

The Logoscope: Semi-automatic Tool for Detecting and Documenting the Contexts of French New Words

Equipe Linguistique, langues,

LILPA

Université de Strasbourg

C. Gérard 28.06.2016

Presentation

- 1 | The Logoscope: What is it?
- 2 | Contexts diversity and its impact on lexical creativity
- 3 | Detecting themes, genres and generic fields

1

The Logoscope: What is it?

The Team:

C. Gérard

D. Bernhard

A. Todirascu

I. Falk

L. Bruneau

A.-L. Rosio

C. Stapleton

Financial support:

University of Strasbourg (2012-2015)

French Ministry of Culture (...2017)

[Ordre alphabétique](#)[Fréquence](#)[Chronologie](#)[Journal](#)[Catégorie gram.](#)[Thème](#)[Procédé](#)[Position](#)[Rechercher](#)

Logoscope

Documentation quotidienne des nouveaux mots français

Projet et Équipe

Guide d'utilisation

Publications

Mentions légales

Obtenir la version experte

lilpa
linguistique, langues, parole

UNIVERSITÉ DE STRASBOURG



Ministère de la Culture
et de la Communication

FranceTerme



Néologisme : [Astuces](#) [recherches avancées](#)

Filtrer par :

Catégories :



Journaux :



Filter des thèmes pour une entrée : *Au moins un* *Seulement ceux sélectionnés*

Thèmes :

- Communications
- Culture
- Défense
- Droit
- Economie
- Education
- Histoire
- International
- Loisirs
- Politique
- Santé
- Sciences
- Société
- Transports
- Technologies

Date : à

[Rechercher](#)

Offline Version



Base des nouveaux mots du français



Tweets 724 Abonnements 190 Abonnés 263 J'aime 199 Moments 1

Éditer le profil

Logoscope

@_logoscope
Twitter du Logoscope, premier site de documentation quotidienne des nouveaux mots du français. *Chaque semaine un mot inédit. **Ajout de néologismes : bimestriel.
Strasbourg, France
logoscope.unistra.fr
Inscrit en décembre 2015

43 Photos et vidéos



Tweets Tweets & réponses Médias

Tweet épinglé

Logoscope @_logoscope · 11 janv.
Nouvelle arrivée de #néologismes dans le Logoscope : logoscope.unistra.fr



1 2

Logoscope @_logoscope · 2 h
Violence du vocabulaire (#loups) : l'arrêté officiel ne dit pas #abattage mais

Suggestions Actualiser Tout afficher

- IRC Intl Rescue Comm @ [Suivre]
- Michel Francard @Michel... Suivi par Zoé Zaam et d'autres [Suivre]
- Ch. Ono-dit-Biot @C... [Suivre]

Trouver des amis

Tendances : Paris Modifier

- #Goulard 2 195 Tweets
- #journeemondialeedesrefugies
- #sportsMW @GallicaBnF, @reseau_canope et 2 autres tweetent à ce propos
- Jean-Marie Le Pen

Database building: 4 stages process



1. Detecting

2. Selecting

3. Describing

4. Integrating

automatic

manual

contextual features

lexical features

=

true neologisms

1st stage

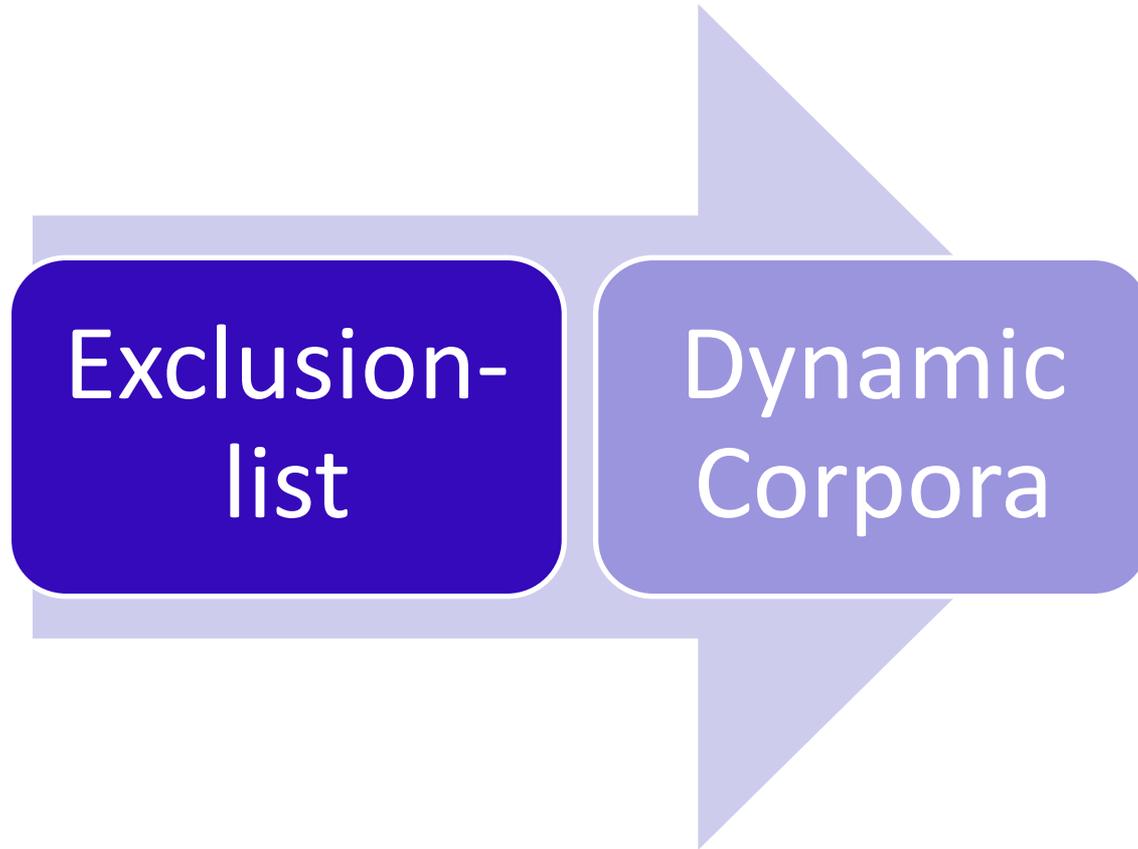
1. Detecting

automatic

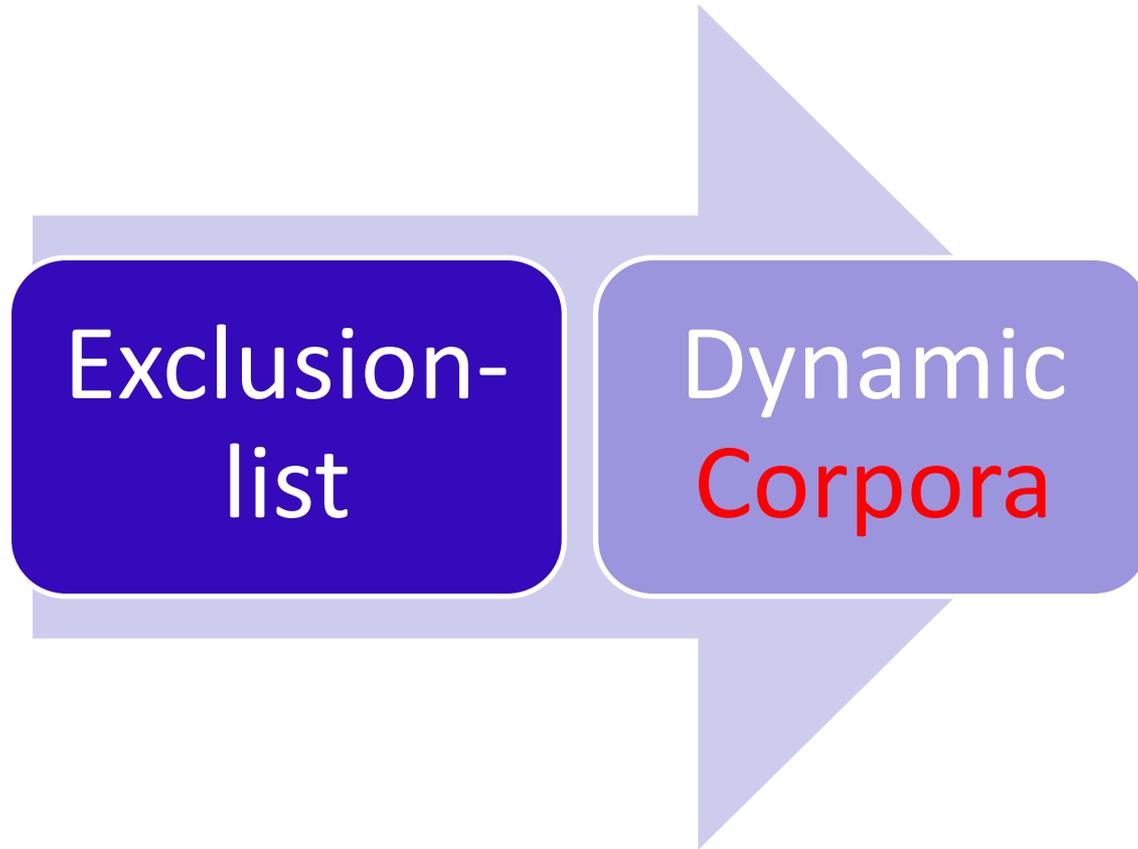
manual



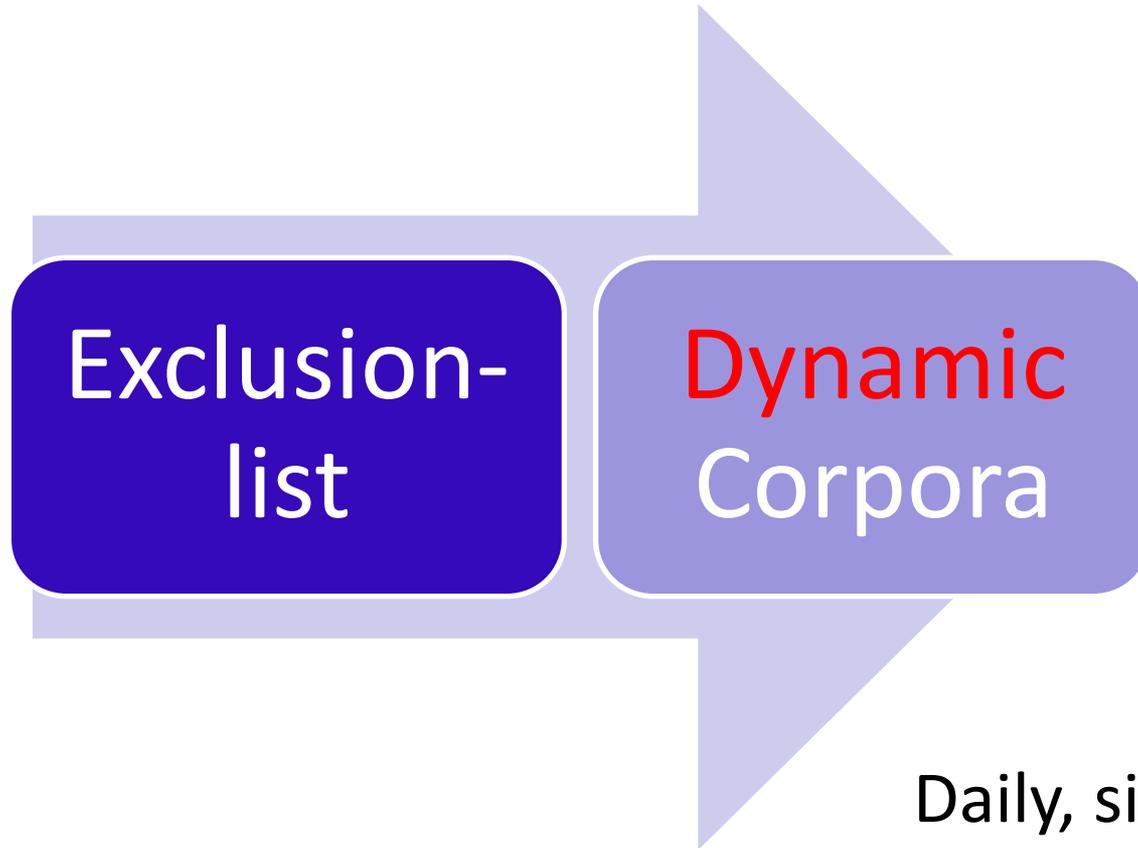
Exclusion-list and dynamic corpora



Exclusion-list and dynamic corpora



Exclusion-list and dynamic corpora



Daily, since 2014...

RSS-feeds

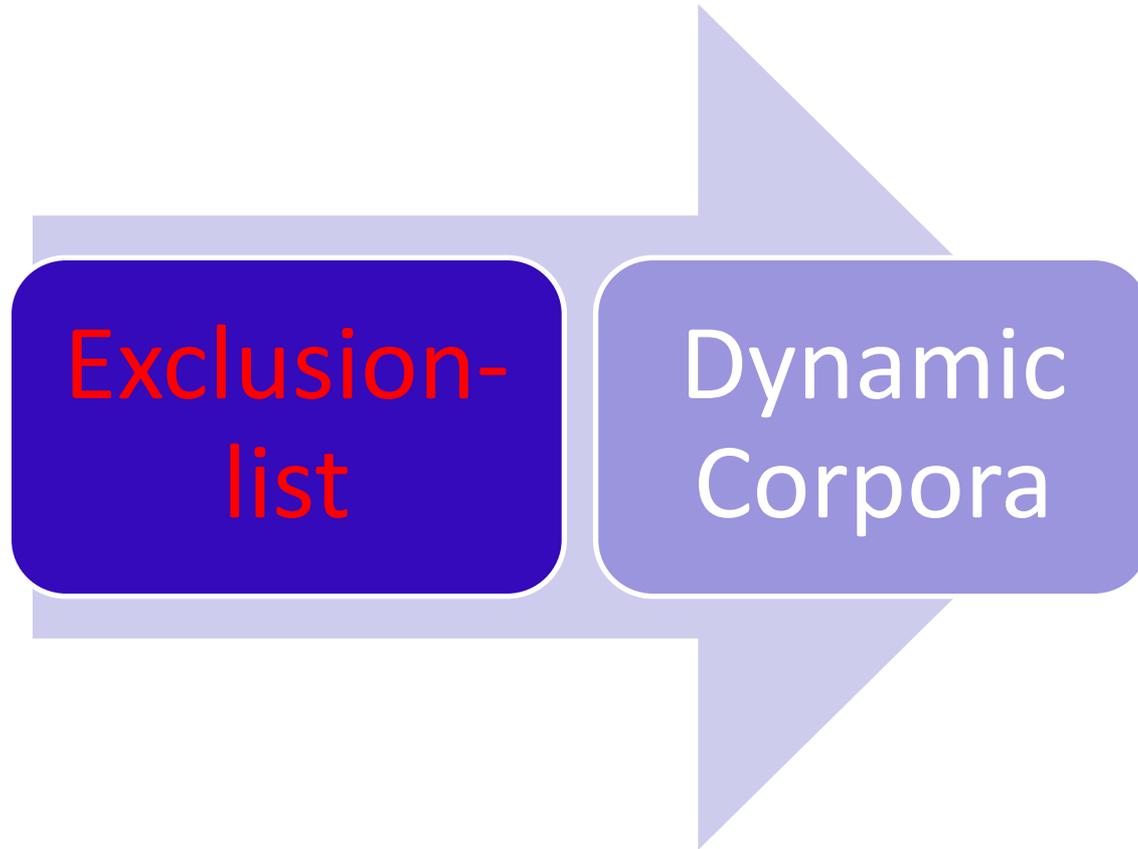
June 19th 2017:

- Number of words: 23248
- Collected articles: 369
- Unknown forms detected: 268

June 20th 2017:

- Number of words: 24330
- Collected articles: 413
- Unknown forms detected: 224

Exclusion-list and dynamic corpora



Exclusion-list:

- Known words / various forms
- 22 megabytes / 2 millions ~~forms~~ ^{XX XXX} (25/11/2015)
- Originally: *fra_mixed_2012* ^{@xxx}

...



CORPORA COLLECTION

UNIVERSITY LEIPZIG

247 corpora corpora basé dictionnaires monolingues pour 222 langues.

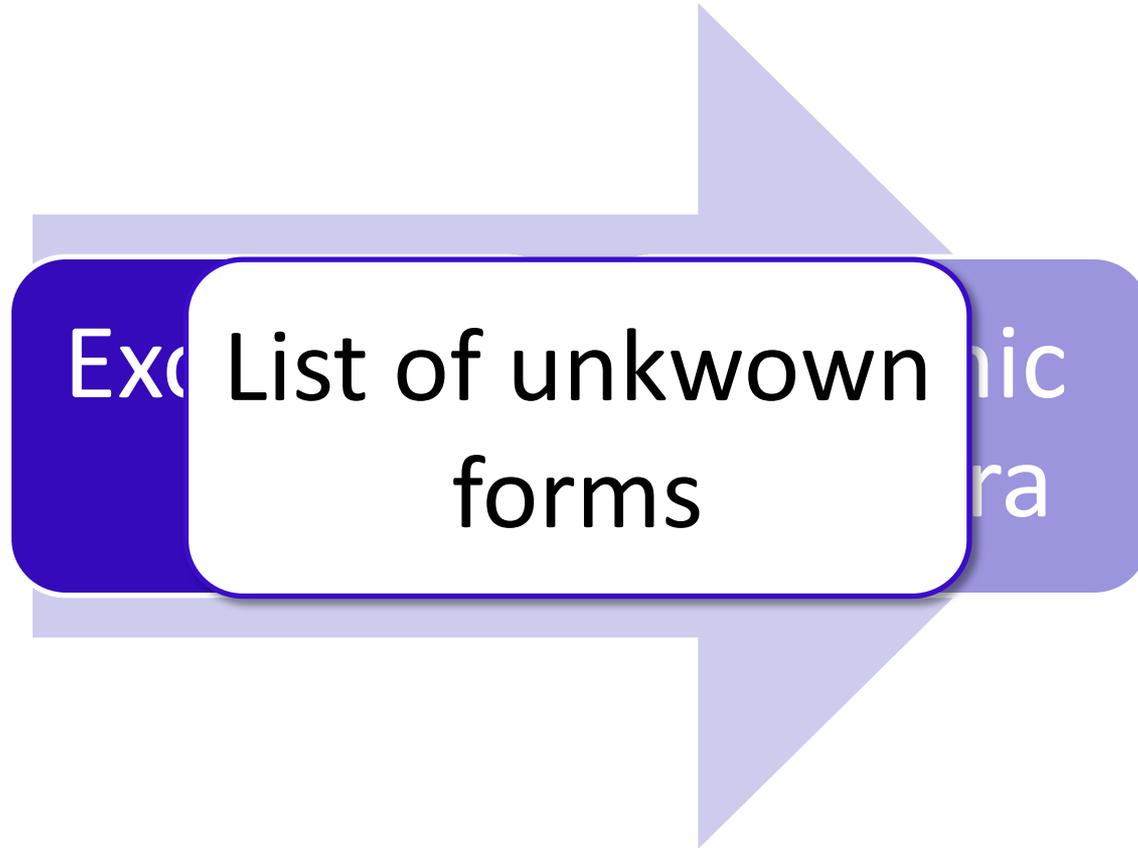
 ?

Corpus: French (fra_mixed_2012)

French mixed corpus based on material from 2012

Phrases: 74,823,426 · Types: 7,873,935 · Mots: 1,468,766,604 [plus...](#)

Exclusion-list and dynamic corpora



June 19th 2017:

- Number of words: 23248
- Collected articles: 369
- Unknown forms detected: 268

June 19th 2017:

...

macronisation (2, 2017-06-20#13-lefigaro.txt)

antispasti (2, 2017-06-20#172-lefigaro.txt)

manuque (2, 2017-06-20#250-lefigaro.txt)

consilio (2, 2017-06-20#250-lefigaro.txt)

germanopatine (2, 2017-06-20#172-lefigaro.txt)

computare (2, 2017-06-20#250-lefigaro.txt)

pub-hôtel (1, 2017-06-20#279-lemonde.txt)

prêt-à-gober (2, 2017-06-20#172-lefigaro.txt)

raditionnellement (2, 2017-06-20#23-lefigaro.txt)

anti-friction (1, 2017-06-20#263-60millions.txt)

germanocratie (2, 2017-06-20#172-lefigaro.txt)

hachiches (1, 2017-06-20#402-liberation.txt)

avion-voiture (2, 2017-06-20#125-lefigaro.txt)

...

Type of neologisms?

[+] Simple lexems (+ loanwords)

[-] Polylexical units (*smartphone zombie*)

[-] Semantic neologism (*troll*)

[-] Variety (frozen ; slang...)

Type of neologisms?

[+] Simple lexems (+ loanwords)

[-] Polylexical units (*smartphone zombie*)

[-] Semantic neologism (*troll*)

~~[-] Variety (frozen ; slang...)~~

4 stages process

1. Detecting

automatic

manual

2. Selecting

true neologisms

June 19th 2017:

...

✓ macronisation

✗ antispasti

✗ manuque

✗ consilio

✓ germanopatine

✗ computare

✗ pub-hôtel

✓ prêt-à-gober

✗ Traditionnellement

✗ anti-friction

✓ germanocratie

✓ Hachiches

✓ avion-voiture

...

✗ = rejected

> exists

> spelling error

June 19th 2017:

...

✓ macronisation

✗ antispasti

✗ manuque

✗ consilio

✓ germanopatine

✗ computare

✗ pub-hôtel

✓ prêt-à-gober

✗ Traditionnellement

✗ anti-friction

✓ germanocratie

✓ Hachiches

✓ avion-voiture

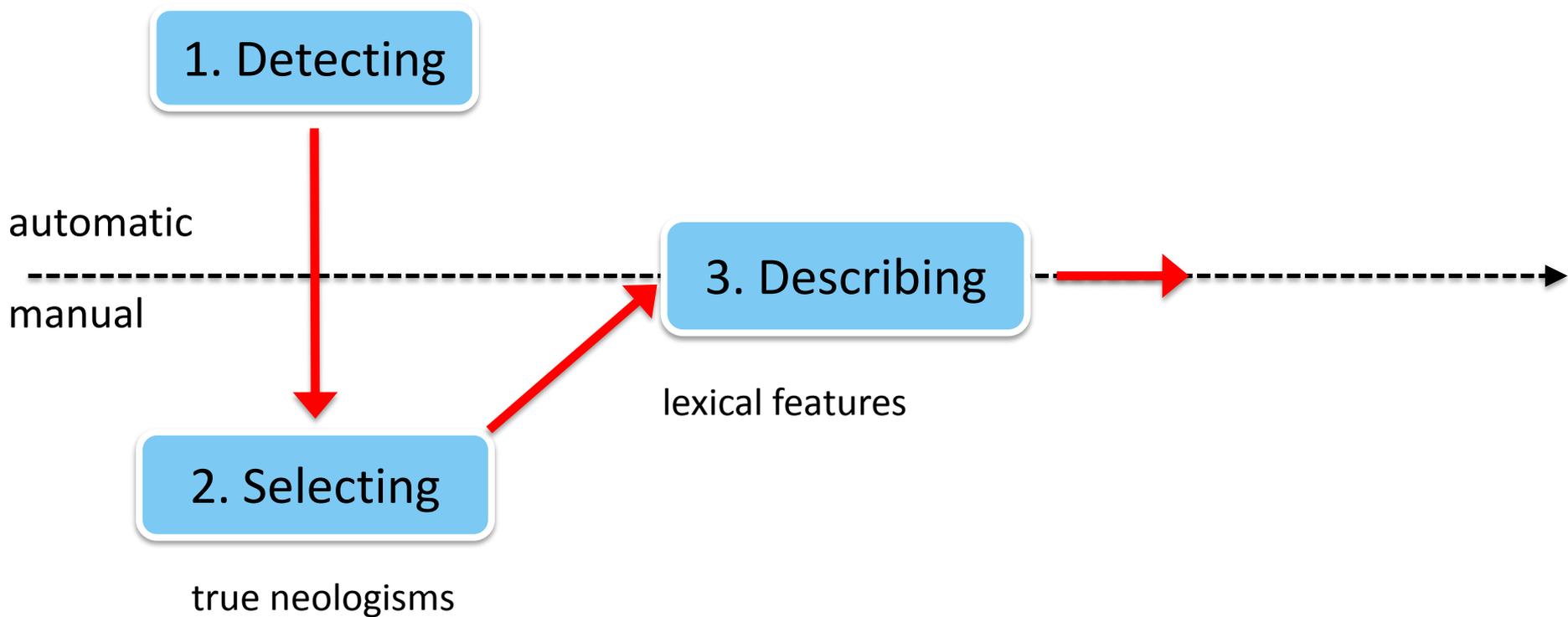
...

✓ = rare words

> google.fr (1>100 occ.)

> Factiva

4 stages process



	POS	Meta-category of creation

	POS	Meta-category of creation
macronisation	NAME	
a-dô-re	VERB	

	POS	Meta-category of creation
macronisation	NAME	MORSEM
a-dô-re	VERB	MORPH

Méta-catégories	Procédés de création
morphosémantique MORSEM	préfixation / suffixation
	flexion
	composition(s)
	onomatopée
	paronymie
syntactico-sémantique SYNSEM	conversion(s)
	combinatoire syntaxique / lexicale
	métaphore
	métonymie
morphologique MORPH	troncation
	siglaison / acronyme
emprunt EMP	emprunt brut / emprunt assimilé

Sablayrolles 2011

	POS	Meta-category of creation
macronisation	NAME	MORSEM
a-dô-re	VERB	MORPH

Simple, but...

saves time

avoids mistakes

 lexicographers!



crowdlending

Sommaire [masquer]

- [+] Français
- [+] Anglais

Français [modifier le wikicode]

Étymologie [modifier le wikicode]

Emprunté à l'anglais *crowdlending*.

Nom commun [modifier le wikicode]

crowdlending \kʁawd.lɛn.diŋ\ masculin

1. (*Anglicisme*) (*Finance*) Financement participatif sous forme de prêt de particuliers.

- *Il existe trois formes de « crowdfunding » qui cohabitent et lèvent, chaque année, de plus en plus de fonds. La version originelle propose aux internautes (les « crowdfunders ») d'effectuer des dons (avec ou sans contrepartie) en ligne pour soutenir des projets d'ordre culturel, artistique ou associatif. Plus récente, la seconde forme concerne des prêts avec intérêt à destination des entreprises, afin de financer leur développement ou de réaliser des opérations particulières. C'est le « **crowdlending** ».* — (Laurence BOCCARA, lesechos.fr, 23 octobre 2016)

Invariable
crowdlending \kʁawd.lɛn.diŋ\

Synonymes [modifier le wikicode]

- prêt participatif
- financement participatif par prêt

Vocabulaire apparenté par le sens [modifier le wikicode]

- crowdfunding

Traductions [modifier le wikicode]

- Anglais : crowdlending ^(en)

Mouse click!

Ajouter une traduction en français : traduction [Ajouter] Plus

Références [modifier le wikicode]

- <https://www.ces-berra-immo.fr/crowdfunding-immobilier-qu-est-ce-que-le-crowdlending-pret-definition-1054.html>
- [Logoscope](#), Université de Strasbourg, 2016 (crowdlending)

- Page d'accueil
- Recherche avancée
- Portails thématiques
- Page au hasard
- Page au hasard par langue
- Poser une question

- Contribuer
- Journal des contributeurs
- La Wikidémie
- Communauté
- Discuter sur IRC
- Modifications récentes
- Faire un don

- Aide
- Aide
- Modèles
- Conventions
- Créer un article
- Télécharger

- Imprimer / exporter
- Créer un livre
- Télécharger comme PDF
- Version imprimable
- Outils
- Pages liées
- Suivi des pages liées
- Importer un fichier
- Pages spéciales
- Adresse permanente
- Information sur la page
- Citer cette page

- Dans d'autres langues

Logoscope

Ordre alphabétique Fréquence Chronologie Journal Catégorie gram. Thème Procédé Position Rechercher

crowdlending Graphique 1 2 3

- Nom Commun - **ECONOMIE-Finance** | **DROIT-Justice et Législation** | **ECONOMIE-Commerce** - Position initiale - Emprunt - 23/10/2015 - Les Echos - LAURENCE BOCCARA - Lien -
Contexte : "Il existe trois formes de « crowdfunding » qui cohabitent et lèvent , chaque année , de plus en plus de fonds . La version originelle propose aux internautes (les « crowdfunders ») d'effectuer des dons (avec ou sans **contrepartie**) en ligne pour soutenir des **projets** d'ordre culturel , artistique ou associatif . Plus récente , la seconde forme concerne des **prêts** avec **intérêt** à **destination** des **entreprises** , afin de **financer** leur **développement** ou de réaliser des **opérations** particulières . C' est le « **crowdlending** » . Enfin , troisième et dernière déclinaison possible : entrer dans le **capital** d' une PME non **cotée** en détenant des **actions** . C' est ce que l' on appelle le « private equity » . "

- Nom Commun - **ECONOMIE-Finance** | **ECONOMIE-Commerce** | **SOCIÉTÉ-Travail et Métiers** - Position médiale - Emprunt - 06/11/2015 - Les Echos - LAURENCE BOCCARA - Lien -
Contexte : "Les taux des **prêts** proposés par ce site évoluent entre 3 , 30 et 6 , 70 % pour des dossiers **remboursables** entre 12 et 60 mois . Jusqu' au 30 novembre prochain , date du lancement de la COP21 à Paris , Lendosphere propose la plus importante **opération** de « **crowdlending** » lancée en France . C' est un **projet** de la **société** Valorem , spécialisée dans les parcs éoliens participatifs . Pour sa troisième **opération** en « crowdfunding » organisée avec ce site , cette **entreprise** girondine espère collecter 500.000 **euros** (110.000 **euros** ont déjà été prêtés) . Elle s' engage à **rembourser** , d' ici à cinq ans , la **somme** avec 5 % d' **intérêts** annuels avec une bonification de 7 % pour les prêteurs riverains de ce parc situé à Saint-Michel-Chef-Chef en Loire-Atlantique . "

- Nom Commun - **SOCIÉTÉ-Travail et Métiers** | **LOISIRS-Tourisme et Voyages** | **ECONOMIE-Finance** - Position finale - Emprunt - 16/11/2015 - Les Echos - GERALDINE DAUVERGNE - Lien -
Contexte : "Paris 17e . Jusqu' alors dédié aux **sociétés cotées** , le salon Actionaria consacre pour la troisième année consécutive un **espace** aux start-ups et PME de **croissance** à la recherche d' **investisseurs** privés . Leader du **secteur** , c' est la plateforme SmartAngels qui encadrera et animera cet **espace** . Chaque **dirigeant** aura quelques minutes pour « pitcher » et tenter de séduire des **investisseurs** . Les visiteurs pourront également se présenter sur leurs stands pour leurs poser des questions et découvrir les meilleurs moyens de participer à leur **financement** . 3 conférences pour découvrir le crowdfunding : crowdequity , crowdlending : deux solutions pour soutenir les PME et les start-up ; investir dans le capital via le crowdfunding ; Dynamiser votre

The image features a stylized logo for 'Britpop 90s'. The word 'britpop' is written in a large, white, lowercase, sans-serif font with a slight slant, set against a black rectangular background. Below it, the word '90s' is written in a smaller, red, lowercase, sans-serif font, also with a slight slant, set against a smaller black rectangular background. The entire logo is centered on a background of red and blue splatters and textures.

britpop

90s

Larousse dictionary, may 2017 !

	POS	Meta-category of creation
macronisation	NAME	MORSEM
a-dô-re	VERB	MORPH

saves time

avoids mistakes

+ a bit pointless



Accueil

Présentation

Recherche

Aides

Contact

Recherche

Regular expressions

Néologisme : anti.*

Astuces



recherches avancées

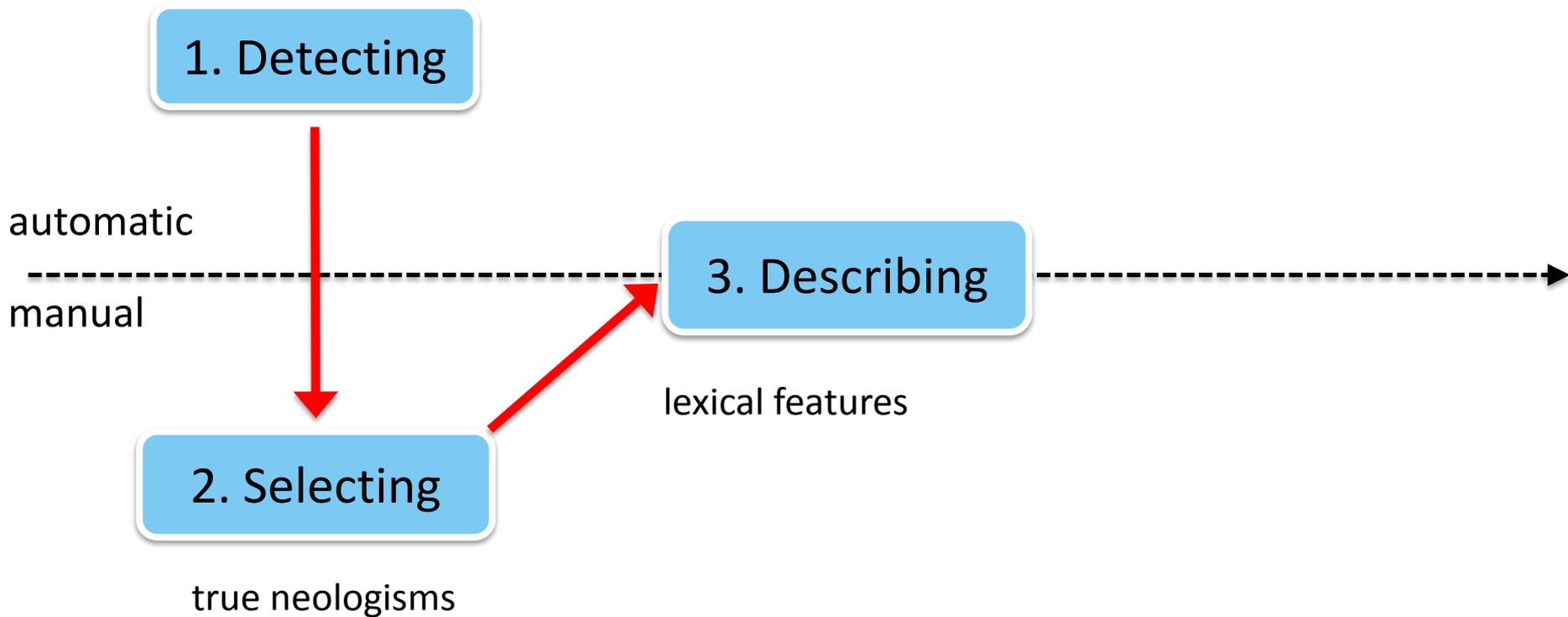
Filtrer par :

Catégories :

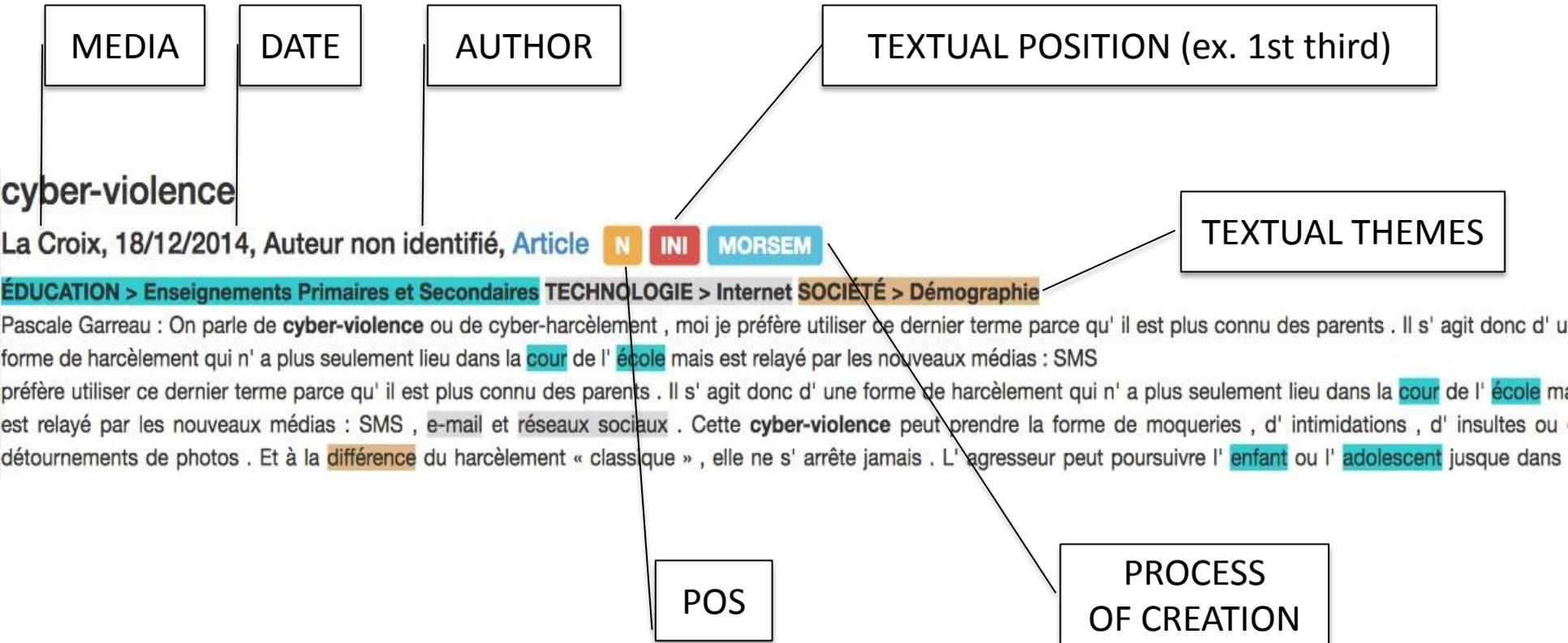
sélection(s)



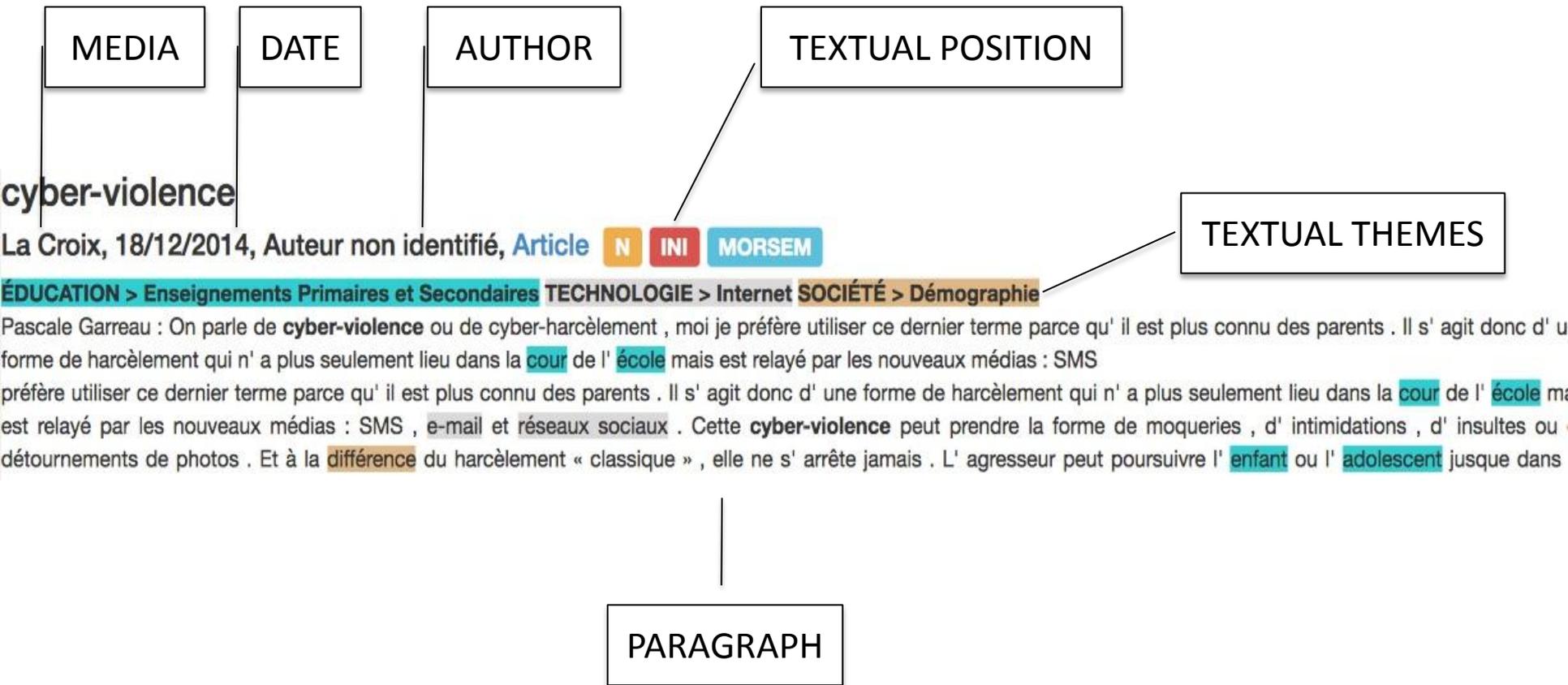
4 stages process



Contextual features



Contextual features



Contextual features

Why are contextual features so important?

How do we manage to implement this?

What remains to be done (technically)?

Contexts diversity and its impact on lexical creativity.

- Schröder, Marianne (1978): « Zum Anteil von Wortbildungskonstruktionen an der Konstitution von Texten »
- Polenz, Peter Von (1980): « Wortbildung ».
- Dederding, Hans-Martin (1983): « Wortbildung und Text (Zur Textfunktion von Nominalkomposita) »
- Erben, Johannes (1995): « Wortbildung und Textbildung »
- Peschel, Corinna (2002): *Zum Zusammenhang zwischen Wortneubildung und Textkonstitution.*
- Elsen, Hilke (2007): « Die Wortbildung der Eigennamen in fiktionalen Texten »

Lexical creation and productivity

Biber (1988) [...] provided detailed evidence that the linguistic resources employed **in speech** differ from those **in writing**, and that within each of these communicative modalities, further systematic differences differentiate the **styles** of more specific **text types**.

(**Baayen**: 2009)

Lexical creation and productivity

Substantial variability:

- Author
- Style
- Register
- Domain
- Readership

(Baayen: **2001**)

Lexical creation and productivity

Plag, Ingo/Dalton-Puffer, Cristiane/Baayen, R. Harald (**1999**),
« Productivity and **register**. », *English Language and Linguistics*, 3, 209-228.

Plag, I., Dalton-Puffer, C., and Baayen, R. H. (**1999**).
« Morphological productivity across **speech** and **writing**. », *Journal of English Language and Linguistics*, 3, 209-228.

Neology and Types of Contexts

Type of context	<i>Exemple</i>
languages	<i>french (XVIIth c.)</i>

Neologie and Types of Contexts

Type of context	<i>Exemple</i>
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>

- Sufficient: one level of observation

Tab. 17: Möglichkeiten der Wortschatzerweiterung im Überblick, in %

Korpus Wortbildungstyp	Chemie	Technik	Linguistik	Jugend	Werbung	Zeitung	Literatur	Kinderlit.
Komposition	37,62	76,24	53,33	17,1	61,64	80,94	64,43	61,38
<i>davon Sonderkomp.</i>	1,19	1,24	0,42	0,4	3,68	1,77	1,11	1,05
Wortgr.lexembildung	13,07	20,96	14,57	0,2	9,63	3,34	0,37	1,8
Derivation	8,12	0,62	16,12	27,24	4,2	5,11	12,22	13,32
<i>davon Zusanbildung</i>	0	0	0	0,6	0	0,59	0,93	1,2
<i>abgel. Rückbildung*</i>	0	0	0	0,4	0	0,59	0,74	0
<i>abgel. Kürzung*</i>	0	0	0	4,17	0	0	0	0
Präfixoidbildung	0	0	0	5,17	0,35	0	1,3	1,95
Suffixoidbildung	0	0	0	1,19	0,18	0	0,19	0
Konversion	0,2	0	1,98	6,76	1,23	5,11	16,11	3,74
Kurzwortbildung	10,69	0,31	2,4	6,56	0,18	1,18	0,19	0,15
Rückbildung	0	0	0	1,39	0	0,59	0,93	0
Zusammenrückung	0	0	0	0	0	0,2	2,22	1,95
Bedeutungsveränd.	1,39	0,78	3,54	28,03	1,4	2,55	1,48	2,54
Fremdw.übernahme	4,36	1,09	7,78	6,96	21,19	0,98	0,37	0,45
Kontamination	6,73	0	0,28	1,39	0	0,39	0,74	4,94
Kunstwortbildung	17,03	0	0	0	0	0,2	0,19	7,63
Verben in neu. synt. Str.	0	0	0	1,59	0	0	0	0
Sondermuster	0,79	0	0	0,99	0	0	0	0,15

*Abgeleitete Rückbildungen und Kurzwörter sind zweimal gerechnet.

Elsen 2004

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>

- Sufficient: one level of observation
- Unsufficient: very global (abstract) level of observation
- Unsufficient: one level among others

Neologie and Types of Contexts

Type of context	<i>Exemple</i>
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>
social group	<i>spanish homosexuals (anglicisms)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>
social group	<i>spanish homosexuals (anglicisms)</i>
editorial tone	<i>the Guardian</i> <i>the Times (Baayen/Renouf 1996)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>
social group	<i>spanish homosexuals (anglicisms)</i>
editorial tone	<i>the Guardian</i>
genre of discourse	<i>slam poetry</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>
social group	<i>spanish homosexuals (anglicisms)</i>
editorial tone	<i>the Guardian</i>
genre of discourse	<i>slam poetry</i>
individual style	<i>John Donne (complete works, XVIIth c.)</i>

Neologie and Types of Contexts

Type of context	Exemple
languages	<i>french (XVIIth c.)</i>
domain	<i>medecine (terminology)</i>
theme/topic	<i>terrorism (« cyber-jihad »)</i>
social group	<i>spanish homosexuals (anglicisms)</i>
editorial tone	<i>the Guardian</i>
genre of discourse	<i>slam poetry</i>
individual style	<i>John Donne (complete works, XVIIth c.)</i>
text	<i>« Jabberwocky » (Carroll, 1971)</i>

To sum up:

Neology studies are very **diverse**

Reflects the **empirical** diversity of contexts

In all societies, **various forms/levels** of communication

To sum up:

Neology studies should **more** focus on **contextual diversity**

Each neologism **only exist** in a particular context

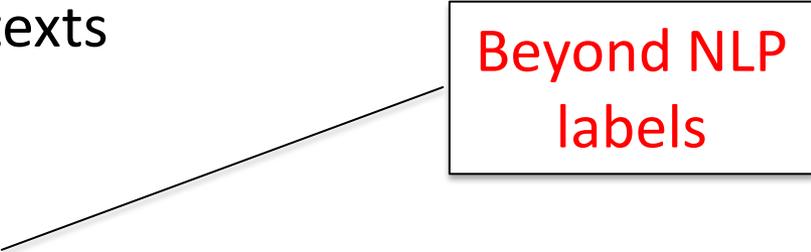
A neologism is never a pure **form**:

A neologism is a form which only **functions in relation to one or many particular context(s)**.

We need:

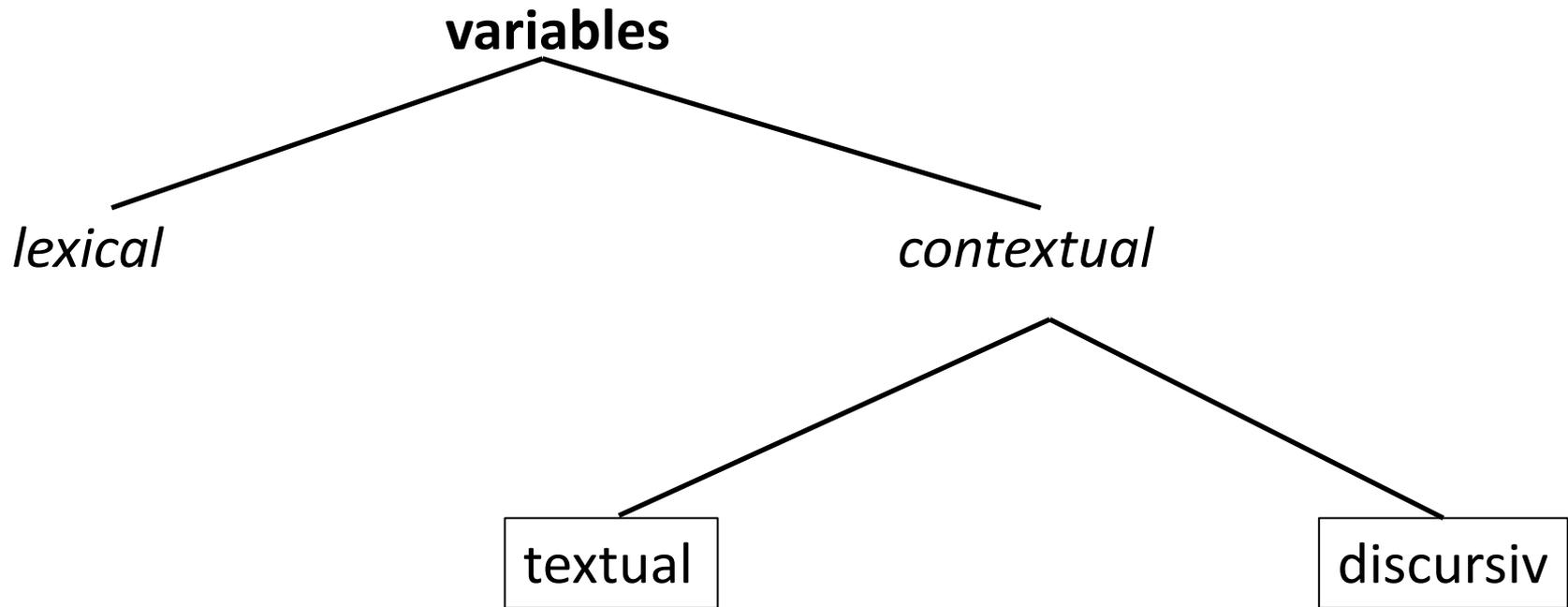
Precise & adapted typology of contexts

Beyond NLP
labels



Clear **conception: creativity-functionality-context**

- Key for interpreting statistics
- Key for corpus building



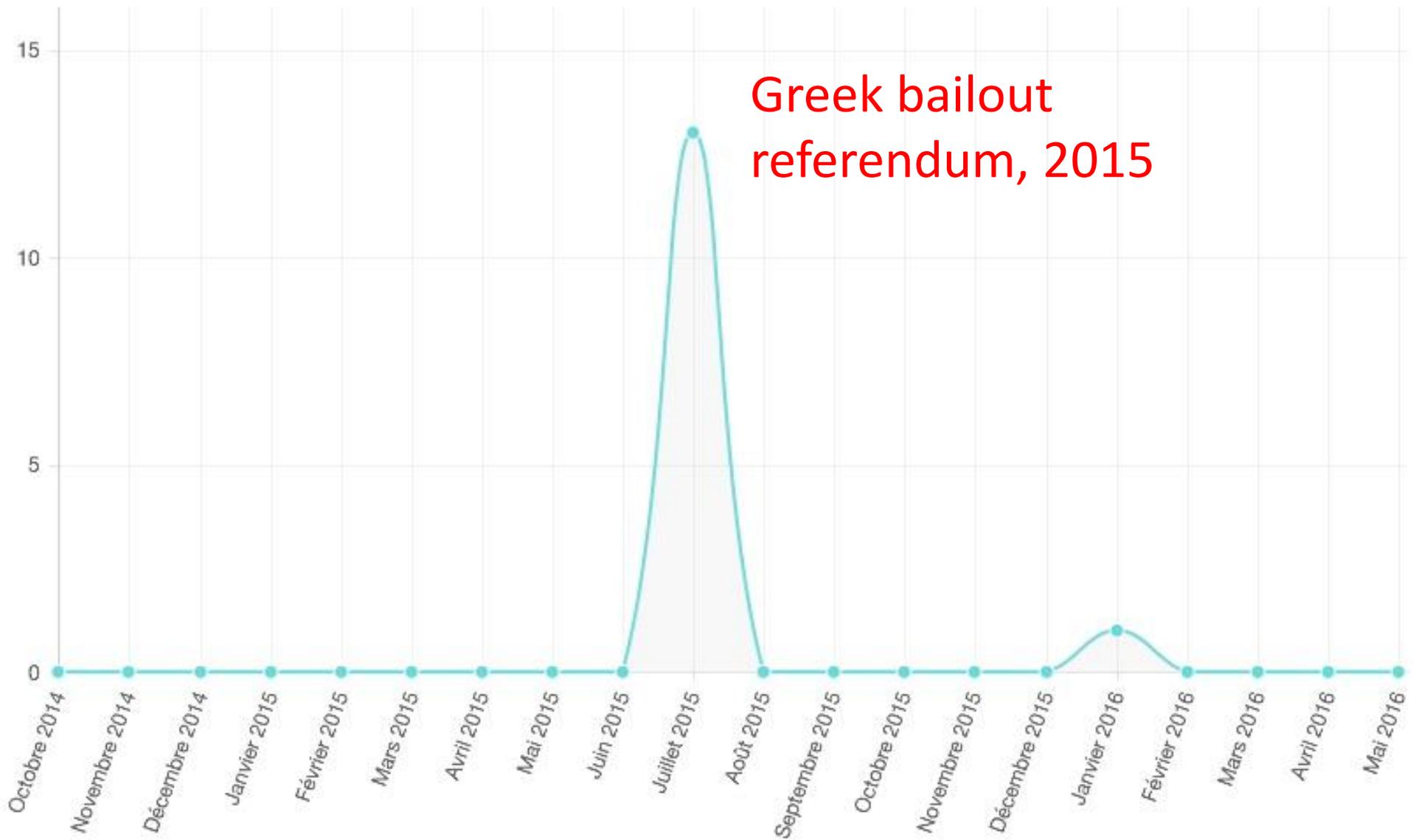
- medium (written/spoken)
- paragraph
- quotemarks & glosses
- titel & position
- textual themes

- date
- author / readership
- genre of discours
- generic field
- collectiv style

Documenting the date.

Context = historical event

Greek bailout referendum, 2015



« oxi »: no(greek)

Documenting paragraphs and themes.

Context = text et content (sense)

Documenting the **paragraph**.

Context = text



crowdlending

Sommaire [masquer]

- [+] Français
- [+] Anglais

MEANING MATTERS!
EXPLAIN / UNDERSTAND

Français [modifier le wikicode]

Étymologie [modifier le wikicode]

Emprunté à l'anglais *crowdlending*.

Nom commun [modifier le wikicode]

crowdlending \kʁawd.lɑ̃.diŋ\ masculin

1. (*Anglicisme*) (*Finance*) Financement participatif sous forme de prêt de particuliers.

- Il existe trois formes de « crowdfunding » qui cohabitent et lèvent, chaque année, de plus en plus de fonds. La version originelle propose aux internautes (les « crowdfunders ») d'effectuer des dons (avec ou sans contrepartie) en ligne pour soutenir des projets d'ordre culturel, artistique ou associatif. Plus récente, la seconde forme concerne des prêts avec intérêt à destination des entreprises, afin de financer leur développement ou de réaliser des opérations particulières. C'est le « **crowdlending** ». — (Laurence BOCCARA, lesechos.fr, 23 octobre 2016)

Invariable
crowdlending
\kʁawd.lɑ̃.diŋ\

Synonymes [modifier le wikicode]

- prêt participatif
- financement participatif par prêt

Vocabulaire apparenté par le sens [modifier le wikicode]

- crowdfunding

Traductions [modifier le wikicode]

- Anglais : crowdlending ^(en)

Ajouter une traduction en français : traduction [Ajouter] Plus

Références [modifier le wikicode]

- <https://www.ces-berra-immo.fr/crowdfunding-immobilier-qu-est-ce-que-le-crowdlending-pret-definition-1054.html>
- Logoscope, Université de Strasbourg, 2016 (crowdlending)

- Page d'accueil
- Recherche avancée
- Portails thématiques
- Page au hasard
- Page au hasard par langue
- Poser une question

- Contribuer
- Journal des contributeurs
- La Wikidémie
- Communauté
- Discuter sur IRC
- Modifications récentes
- Faire un don

- Aide
- Aide
- Modèles
- Conventions
- Créer un article
- Télécharger

- Imprimer / exporter
- Créer un livre
- Télécharger comme PDF
- Version imprimable

- Outils
- Pages liées
- Suivi des pages liées
- Importer un fichier
- Pages spéciales
- Adresse permanente
- Information sur la page
- Citer cette page

- Dans d'autres langues

Detected neologism: 90% hapax legomena

Hapax legomena: **textual/rhetorical function**

PESCHEL (2002) : *Zum Zusammenhang zwischen Wortneubildung und Textkonstitution*

OMROD (2004) : « Creation and subsequent usage of terms for discourse purposes in the scientific research article »

GERARD (2011) : « Création lexicale, sens et textualité: théories et analyses ».

GERARD (2014) : « Sémiotique interprétative des créations verbales ».

« Gabriel, le personnage pivot de "l'Exposition coloniale", narrateur désolé de sa vie, est un homme **caoutchouc** ou du moins qui a la maladie du **caoutchouc**, qui ne cesse de rebondir comme une balle, [...] et dont la vie privée est aussi **caoutchouteuse** que possible, entre son père loufdingue qui rêve des colonies, ses deux épouses entre lesquelles il **caoutchouchoutera** toute sa vie [...] ». (*Nouvel Observateur*, 02/09/1988).

« caoutchouc » = natural rubber

Documenting textual themes.

Context = content...

1. EDUCATION

2. TECHNOLOGIE

3. SOCIETY

cyber-violence

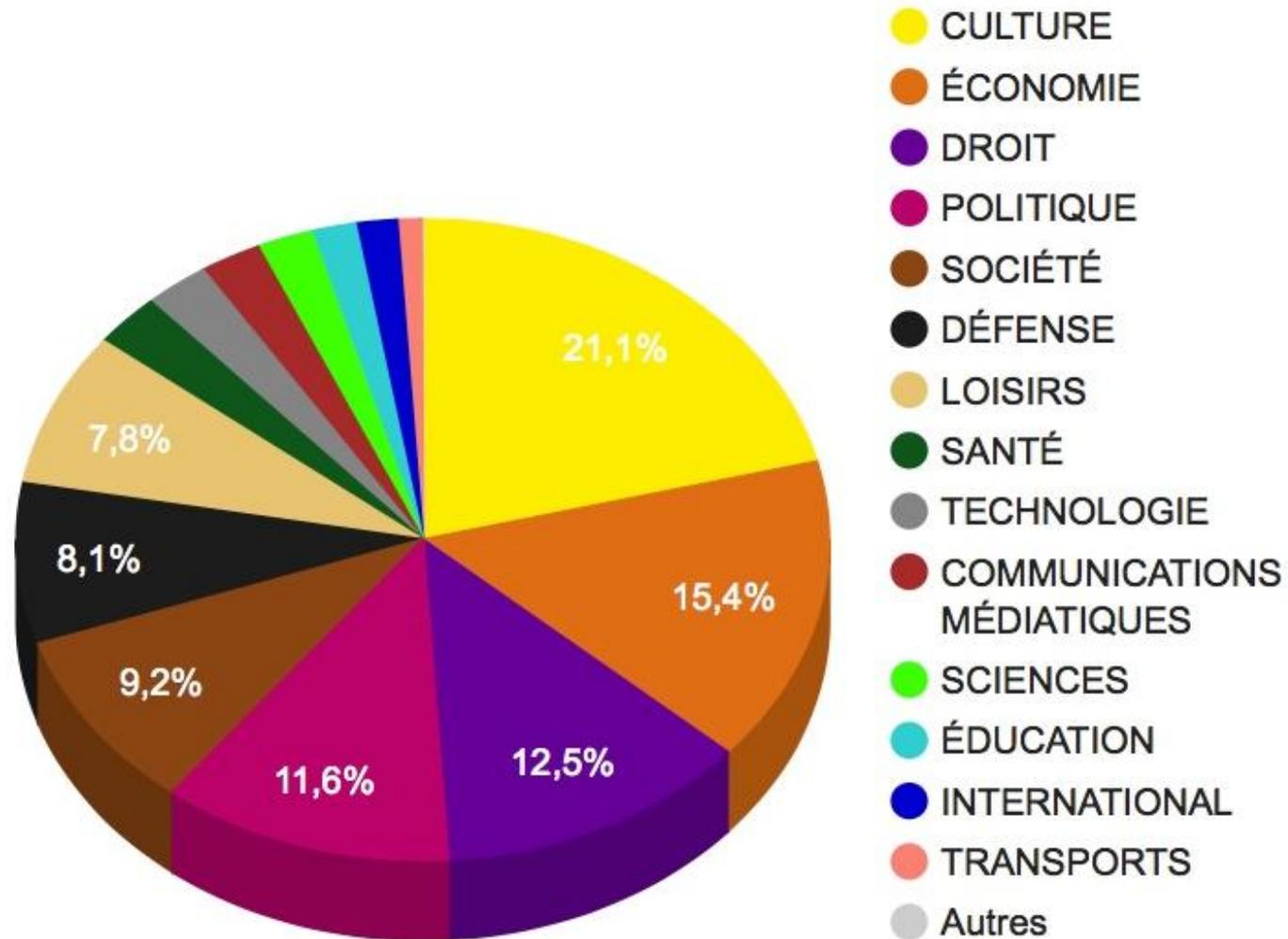
La Croix, 18/12/2014, Auteur non identifié, Article N IM MORSEM

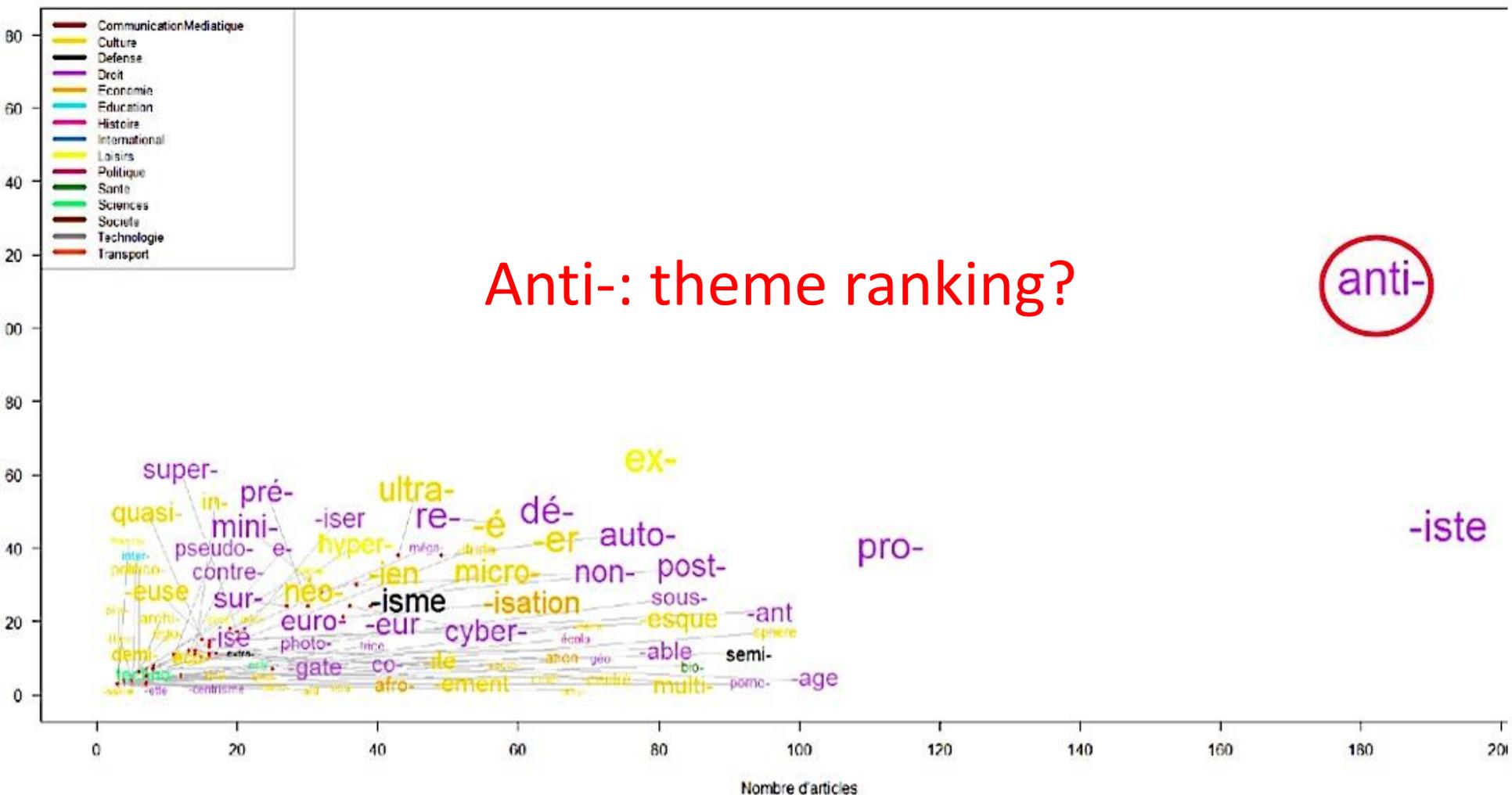
ÉDUCATION > Enseignements Primaires et Secondaires TECHNOLOGIE > Internet SOCIÉTÉ > Démographie

Pascale Garreau : On parle de **cyber-violence** ou de cyber-harcèlement , moi je préfère utiliser ce dernier terme parce qu' il est plus connu des parents . Il s' agit donc d' une forme de harcèlement qui n' a plus seulement lieu dans la **cour** de l' **école** mais est relayé par les nouveaux médias : SMS
préfère utiliser ce dernier terme parce qu' il est plus connu des parents . Il s' agit donc d' une forme de harcèlement qui n' a plus seulement lieu dans la **cour** de l' **école** mais est relayé par les nouveaux médias : SMS , **e-mail** et **réseaux sociaux** . Cette **cyber-violence** peut prendre la forme de moqueries , d' intimidations , d' insultes ou de détournements de photos . Et à la **différence** du harcèlement « classique » , elle ne s' arrête jamais . L' agresseur peut poursuivre l' **enfant** ou l' **adolescent** jusque dans

Text: multi-thematic unit

French national press





POLITIQUE > Etat et Gouvernement

247

DROIT > Justice et Législation

232

DÉFENSE > Militaire

151

DÉFENSE > Conflits Armés Contemporains

104

POLITIQUE > Election

72

SOCIÉTÉ > Conflits Sociaux

9

DÉFENSE > Sécurité Civile et Catastrophes

7

DROIT > Prison

7

ÉDUCATION > Enseignements Primaires et

7

ÉDUCATION > Enseignements Supérieurs

7

Top of the list

Bottom of the list

Author's style & readership.

Give
liberties / constraints
on lexical creativity.

Neologenic / neolophoric

	+ néologène	- néologène
+ néolophore	H. Barbusse <i>(Le feu, 1916)</i>	M. Genevoix <i>(Ceux de 14, 1916-23)</i>
- néolophore	M. Boasson <i>(Au soir d'un monde, 1926)</i>	G. Duhamel boche, poilu, cagna, singe, cagna crapouillot, guitoune.

Gérard/Lacoste (2017) : novelists (WW1)

Neologenic / neolophoric

Shakespeare:

neologiser = creator

Thomas Elyot:

neologiser = adopter

HUFFPOST

- Most neologenic (Néoveille)
- Unpaid bloggers (less formal style)
- Left-wing authors (« *Pegidoof* »)
- « Borrow » informations...

Differences, similarities?



women's press

Kinds of readership

gender / sexuality (**women**/men, feminist, gay, hot...)

Kinds of readership

gender / sexuality (**women**/men, feminist, gay, hot...)

education (working-class, middle-class...)

age (child, **teen**, elderly...)

politics (left-wing, right-wing...)

religion (protestant, islamic, catholic...)

Collectiv style, generic field and genre.

Give
liberties / constraints
on lexical creativity.

Differences, similarities?



Editorial tone:

« glamour and investigation »



Editorial tone:

« Mode, beauty, trends, celebrity, love »

No. 1443
5 May –
18 May 2017
£1.80

PRIVATE EYE

ELECTION 2017

IT'S THERESA MAYNIA!





Ordre alphabétique

Fréquence

Chronologie

Journal

Catégorie gram.

Thème

Procédé

Position

Rechercher

Logoscope

Documentation quotidienne des nouveaux mots français

Projet et Équipe

Guide d'utilisation

Publications

Mentions légales

Obtenir la version experte

lilpa
linguistique, langues, parole

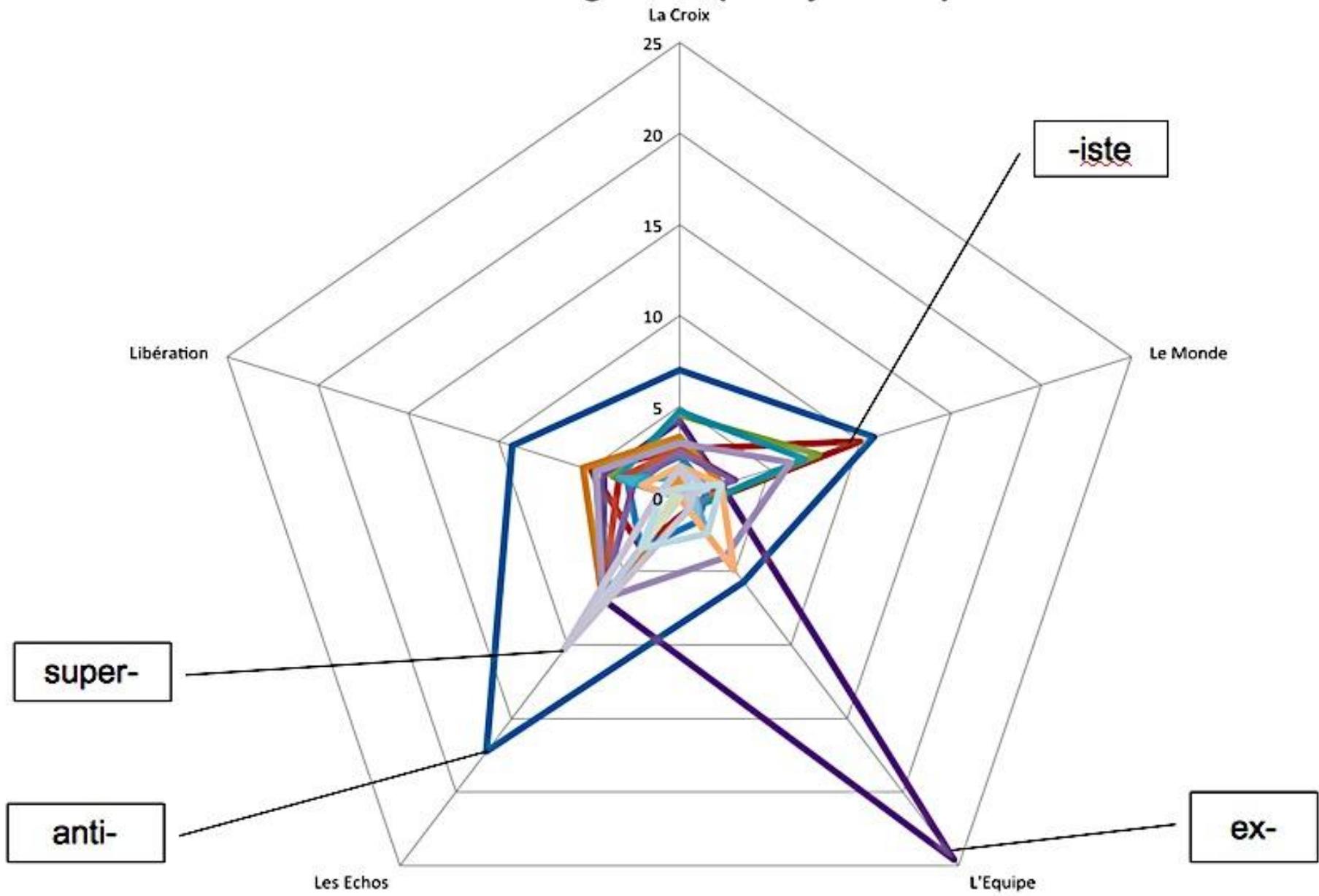
UNIVERSITÉ DE STRASBOURG



FranceTerme



usages en % (mars-juin 2015)



Global
constraints

Collectiv style

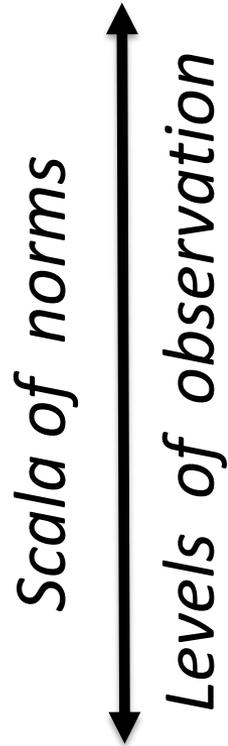
?

?

Author's style / Readership

Local
constraints

Textual cohesion/progression



Global
constraints

Collectiv style

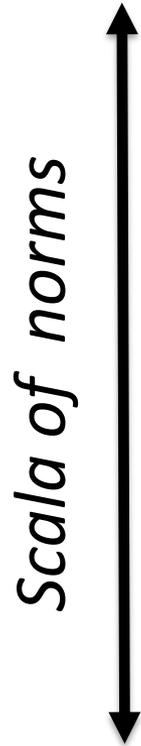
Generic field

Genre of discourse

Author's style / Readership

Local
constraints

Textual cohesion/progression





INFORMATION

Generic field:

Group of genres which share
a communicational function,
inside a particular domain.

Ulrich Gross (2001) : fields and genres / François Rastier (2001) : « champ générique »

...
Hard/soft news
Notice
Report
...

INFORMATION

...
Analysis
Interview
...

**INTERMEDIATE
FIELD**

...
Editorial
Chronicle
Review
...

OPINION

...
Horoscope
Recipe
...

ADVICE

...
Crossword
Sudoku
...

ENTERTAINMENT

Ulrich Gross (2001) : fields and genres / François Rastier (2001) : « champ générique »

...
Hard/soft news
Notice
Report
...

INFORMATION

...
Analysis
Interview
...

**INTERMEDIATE
FIELD**

**Most
neologenic
press field**

OPINION

...
Horoscope
Recipe
...

ADVICE

...
Crossword
Sudoku
...

ENTERTAINMENT

Ulrich Gross (2001) : fields and genres / François Rastier (2001) : « champ générique »

Global
constraints

Collectiv style

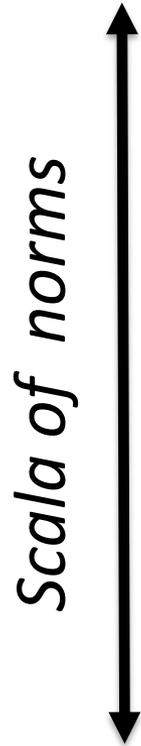
Generic field

Genre of discourse (*novel, CV,...*)

Author's style / Readership

Local
constraints

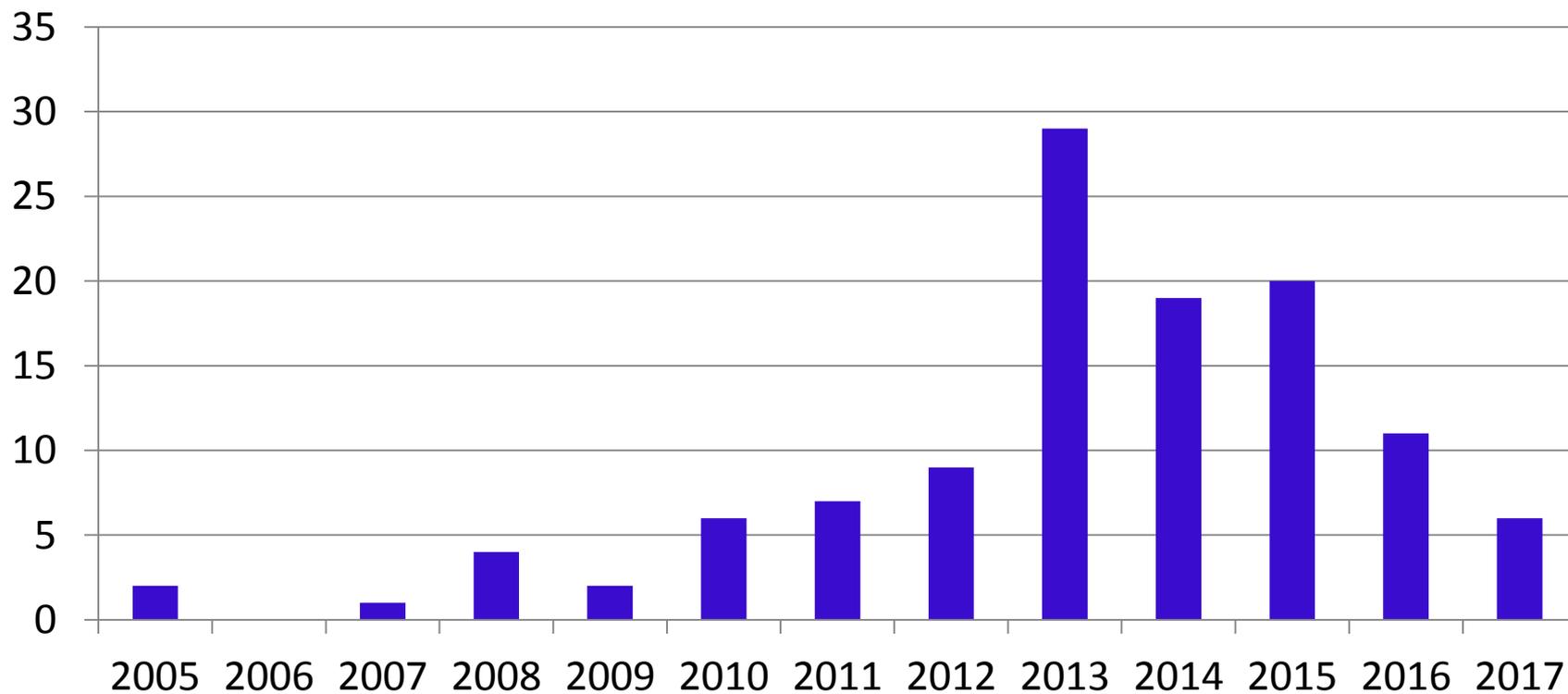
Textual cohesion/progression



über-masculin
über-glam
über-tacticiens
über-trendy
über-kawaii
überdrogué
über-virile
über
über
über
über
über
über-crédible
über-grossièreté
über-sexy
über-dude
über-femmes
über-modeux
über-chic

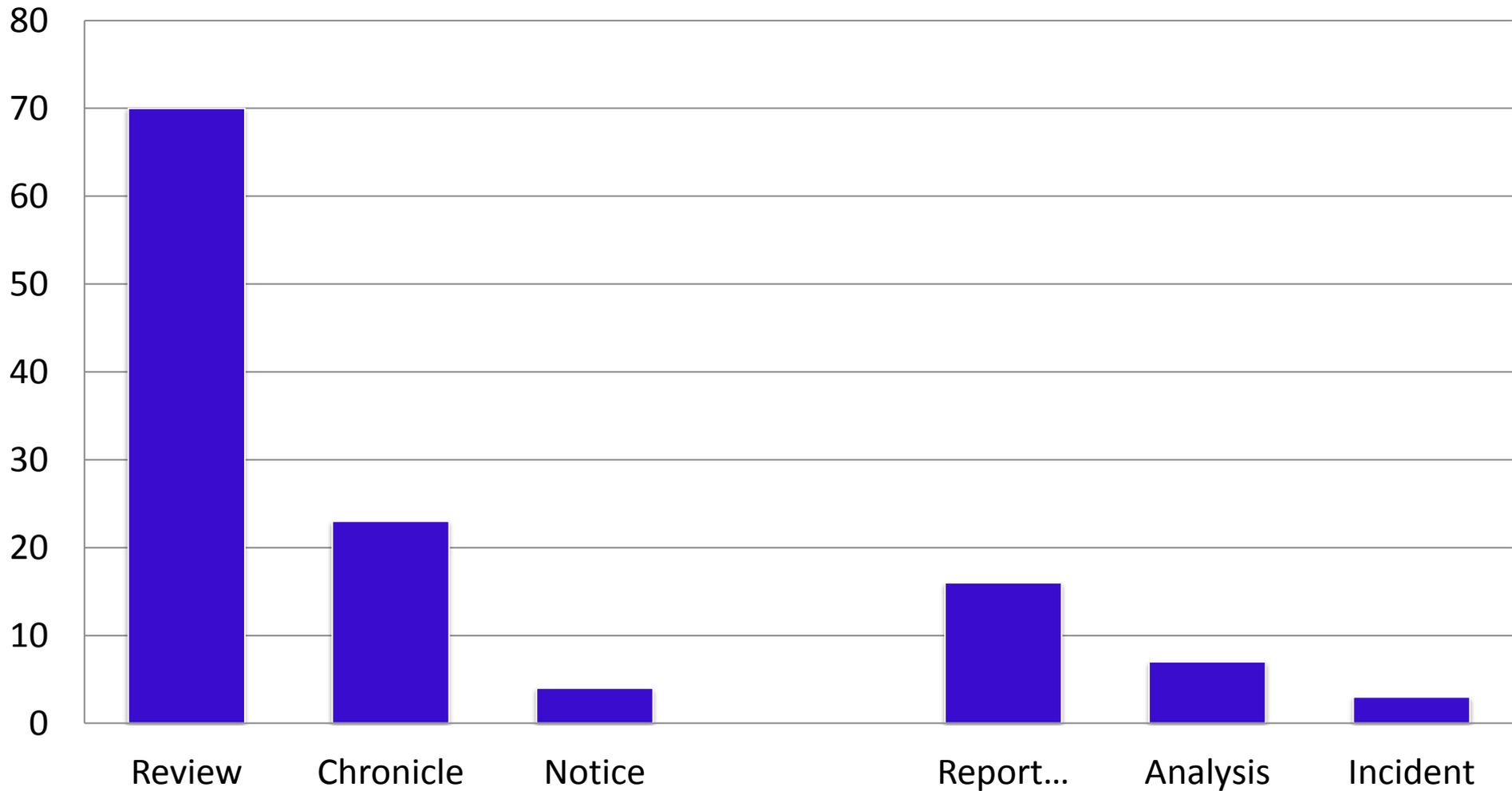
über-riches
über-médiatique
über-seniors
über-cancer
über-féminine
über-catchy
über-poule
über-talentueux
über-information
über
über-local
über-bisou
über-cuisinier
über-luxe
über-sexuel

Where? Which texts allow this kind of creation?



Use of **über-** in french newspapers
(114 forms / base Factiva)

Use of über- / genre (Factiva : 2005-2017)



blockchain
anti-daech
alsace-lorraine-champagne-ardenne
startupper
uberisation
hauts-de-france
alsace-lorraine-champagne
aquitaine-limousin-poitou
zadistes
ex-sncm
crowdlending
pro-brexit

« Op-ed »	Interview	Review	Editorial	Recipe
zadiste blockchain adblockers fintech(s) uberisé(e)(s) hollandisme macronisme youtubeur startupper(s) convivialiste sans-dents Hollandisé hollandienne convivialiste immeuble-	zadiste(s) blockchain brexit adblockers Fintech uberise uberisation ex-vins padel dieselgate fablab	— blockchain brexit tweetos youtubeuse zika	zadistes — brexit fintech hollandisme pseudo-califat	bibimbap combawa

Neologenic? Chronicle or Announcement of death?

	+ néologène	- néologène
+ néolophore	<p><i>chronique</i> / <i>billet d'humeur</i></p> <p><i>critique</i></p> <p><i>éditorial</i></p> <p><i>courrier des lecteurs</i></p> <p><i>tribune libre</i></p> <p><i>reportage</i></p>	<p><i>brève / compte rendu / agenda</i></p> <p><i>analyse</i></p> <p><i>interview dialoguée</i></p> <p><i>recette</i></p> <p><i>mots-croisés</i></p> <p><i>petites annonces / offres d'emploi</i></p>
- néolophore	<p><i>marronnier</i></p> <p><i>nécrologie</i></p> <p><i>portrait</i></p>	<p><i>avis d'obsèques</i></p> <p><i>horoscope</i></p> <p><i>bulletins (circulation / météo)</i></p> <p><i>faits-divers</i></p>

Global
constraints

Collectiv style

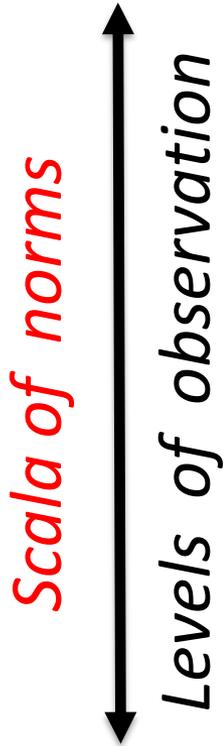
Generic field

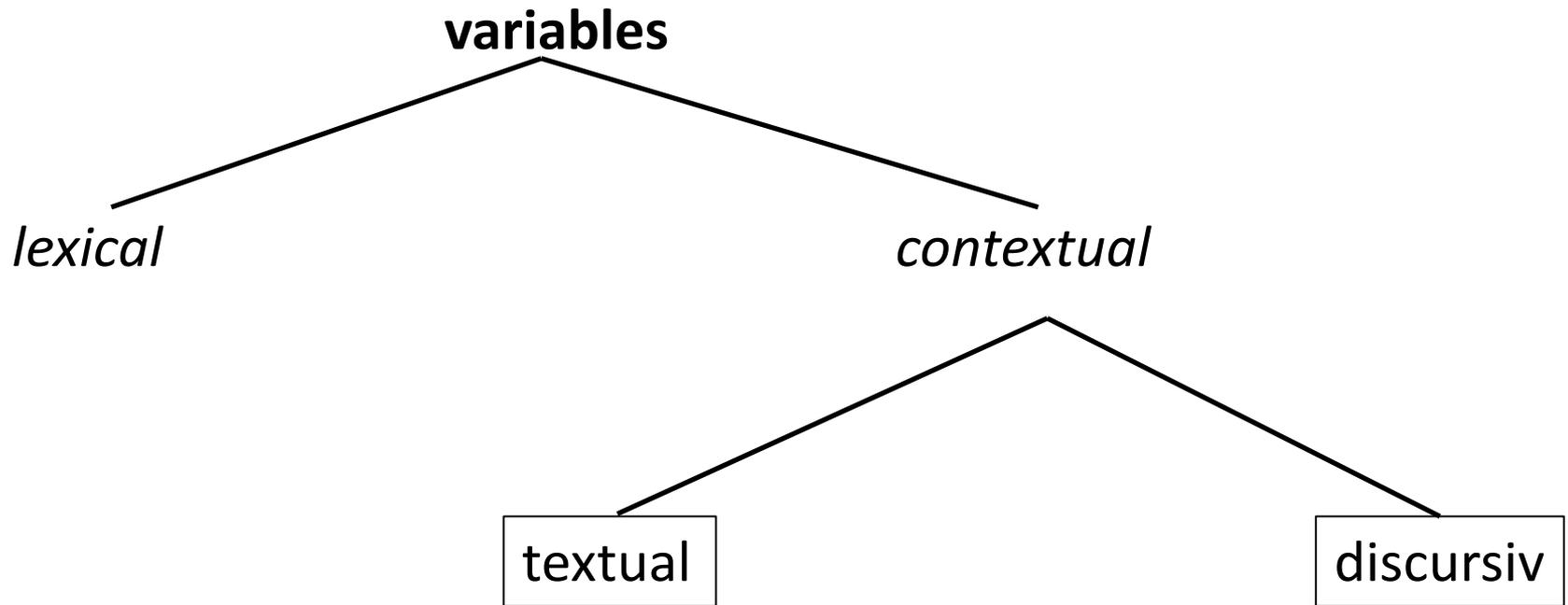
Genre of discourse

Author's style / Readership

Local
constraints

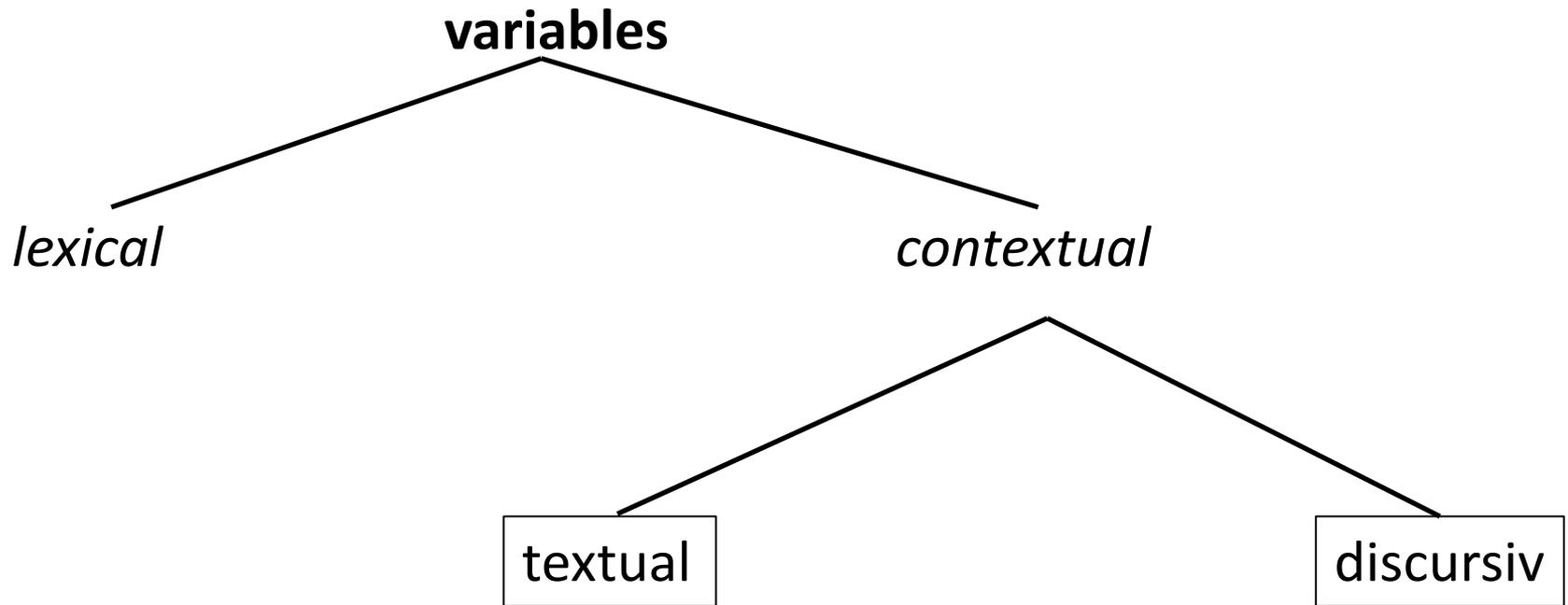
Textual cohesion/progression





- medium (written/spoken)
- paragraph
- quotemarks & glosses
- **titel** & position
- textual themes

- date
- author / **readership**
- **genre of discours**
- **generic field**
- editorial tone



- medium (written/spoken)
- paragraph
- quotemarks & glosses
- titel & position
- **textual themes**

- date
- author / readership
- **genre of discours**
- **generic field**
- editorial tone

3

Detecting Automatically Texual Themes, Genres and Generic Fields.

1. EDUCATION

2. TECHNOLOGIE

3. SOCIETY

cyber-violence

La Croix, 18/12/2014, Auteur non identifié, Article **N** **IM** **MORSEM**

ÉDUCATION > Enseignements Primaires et Secondaires **TECHNOLOGIE** > Internet **SOCIÉTÉ** > Démographie

Pascale Garreau : On parle de **cyber-violence** ou de cyber-harcèlement , moi je préfère utiliser ce dernier terme parce qu' il est plus connu des parents . Il s' agit donc d' une forme de harcèlement qui n' a plus seulement lieu dans la **cour** de l' **école** mais est relayé par les nouveaux médias : SMS
préfère utiliser ce dernier terme parce qu' il est plus connu des parents . Il s' agit donc d' une forme de harcèlement qui n' a plus seulement lieu dans la **cour** de l' **école** mais est relayé par les nouveaux médias : SMS , e-mail et réseaux sociaux . Cette **cyber-violence** peut prendre la forme de moqueries , d' intimidations , d' insultes ou de détournements de photos . Et à la **différence** du harcèlement « classique » , elle ne s' arrête jamais . L' agresseur peut poursuivre l' **enfant** ou l' **adolescent** jusque dans

Text: multi-thematic unit

ANALYSE

PORTRAIT Spécialiste de la guerre 1914-1918, cet historien revendique une identité française mouvante et souhaite des débats citoyens.

Canada, Chine, Portugal, Italie, Allemagne... ce n'est pas la tournée d'une star internationale, mais les voyages d'un

historien, spécialiste de la Première Guerre mondiale en cette année de commémorations de la Grande Guerre. Nicolas

Offenstadt, 47 ans, maître de conférences à l'université Paris-I, arpente colloques, expositions et manifestations depuis le

début de 2014. Boulimique de travail, obsessionnel, monomaniacque, l'historien à lunettes et au sourire malicieux a la

mémoire dans la peau. Le hardi est de toutes les batailles et sur tous les fronts.

Le Front populaire. L'universitaire défend la position d'un «historien dans la cité» et prône une histoire «hors les murs»

s'appuyant sur les lieux et à destination de tous les publics. Près de l'église Saint-Eustache, on a droit à une démonstration :

«Nous sommes au coeur du Paris médiéval, ça ne se voit pas, mais les lieux sont encore habités. On peut, ici, reconstituer

le Paris des Armagnacs et Bourguignons au XVe siècle.» Le centenaire de la Grande Guerre est pour lui l'occasion de

répondre à une demande sociale forte, surtout pour cette période de l'histoire dont les Français sont friands. «Les

historiens doivent investir l'espace public, mais à condition d'abaisser la chaire et d'avoir un peu de modestie dans la

parole.» Abaisser la chaire d'accord, mais il tient tout de même à faire valoir son statut d'intellectuel «spécifique», qui

intervient dans ces domaines de compétences (1914-1918, le Moyen Age et l'historiographie). Une pierre dans le jardin des

vulgarisateurs, Lorànt Deutsch et Stéphane Bern, qu'il épingle régulièrement sur Twitter. Loin de voir en eux une porte

d'entrée populaire, il estime qu'ils sont nocifs : «Derrière leur posture de conteurs, ces histrions réactionnaires sont

beaucoup plus idéologiques qu'ils ne le prétendent. L'histoire est une matière pour réfléchir ensemble et non un récit figé

du roman national.»

La guerre de cinq ans. Dès 2007, contre Nicolas Sarkozy et son «effrayant usage identitaire de la France», l'historien est

80 themes

CommunicationMediatique-Edition.txt
CommunicationMediatique-Internet.txt
CommunicationMediatique-Presse.txt
CommunicationMediatique-Radio.txt
CommunicationMediatique-Television.txt
Culture-ArtsPlastiques.txt
Culture-Cinema.txt
Culture-Danse.txt
Culture-Festivals.txt
Culture-Gastronomie.txt
Culture-Litterature.txt
Culture-Mode.txt
Culture-Musique.txt
Culture-Scene.txt
DefenseMiliatire- ConflitsArmes.txt
DefenseMilitaire-Militaire.txt
Droit-Justice.txt
Droit-Legislation.txt
Droit-Prison.txt
Economie-Agriculture.txt
Economie-Commerce.txt
Economie-Finance.txt
Economie-Industrie.txt
Education-EnseignementPrimaireSecondaire.txt
Education-EnseignementSuperieure.txt
Histoire-Archeologie.txt
Histoire-GuerresHistoriques.txt

CommunicationMediatique-Edition.txt
CommunicationMediatique-Internet.txt
CommunicationMediatique-Presse.txt
CommunicationMediatique-Radio.txt
CommunicationMediatique-Television.txt
Culture-ArtsPlastiques.txt
Culture-Cinema.txt
Culture-Danse.txt
Culture-Festivals.txt
Culture-Gastronomie.txt
Culture-Litterature.txt
Culture-Mode.txt
Culture-Musique.txt
Culture-Scene.txt
DefenseMiliatire- ConflitsArmes.txt
DefenseMilitaire-Militaire.txt
Droit-Justice.txt
Droit-Legislation.txt
Droit-Prison.txt
Economie-Agriculture.txt
Economie-Commerce.txt
Economie-Finance.txt
Economie-Industrie.txt
Education-EnseignementPrimaireSecondaire.txt
Education-EnseignementSuperieure.txt
Histoire-Archeologie.txt
Histoire-GuerresHistoriques.txt

International-Afrique.txt
International-Ameriques.txt
International-Asie.txt
International-Europe.txt
International-MoyenOrient.txt
International-Orient.txt
Loisirs-Jeux.txt
Loisirs-Sport.txt
Loisirs-Tourisme.txt
Medical-Addictions.txt
Medical-Ethique
Medical-Medecine.txt
Politique-Election.txt
Politique-EtatGouvernement.txt
Politique-Immigration.txt
Politique-RegimePolitique.txt
Sciences-Astronomie.txt
Sciences-Biologie.txt
Sciences-Biosphere.txt
Sciences-CatastrophesNaturelles.txt
Sciences-PhysiqueChimie.txt
Societe-ConflitsSociaux.txt
Societe-Discrimination.txt
Societe-Ecologisme.txt
Societe-Emploi.txt
Societe-Famille.txt
Societe-Religion.txt
Societe-Urbanisme.txt

International-Afrique.txt
International-Ameriques.txt
International-Asie.txt
International-Europe.txt
International-MoyenOrient.txt
International-Orient.txt
Loisirs-Jeux.txt
Loisirs-Sport.txt
Loisirs-Tourisme.txt
Medical-Addictions.txt
Medical-Ethique
Medical-Medecine.txt
Politique-Election.txt
Politique-EtatGouvernement.txt
Politique-Immigration.txt
Politique-RegimePolitique.txt
Sciences-Astronomie.txt
Sciences-Biologie.txt
Sciences-Biosphere.txt
Sciences-CatastrophesNaturelles.txt
Sciences-PhysiqueChimie.txt
Societe-ConflitsSociaux.txt
Societe-Discrimination.txt
Societe-Ecologisme.txt
Societe-Emploi.txt
Societe-Famille.txt
Societe-Religion.txt
Societe-Urbanisme.txt

Theme-Culture-Gastronomie: 1300 words

beurre_nc

beurre_nc clarifier_v

beurre_nc en_prp pommade_nc

bière_nc

bifteck_nc

bijane_nc

biller_v

bio_adj

végan_adj

véganisme_nc

végan_nc

végétalien_adj

végétalien_nc

végétalisme_nc

végétarien_adj

végétarien_nc

végétarisme_nc

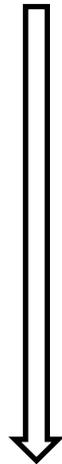
Macron trolls Trump with 'Make our Planet Great Again' website

Initiative encourages scientists and entrepreneurs to move to France to work on climate change.

By **NICHOLAS VINO CUR** | 6/8/17, 7:51 PM CET | Updated 6/8/17, 8:08 PM CET

POLITICS

President, Emmanuel, Macron, government...



INTERNET

President, Emmanuel, Macron, government...



ECOLOGY

ecology, planet, water, oil, lobbies, wild life...

Macron trolls Trump with 'Make our **Planet** Great Again' **Website**. **Initiative** encourages **scientists** and entrepreneurs to move to work on **climate change**.

Identifying the **subject matter**:

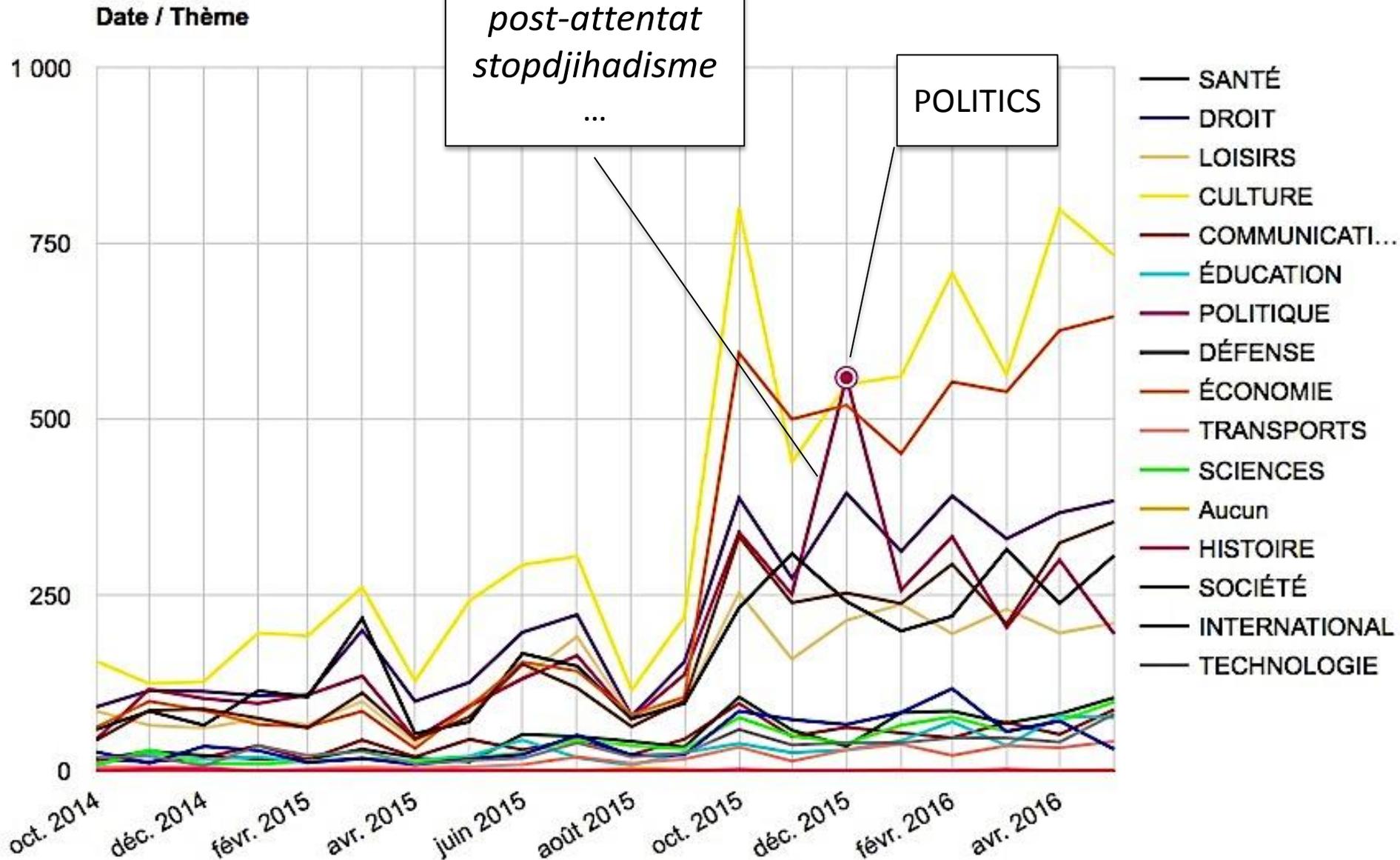
French President Emmanuel **Macron** launched a **WEBSITE** Thursday encouraging anyone concerned about climate change to emigrate to France.

ITS address — www.MakeOurPlanetGreatAgain.fr — uses the same riff on Donald Trump's campaign slogan that **Macron** unveiled the day the U.S. president announced his decision to withdraw from the Paris climate deal.

THE SITE invites researchers, entrepreneurs and NGO workers from around the world (although an aide chose to feature a researcher from California during a demonstration for reporters) to state their interest in climate change and their specialist field, then apply for positions and immigration documents to come to France.

Thematic events:

November 2015 Paris attacks



...
Hard/soft news
Notice
Report
...

INFORMATION

...
Analysis
Interview
...

INTERMEDIATE FIELD

...
Editorial
Chronicle
Review
...

OPINION

...
Horoscope
Recipe
...

ADVICE

...
Crossword
Sudoku
...

ENTERTAINMENT

...
Hard/soft news
Notice
Report
...

...
Analysis
Interview
...

...
Editorial
Chronicle
Review
...

...
Horoscope
Recipe
...

Automatic
recognition?

...
Crossword
Sudoku
...

Corpus

Genre	texts
Weather/Traffic	164
Interview	138
Obituary	133
Recipe	112
Op-ed	120
Review	103
Editorial	100

Analysis based on 50 features:

Stylometric: length of words / sentences / texts

Lexical: opinion verbs, obituary vocabulary, profession names...

Morphosyntactic: POS, tense...

Syntactic: type of sentence, type of subject...

Genre	Precision
Weather/Traffic	
Interview	
Obituary	
Recipe	
Op-ed	
Review	
Editorial	

Genre	Precision
Weather/Traffic	
Interview	
Obituary	
Recipe	
Op-ed	
Review	
Editorial	

Global accuracy: **81.41%**

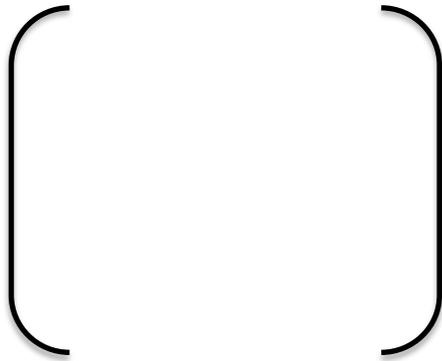
Genre	Precision
Weather/Traffic	0,87
Interview	0,96
Obituary	0,87
Recipe	1
Op-ed	
Review	
Editorial	

Opinion
texts?

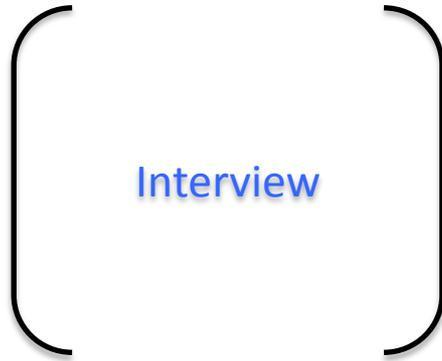
Global accuracy: **81.41%**

Genre	Precision
Weather/Traffic	0,87
Interview	0,96
Obituary	0,87
Recipe	1
Op-ed	0,53
Review	0,69
Editorial	0,58

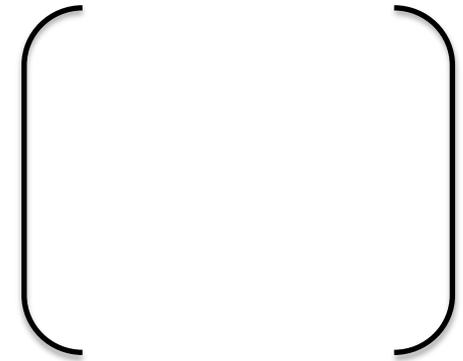
Global accuracy: **81.41%**



INFORMATION



INTERMEDIATE
FIELD



OPINION

Automatic recognition score:

91.3155%

Conclusion

Particularities of the Logoscope

Necessity to conceive a scala of norms

For corpus building / cultural understanding

Remaining problem: genre detection

Contact

Christophe Gérard

Maître de conférence

Equipe Linguistique, langues,

Université de Strasbourg

christophegerard@unistra.fr