



HAL
open science

Le Logoscope : observatoire des innovations lexicales en français contemporain

C Gérard, L Bruneau, Ingrid Falk, Delphine Bernhard, A.-L Rosio

► To cite this version:

C Gérard, L Bruneau, Ingrid Falk, Delphine Bernhard, A.-L Rosio. Le Logoscope : observatoire des innovations lexicales en français contemporain. Joaquín García Palacios, Goedele de Sterck, Daniel Linder, Jesús Torre del Rey, Miguel Sánchez Ibanez et Nava Maroto García. La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones, Peter Lang, 2017, 9783631698624. hal-01388255

HAL Id: hal-01388255

<https://hal.science/hal-01388255>

Submitted on 26 Oct 2016

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.

Le Logoscope : observatoire des innovations lexicales en français contemporain

C. Gérard, L. Bruneau, I. Falk, D. Bernhard, A.-L. Rosio (Université de Strasbourg)

À paraître dans Joaquín García Palacios et alii (dir.), Actes du colloque Cinéo III, *La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones*, Peter Lang.

Le Logoscope est un observatoire linguistique qui documente, journalièrement, les nouveaux mots employés dans la presse en ligne française. Il est actuellement le seul veilleur de néologie semi-automatique qui non seulement informe sur le contexte thématique des innovations lexicales, mais aussi offre une interface web permettant à l'utilisateur une recherche multicritères et enrichie par des représentations statistiques. Cet article présente les principales fonctions du Logoscope : *détecter* et *documenter* les mots nouveaux, mais aussi *assister* leur étude quantitative et qualitative.

Création lexicale ; TAL ; contexte ; thématique ; textualité ; ressource linguistique ; interface utilisateur.

Développé pendant trois ans à l'Université de Strasbourg¹, le Logoscope ne se résume pas à un module de détection de mots nouveaux pour le français, comme il en existe déjà (Torres Rivera 2015), puisque non seulement la détection se lance chaque jour d'elle-même, sans intervention humaine, mais aussi parce que le Logoscope dispose d'une interface web donnant librement accès à sa ressource néologique. De plus, cette ressource se distingue des bases francophones existantes², au sens où d'une part nous fournissons, notamment, des informations à ce jour inédites sur le contexte thématique des mots nouveaux, et au sens où, d'autre part, l'utilisateur bénéficie de graphiques lui permettant une visualisation immédiate de divers phénomènes quantitatifs et qualitatifs liés à l'usage des mots nouveaux.

Le but de cet article est d'exposer les marques d'originalité de cet observatoire linguistique, puis de définir et d'illustrer les variables contextuelles qu'il intègre dans sa documentation des mots nouveaux, avant de donner un aperçu des possibilités de son interface-utilisateur.

1 Détection semi-automatique d'innovations lexicales

Parce que « Le langage est une activité libre ou créatrice » (Coseriu 2001 : 433), l'exercice de cette faculté par les individus parlants donne constamment lieu à ce qu'il convient d'appeler des *innovations*, c'est-à-dire tout ce qui, dans une production orale ou écrite, s'éloigne de modèles langagiers existants (i.e. Coseriu 1958, chap. III : 6). En ce sens large, une innovation peut concerner soit le niveau historique des langues/dialectes (ex. création d'un nouvel affixe, d'un nouveau lexème, d'une nouvelle locution verbale³), soit le niveau historique des traditions

¹ Financement IDEX-Attractivité 2012-2015.

² Bornéo (www.atilf.fr/borneo/) ou Neologia (Cartier et Sablayrolles 2008 ; Sablayrolles 2011 ; Cartier 2011).

³ Ce dernier phénomène créatif est nommé « détournement » par Sablayrolles (à paraître).

discursives⁴ (ex. création d'une expression phraséologique inédite, invention d'un genre de discours, etc.).

En ce qui concerne le Logoscope, les innovations linguistiques recueillies sont essentiellement d'ordre *lexical*, mais pas exclusivement puisqu'il arrive de repérer des innovations grammaticales⁵. La détection semi-automatique de ces innovations suit un processus en plusieurs étapes, que nous décrirons comme suit, sans entrer ici dans la technicité des solutions informatiques :

1. Collection automatique des textes. Chaque jour, notre automate recueille entre 250 et 500 textes provenant des flux RSS de six journaux français disponibles en ligne : *La Croix*, *Les Échos*, *Libération*, *Le Monde*, *Le Figaro* et *L'Équipe*.⁶
2. Extraction automatique de mots-candidats. Les articles journalistiques collectés sont ensuite examinés au moyen d'un programme d'extraction lexicale (Falk/Bernhard/Gérard 2014) dont la tâche consiste à trouver toutes les unités inconnues, c'est-à-dire toute unité qui ne figure pas dans notre liste d'exclusion. Cette liste d'exclusion, qui au départ de notre projet a été constituée à partir du *Corpus français* de la base Wortschatz (projet Leipzig Corpora Collection)⁷, comporte aujourd'hui plus de 2 millions de formes (linguistiques ou non), constamment enrichie par tout mot nouveau détecté. Au terme du processus d'extraction, le programme génère un fichier qui, selon les jours, comprend entre 100 et 300 mots-candidats, parmi lesquels se trouve un ensemble de mauvais candidats (non-mot, mauvaise orthographe, nom propre, mots étrangers en contexte hétérologue, etc.), comme le montre cet extrait d'une liste datée du 23/10/2015 : *ggg*, *Royaume-Uniet*, *lumièreedes*, *Añatuya*, *Milleret-Godet*, *precautiones*, *refugios*. Il est clair que cette méthode ne permet pas d'accéder à certains types d'innovation lexicale, comme la « néologie sémantique » (Gérard/Kabatek 2011) et, dans le champ de la « néologie formelle », les unités polylexicales (du type *selfie stick*, *réalité augmentée*, *capitaine de soirée*, etc.), dont le repérage demeure un défi majeur du traitement automatique de la langue.
3. Identification manuelle des innovations lexicales. Cette troisième phase consiste à effectuer un tri, parmi les données obtenues lors de la phase précédente, en isolant en premier lieu les mauvais mots-candidats. Clairement interprétative, cette tâche fait appel à divers savoirs linguistiques, différemment mobilisés selon le cas : le cotexte du mot observé, son nombre d'occurrences ou encore sa relative récence (le moteur de recherche *Google* et la base *Factiva*⁸ s'avérant ici des sources utiles d'aide à la décision). Les bons candidats, ou véritables innovations lexicales, font quant à eux l'objet d'une annotation manuelle, afin de documenter leurs principales variables lexicales (infra 2.1).
4. Suivi automatique de la diffusion des innovations. Lorsque la détection semi-automatique des innovations lexicales d'une journée particulière est accomplie, c'est-à-dire lorsque la

⁴ Sur cette notion-clé de la romanistique allemande contemporaine, englobant celle de genre discursif, voir Koch/Oesterreicher 2001 et Kabatek 2015.

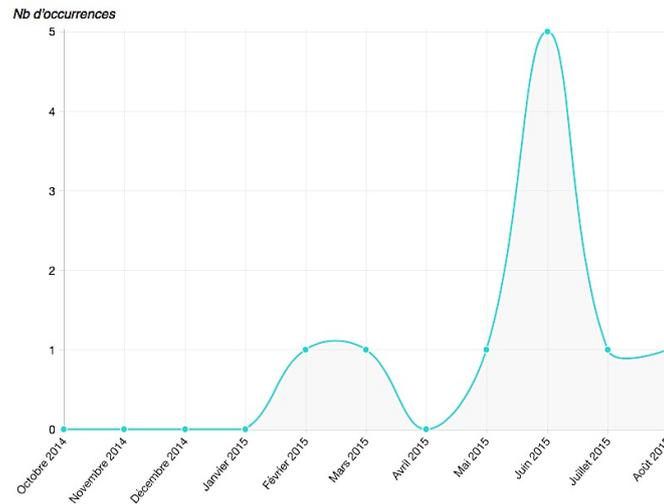
⁵ Ex. « cette personne est-elle insexuelle ? Est-ce ainsi que **lui-elle-même** se définit ? » (*Libération*, 10/06/2015).

⁶ Bientôt augmentés, sous réserve de faisabilité technique, grâce au soutien financier de la DGLFLF (Délégation générale à la langue française et aux langues de France), des journaux *La Recherche*, *60 millions de consommateurs*, *Science et Vie* et *La Tribune*.

⁷ Pour une description : wortschatz.uni-leipzig.de/ws_fra.

⁸ Site : www.dowjones.com/products/product-factiva/.

phase d'identification manuelle parachève la phase d'extraction automatique, un nouvel automate prend le relais, sur le long terme. Il a en effet pour tâche de documenter toutes les reprises ultérieures de chaque innovation détectée, en indiquant par exemple que telle innovation est réapparue une nouvelle fois le lendemain, mais dans un autre journal, ou qu'elle a été réutilisée trois mois plus tard dans plusieurs journaux, de manière massive. Le Logoscope procure ainsi des indices sur l'histoire contemporaine du lexique, comme en témoigne par exemple le cas du récent *uberisation*⁹ :



Parallèlement au processus de détection des mots nouveaux, le travail d'observation du Logoscope intègre un processus de documentation, également semi-automatique, qui consiste à apporter des informations sur des variables lexicales et des variables contextuelles.

2 Documenter les variables contextuelles

2.1 Documentation automatique des variables contextuelles

Les veilleurs de néologie¹⁰ doivent faire appel à un ensemble de variables lexicales (catégorie grammaticale, valence verbale, type de création, etc.) et contextuelles (date d'apparition, contexte, présence de guillemets, etc.) pour préciser les caractéristiques des mots constituant leur base (Gérard/Falk/Bernhard 2014). En ce qui concerne le Logoscope, la documentation des variables lexicales est volontairement réduite à l'essentiel¹¹, puisqu'il s'agit seulement de préciser la catégorie grammaticale du mot nouveau et une ou des méta-catégorie(s) de création lexicale (d'après Sablayrolles 2011) :

⁹ « (*Néologisme*) adoption d'un modèle de commerce consistant à mettre des ressources à disposition des clients depuis leurs smartphones, à tout moment et sans délai. » (fr.wiktionary.org). Voir aussi infra 3.

¹⁰ WORTWARTE (Lemnitzer 2010), OBNEO (Cabré et alii, 2003 ; Estopà / Cabré 2004), NEOLOGIA (Cartier / Sablayrolles 2008 ; Sablayrolles 2011 ; Cartier 2011).

¹¹ Cette position minimaliste se justifie ainsi : a) le Logoscope n'a aucune d'ambition lexicographique (aucune définition des mots n'est proposée) ; b) l'information lexicale est suffisante pour que l'utilisateur puisse approfondir par lui-même, selon ses objectifs, l'analyse des innovations de la ressource (ainsi les bases et les affixes sont directement accessibles en appliquant les expressions régulières) ; c) compte tenu du nombre d'innovations détectées chaque jour, le temps de traitement manuel nécessite absolument d'être réduit.

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

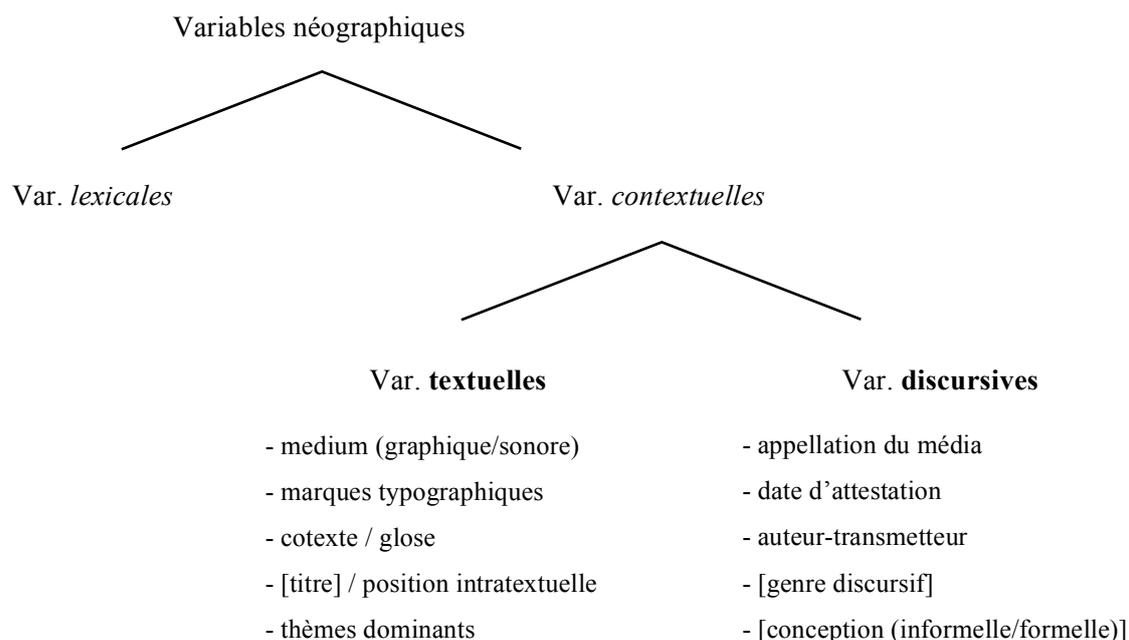
En effet, les innovations sont globalement annotées, sans réflexion sur le procédé de création proprement dit. Aussi, par exemple, *songwri-teuse*¹² sera simplement annoté /nom commun/ et /emprunt + morphosémantique/, mais pas /emprunt assimilé + suffixation/. Le néographe gagne ainsi en temps en minimisant son taux d'erreur, et en délivrant une information qui reste néanmoins pertinente pour une recherche lexicologique plus approfondie.

En revanche, une des particularités distinctives du Logoscope concerne la documentation automatique des variables contextuelles. La raison de cette focalisation est le fait qu'il n'existe pas d'innovation lexicale qui soit indépendante de la situation de communication où elle fonctionne et du texte où elle est interprétée. Mieux, non seulement le sens d'une innovation, mais aussi son identité variationnelle, les valeurs linguistiques et extralinguistiques qu'on lui attribue, ainsi que ses fonctions rhétoriques, voire l'identification même du signifiant, dépendent d'interprétants trouvés dans la situation en cours (Gérard 2014). De fait, l'approche privilégiée par le Logoscope s'inscrit dans une problématique du texte (Gérard 2011), fortement développée dès les années 1980 par la germanistique allemande (Peschel 2002), et se traduisant en particulier aujourd'hui par des études portant sur des genres discursifs particuliers (e.g. Siebold 2000 ; Vorger 2011 ; Lacoste/Gérard 2016¹³).

Pratiquement, les variables contextuelles se divisent en deux ordres, les variables *textuelles* (indissociables de la matérialité sémiotique du texte) et les variables *discursives* (témoignant plus largement de la situation de communication et des normes socio-discursives en cours) :

¹² « L'histoire officielle raconte que Courtney Barnett a 26 ans et qu'elle fut serveuse à Melbourne. Pêchue, acidulée, tonique, c'est surtout la **songwri-teuse** de ce printemps » (*Libération*, 23/03/2015).

¹³ Citons de même, lors de CINEO 2015, les communications de M. Roche et alii (« La néographie dans un grand corpus de SMS français : 88milSMS »), de C. Détrie (« Produire du sens en textotant : de quelques innovations lexicales et lexico-morphologiques/syntaxiques dans les SMS »), de A. Napieralski (« Les néologismes dans les commentaires de la presse sportive ») ou encore de M. Casado-Velarde (« Innovación léxica y discurso: la columna de opinión de Ignacio Camacho »).



Pour le Logoscope, pour des raisons trop longues à développer ici, seuls le titre de l'article, le genre discursif de l'article (éditorial, analyse, interview, etc.)¹⁴ et la conception de la communication¹⁵ échappent aujourd'hui à une identification automatique. En dehors de ces variables, la documentation contextuelle s'effectue sans intervention humaine, des marques typographiques (guillemets, parenthèses) à la position intratextuelle de l'innovation (début, milieu ou fin de texte), en passant par les thèmes dominants du texte d'accueil. Dans le Logoscope, toutes ces informations contextuelles servent à générer automatiquement de nombreuses statistiques sur la création lexicale en situation médiatique (infra 3).

2.2 Thematic : un outil de détection automatique des thèmes textuels

Concernant la documentation contextuelle des innovations lexicales, l'apport le plus original du Logoscope concerne indubitablement la variable thématique, avec le développement par notre équipe d'un outil — nommé Thematic — qui automatise la détection des thèmes de textes issus de la presse écrite francophone. Ayant progressé depuis nos premières recherches (Gérard/Falk/Bernhard 2014), l'outil Thematic a atteint sa maturité, puisqu'il parvient en quelques secondes à obtenir les résultats suivants (article du *Monde*, 25/12/2014) :

¹⁴ À notre connaissance, OBNEO est la seule base à documenter, manuellement, la variable du genre discursif, en notant les différents genres radiophoniques et télévisés de son corpus oral.

¹⁵ Certains veilleurs de néologie notent la qualité « orale » ou « écrite » des innovations (ex. dans la base BORNEO des mots comme *craignoss*, *trip musique*, *à chier* ou *méga-salle* sont documentés comme étant de nature « orale »). En effet, l'opposition oral/écrit ne se réduit pas à l'opposition médiale phonique/graphique. Mais les marques de variation diaphasique ne résument pas à elles seules l'opposition oral/écrit, qui ne se conçoit pleinement sans intégrer la notion de conception communicationnelle : les locuteurs comprennent que leur comportement communicationnel doit être conforme à la situation en cours, et que cette conformité passe par la réalisation d'un « style » plus ou moins libre, informel et affectif, ou d'un style plus ou moins formel, académique et distancié (Koch/Oesterreicher 2001).

-- Le pape condamne la " persécution brutale " des " minorités ". La préoccupation du pape pour le sort des chrétiens d' Orient s' est exprimée tout au long de cette semaine de Noël . Lors de la bénédiction *urbi et orbi* qu' il a prononcée à midi , jeudi 25 décembre , place Saint-Pierre à Rome , à l' occasion de la fête de la nativité , François a dénoncé les " persécutions brutales " dont sont victimes les chrétiens d' Irak et de Syrie , avec ceux " qui appartiennent à d' autres groupes ethniques et religieux " . Il a évoqué " les nombreuses personnes déplacées , dispersées et réfugiées [...] de la région et du monde entier " et a demandé qu' elles " puissent recevoir les aides humanitaires nécessaires pour survivre à la rigueur de l' hiver et revenir dans leur pays " . Lundi , le pape argentin avait adressé une lettre aux chrétiens du Moyen-Orient pour les " encourager " et leur dire " combien [leur] présence et [leur] mission sont précieuses en cette terre " où " est né et où s' est répandu le christianisme " . Comme à son habitude , il n' y citait pas nommément l' Etat islamique mais il y dénonçait une " organisation terroriste " qui " commet toutes sortes d' abus et de pratiques indignes de l' homme , en frappant de manière particulière certains d' entre vous qui ont été chassés de façon brutale de leurs propres terres , où les chrétiens sont présents depuis les temps apostoliques " . TROP D' ENFANTS " VICTIMES D' ABUS ET EXPLOITÉS " Jeudi 25 décembre , le pape François a prononcé le deuxième discours de Noël de son pontificat . Jeudi , devant les dizaines de milliers de personnes rassemblées devant la basilique Saint-Pierre , le pape a demandé un large " soutien aux efforts de ceux qui s' engagent efficacement pour le dialogue entre Israéliens et Palestiniens " . Il a aussi mentionné l' Ukraine , le Nigeria , où " à nouveau du sang est versé et trop de personnes sont injustement soustraites à l' affection de leurs proches et tenues en otage ou massacrées " , ainsi que d' autres pays africains : la Libye , le Soudan du Sud , la République centrafricaine et la République démocratique du Congo . Dans une bénédiction où , comme il l' a souligné , étaient présentes " tant de larmes en ce jour de Noël " , le souverain pontife a également dénoncé les mauvais traitements infligés à des enfants , victimes de violences , de trafics , de traite des personnes ou contraints à être soldats . Il y a trop d' enfants " victimes d' abus et exploités sous nos propres yeux et avec notre silence complice " , a -t-il déclaré . Il a cité " les enfants massacrés sous les bombardements , y compris là où est né le fils de Dieu " , c' est-à-dire en terre sainte , et leur " silence impuissant qui crie sous l' épée " . Il a aussi mentionné les enfants tués à Peshawar , au Pakistan , dans l' attentat contre une école la semaine dernière , ainsi que ceux victimes du virus Ebola au Liberia , en Sierra Leone et en Guinée . Dans une référence à l' avortement , il a aussi cité les " enfants tués avant de voir la lumière " .

Thèmes Spécifiques :						
Thèmes Principaux	Tokens	Types	Poids du Thème	Poids/Tokens du Thème	% mot colorié sur le total de mot du Thème	% mot colorié sur le total de mot de l'article
Mythologie-Religion et Croissance	24	14	22.58	94.10	1.37%	1.96%
Conflits Armés Contemporains	18	13	9.28	51.57	4.63%	1.47%
Afrique	8	8	8.00	100.00	1.97%	0.65%
Thèmes Secondaires	Tokens	Types	Poids du Thème	Poids/Tokens du Thème	% mot colorié sur le total de mot du Thème	% mot colorié sur le total de mot de l'article
Moyen-Orient	7	7	7.00	100.00	3.32%	0.57%
Démographie	10	5	6.00	60.00	2.15%	0.81%

Thèmes Génériques :						
Thèmes Génériques	Tokens	Types	Poids du Thème	Poids/Tokens du Thème	% mot colorié sur le total de mot du Thème	% mot colorié sur le total de mot de l'article
Société	38	22	32.58	85.75	0.85%	3.10%
Défense	22	17	11.62	52.80	0.81%	1.79%
International	18	17	18.00	100.00	0.99%	1.47%
Santé	9	9	6.25	69.44	0.74%	0.73%
Politique	7	5	5.00	71.43	0.71%	0.57%
Culture	6	5	3.92	65.28	0.07%	0.49%
Économie	6	5	2.83	47.22	0.45%	0.49%
Droit	6	4	3.00	50.00	0.37%	0.49%
Histoire	2	2	1.00	50.00	0.10%	0.16%

En somme, Thematic : 1) détecte les thèmes de n'importe quel texte journalistique, à partir d'un répertoire de 68 thèmes¹⁶ ; 2) hiérarchise les thèmes détectés en distinguant les thèmes dominants au sein du texte analysé, à la fois sur un plan générique (ici Société, Défense, International) et sur un plan spécifique (Mythologie-Religion-Croyances, Conflits-armés-contemporains, Afrique) ; 3) désambiguïse, le cas échéant, les mots indexant a priori différents thèmes, comme ici *victime* qui contribue ici à lexicaliser le thème Conflits-armés (mais qui pourrait dans un autre texte lexicaliser la Médecine ou les Conflits sociaux) ou *virus* qui ici lexicalise le thème de la Santé (mais qui pourrait ailleurs lexicaliser l'Informatique) ; 4) tient compte, pour l'identification thématique, des unités polylexicales (ici *organisation terroriste*, *urbi et orbi*, *République démocratique du Congo*, etc.).

Pour le Logoscope, les enjeux de la détection thématique sont multiples. D'ordre pratique, un premier enjeu concerne les pratiques qui distinguent entre eux les utilisateurs potentiels du Logoscope. En effet, notre ressource n'a pas été seulement conçue pour être utile aux lexicographes, aux terminologues, aux traducteurs ou aux linguistes en général, mais également pour répondre aux besoins de communautés d'utilisateurs pour qui le langage en lui-même importe moins que les réalités dont il témoigne. Ainsi, l'historien, le sociologue ou le journaliste peuvent interroger le Logoscope pour connaître les changements ou les événements qui se sont produits pour un sujet d'intérêt particulier, pendant une période donnée. Par exemple, une recherche sur les événements liés au thème de la Finance, entre mars et juin 2015, donne les résultats suivants :

¹⁶ Dont : presse, radio, architecture, cinéma, musique, sécurité civile, justice, gastronomie, finance, commerce, mode, Afrique, Europe, enseignement scolaire, sports, jeux, élection, médecine, astronomie, zoologie, écologie, famille, immigration, conflits sociaux, internet, informatique, transports aériens, etc.

Grexitent

Libération, 22/03/2015 **N** **MIL** **MORSEM**

Droit - Justice et Législation **Politique - Etat et Gouvernement** **Economie - Finance**
plusieurs responsables politiques ont même évoqué froidement « un Grexitent » ,

désatlantiser

La Croix, 24/03/2015 **V** **MIL** **MORSEM**

Economie - Finance **Droit - Justice et Législation** **Economie - Commerce**

le silence européen est assourdissant . Or , l'absence de réaction rapide et proportionnée européenne face à l'initiative d' Obama vaudrait aveu de vassalité morale et politique de l'Union européenne à l'égard des États-Unis , alors même que l'Europe devrait de plus en plus se « désatlantiser » et construire sa propre vision , émancipée et anticipative , d'un monde multipolaire mieux équilibré et plus juste .

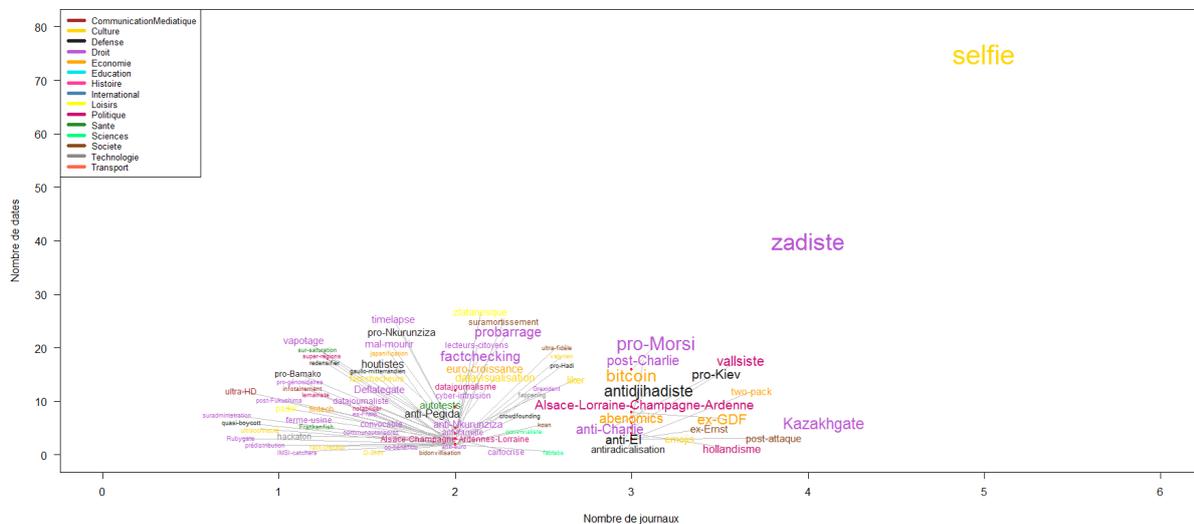
afro-optimisme

La Croix, 25/03/2015 **N** **MIL** **MORSEM**

Economie - Finance **Droit - Justice et Législation** **Sciences - Mathématiques**

À en croire l'étude que le laboratoire d'idées dédié aux pays émergents a réalisée avec Havas Horizons , nouveau service pour les pays émergents , l'afro-optimisme est en tout cas de rigueur dans les milieux d'affaires internationaux . Sur les 43 grandes institutions financières et bancaires interrogées entre janvier et février , 100 % des investisseurs estiment que l'économie africaine a de beaux jours devant elle à l'horizon 2020 , 53 % d'

Un second enjeu de la détection thématique est d'ordre scientifique. En effet, les études de néologie n'ont généralement pas encore reconnu toute l'attention que méritent les rapports entre néologie et thématique. Or la prise en compte des thèmes textuels conduit à mieux comprendre le phénomène de la création lexicale, qu'il s'agisse de l'implication des innovations lexicales dans la progression thématique (Gérard 2011 et 2014), ou qu'il s'agisse de leur diffusion sociale (Bernhard et alli, à paraître). Par exemple, *selfie* s'est rapidement et massivement diffusé dans beaucoup de domaines d'expérience (politique, droit, science, etc.), mais son domaine d'usage privilégié reste la culture, du moins dans la presse quotidienne française :



Un troisième enjeu, technologique cette fois, intéresse la détection de la néologie sémantique. En effet, la possibilité de détecter automatiquement les thèmes caractéristiques d'un texte ouvre la porte à la détection automatique des innovations de type métaphorique (Falk/Bernhard/Gérard 2014). Par exemple, depuis une dizaine d'années, le mot *quenelle* n'est plus seulement associé au domaine de la gastronomie (plat cuisiné à base de poisson, depuis 1750), mais au domaine de la politique, parce qu'il acquiert en 2005 la signification d'un geste polémique¹⁷. Au sein des textes, cela se traduit par un emploi de la forme *quenelle* qui ne s'accompagne plus du vocabulaire

¹⁷ Geste popularisé par l'humoriste français Dieudonné M'Bala M'Bala (fr.wikipedia.org/wiki/Dieudonné).

traditionnel attendu (beurre, farine, poisson, cuire, mélanger, four, etc.), mais d'un environnement thématique inédit (les thèmes de la justice, des médias, du spectacle ou de la discrimination). Ainsi, savoir caractériser automatiquement l'environnement thématique des mots constitue une grande avancée dans la détection automatique de ce type de changement sémantique.

3 Deux interfaces : des requêtes assistées et expertes

Pour accéder aux innovations et interroger leurs propriétés, le Logoscope dispose actuellement de deux interfaces aux fonctionnalités complémentaires. La première interface, qui correspond au site mis en ligne, est destinée à une utilisation simple : elle ne permet d'interroger qu'une variable néographique à la fois, sans possibilité d'en combiner plusieurs. On peut ainsi choisir d'interroger la ressource « par néologisme », « par date », « par journal », etc. :

The screenshot shows the Logoscope website interface. At the top, there is a search bar with the text "Logoscope" and a magnifying glass icon. Below the search bar, there is a navigation menu with several tabs: "Accueil", "Par Néologisme", "Par Date", "Par Journal", "Par Catégorie Grammaticale", "Par Thème", "Par Procédé", and "Par Position". A search input field contains the text "Google" and a green checkmark icon. Below the navigation menu, there is a keyboard layout with letters A through Z, Ç, É, Î, Œ, and Ü. A "Prev" button and a "Next" button are also visible. The main content area displays a list of neologisms under the letter "U".

U				
u2p	uberisation	ubérisation	uberise	uberisent
uberiser	ubériser	uber-mobilisation	uber-pop	uls
ultra-antisémite	ultracaméra	ultra-ciselés	ultra-commerçant	ultracomplotiste
ultra-concentrés	ultra-conformiste	ultra-connecté	ultraconnecté	ultraconsommation
ultra-contemporain	ultra-déterminée	ultra-droitier	ultra-égalitariste	ultra-fidèle
ultra-fidèles	ultragraphique	ultra-hd	ultra-immersive	ultra-imprévisible
ultrakiffagerie	ultra-limitées	ultramaniabiles	ultra-marketé	ultramobilisés
ultra-naïve	ultra-noires	ultra-patriotes	ultra-périssable	ultra-polyvalent

Si l'utilisateur clique sur l'onglet « par néologisme », il peut alors sélectionner un mot particulier (ex. *uberisation*) pour en obtenir les occurrences rangées par ordre chronologique d'apparition, ainsi que son graphique de diffusion (supra 1) :

Logoscope : Recherche uberisation (1/1)

Logoscope

Base des nouveaux mots contemporains

Accueil | Par Néologisme | Par Date | Par Journal | Par Catégorie Grammaticale | Par Thème | Par Procédé | Par Position

uberisation Graphique 1

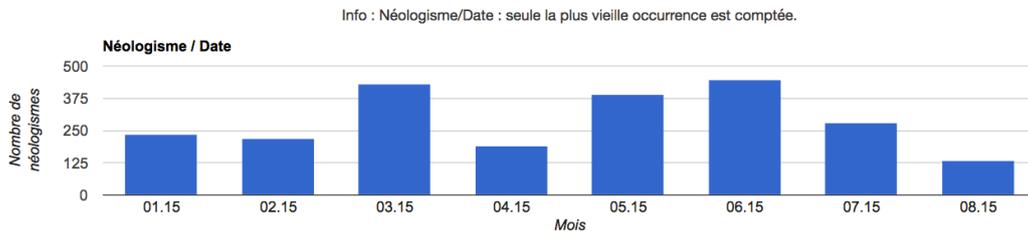
- Nom Commun - **ECONOMIE-Commerce** | **ECONOMIE-Industrie** | **ECONOMIE-Finance** - Fin - Aucun - 24/02/2015 - Les Echos -
Contexte : "L' « **uberisation** » de l' **économie** et le grand vertige des élites "

- Nom Commun - **INTERNATIONAL-Europe** | **SOCIÉTÉ-Mythologie, Religion et Croyances** - Fin - Aucun - 16/03/2015 - Les Echos -
Contexte : "L' « **uberisation** », nouvelle mythologie **française** "

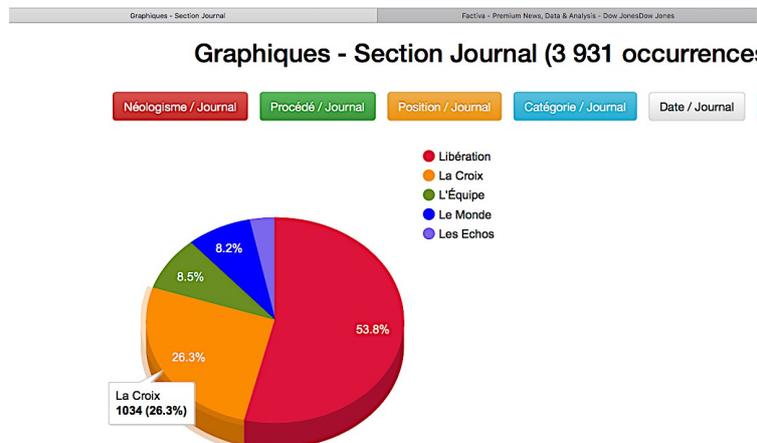
- Nom Commun - **ECONOMIE-Finance** | **ECONOMIE-Commerce** | **DROIT-Justice et Législation** - Milieu - Morphosémantique, Emprunt - 07/05/2015 - Libération -
Contexte : "Fin des **banques** ? (éditions Eyrolles , février 2015) . On est dans un **cas** typique " d' **uberisation** " . » L' avant-gardiste de la fintech serait PayPal qui , en 1998 , a lancé son **service** de "

- Nom Commun - **POLITIQUE-Élection** | **SANTÉ-Médecine** | **INTERNATIONAL-Europe** - Fin - Aucun - 02/06/2015 - Les Echos -
Contexte : "Vivement « l' **uberisation** » de la **vie politique française** ! "

La valeur ajoutée de cette interface réside dans sa grande stabilité technologique (on résout ainsi les problèmes de maintenance de la base et du serveur informatique), mais surtout dans la possibilité d'obtenir des graphiques statistiques sur toutes les dimensions principales de notre ressource. À titre d'exemple, il est possible de visualiser immédiatement l'évolution annuelle des innovations, en comptant les premières attestations repérées dans la presse :



De même, la productivité néologique des journaux observés par le Logoscope est présentée par un graphique comparatif, accessible au moyen d'un seul clic :



En tout, le Logoscope génère plus d'une dizaine de graphiques différents qui, à chaque mise à jour mensuelle de la base, sont tous automatiquement réédités. La recherche de l'utilisateur peut ainsi être dite assistée, dans le sens où de tels graphiques procurent une information propre à donner une vue d'ensemble de la ressource utilisée.

La seconde interface est librement téléchargeable depuis le site du Logoscope. Le dossier téléchargé au format .zip s'installe localement sur n'importe quel ordinateur et autorise ainsi des recherches hors ligne :

The screenshot shows the 'Recherche' (Search) interface of the Logoscope website. At the top, there is a navigation bar with 'Accueil', 'Présentation', 'Recherche', 'Aides', and 'Contact'. The main search area includes a search bar with the placeholder 'Néologisme : forme et/ou expression(s) régulière(s)'. Below this, there are several filter sections: 'Filtrer par' with options for 'Guillemets' (Oui, Non, Sans importance) and 'Suivi d'un commentaire'; 'Position' with options (Début, Milieu, Fin, Aucune position précise); 'Catégories' and 'Journaux' dropdown menus; 'Procédés de création' with checkboxes for 'Emprunt', 'Morphologique', 'Morphosémantique' (checked), 'Syntacticosémantique', and 'Non identifié'; 'Procédé morphosémantique' with 'Affixation' (préfixe(s), suffixe(s)) and 'Onomatopée'; 'Filtre des thèmes pour une entrée' with 'Au moins un' (checked) and 'Seulement ceux sélectionnés'; and 'Thèmes' with 'ajout(s)' and 'exclusion(s)' buttons. A date range filter is also present. At the bottom, there is a 'Rechercher' button and a grid of topic tags including Communications, Culture, Défense, Droit, Economie, Education, Histoire, International, Loisirs, Politique, Santé, Sciences, Société, Transports, and Technologies.

Il s'agit en quelque sorte de la version « expert » du Logoscope, dans la mesure où cette interface offre une plus grande précision et liberté de recherche. Entre autres, l'utilisateur peut :

1. Combiner entre elles toutes les variables néographiques, en choisissant plusieurs « valeurs » d'une même variable, et en définissant n'importe quel intervalle chronologique. On peut, par exemple, souhaiter obtenir seulement les emprunts nominaux, apparus entre le 01/03/2015 et le 30/06/2015, dans *La Croix* et *Le Monde*, et en position finale de texte :

chepeceletas
 La Croix, 13/03/2015 **N** **FIN** **EMP**
 Droit - Justice et Législation Culture - Musique Défense - Militaire
 Des ateliers de construction de vélos en bambou au Ghana aux promenades à vélo à San José , au Costa Rica , avec les « chepeceletas » (1) qui tissent aussi d' autres relations entre ces promeneurs , une foule d' initiatives se mettent en place . Comment dire avec eux ce qui nous est le plus cher : notre exigence de justice et d' égalité qui nourrisse des relations interculturelles et intergénérationnelles qui

big-pharma
 La Croix, 28/03/2015 **N** **FIN** **EMP**
 Droit - Justice et Législation Santé - Médecine Economie - Finance
 Tout en collaborant à la mise au point de nouveaux « candidats » médicaments avec les « big-pharma » comme Sanofi , Novartis , GSK ou AstraZeneca , l' Inserm se focalise aussi sur les médicaments d' origine biologique comme les protéines thérapeutiques , les anticorps monoclonaux , les vaccins et les thérapies géniques et cellulaires . Ainsi , pas plus tard que cette semaine , l' Inserm

- Interroger, dans la barre de recherche, les innovations elles-mêmes au moyen d'expressions régulières. Un tel moyen autorise facilement des requêtes morphologiques sur les affixes, mais aussi sur n'importe quelle composante alphabétique du mot. Par exemple, la requête « *twi.** » donne pour résultat :

twittomélie

La Croix, 06/03/2015 **N** **MIL** **MORSEM**

Société - Mythologie et Religion Culture - Musique Technologie - Internet

Car à ses deux mitres, il faut encore ajouter une casquette qui l'a rendu célèbre auprès des journalistes et des passionnés du numérique : celle de président du conseil épiscopal pour la communication. L'homme des « **twittomélie** » – méditations en 140 caractères postées chaque jour sur le réseau social Twitter – encourage les autres évêques à être présents sur le « continent digital ».

Twittclash

Libération, 06/03/2015 **N** **FIN** **MORSEM** **EMP**

Droit - Justice et Législation Défense - Militaire Politique - Election

de Safran. Pour lui, les mots « rouge brun », c'est un « **Twittclash** à deux balles ». Donc, pour développer ma pensée, j'obtempère avec l'aide de Toral dans l'article ci-dessous. Mais comme le débat risque de tourner à son

twittos-parisiens-qui-ne-votent-pas-mais-font-des-blaques-pour-participer-quand-même

Libération, 22/03/2015 **N** **FIN** **MORSEM**

Politique - Election Droit - Justice et Législation Défense - Militaire

Pause loi (2). Suite du concours des twittos-parisiens-qui-ne-votent-pas-mais-font-des-blaques-pour-participer-quand-même.

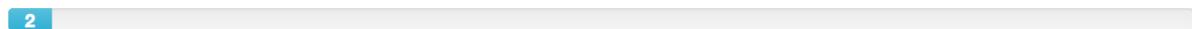
- Sélectionner les innovations marquées par des guillemets ou accompagnées d'une glose (introduite par une parenthèse). Ces marques témoignent généralement d'un grand degré de nouveauté.
- Télécharger les résultats obtenus pour les traiter dans un tableur de type Excel, afin de réaliser ses propres diagrammes statistiques.

Enfin, cette version « portable » du Logoscope dispose elle aussi de graphiques délivrant une synthèse des résultats, en nombre d'occurrences, pour toutes les variables néographiques. Ainsi, un graphique donne le nombre d'innovations créées au moyen de multiples procédés de création :

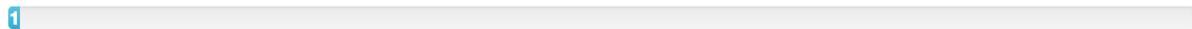
Morphosémantique, Emprunt



Morphosémantique, Syntacticosémantique



Morphologique, Emprunt



Au moyen de ses deux interfaces, le Logoscope assiste ainsi la recherche de l'utilisateur de manière complémentaire.

4 Directions de recherche

En guise de conclusion, nous aimerions évoquer deux directions de recherche que l'état actuel du Logoscope nous incite à envisager, mais qui intéressent également les autres veilleurs de néologie.

Une première direction est un défi technologique bien connu : elle concerne la détection automatisée des unités polylexicales (ex. *selfie stick*, *réalité augmentée*, etc.) et des différents types de

néologie sémantique. Sur ce dernier point, la littérature témoigne de très intéressantes réalisations, et nous avons pour notre part indiqué dans cet article comment l'outil Thematic peut apporter sa contribution. Mais ces recherches demeurent expérimentales et, surtout, isolées, et nous ne pouvons aujourd'hui qu'appeler de nos vœux une collaboration entre tous ces travaux.

Une seconde direction de recherche concerne les aspects sociolinguistes, et en particulier variationnels, de la néologie. En effet, les veilleurs de néologie n'incluent généralement pas dans leurs variables de documentation, au-delà de la langue du néologisme, la variation idiomatique (diatopique, diastratique, diaphasique). Or, par exemple, la francophonie mériterait un projet d'envergure documentant sa vitalité lexicale dans différentes régions, notamment pour examiner l'émergence supposée des différents français d'Afrique (Calvet 2010, chap. 5). Par ailleurs, outre la question de la variation idiomatique, se pose la question de la *variation discursive*. De fait, la création néologique varie pas seulement selon les paramètres sociolinguistiques classiques (diatopie, diastratie, diaphasie), mais aussi selon le *genre de discours* dans lequel on s'exprime.¹⁸ Ainsi, en quantité comme en qualité, la création lexicale est différemment conditionnée dans le cas d'une néologie et dans le cas du courrier des lecteurs, genre très néologène et ludique. Pour étudier, à grande échelle, les rapports entre genre discursif et néologie, nous manquons toutefois de données ; des données que laissent espérer les projets de recherche sur l'identification automatique des genres. Ajoutée à la documentation thématique, cette documentation générique ferait du Logoscope une ressource représentant pleinement les variables contextuelles de la création lexicale.

Bibliographie

Bernhard, Delphine. / Bruneau, Lauren. / Falk, Ingrid. / Gérard, Christophe. « Création lexicale et exploitation des données contextuelles. L'exemple de la presse quotidienne. ». Actes des Journées du réseau LTT, « La création lexicale en contexte : texte, genre, culture », à paraître.

Cabré, Maria-Térésa. / Domènech, Ona. / Estopà, Rosa. / Freixa, Judith. / Solé, Elisabeth. « L'observatoire de néologie : conception, méthodologie, résultats et nouveaux travaux ». Jean-François Sablayrolles (éd.), *L'innovation Lexicale*. Paris : Honoré Champion, 2008, pp. 125–147.

Calvet, Louis-Jean. *Histoire du français en Afrique*. Paris : Éditions Écriture, 2010.

Cartier, Emmanuel. « Néologie et description linguistique pour le TAL ». *Langages*, 183. 2011, pp. 105–117.

Cartier, Emmanuel. / Sablayrolles, Jean-François. « Néologismes, dictionnaires et informatique ». *Cahiers de Lexicologie* 93, pp. 175 — 192.

Coseriu, Eugenio. *L'homme et son langage*. Peeters, 2001.

Coseriu, Eugenio. *Sincronía, diacronía e historia*, 2a ed. (3a ed, 1978). Traduit en français par Thomas Verjans, [1958] 1973. [http://www.revue-texto.net/Parutions/Livres-E/Coseriu_SDH/Sommaire.html.]

¹⁸ Sur les différences, du point de vue néologique, entre le roman de guerre, la lettre aux proches et le journal privé, voir Gérard/Lacoste 2016.

Estopà, Rosa / Cabré Castellví, Maria-Térésia. « Metodologia del treball en neologia: criteris, materials i processos », *Document de treball*, Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. 2004. [<http://hdl.handle.net/10230/1304> (consulté le 23/11/2013)].

Falk, Ingrid. / Bernhard, Delphine. / Gérard, Christophe. « From Non Word to New Word: Automatically Identifying Neologisms in French Newspapers ». LREC —Proceedings of the International Conference on Language Resources and Evaluation. Reykjavik : 2014. [<https://hal.inria.fr/hal-00959079>].

Gérard, Christophe. « Création lexicale, sens et textualité : théories et analyses », *PhiN, Philologie im Netz*, 2011. [<http://web.fu-berlin.de/phn/>].

Gérard, Christophe. « Sémiotique interprétative des créations verbales ». Driss Ablali / Sémir Badir / Dominique Ducard (éd.). *Documents, textes, œuvres. Perspectives sémiotiques*. Presses universitaires de Rennes, 2014, pp. 333-350.

Gérard, Christophe. / Kabatek, Johannes. « Introduction : la néologie sémantique en questions », *Cahiers de lexicologie, Revue internationale de lexicologie et de lexicographie*. Éditions Classiques, 2012, pp. 11-36.

Gérard, Christophe. / Falk, Ingrid. / Bernhard, Delphine. « Traitement automatisé de la néologie : pourquoi et comment intégrer l'analyse thématique ? ». *Actes du 4e Congrès Mondial de Linguistique Française (CMLF 2014)*. Berlin : Web of Conferences, 2014. [http://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01208.pdf].

Gérard, Christophe. / Lacoste, Charlotte. « “Mots de la Grande Guerre” : créations inaperçues et usages réels dans les écrits de combattants. Étude de lexicologie textuelle ». Gilles Siouffi / Odile Roynette (éd.). *La première guerre mondiale et la langue*. Rennes : PUR, 2016, à paraître.

Kabatek, Johannes. « Genre textuel et traditions discursives ». Christophe Gérard / Régis Missire (éd.). Eugenio Coseriu aujourd'hui, Linguistique et philosophie du langage. Limoges : Lambert-Lucas, 2015, pp. 195-206.

Koch, Peter / Oesterreicher Wulf. « Langage parlé et langage écrit ». Günter Holtus / Michael Metzeltin / Christian Schmitt (éd.), *Lexikon der romanistischen Linguistik*, Tübingen : Niemeyer Verlag, 2001, pp. 601-604.

Lemnitzer, Lothar. « Neologismenlexikographie und das Internet ». *Lexicographica*, 26, 2010, pp. 65-78.

Peschel Corinna. *Zum Zusammenhang zwischen Wortneubildung und Textkonstitution*. Tübingen : Niemeyer, 2002.

Sablayrolles, Jean-François. « Quelques remarques sur une typologie des néologismes : Amalgamation ou télescopage : un processus aux productions variées (mots valises, détournements...) et un tableau hiérarchisé des matrices », Actes de CINEO II, São Paulo, 5-8 décembre 2011. À paraître.

Sablayrolles, Jean-François. « Neologia : un dictionnaire néologique sous forme de base de données ». S. A. Marcelino Cardoso / S. Mejri / J. A. Mota (éd.), *Os di.ci.o.na.rios, fontes, métodos et novas tecnologias*, Instituto de Letras da Universidade federal da Bahia, Brésil, 2011, pp. 221-235.

Siebold, Oliver. *Wort - Genre - Text. Wortneu-bildungen in der Science Fiction*. Tübingen : Gunter Narr Verlag, 2000.

Spitzer, Leo (1910) : *Die Wortbildung als stilistisches Mittel. Exemplifiziert an Rabelais. Nebst einem Anhang über die Wortbildung bei Balzac in seinem "Contes Drolatiques"*, Halle (Beihefte zur Zeitschrift für Romanische Philologie XXIX).

Torres Rivera, Andrés. « Herramientas de detección y extracción de neología: estado de la cuestión ». Treball de fi de màster Recerca. Universitat Pompeu Fabra, 2015.

Vorger, Camille. *Poétique du slam : de la scène à l'école. Néologie, néostyle et créativité lexicale*, Thèse de doctorat, Université de Grenoble, 2011.