



**HAL**  
open science

## La créativité lexicale : des pratiques sociales aux textes

Sandrine Ollinger, Mathieu Valette

► **To cite this version:**

Sandrine Ollinger, Mathieu Valette. La créativité lexicale : des pratiques sociales aux textes. CI-NEO'08, May 2008, Barcelone, Espagne. 2008. hal-01806090

**HAL Id: hal-01806090**

**<https://hal.science/hal-01806090>**

Submitted on 1 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# La créativité lexicale : des pratiques sociales aux textes

## Objectifs généraux :

- Une plateforme de veille lexicale semi-automatisée pour la production de ressources lexicographiques (attestations, mesures, contextes et sources)
  - matériau pour les lexicographes (dictionnaire) ■ propositions pour de nouvelles pratiques lexicographiques
- Un outil pour l'étude de la néologie (observations rapportées aux discours et aux genres textuels)

### Corpus « pouvoir d'achat »

Discours	Politique			Journalistique			Blog		
	Sources	Elysée	Marianne	Le Point	Nouvel Observateur	Couillies Bruxelles, UE	Versac	Plume de Presse	
Nb de textes	87	108	170	128	11	20	30		
Occurrences	325 135	108 753	86 772	109 202	110 633	108 475	106 883		
Formes	18 871	14 593	9 030	9 860	14 669	14 449	15 170		

- Discours politique : 87 textes de propagande, 325 135 occurrences, 18 871 formes
- Discours journalistique : 406 articles, 304 727 occurrences, 21 825 formes
- Discours blog : 61 posts (commentaires compris), 325 991 occurrences, 31 015 formes

### poamo

- **1 table de candidats / type**
  - Id\_candidat
  - Forme graphique
  - Hypothèse d'étiquette
  - Hypothèse de lemmatisation
- **1 table d'attestations / type**
  - Id\_candidat
  - Id\_document
  - Extrait étiqueté
  - Localisation
- **1 table de description des documents**
  - Id\_document
  - Noms corpus et sous-corpus
  - Titre, auteur, date
  - Typologie (discours, domaine, genre)
  - Nombre de mots, phrases, paragraphes
  - Taux de Néologie globale, formelle, catégorielle

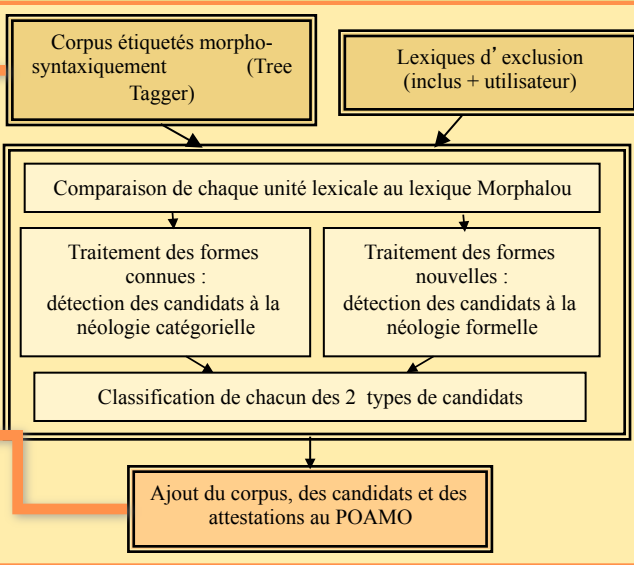
- **Observatoire de créativité lexicale**
  - Base de données relationnelle MySQL
  - Entrée : sortie de POMPAMO
  - Sortie : interface d'interrogation PHP
- **Interrogations croisées**
  - Requêtes sur méta-données
  - Requêtes sur formes et expressions régulières
  - Calculs de fréquences
- **Caractérisation des candidats et sélection**
  - Enrichissement lexicométrique
  - Évolution diachronique
  - Répartition entre types de corpus

Diagramme de données montrant la structure du corpus et des attestations.

**corpus** → sous-corpus → description texte

**candidat** → attestations

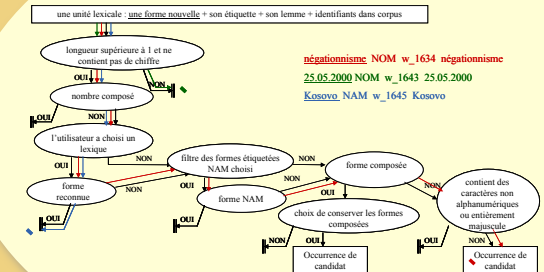
Exemple de données : `Allemagne : Textième gauche montre la voie - Plume de presse - 2008-01-01 - Olivier Bismet`



### Lexiques d'exclusion

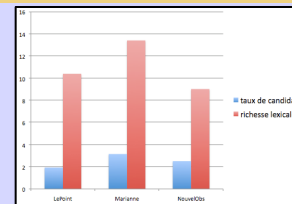
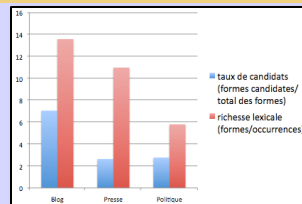
- **Lexique principal de formes flechées du français : MORPHALOU 2.0**
- **Validité linguistique (Nomenclature TLF)**
  - Large couverture (524 725 formes flechées 95 810 lemmes)
  - Accès libre au format XML - LMF
- **Lexiques supplémentaires inclus**
  - 70 438 Noms propres
  - 6 903 Adjectifs toponymiques et gentils
  - 140 nombres composés
- **Lexiques supplémentaires utilisateur**

### Filtrage des formes nouvelles



## Remarques :

- Tous les discours sont *a priori* créatifs
- Mais les genres (ensembles de normes de production des textes), présentent des potentiels néologiques variables.
- Ex. Marianne (pamphlétaire). Taux de créativité : 3,16
- Le Point (information générale). Taux de créativité : 1,94



	candidats	formes
LePoint	175	9030
Marianne	461	14593
NouvelObs	248	9860

**Le genre polémique dans Marianne :** la néologie dérivationnelle et suffixale comme mode d'expression polémique

« bobo TGV », « tactico-politicien », « ultrapopuliste », « Sarkorigolo », « starisation », « centristo-bayrouiste », etc.

➔ Vers une lexicologie textuelle