



HAL
open science

NEOLOGIA, une base de donnée pour la gestion des néologismes

Emmanuel Cartier, Jean-François Sablayrolles

► **To cite this version:**

Emmanuel Cartier, Jean-François Sablayrolles. NEOLOGIA, une base de donnée pour la gestion des néologismes. Congrès International de Neologia de les Llengües Romàniques, May 2008, Barcelone, Espagne. pp.759-767. halshs-00608876

HAL Id: halshs-00608876

<https://shs.hal.science/halshs-00608876>

Submitted on 28 Sep 2012

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.

Sablayrolles Jean-François (Paris 13 et LDI UMR 7187) et Cartier Emmanuel (Paris 13 et LDI UMR 7187)

« Neologia, une base de données pour la gestion des néologismes »,

Actes del I Congrés Internacional de neologia de les Llengües Romàniques (Barcelone (7-10 mai 2008), M. Teresa Cabré, Ona Domènech, Rosa Estopà, Judit Freixa, Mercè Lorente eds., Barcelone, Université Pompeu Fabra, IULA, sèrie activitats 22, p. 757-766 (759-767 en fait).

Résumé

La base Neologia, élaborée au LDI par Emmanuel Cartier en fonction des demandes de Jean-François Sablayrolles, a pour objectif de collecter les néologismes mais surtout de servir d'outil d'analyse de la créativité lexicale (et de l'évolution du lexique) du français hexagonal contemporain. Après la présentation de son architecture et de ses fonctionnalités sont exposés les résultats de quelques premières recherches comme illustration des objectifs de l'équipe.

Mots-clés : néologie, base de données, analyse de la néologie

NEOLOGIA, UNE BASE DE DONNEES POUR LA GESTION DES NEOLOGISMES

Un petit rappel historique en guise d'introduction

Giovanni Adamo en avait évoqué l'idée lors de la conclusion du colloque de Limoges sur l'innovation lexicale en 2001, nous en avons reparlé lui, Térésa Cabré et moi (Jean-François Sablayrolles, JFS désormais) en juin 2002 à Rome, Térésa l'a fait en créant le réseau d'observatoires des néologismes des langues romanes, NEOROM. La réunion fondatrice de Barcelone (en 2004), à laquelle ont participé JFS et John Humbley pour le français hexagonal, a été un moment intellectuel fort, suscitant beaucoup d'enthousiasme et ouvrant de nombreuses perspectives.

L'équipe à laquelle JFS appartenait alors, le CIEL, Centre Interlangue d'Études en Lexicologie, de Paris 7, n'a malheureusement pas eu les moyens humains, techniques et financiers nécessaires pour réaliser sa part du projet commun. Une base de données a bien été construite, mais elle l'a été tardivement et elle était peu maniable. Elle n'a pas vraiment pris son essor avant que le CIEL ne soit conduit à se fondre dans le CLILLAC, équipe de recherches constituée essentiellement d'anglicistes.

Une nouvelle étape s'est alors ouverte avec l'entrée de JFS au LLI (Laboratoire de Linguistique Informatique devenu depuis laboratoire LDI, Lexiques, Dictionnaires et Informatique) et à Paris 13. Après une longue période de réflexion collective des membres de l'équipe néologie du laboratoire, une base de données, Neologia, a été construite par Emmanuel Cartier (EC). La phase de rodage vient juste de s'achever et des mises au point doivent être effectuées prochainement.

Une présentation de l'architecture de la base et de ses fonctionnalités précède l'exposé de quelques résultats comme exemples d'utilisation possible. Ceux-ci sont à prendre avec prudence du fait, d'une part, du nombre limité de néologismes entrés effectivement (405 mais des milliers ont été relevés qui attendent d'être traités), et du fait, d'autre part, que les fiches n'ont pas pu encore être revues et validées collectivement.

1. La présentation de la base Neologia

1.1. Ses objectifs et son contenu

La base Neologia est un outil de travail¹ qui a été mis au point conjointement par les deux auteurs, linguiste (JFS) et linguiste-informaticien (EC), membres du laboratoire LDI, qui renoue ainsi, après une interruption, avec une tradition d'études néologiques françaises dans laquelle Bernard Quemada (1993 entre autres) a joué un rôle important.

Cette base a pour objectifs l'enregistrement et le traitement des néologismes que la veille néologique, en partie automatisée mais encore en grande partie manuelle, nous conduit à identifier. La constitution de la base avec sa vingtaine de champs et les types d'informations qu'ils contiennent ont pour mission de nous permettre de mieux connaître, d'une part, la créativité lexicale des Français contemporains (en relevant leurs créations indépendamment de leur avenir, toujours incertain) et, d'autre part, l'évolution du lexique français, en ne prenant en compte que les néologismes qui se diffusent largement et durablement.

Fruit de la recherche de JFS depuis une vingtaine d'années et assise sur le modèle des classes d'objets du laboratoire LDI, cette base comporte cinq séries de champs, outre la page d'entrée qui présente essentiellement le néologisme sous sa forme lemmatisée et sa définition.

Des informations morpho-syntaxiques comme l'indication de la partie du discours et de catégories grammaticales associées constituent la première série d'informations, incontournables.

La seconde série note en revanche des informations syntactico-sémantiques souvent omises et propres au modèle des classes d'objets. On indique si le néologisme est un prédicat, un argument ou un actualisateur². Suivent des champs indiquant le schéma argumental des prédicats ou l'appartenance à une classe d'objets pour les arguments, etc.

Le troisième ensemble de champs est incontournable puisqu'il consigne des propriétés néologiques proprement dites, mais certaines d'entre elles font rarement l'objet d'études systématiques. Le premier champ est réservé bien sûr à la matrice néologique. Nous utilisons une typologie très hiérarchisée, inspirée de celle de J. Tournier (1985, pour l'anglais) et exposée dans Sablayrolles (2000) et Pruvost-Sablayrolles (2003) avec quelques remaniements de détail apportés récemment³. On note également un éventuel changement de catégorie⁴,

¹ Elle n'a donc pas vocation à être accessible au grand public (pénètre-t-il dans les cuisines des restaurants ?), mais un statut invité est prévu selon certaines conditions pour des collègues qui seraient intéressés et qui nous en feraient la demande. L'adresse et un mot de passe leur seront alors communiqués.

² Pour l'instant nous n'avons pas relevé de néologismes actualisateurs. L'analyse en prédicats et arguments doit par ailleurs être reprise systématiquement du fait d'une modification de notre approche à la suite de la phase de rodage : les premières analyses se sont faites à partir de l'occurrence en contexte, alors que les quatre premières séries de champs concernent le néologisme en général et non le néologisme en contexte. Un certain nombre de noms d'action sont, comme tels, des prédicats, mais ils peuvent être utilisés comme des arguments. Ainsi *boboisation* « fait de devenir / rendre bobo » dans *l'évolution sociologique et la boboisation ne sont pas a priori favorables à la droite* ou *reliassage* 'action de relisser' dans *Pour les peaux matures qui ont des rides et des taches, on parle de relissage ou de resurfacing* où les prédicats sont *favorable* et *parler de*. Nous avons donc décidé de traiter de tels cas de prédicats en position d'arguments comme des prédicats et non comme des arguments.

³ On distingue ainsi deux types de conversions : conversion simple (quand une unité change de partie du discours sans ajout ni suppression d'affixe) et conversion verticale quand un syntagme devient une unité lexicale. À la suite de Fabienne Cusin-Berche (2003) on distingue aussi les mots valises et les compocations où aucun segment phonétique n'est commun aux unités qui s'amalgament en une seule.

⁴ Indépendamment de la matrice qui opère ce changement : préfixation (*vape* -> *envapé*), suffixation (*mordre* -> *mordeur*), dérivation inverse (*prestation* -> *prester*), conversion (*gamelle* -> (*se*) *gameller*)...

l'éventuelle présence d'un nom propre comme base, la configuration en syllabes phoniques, ouvertes (O) et fermées (F), en vue d'une comparaison avec l'aspect morphophonologique des mots conventionnels tel que Dauzat (1943) ou des études statistiques plus récentes l'ont pris analysé. Nous accordons aussi une importance particulière à la configuration morphologique, avec l'identification de tous les constituants (préfixe(s), radical/aux, suffixe(s) essentiellement), puisqu'il ne faut pas confondre, comme on le fait parfois⁵, cette analyse avec l'identification de la matrice. On note également l'éventuelle influence d'une langue étrangère et son mode d'influence pour des néologismes qui ne sont pas des emprunts directs mais dont l'existence est dépendante de mots de langues étrangères : des calques (tel l'ancien *gratte-ciel* pour *skyscraper*), des équivalents (*voiture-sandwich* pour *car wrapping*), des emprunts de formants (*girl* et *boy* dans *Sarkogirl* ou *Bouygueboy*) ou de structures (l'ordre régressif de *Royale attitude*), ou des emprunts sémantiques (tel l'ancien *réaliser* « comprendre »).

Le champ « relations sémantiques » (qui peut recevoir plusieurs relations) indique des cas de synonymie, d'antonymie, d'hyponymie, etc. avec des mots de la langue française et aussi des cas de traduction. Cette rubrique, qui existe dans d'autres dictionnaires du LDI, s'avère intéressante pour l'étude de l'insertion des néologismes au sein du lexique de la langue.

Enfin dans la zone « contexte » sont consignés le cotexte linguistique (au moins une phrase, souvent plus), éventuellement des circonstances nécessaires à la compréhension du néologisme, (le contexte au sens large), puis les sources (écrites / orales ; presse, affiche, etc.), la date et diverses informations sur la typographie, sur les créateurs et les réémetteurs-transmetteurs, etc. Cette zone est démultipliable pour consigner des attestations antérieures ou des contextes particulièrement intéressants. Un champ est désormais prévu pour la mesure de la diffusion des néologismes.

1.2. Ses fonctionnalités

1.2.1. Accès internet

Pour que les utilisateurs puissent travailler de n'importe quel lieu et quel que soit leur environnement informatique, la base est installée sur un serveur Web et est accessible via n'importe quel navigateur Web.

1.2.2. Gestion des utilisateurs

Trois niveaux d'utilisateurs ont été initialement définis (administrateur, auteur, invité) correspondant à différents pouvoirs sur la base : les invités ont la capacité de visualiser les différentes entrées validées par les administrateurs, ainsi que d'utiliser le moteur de recherche associé. Les auteurs ont les droits attachés aux invités, ainsi que la capacité de créer de nouvelles entrées, de les modifier et/ou de les supprimer. Les auteurs contrôlent ainsi leurs propres entrées, mais n'ont aucun droit autre que celui d'invité sur les autres entrées. Les administrateurs, enfin, ont les droits attachés aux auteurs, ainsi que la capacité de valider ou

⁵ L'analyse morphologique et morphématique est conduite selon les travaux de D. Corbin et de C. Touratier. *Détatouage* est, structurellement, un mot préfixé et suffixé (préfixe *dé*, radical *tatou* et suffixe *-age*). Pour des raisons tant morphologiques que sémantiques, ce néologisme est créé par la suffixation et non par la préfixation : morphologiquement le préfixe *dé-* sert fondamentalement à construire sur des bases verbales des verbes (qui peuvent ensuite être nominalisés par suffixation le plus souvent) et sémantiquement c'est l'« action de ^odétatouer ». ^o*Détatouer* est un « mot possible non attesté », dans la terminologie de D. Corbin, qui utilise un petit cercle placé en exposant devant pour indiquer ce statut. Nombre de néologismes sont formés sur de telles bases.

d'invalider les entrées proposées par les auteurs. Leur rôle principal est donc de valider ou d'invalider les entrées proposées par les auteurs, et de les faire passer ainsi dans le domaine ouvert aux invités. Mais, lors de la réunion de mise au point à l'issue de la période de rodage, un problème est apparu : les auteurs ne peuvent pas ajouter des contextes d'occurrences plus anciennes ou particulièrement intéressantes pour des néologismes dont ils n'ont pas eux-mêmes créé la fiche. Afin de remédier à cet inconvénient, il a été décidé que tous les auteurs auraient le statut d'administrateur mais que chacun s'engagerait à ne pas intervenir directement sur le contenu des fiches d'autrui. Une boîte à idées permettra de faire des remarques, de proposer des analyses, de poser des questions, etc. Des séances de validation collective auront lieu régulièrement.

1.2.3. Moteur de recherche

L'accès aux entrées de Neologia se fait via un moteur de recherche permettant de filtrer les entrées selon l'un ou plusieurs de tous les champs utilisés pour décrire une entrée. Par défaut, le moteur de recherche ne comporte aucun filtre, et l'ensemble des entrées est présenté sous forme de tableau synthétique triable. L'interface se compose d'une zone « recherche » et d'une zone « résultat ». La zone « recherche » permet aux utilisateurs de filtrer les résultats selon un ou plusieurs des divers champs renseignés pour chaque néologisme. La zone « résultat » présente, sous forme de tableau, les différents items répondant aux critères de recherche. Les deux boutons à gauche de chaque ligne permettent respectivement de visualiser la fiche complète du néologisme, et de la supprimer. Enfin, les champs de ce tableau sont triables et il est possible d'afficher un nombre variable d'items par page.

1.2.4. Visualisation et ajout, modification, suppression de fiches

Le second mode de visualisation des entrées est la « fiche » qui permet de visualiser l'ensemble des détails liés à une entrée. Lorsque l'utilisateur est invité, il ne pourra que visualiser la fiche, dans les autres cas, il pourra accéder totalement aux fonctionnalités d'ajout, de suppression ou de modification des fiches.

1.2.5. Paramétrage des informations

L'outil est par ailleurs évolutif, avec la possibilité d'ajouter ou de retirer certains champs, ou de modifier les valeurs dans un menu déroulant. La première révision doit être effectuée incessamment après une phase de rodage de plusieurs mois qui nous a permis de lister un certain nombre de modifications à apporter.

1.2.6. Interactions de la base avec l'extérieur

Neologia a deux liens avec le monde extérieur : les dictionnaires généraux du LDI et un extracteur de néologismes à partir de corpus Web.

La néologicit  est un  tat transitoire : soit le n ologisme tombe dans l'oubli et reste un hapax, soit il se r pand et entre dans les usages linguistiques avec un sens fixe. Il doit alors faire partie d'un dictionnaire g n ral. La base de donn es Neologia pr voit ce cas de figure et permet aux administrateurs, lorsqu'ils consid rent que le n ologisme a perdu ce statut, de le reverser dans les dictionnaires g n raux d velopp s au LDI : Morfetik et Semantik. L'exportation des termes vers les dictionnaires g n raux du LDI est assur e par une fonctionnalit  informatique. Celle-ci permet de r cup rer non seulement l'information

orthographique, mais également d'autres informations pertinentes dans un dictionnaire général, notamment les couches descriptives « morphologiques » et « sémantiques » décrites plus haut.

La récupération automatique de candidats néologismes est l'autre fenêtre du logiciel vers l'extérieur. Il s'agit de pouvoir récupérer, de façon automatique, des candidats néologismes, à partir de corpus. Cette intégration est en cours de réalisation avec Telanaute, mais les premiers tests montrent que cette piste est très prometteuse, car elle permet de repérer de manière systématique des candidats néologismes formels sur des masses de données que ne saurait traiter un humain. Cependant, à l'heure actuelle, le système mis en place, qui utilise le dictionnaire Morfetik comme corpus d'exclusion, souffre de limitations : d'une part, le processus récupère toutes sortes de mots (abréviations, noms propres etc.) qui, par nature, ne sauraient entrer dans un dictionnaire général, mais qui ne sont pas non plus des néologismes. D'autre part, les textes, notamment Internet, regorgent de fautes typographiques qui rendent le tri humain postérieur actuellement un peu lourd. Enfin, les néologismes syntaxiques ou sémantiques, pas plus que les néologismes homonymiques ne peuvent être repérés de cette façon. Les deux premières difficultés sont actuellement en cours de résolution. Par contre, la dernière limitation ne pourra être levée dans le système actuel. Mais des recherches sont conduites au laboratoire pour repérer automatiquement les innovations dans la distribution d'unités lexicales conventionnelles, à partir d'écarts par rapport à leur schéma argumental canonique⁶.

2. Quelques premiers résultats en guise d'exemples

L'étude des néologismes donne lieu à des requêtes simples (par catégories grammaticales, par matrices lexicales...) et surtout croisées avec plusieurs paramètres pris simultanément afin de discerner les interrelations entre les types de néologismes, les domaines, les conditions d'énonciation, les types d'énonciateurs, etc. Voici les résultats de deux requêtes simples (parties du discours et base nom propre) et de deux requêtes croisées (matrices et parties du discours, puis transcatégorisation et matrices) portant sur les 405 néologismes analysés au moment de la préparation de cette communication.

2.1. Requêtes simples

2.1.1. Par parties du discours

Nom	285	70,4%
Adj	79	19,5 %
Vb	39	9,6%
Adv	1	0,25%
Onomatopée	1	0,25%

2.1.2. Nombre et proportions de Noms propres comme bases de néologismes

48 sont construits sur un anthroponyme	11,9%
10 sont construits sur un toponyme	2,5%
10 sont construits sur un nom de marque	2,5%
68 sont donc construits sur un nom propre	16,8%

⁶ Voir « néologismes, classes d'objets et extraction automatique » de JFS dans les actes de ce congrès.

2.2. Requêtes croisées

2.2.1. Matrices et parties du discours⁷

	Nombre	%	Nom	Adjectif	Verbe	Autres
Préfixation	29	7,2%	14	11	4	-
Suffixation	87	21,5%	46	32	8	1 (adverbe)
Dérivation inverse	3	0,75%	-	-	3	-
Parasynthétique	5	1,25%	3	1	1	-
Flexion ⁸	3	0,75%	3	-	-	-
Composition ⁹	72	17,7%	57	14	1	-
Synapsie	10	2,5%	10	-	-	-
Composition savante	3	0,75%	2	1	-	-
Composition hybride	39	9,6%	34	2	2	-
Mot-valise	21	5,2%	17	3	1	-
Compoction	6	1,5	5	-	1	-
Onomatopée	1	0,25%	-	-	-	1 interjection
Conversion	12	3%	3	-	9	-
Conversion verticale	2	0,5%	-	2	-	-
Déflexivation ¹⁰	Ø					
Combinatoire sémantico syntaxique ¹¹	3	0,75%	-	-	3	-
Extension /restriction de sens	2	0,5%	-	2	-	-
Métaphore	1	0,25%	1	-	-	-
Métonymie	1	0,25%	-	1	-	-
Antonomase	2	0,5%	2	-	-	-
Autres figures	Ø	-	-	-	-	-
Troncation	5	1,25%	5	-	-	-
Siglaision/acronymie	3	0,75%	3	-	-	-
Détournement ¹²	11	2,7%	7	2	2	-
Emprunt	84	20,7%	72	8	4	-

⁷ L'ordre des matrices correspond au système hiérarchisé évoqué précédemment avec des références bibliographiques.

⁸ Un néologisme est flexionnel quand apparaît une forme fléchie inexistante auparavant pour ce mot : le féminin *leadeuse* par exemple.

⁹ À la suite de Benveniste sont distingués plusieurs grands types de composition : composition savante (distinguée en outre de la composition hybride quand tous les éléments n'appartiennent pas au même système linguistique), synapsie et tous les autres types de composition.

¹⁰ La déflexivation ressemble à la conversion —avec laquelle elle est souvent confondue— mais elle s'en distingue par le fait qu'elle garde les marques flexionnelles de sa catégorie d'origine (qui perdent leur valeur, d'où le nom de déflexivation). *Le boire et le manger* en sont des exemples anciens.

¹¹ La néologie par combinatoire sémantico-syntaxique correspond aux infractions aux schémas prédicatifs canoniques telles qu'elles sont présentées dans la communication de la table ronde « Néologismes, classes d'objets et extraction automatique ».

¹² Les détournements sont des défigements de lexies figées qui ne peuvent être interprétés que par le recours à la lexie figée originelle : la *planche à promesses* ne peut pas avoir de sens en dehors de sa mise en relation avec la *planche à billets* de *faire marcher la planche à billets* pour une politique de fuite en avant qui masque sans les résoudre des problèmes de fond qui resurgiront plus tard.

2.2.2. Transcatégorisation et matrices (90 transcatégorisations = 22,2% du total)

Adj → Adv	1	1 suffixation
Adj → N	10	5 suffixations + 3 synapsies + 1 parasynthétique + 1 conversion
Adj → V	2	1 suffixation + 1 dérivation inverse
GPrep → Adj	1	1 conversion verticale
Adv → N	1	1 conversion
N → Adj	29	23 suffixations + 2 préfixations + 1 parasynthétique + 1 emprunt + 1 conversion verticale + 1 composition
N → V	17	9 conversions + 5 suffixations + 2 dérivations verticales + 1 parasynthétique
V → Adj	3	3 suffixations
V → N	26	19 suffixations + 2 compositions + 2 emprunts + 2 synapsies + 1 conversion

Bien d'autres requêtes sont à effectuer, en particulier sur le nombre assez important de noms à nombre contraint ou pour le nombre plus faible qu'attendu de présence dans les titres. Mais ce sont surtout des requêtes mettant aussi en action les types d'émetteurs et les domaines qui seront les plus instructives sur la néologie¹³ du français contemporain.

Conclusion

Nos perspectives de travail sont donc nombreuses, ambitieuses et ne pourront se concrétiser que sur la longue durée. Les premiers résultats sont prometteurs, même si l'on sait dès maintenant que notre tâche n'aura jamais de fin puisque notre matériau ne cesse de se renouveler, mais c'est le lot de toutes les équipes de néologie. Bien qu'elle sorte à peine de sa période de rodage, la base Neologia a déjà servi pour extraire des néologismes par emprunt pour une étude comparative des néologismes par emprunt en grec moderne et en français, conduite en collaboration avec l'équipe de néologie du grec moderne d'Anna Anastasyadi-Symeonydi de l'Université Aristote de Thessalonique. Certains des champs de la base Neologia serviront également à alimenter la part de la néologie française hexagonale de la plate-forme NEOROM, pour la comparaison de la néologie des langues romanes.

Indications bibliographiques

- BEN HARIZ-OUENNICHE Soundous, GROUD Claudette, SABLAYROLLES Jean-François et YAHMED Samia (à paraître en 2009), « Vers une homogénéisation des incorporations des néologismes » *Neologica* n 3.
- CORBIN Danielle (1987), *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Max Niemeyer Verlag.
- CUSIN-BERCHE Fabienne (2003), *Les mots et leurs contextes*, Presses Sorbonne Nouvelle.
- DAUZAT Albert (1943), *Le génie de la langue française*.
- GROSS Gaston (1994), « Classes d'objets et description des verbes », *Langages*, 115, pp. 15-30.
- PRUVOST Jean et SABLAYROLLES Jean-François (2003), *Les néologismes*, Que sais-je ?, PUF.
- QUEMADA Bernard (1993), (sous la direction de), *Mots nouveaux contemporains 1*, Matériaux pour l'histoire du vocabulaire français, CNRS, Paris, Klincksieck.

¹³ Dans une des acceptions récentes du mot évoquées par Bernard Quemada lors de la séance inaugurale.

- SABLAYROLLES Jean-François (2000), *La néologie en français contemporain, examen du concept et analyse de productions néologiques récentes*, coll. Lexica Mots et Dictionnaires, Champion.
- SABLAYROLLES Jean-François (2002), « Fondements théoriques des difficultés pratiques du traitement des néologismes », *Revue française de linguistique appliquée*, vol. VII-1. / juin 2002 « Lexique : recherches actuelles », pp. 97-111.
- SABLAYROLLES Jean-François, « Néologismes, classes d'objets et extraction automatique » Actes du congrès CINEO.
- TOURATIER Christian (2002), Morphologie et morphématique, Analyse en morphèmes, *Langue et langage* n°8, Publications de l'Université de Provence, 2002.
- TOURNIER Jean (1985), *Introduction descriptive à la lexicogénétique de l'anglais contemporain*, Paris-Genève, Champion-Slatkine.