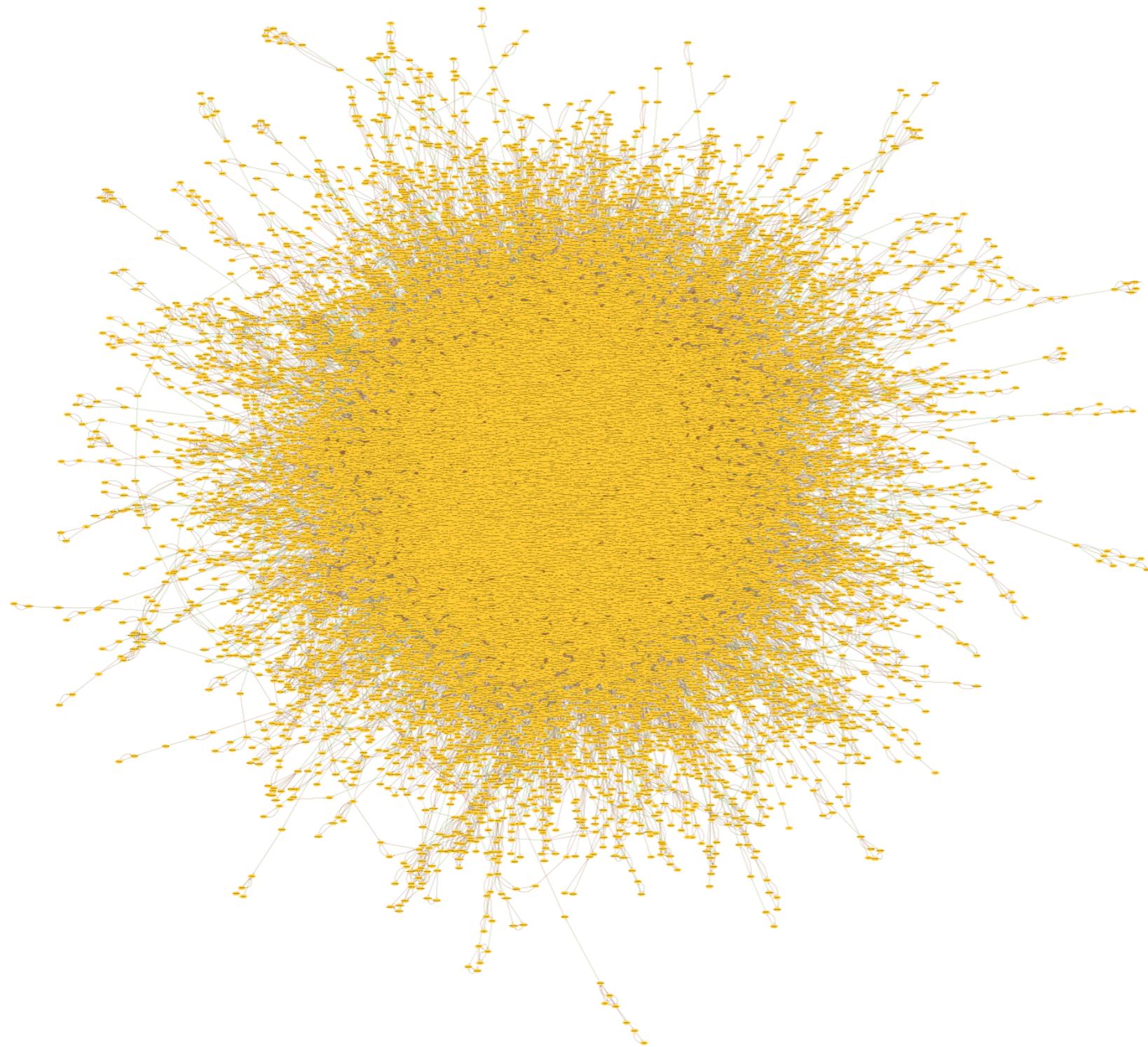


Lexicographie des réseaux lexicaux : le pourquoi et le comment

Sandrine Ollinger — Alain Polguère
ATILF, Université de Lorraine, CNRS

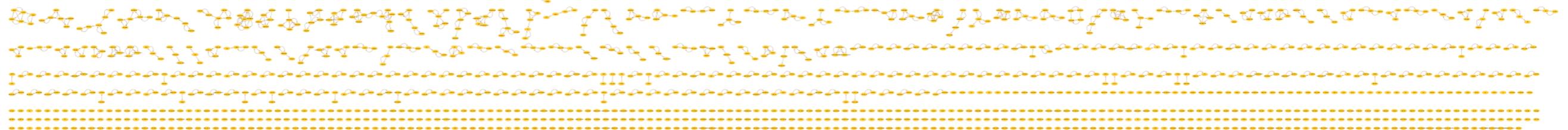
16 décembre 2021



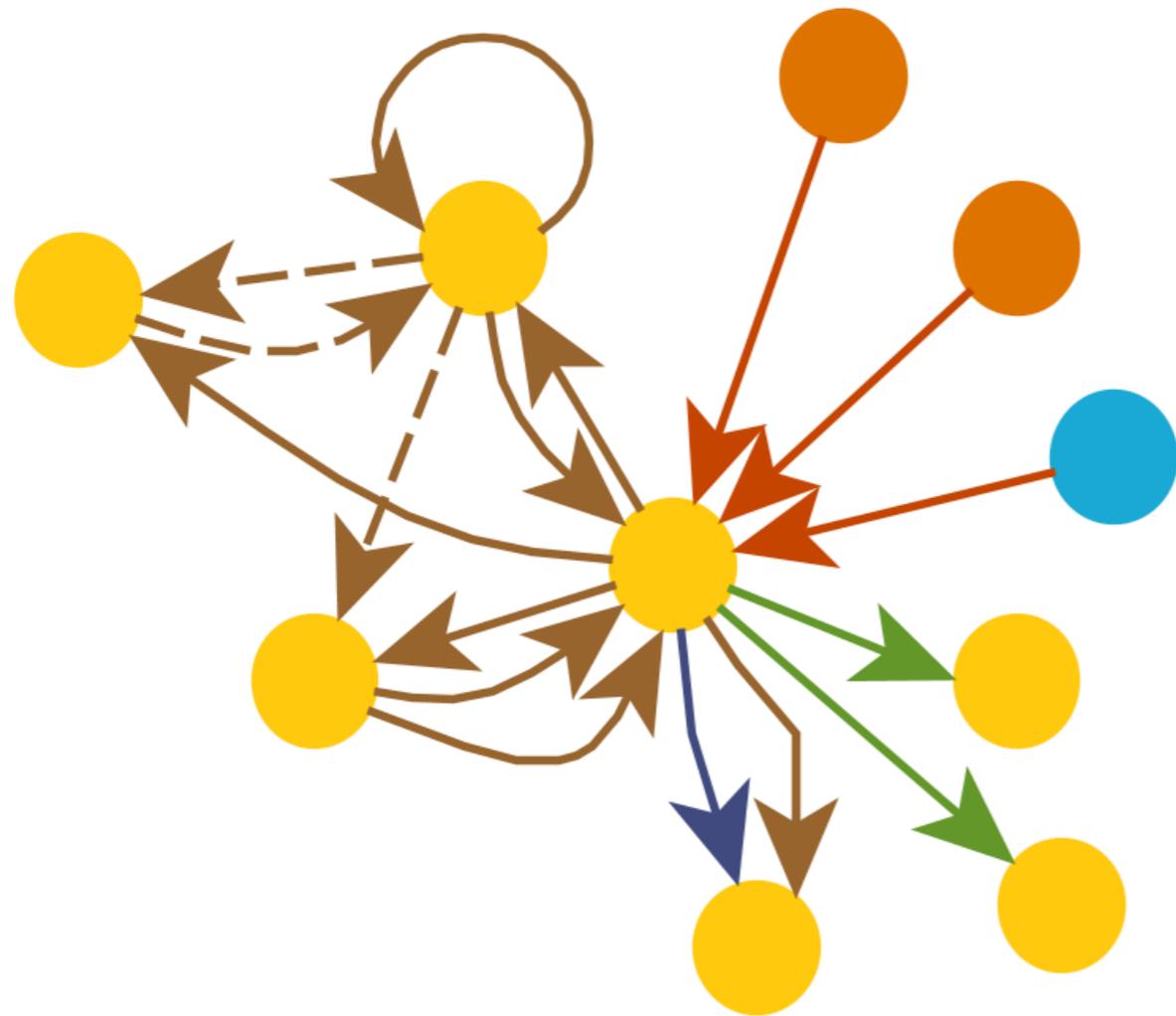


RL-fr

- 29 314 sommets
- ▶ 80 474 arcs



Multigraphe

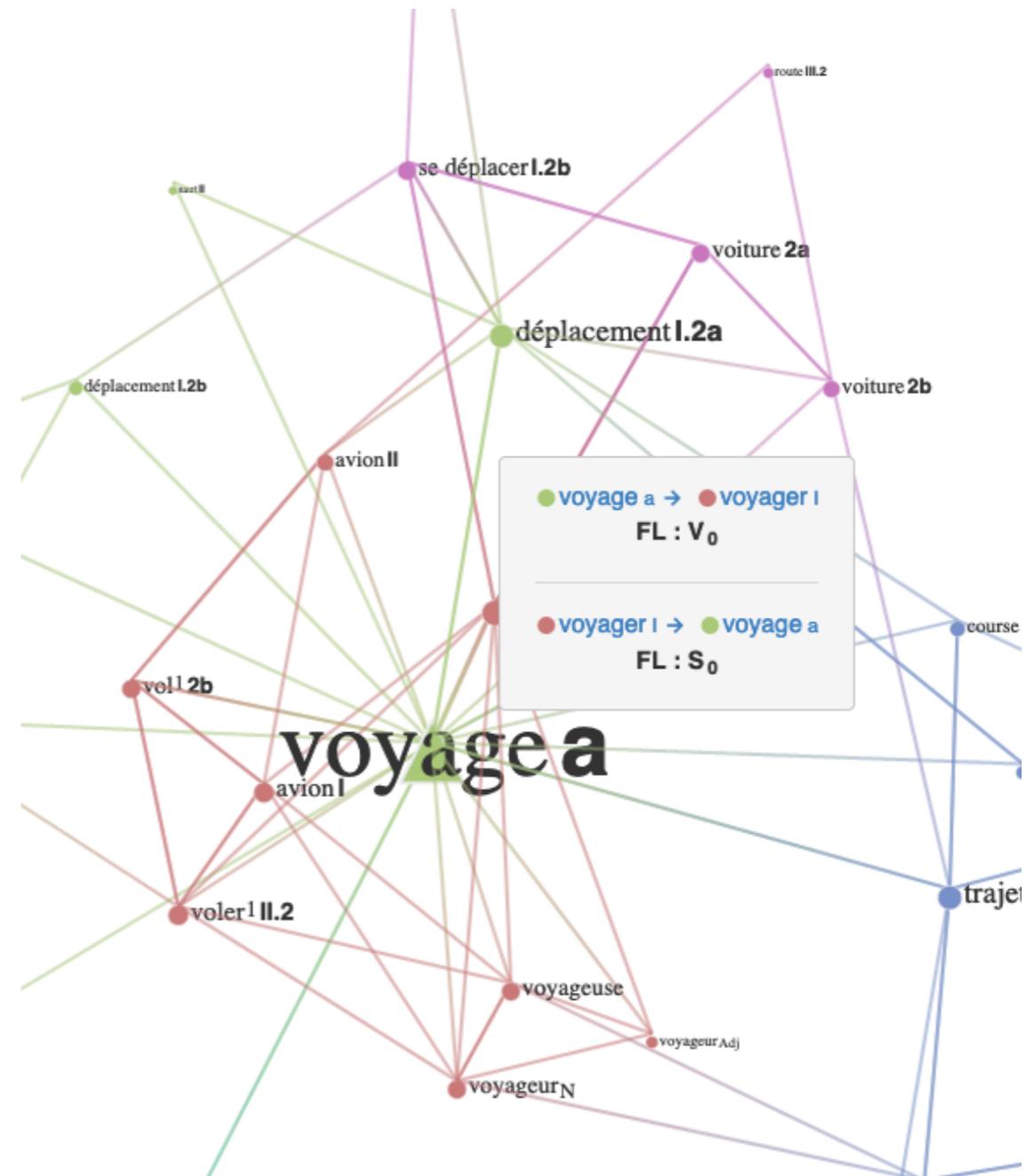


●	lexèmes	88 %
●	locutions	12 %
●	autres	< 1 %

▶	fonctions lexicales	78 %
▶	copolysémie	12 %
▶	inclusion formelle	10 %
▶	inclusion définitionnelle	< 0,5%

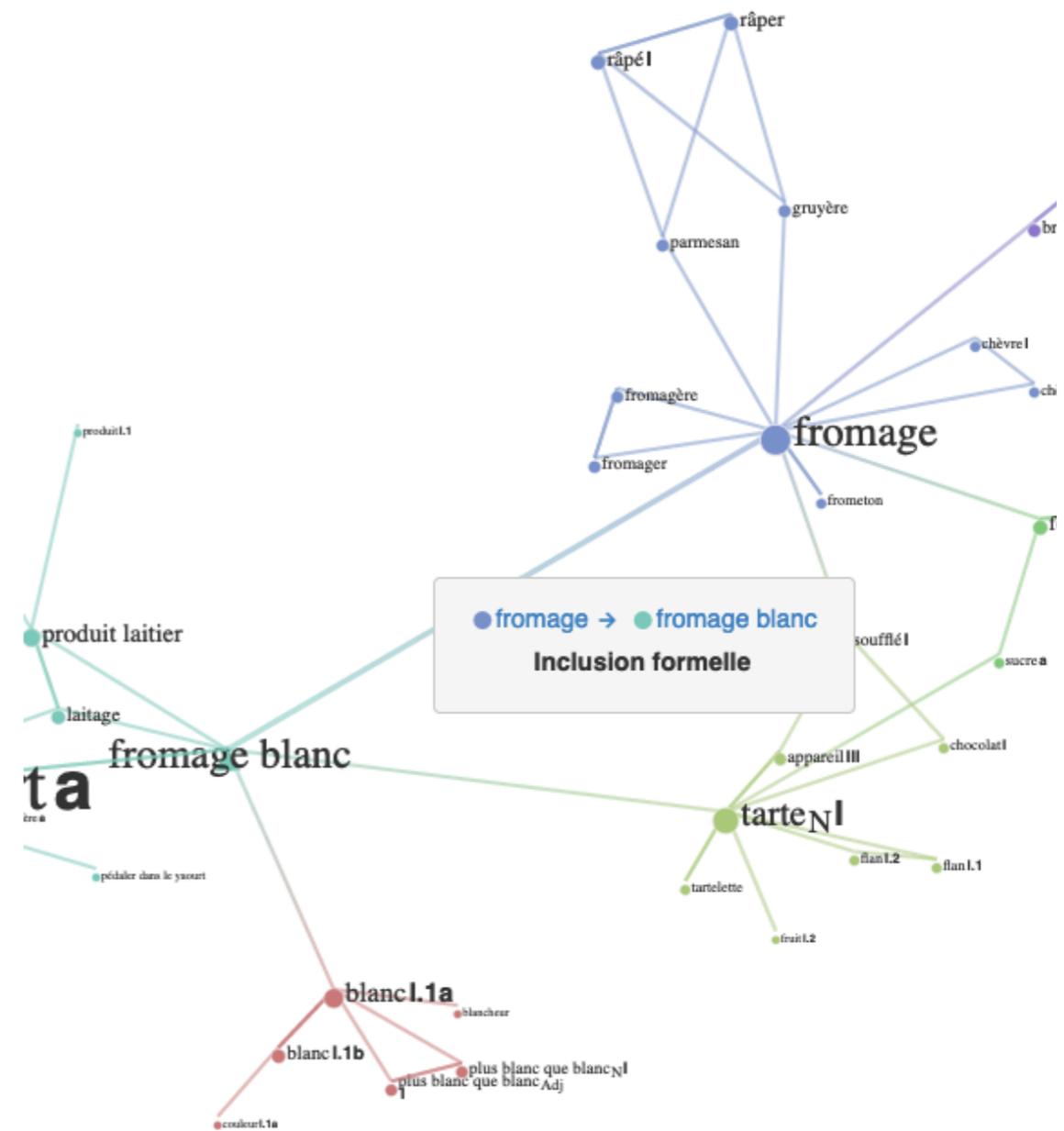
Fonctions Lexicales

- 694 FL réparties en 95 familles
- 295 FL paradigmatiques (Syn, V_0, \dots)
- 399 FL syntagmatiques (Oper₁, Real₁, ...)
- Poids sémantique
0 (Oper₁, ...)
1 (Real₁, ...)
2 (Syn, Anti, V_0, \dots)
- Structure lexicale
Zones denses (Syn, V_0, \dots)
Sommets fortement liés (Oper₁, ...)



Inclusion formelle

- Sans types
- Patrons syntaxiques
- Poids sémantique 0
- Structure lexicale
Dimension culturelle
Espaces lexicaux



Inclusion définitionnelle



x est un individu
y est un individu
x critique y à propos de Ω_1 devant Ω_2
de façon malintentionnée
de façon argumentée
ce qui a pour conséquence de nuire à la réputation de y auprès de Ω_2

Graphe petit monde

Faible densité

- 29 314 sommets
- ▶ 80 474 arcs
- ❖ $n \log(n)$ (Gaume 2004)
= $80\,474 \log(80\,474)$
 $\approx 130\,948$

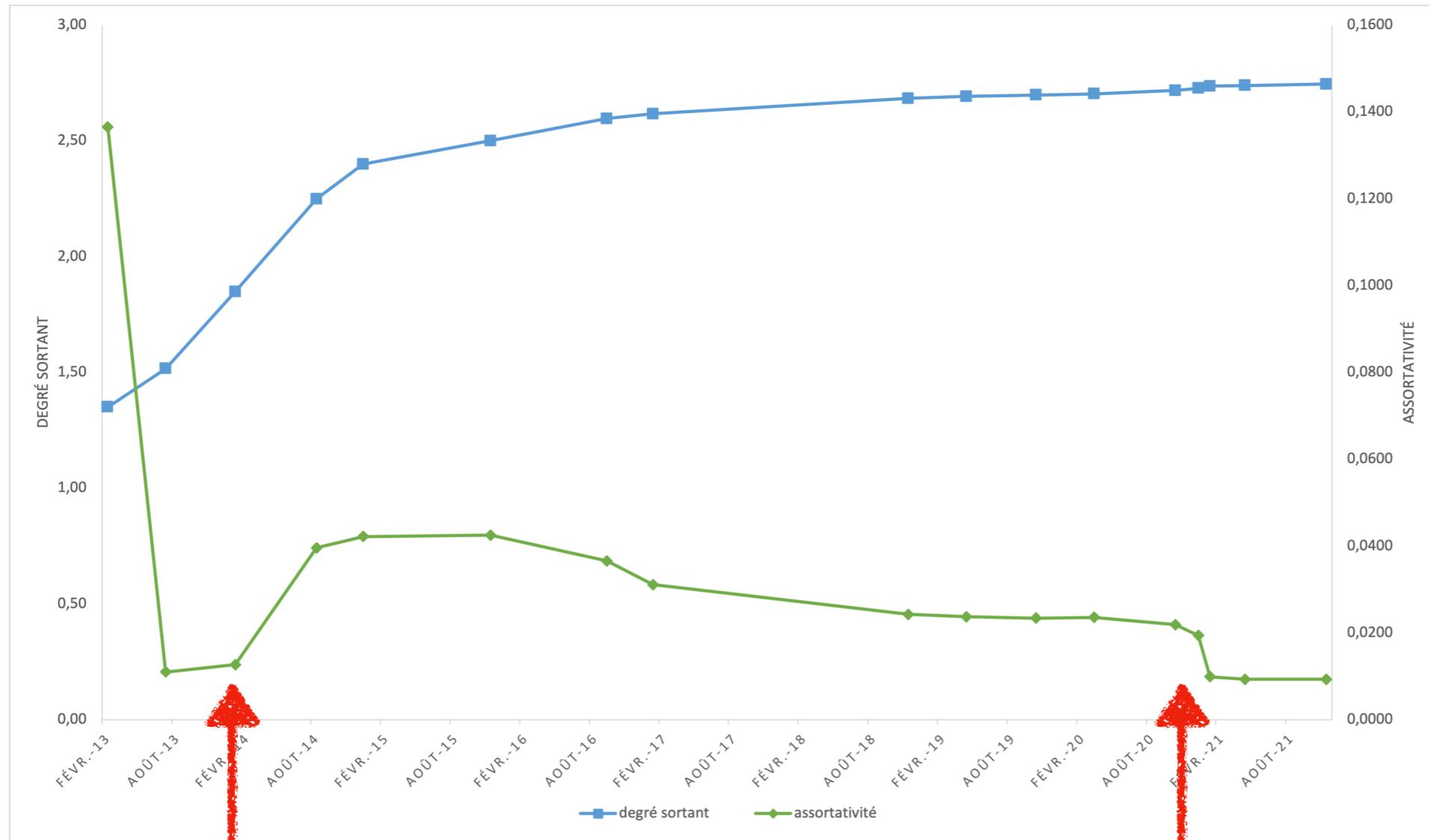
Coefficient d'agrégation élevé

- * $C \approx 0,11$
- * $2m/n^2$ (Delahaye 2003)
= $2 \times 80\,474 / 29\,314^2$
 $\approx 0,00019$

Faible moyenne des + courts chemins

	2014	2021	évolution
<i>sommets</i>	23 599	29 314	+ 5 715
<i>arcs</i>	53 081	80 474	+ 27 393
<i>degré moyen</i>	2,25	2,75	+ 0,50
<i>L</i>	13,31	12,20	- 1,11

Hiérarchique ?



Copolysémie

nuage I - métaphore -> nuage II.2

Gener

marguerite - Gener -> fleur I.2

Sans

	n	m	C	single	ncc	<k> OUT	L	rho
RL-fr complet	29 314	80 474	0,1089	583	10 762	2,75	12,20	0,0093
Sans Gener	29 314	77 706	0,1127	746	11 224	2,65	12,41	0,0301
Sans CP	29 314	70 996	0,1172	1038	13 887	2,42	14,59	0,0040
RL-fr complet 2014	24 217	57 547	0,1240	663	12 096	2,38	13,17	0,0417

Pourquoi et comment ?

The screenshot displays the MvsDicet software interface, which is used for managing a lexical database. It is divided into four main windows:

- Fonctions lexicales - RLF (ATILF) - sandoll (ln_fr/mysql/lexsys.atilf.fr:3306):** This window shows the configuration for a lexical function. It includes fields for 'Identifiant' (20), 'Statut' (0 : Validée), and 'Semantique' (2 : Forte). It also displays 'Type de lien' (paradigmatic), 'Type de standard' (simple standard), and 'Famille' (Gener). A tree view on the left shows a hierarchy of functions, with 'Gener' selected.
- Caractéristiques grammaticales - RLF (ATILF) - sandoll (ln_fr/mysql/lexsys.atilf.fr:3306):** This window shows the configuration for grammatical characteristics. It includes fields for 'Identifiant' (20) and 'Type' (2 : Partie du discours). A tree view on the left shows a hierarchy of grammatical categories, with 'nom' selected under 'Partie du discours'. A table of 'Structures syntaxiques' is visible, listing various grammatical structures and their associated IDs and statuses.
- nuage - RLF (ATILF) - sandoll (ln_fr/mysql/lexsys.atilf.fr:3306):** This window shows the configuration for a 'nuage' (cloud) function. It includes a tree view on the left showing a hierarchy of 'nuage' functions. A diagram in the center shows a 'nuage I' node connected to several other nodes. Below the diagram, there are fields for 'Identifiant', 'Numéro', and 'Statut', and a 'Probabilité' slider set to 100%.
- Nomenclature - RLF (ATILF) - sandoll (ln_fr/mysql/lexsys.atilf.fr:3306):** This window shows the configuration for a nomenclature function. It includes a list of words on the left, with 'nuage' selected. The main area shows the configuration for 'nuage', including its ID (28521), status (En cours de traitement (2)), and activity (Pris en charge (2)). It also shows the 'SENS NON Créés' (non-created senses) and the 'Modèle' (Model) for the function.