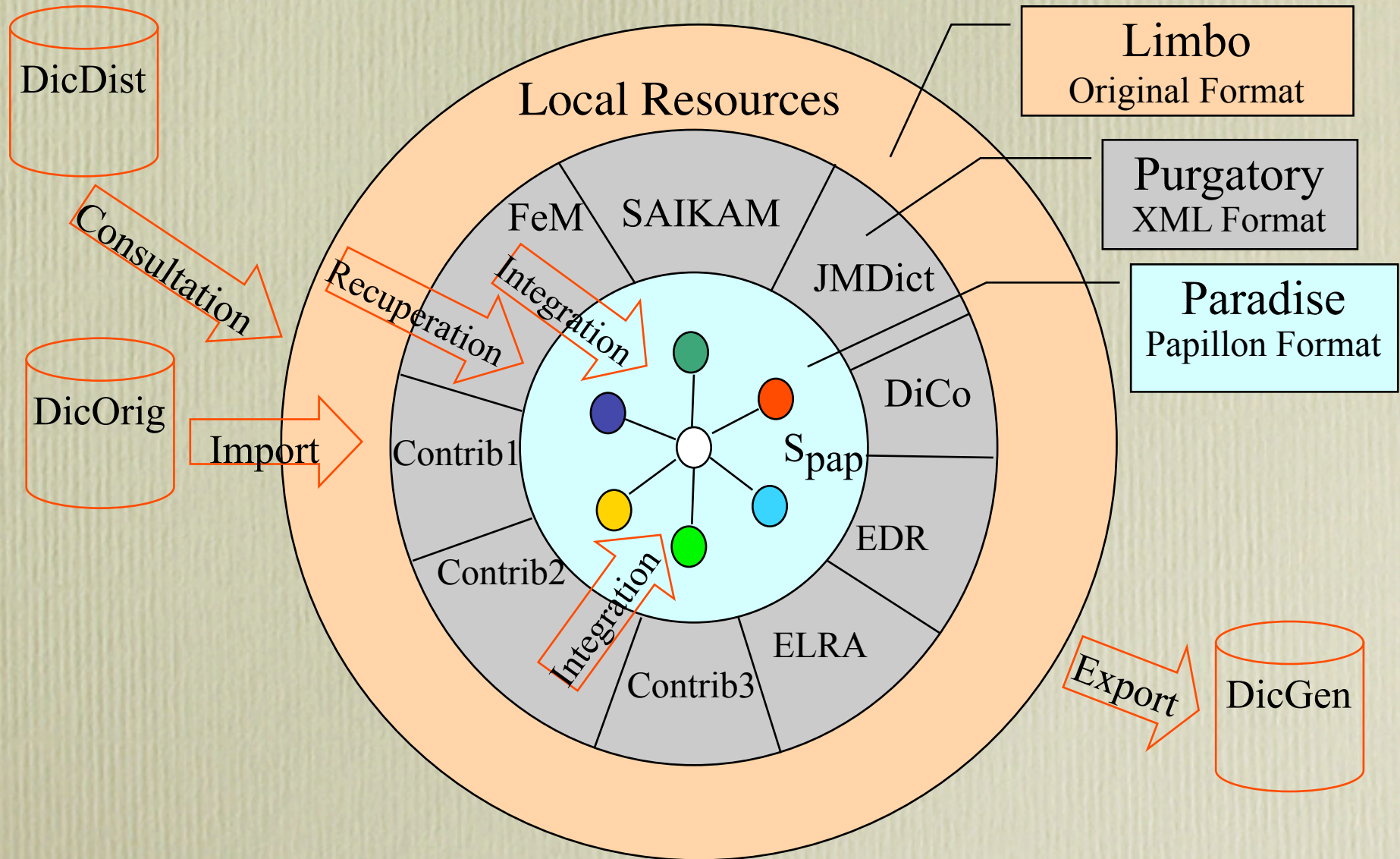
# Entry in XML

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

# Outline

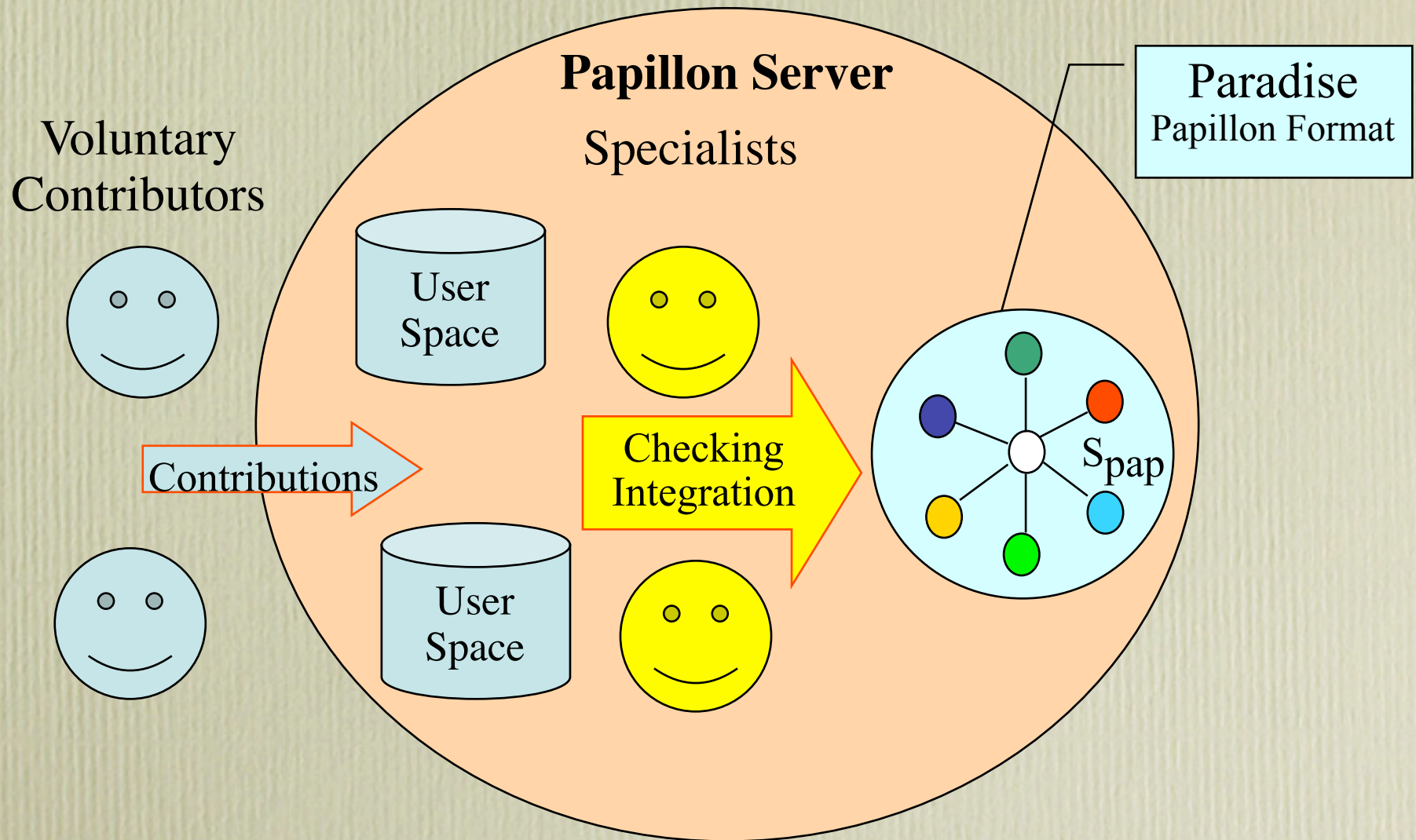
- Motivations, Goals & Status of Papillon Project
- Macrostructure of the Dictionary
- Microstructure of the Entries
- Construction & Edition of the Data
- Argument Structure Divergence Issue
- Proposed Solution

# Preparation of Existing Data





# Contributions & Checking



# Outline

- Motivations, Goals & Status of Papillon Project
- Macrostructure of the Dictionary
- Microstructure of the Entries
- Construction & Edition of the Data
- Argument Structure Divergence Issue
- Proposed Solution

# Argument Position Shift

- fra: lexie MANQUER « *individu X manque à individu Y* »
- eng : lexie MISS « *individu Y miss individu X* »
- X and Y arguments are shifted between French and English.

# Coalescence Divergence

For the verbs:

- - fra: rêver (eng: to dream)
  - jpn: yume wo miru (eng: to see a dream)  
[yume = dream; miru = to see]
- - fra: peser (eng: to weight)
  - jpn: omosa wo hakaru (eng: to measure the weight)  
[omosa = weight; hakaru = to measure]

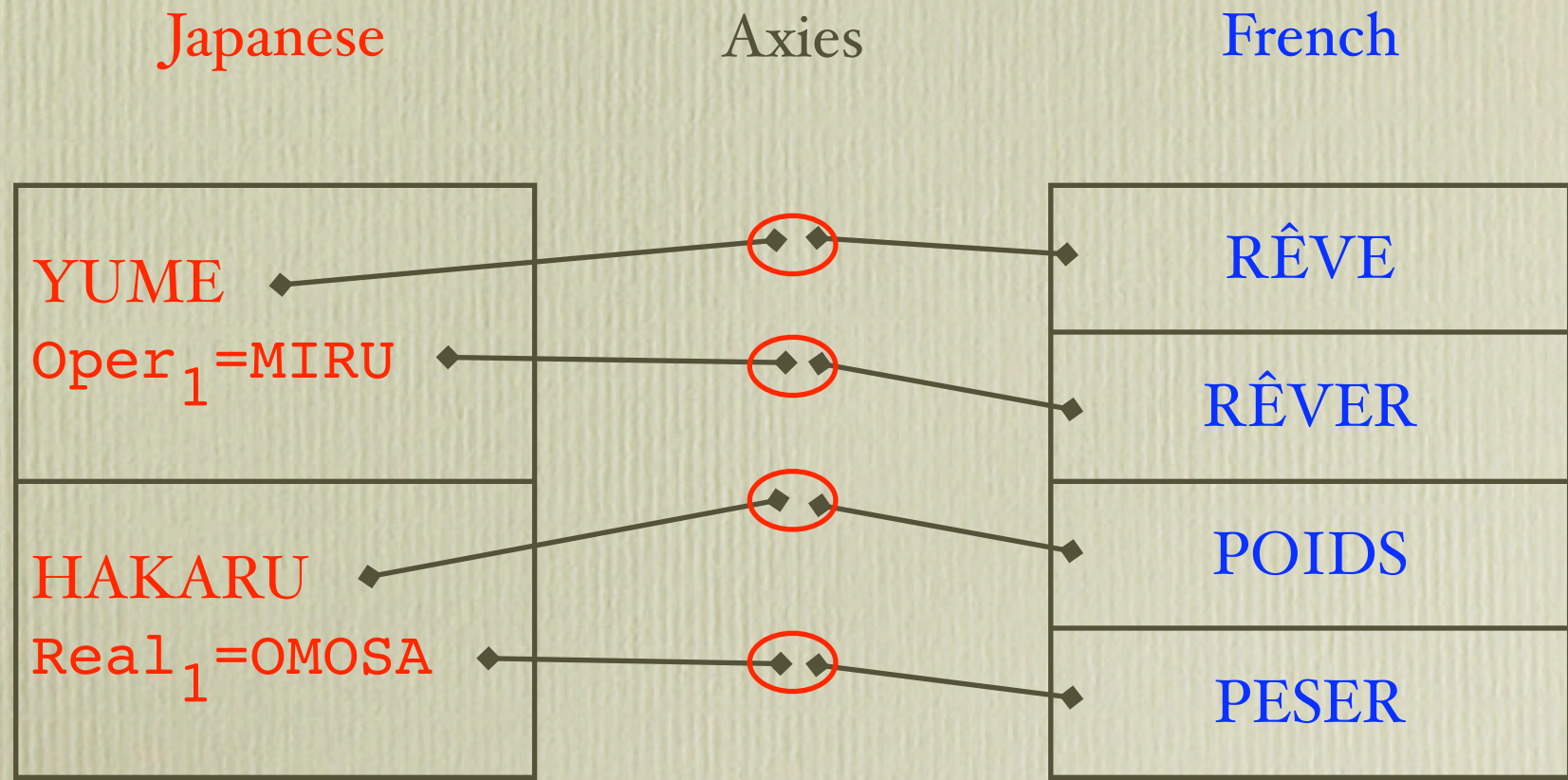
# Outline

- Motivations, Goals & Status of Papillon Project
- Macrostructure of the Dictionary
- Microstructure of the Entries
- Argument Structure Divergence Issue
- Construction & Edition of the Data
- Proposed Solution

# Proposed Solution

- link lexie headwords instead of entire lexies
- declare linkable these information units
  - headword
  - lexical functions
  - examples
  - full idioms
- link all the different information units with the same link type: an axie
- link heterogeneous information units with only one axie

# Coalescence Examples



# Another Solution

- $V = \text{Oper}_I(S_O(V)) + S_O(V)$ 
  - un verbe peut toujours être remplacé par nom + verbe support
- (fra) RÊVE ==> (jpn) YUME
- $S_O(\text{RÊVE}) = \text{RÊVE}$  &  $\text{Oper}_I(\text{YUME}) = \text{"[- wo] miru"}$
- $\text{RÊVER} = \text{Oper}_I(S_O(\text{RÊVER})) + S_O(\text{RÊVER}) =$   
 $\text{Oper}_I(\text{RÊVE}) + \text{RÊVE} ==>$ 
  - (jpn)  $\text{Oper}_I(\text{YUME}) + \text{YUME} = \text{"yume wo miru"}$

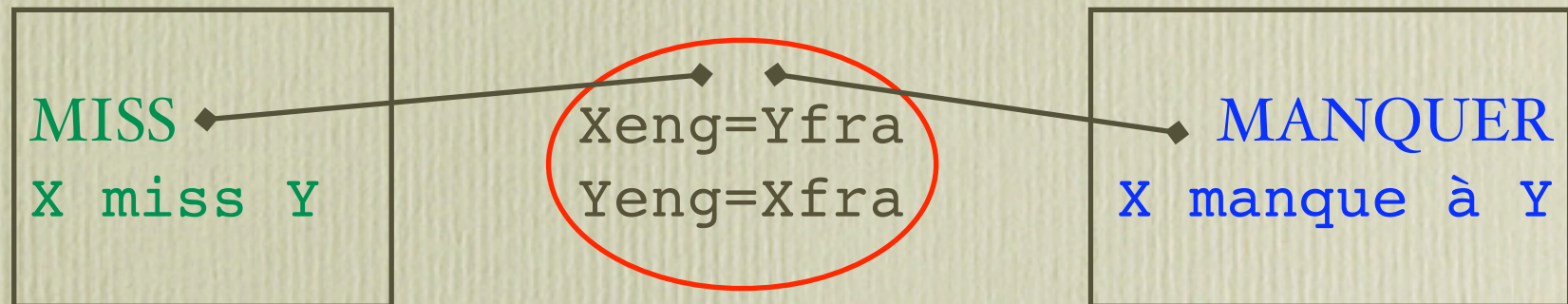


# Position Shift Example

English

Axies

French



# Conclusion

- For Papillon Project
  - Research work remaining (integration, links, )
  - Cannot succeed without public collaboration
- For the Argument Divergence Issue
  - Note the arguments position in axes

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