

Regroupement de structures de dérivations lexicales par raisonnement analogique

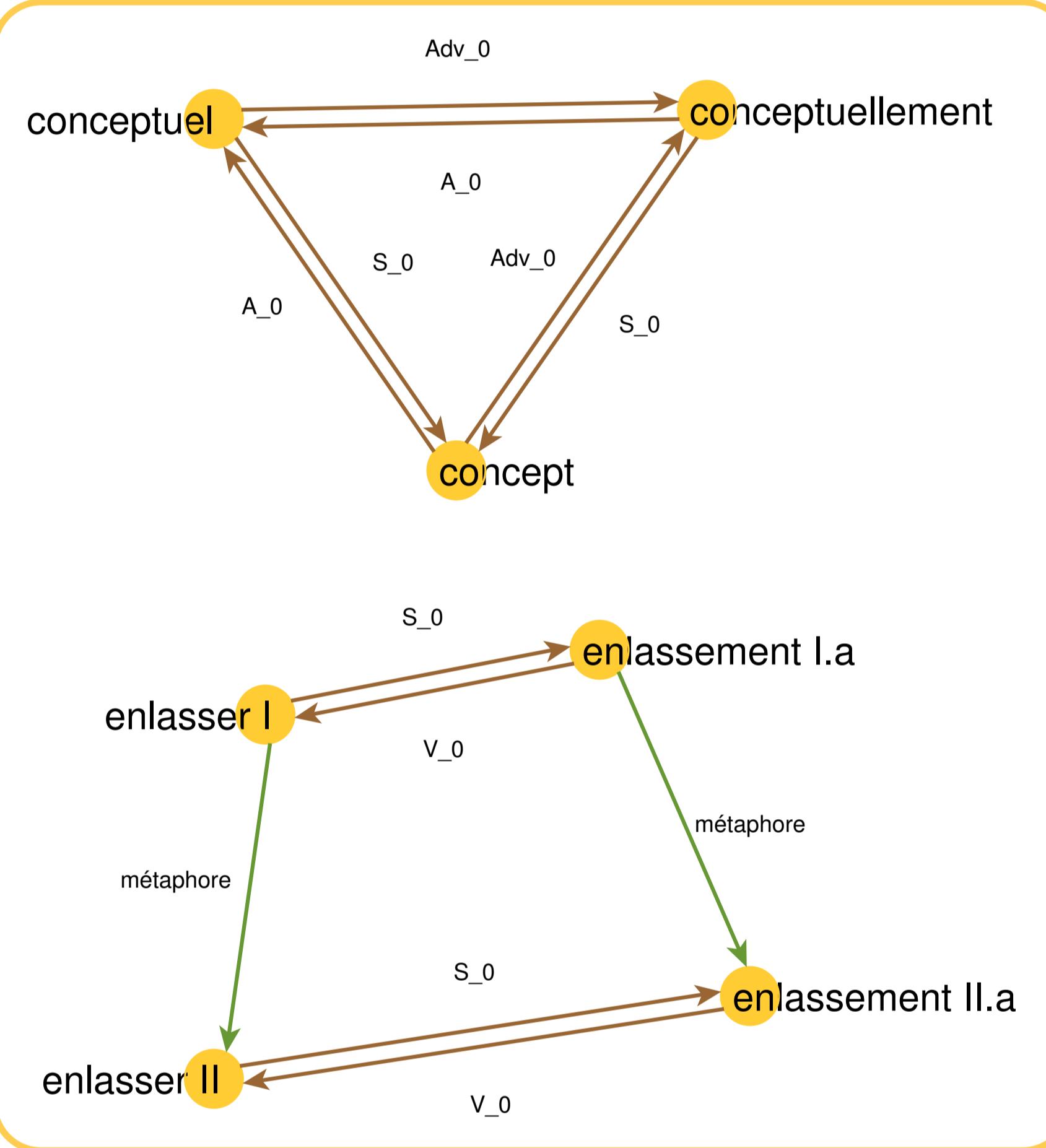
Sandrine Ollinger

ATILF, CNRS, Université de Lorraine, Nancy

Objectifs généraux

- Étude du lexique du français
 - prospection de phénomènes analogiques ;
 - définition de configurations de dérivation lexicales.
- Aide à la rédaction lexicographique
 - proposition de patrons de description ;
 - enrichissement semi-automatique ;
 - vérification de la cohérence de la ressource.

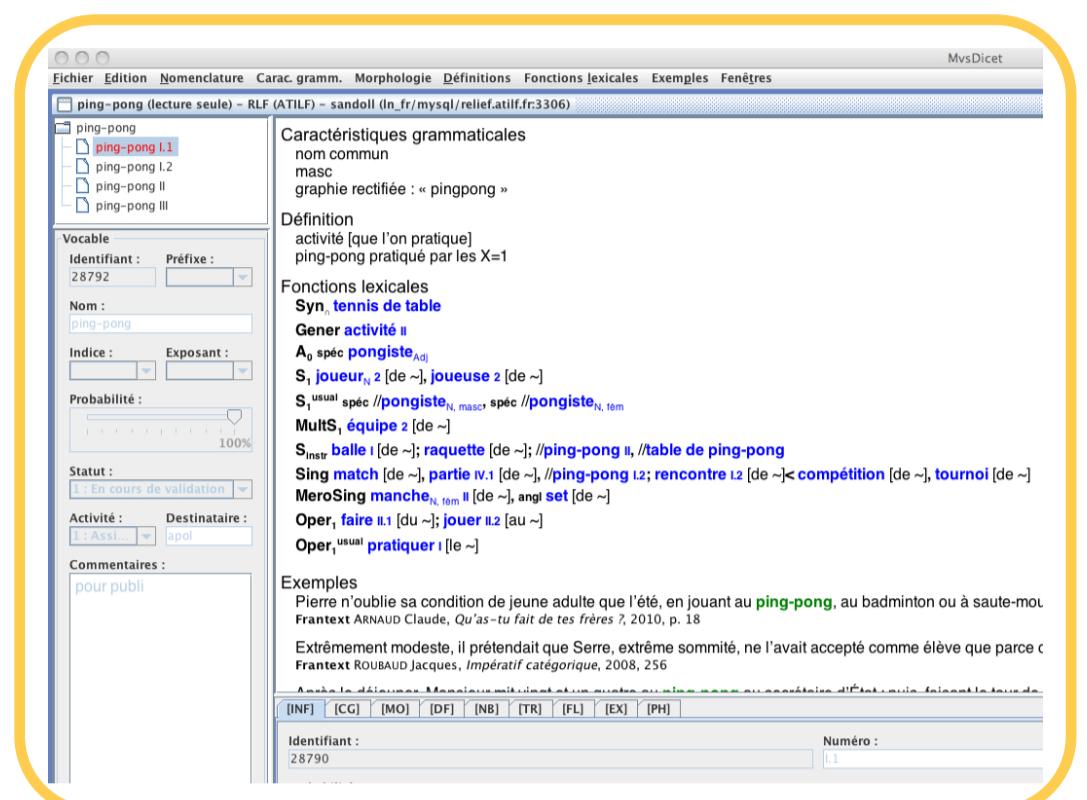
Configuration de dérives lexicales



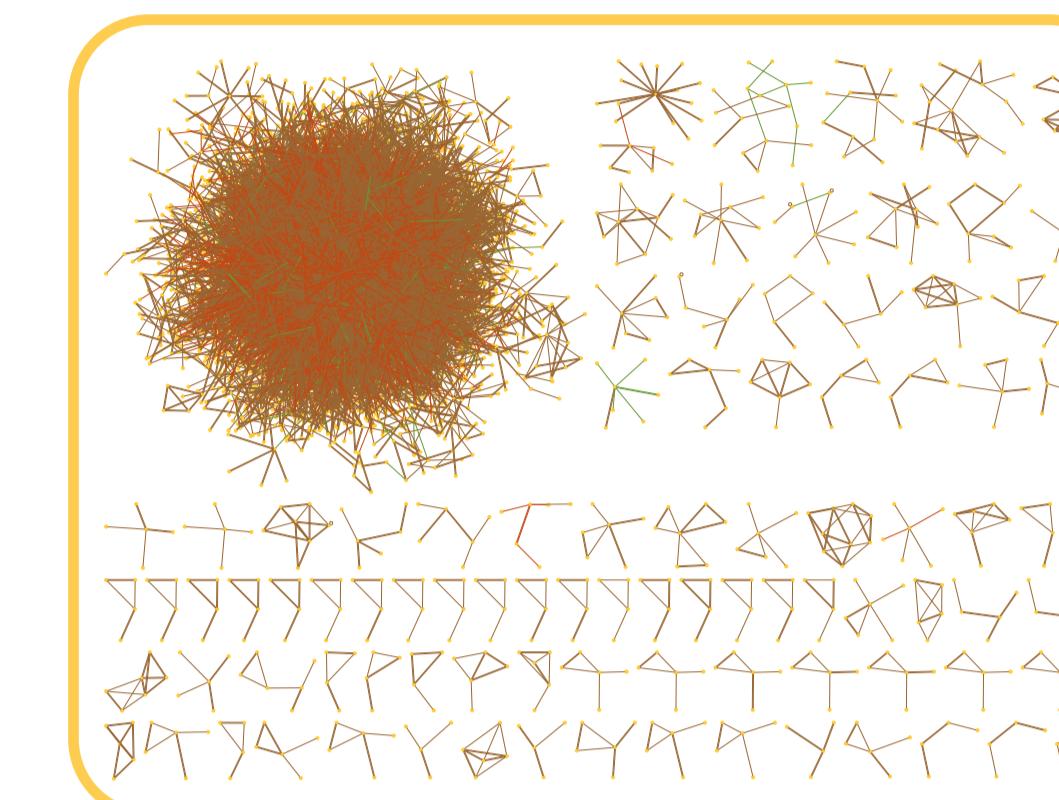
Bibliographie sélective

- Dred Gentner.
Structure-mapping : A theoretical framework for analogy.
Cognitive Science, 7(2) :155–170, April 1983.
- Yves Lepage.
De l'analogie rendant compte de la commutation en linguistique.
PhD thesis, Université Joseph-Fourier – Grenoble I, May 2003.
- Igor Mel'čuk, André Clas, and Alain Polguère.
Introduction à la lexicologie explicative et combinatoire.
Duculot, Paris/Louvain-la-Neuve, 1995.
- Alain Polguère.
Lexical systems : graph models of natural language lexicons.
Language Resources and Evaluation, 43(1) :41–55, 2009.
- Peter D. Turney.
Similarity of semantic relations.
Computational Linguistics, 32(3) :379–416, September 2006.
- Sebastian Wernicke.
Efficient detection of network motifs.
IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 3(4) :347–359, 2006.

Entrée : Réseau Lexical du Français

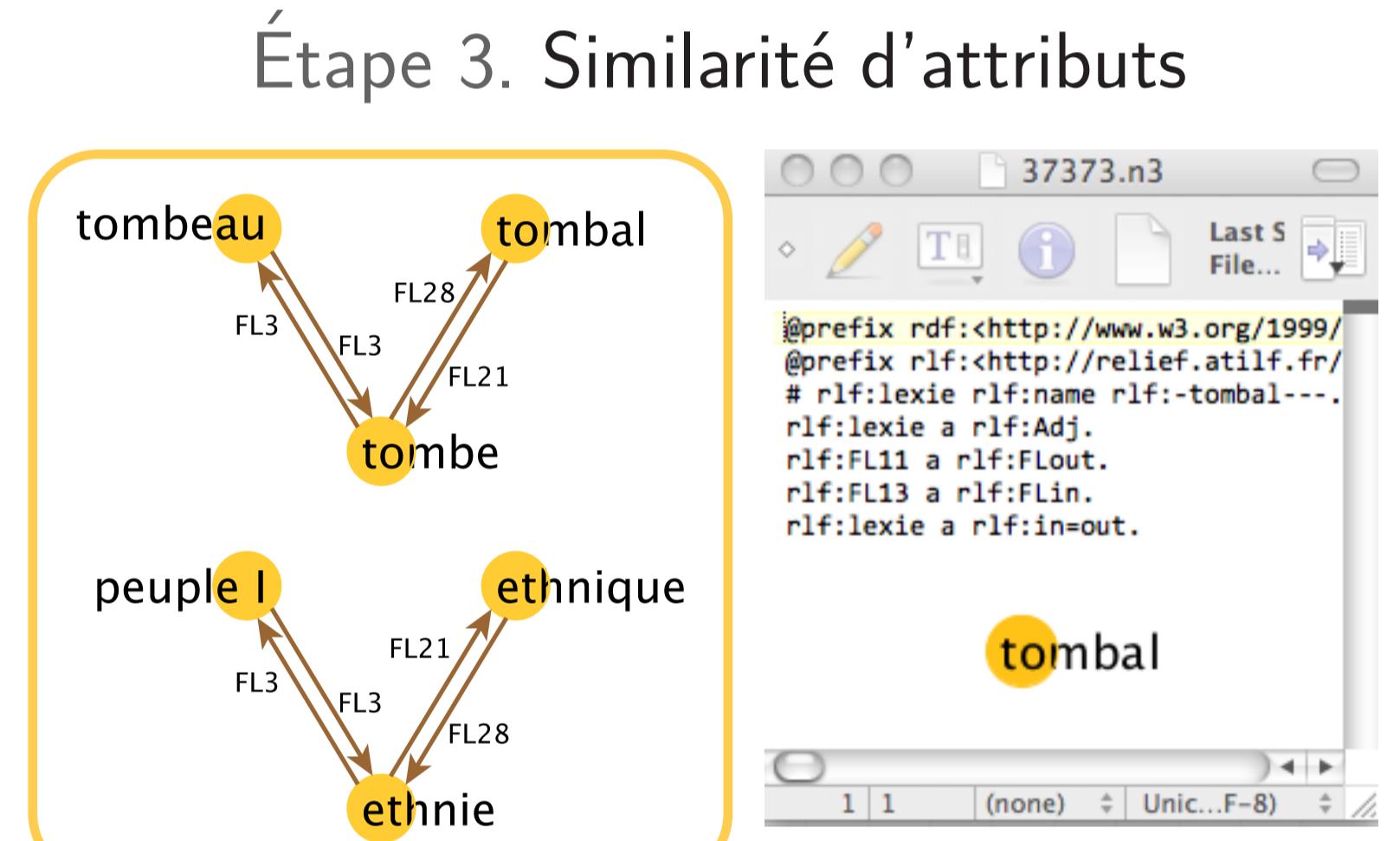
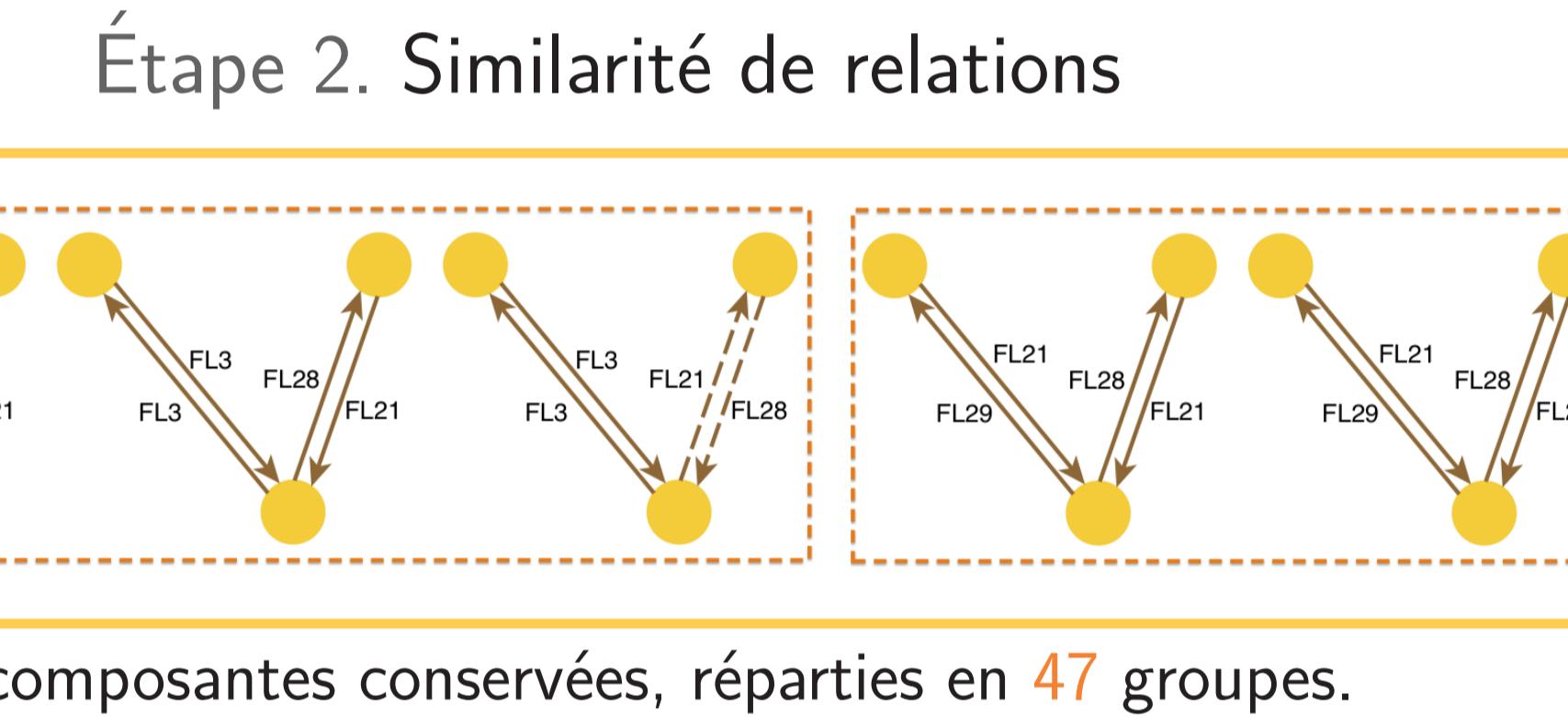
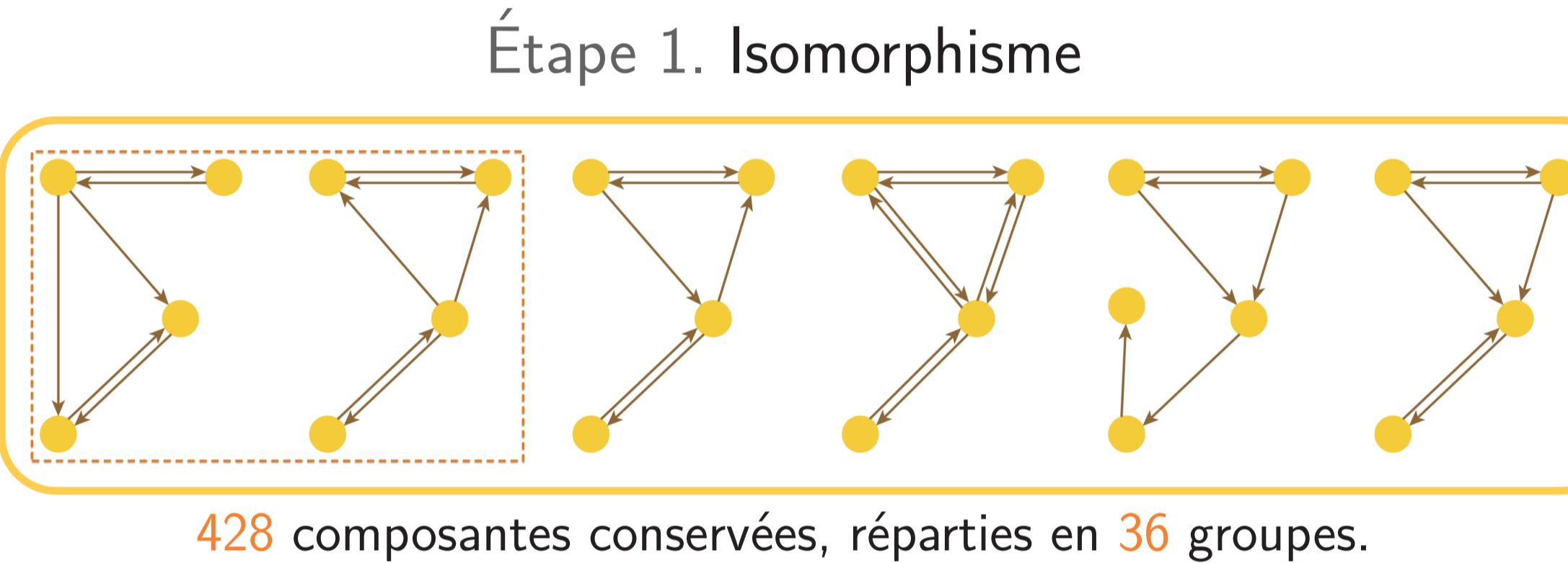


- Lexèmes
- Locutions
- Clichés linguistiques
- Liens de fonctions lexicales
- Liens de copolysémie
- Liens d'inclusion formelle



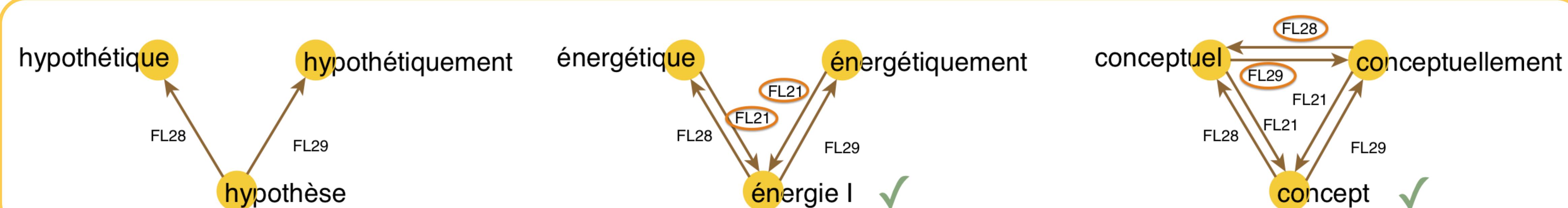
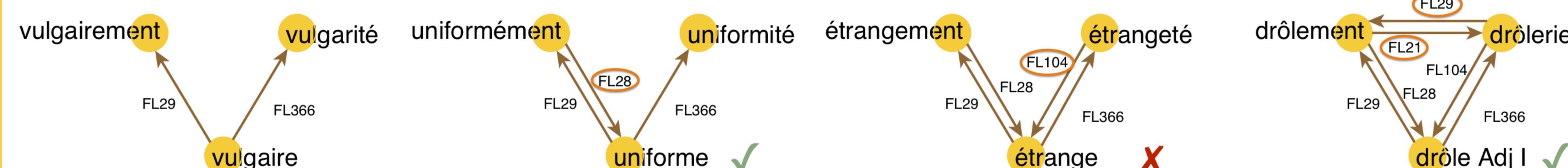
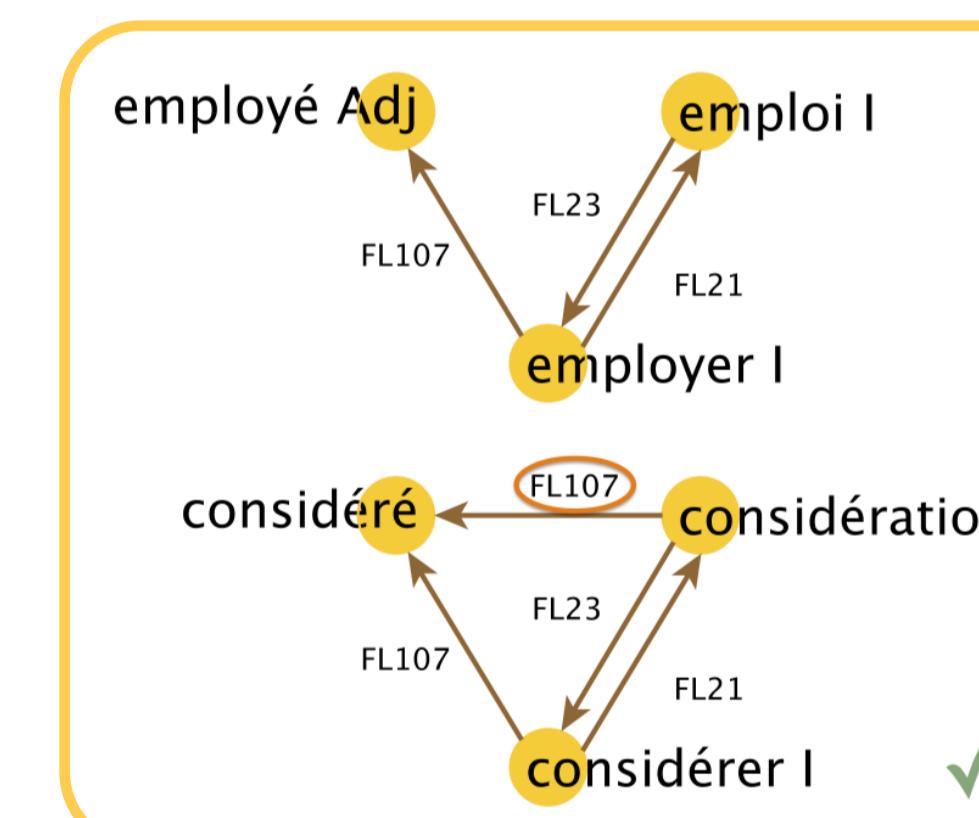
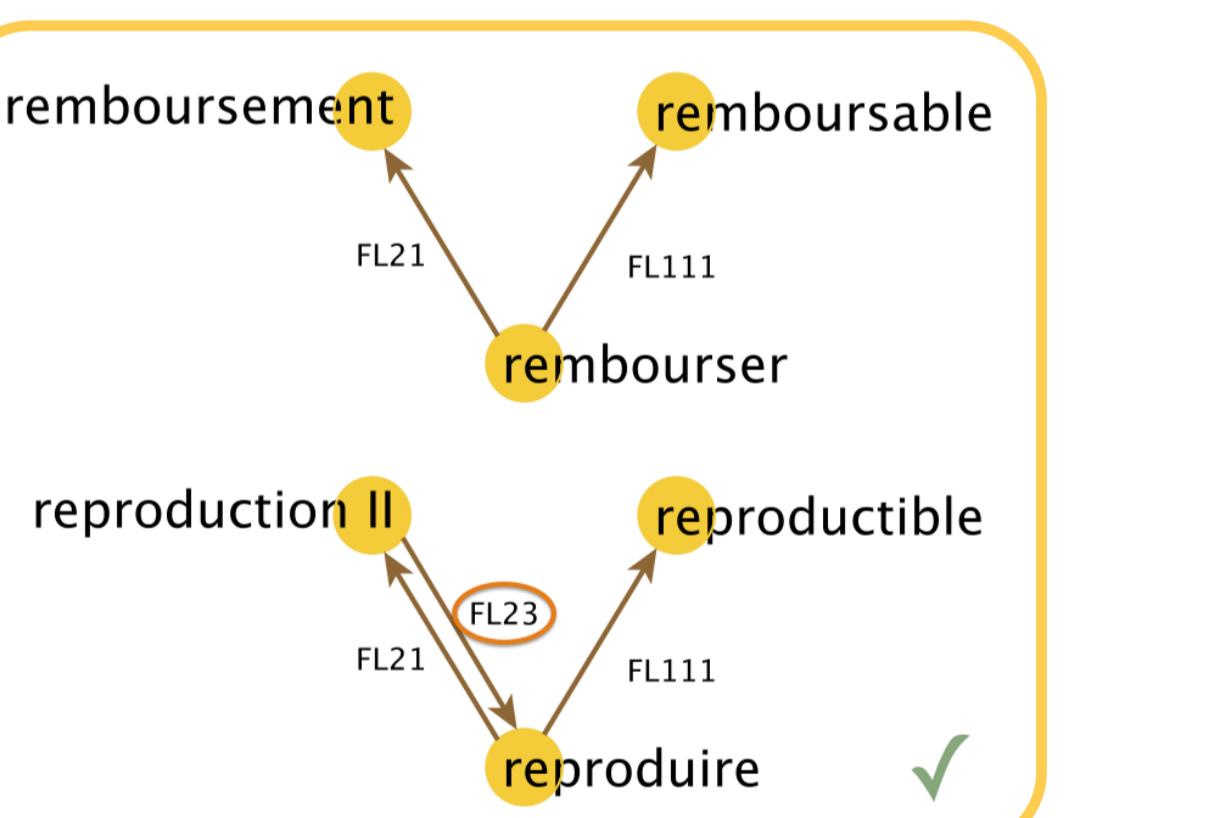
sommets	21 992	coefficient d'agrégation	0,1327
arcs	42 626	Distribution des degrés entrants	
degré sortant moyen	1,9383	a	-2,3977
boucles	36	r ²	0,9397
arcs multiples	577	Plus grande composante connexe	
arcs symétriques	19 906	sommets	15 302
sommets isolés	3 226	arcs	38 274
composantes connexes	4 311	L	13,0402

Regroupement des composantes faiblement connexes analogues



Sortie : Groupes de composantes analogues

- 92 composantes de 3 lexies
- lexies peu décrites
- liens de fonctions lexicales
- analogie confirmée
- observation de structures imbriquées



$$sim_a(Lexie_1, Lexie_2) = \frac{\text{nbr d'attributs communs}}{\text{nbr d'attributs Lexie}_1 + \text{nbr d'attributs Lexie}_2}$$

- nbr $sim_a = 1 <$ nbr de lexies: non analogues
 nbr $sim_a = 1 =$ nbr de lexies: analogues
 nbr $sim_a = 1 >$ nbr de lexies: analogues ?

92 composantes réparties en 24 groupes d'analogues,
 80 composantes réparties en 20 groupes d'analogues ?,
 20 composantes isolées.

Perspectives

- exploitation les motifs locaux du graphe lexical
- exploitation l'état de description des lexies
- développement d'une méthode d'abstraction
- évaluation des résultats par les lexicographes
- granularité des configurations
- risques de propagation d'erreurs