



**HAL**  
open science

# Neologia : un dictionnaire néologique sous forme de base de données

Jean-François Sablayrolles

## ► To cite this version:

Jean-François Sablayrolles. Neologia : un dictionnaire néologique sous forme de base de données. Journée des dictionnaires : Les Dictionnaires : ressources, méthodes et nouvelles technologies, Nov 2010, Salvador, Bahia, Brésil. pp.221-235. halshs-00735693

**HAL Id: halshs-00735693**

**<https://shs.hal.science/halshs-00735693>**

Submitted on 26 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.

*Os di.ci.o.na.rios, fontes, métodos et novas tecnologias*, Suzana Alice Marcelino Cardoso, Salah Mejri et Jacyra Andrade Mota eds, Instituto de Letras da Universidade federal da Bahia, Brésil, 2011, p. 221-235.

## Résumé

Longtemps la veille néologique s'est contentée de relever des mots paraissant nouveaux avec la constitution d'une fiche manuelle pour chacun d'eux intégrant quelques informations, nécessairement peu nombreuses, à leur sujet. Le laboratoire LDI possède, à Cergy, de nombreux fichiers, fruit de l'énorme travail de veille néologique de l'INaLF (conduit sous l'impulsion de Bernard Quemada et avec la collaboration active de Gabrielle Quemada, récemment disparue et à laquelle nous voulons rendre hommage). Ces dépouillements ont donné lieu à la parution, en 1993, de *Mots nouveaux contemporains I* qui est resté malheureusement sans suite. Des collègues de Paris 7 ont aussi fait paraître quelques intéressants recueils de néologismes également au début des années 90 mais sans qu'il y ait une exploitation approfondie de ces collectes.

Le développement de l'informatique a changé en profondeur le travail. Outre l'existence d'extracteur (semi)automatique de néologismes (essentiellement formels pour l'instant, entre autres à l'aide de l'outil Telanaute mis au point par Fabrice Issac au LDI, mais diverses réflexions sur l'extraction semi-automatique de néologismes sémantiques sont en cours), on peut aller facilement au-delà de la simple collecte et on peut constituer le répertoire des néologismes retenus comme un outil de travail et d'analyse.

Ainsi la base Neologia, dictionnaire de néologismes du français contemporain, élaboré au LDI avec Emmanuel Cartier, comprend une trentaine de champs d'analyse pour chaque néologisme. Des requêtes simples ou des requêtes croisant plusieurs critères permettent d'étudier tel ou tel aspect de la néologie du français contemporain, avec le nombre et la liste des items qui correspondent aux requêtes formulées. On peut ainsi, par exemple, si on travaille sur les adjectifs, avoir le nombre d'adjectifs et calculer la proportion qu'ils représentent par rapport à l'ensemble des néologismes collectés. On peut ensuite savoir immédiatement combien sont obtenus par « transcatégorisation » à partir d'un mot de base appartenant à une catégorie grammaticale différente (combien sont formés à partir de noms, de verbes, etc.), combien d'entre eux sont formés par préfixation, suffixation, conversion, etc. En quelques clics, on peut ainsi radiographier l'état de la néologie adjectivale en français. C'est un progrès énorme pour les études néologiques. Un perfectionnement en cours d'élaboration consistera dans l'indication automatique du nombre d'occurrences des néologismes dans certaines sources (non encore arrêtées), ce qui permettra de mesurer leur circulation et leur diffusion. C'est donc une petite révolution dans la création et l'exploitation de dictionnaires de néologismes qui a été rendue possible par le développement des nouvelles technologies.

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

# NEOLOGIA : UN DICTIONNAIRE NEOLOGIQUE SOUS FORME DE BASE DE DONNEES

## Introduction

Longtemps la veille néologique s'est contentée de relever des mots paraissant nouveaux avec la constitution pour chacun d'eux d'une fiche manuelle intégrant quelques informations, nécessairement peu nombreuses, à leur sujet. Le laboratoire LDI possède, à Cergy, de nombreux fichiers, fruit de l'énorme travail de veille néologique de l'INaLF<sup>1</sup>. Ces dépouillements ont donné lieu à la parution, en 1993, de *Mots nouveaux contemporains I* qui est malheureusement restée sans suite. Des collègues de Paris 7 (GRIL) ont aussi fait paraître quelques intéressants recueils de néologismes également au début des années 90 mais sans qu'il y ait une exploitation linguistique approfondie de ces collectes. La recherche des données et la constitution de fiches manuelles ou mécanographiques ou même des premières fiches informatisées absorbaient la plus grande part du temps disponible des chercheurs et surtout la veille néologique constituait un objectif en soi. Le développement de l'informatique a changé en profondeur le travail. Outre l'existence d'extracteur (semi)automatique de néologismes<sup>2</sup>, on peut aller facilement au-delà de la simple collecte et on peut constituer le répertoire des néologismes retenus comme un outil de travail et d'analyse. Ainsi la base Neologia, dictionnaire de néologismes du français contemporain, élaboré au LDI avec Emmanuel Cartier, comprend une trentaine de champs d'analyse pour chaque néologisme. Des requêtes simples ou des requêtes croisant plusieurs critères permettent d'étudier tel ou tel aspect de la néologie du français contemporain, avec le nombre et la liste des items qui correspondent aux requêtes formulées.

## **1. L'architecture de la base Neologia**

La base Neologia, nommée dictionnaire de néologismes par l'informaticien qui en a assuré la création concrète, Emmanuel Cartier, contient plus de 1500 néologismes à l'heure actuelle. Elle se distingue des dictionnaires sur un point important, qui est l'ordre des entrées : ce n'est pas l'ordre alphabétique traditionnel de la lexicographie qui a été adopté, mais on a préféré l'ordre dans lequel les néologismes sont entrés : c'est plus significatif (a priori, ce qui apparaît en dernier est plus récent, ou du moins a été relevé plus tard) et c'est plus facile pour suivre l'évolution de la base : on regarde les dernières pages créées.

---

<sup>1</sup> Ces travaux ont été conduits sous l'impulsion de Bernard Quemada, avec la collaboration active de Gabrielle Quemada, récemment disparue et à laquelle nous voulons rendre hommage.

<sup>2</sup> Néologismes essentiellement formels pour l'instant, entre autres à l'aide de l'outil Telanaute mis au point par Fabrice Issac au LDI, mais diverses réflexions sur l'extraction semi-automatique de néologismes sémantiques sont en cours.

## 1.1. Écran d'accueil

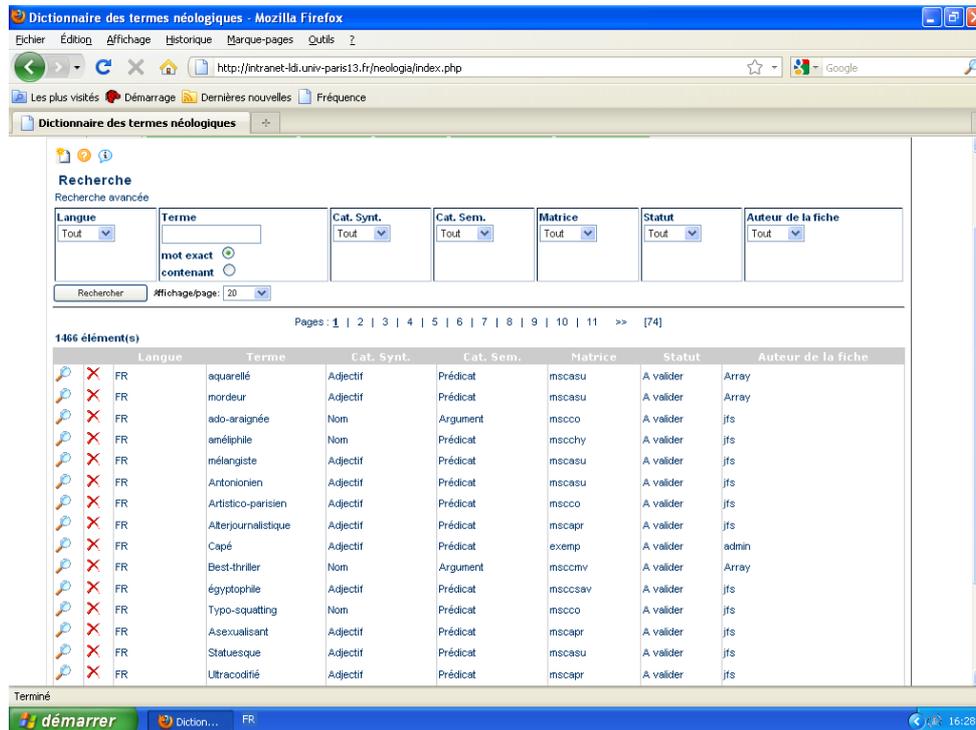


fig 1

L'écran d'entrée se présente comme une interface avec une zone recherche et une zone résultats. Aucun filtre du moteur de recherche n'étant activé lors de l'ouverture de la base, la zone résultats présente la première page des néologismes entrés et analysés. On peut alors faire défiler, page par page, la totalité des néologismes jusqu'au dernier qui y a été entré. La présentation des résultats se fait sous forme de tableau avec une sélection de champs d'analyse, tous ne pouvant pas apparaître ensemble à l'écran. Il suffit de cliquer sur un bouton (la loupe à gauche) pour faire apparaître la fiche entière de tel ou tel néologisme.

## 1.2. Page d'ouverture d'une fiche

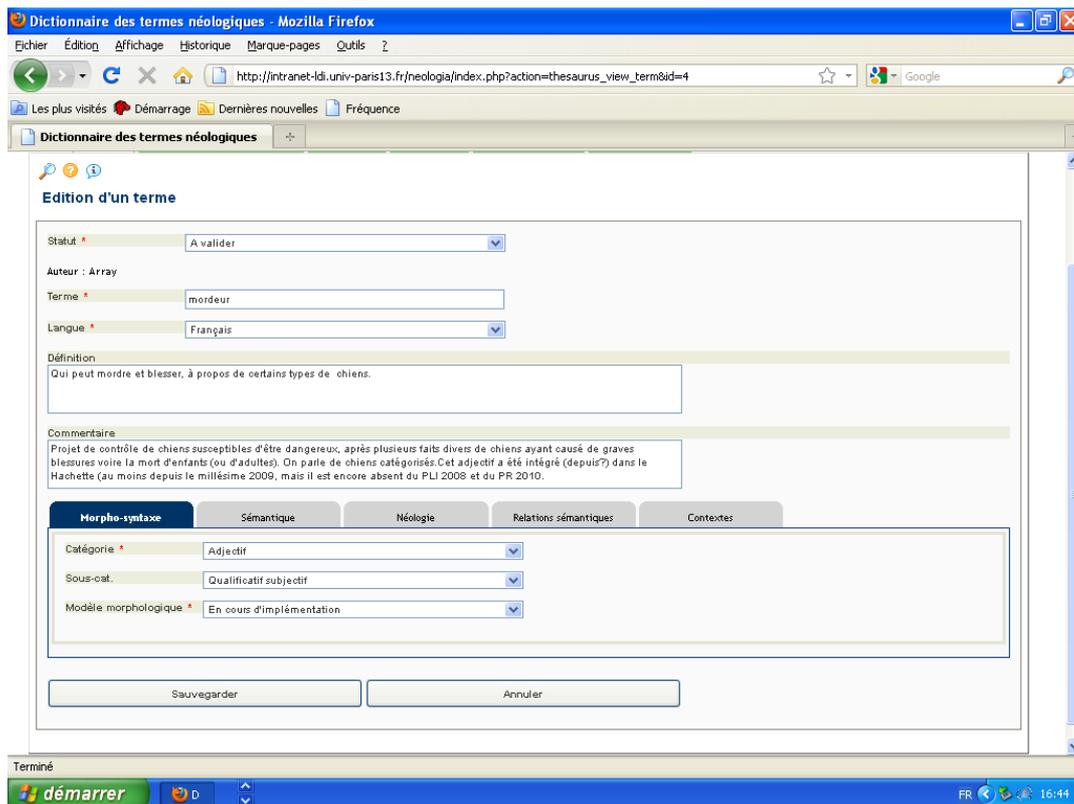


fig 2

Celle-ci comporte une zone « principale », avec le néologisme sous forme lemmatisée ainsi que sa définition et, dans la zone « secondaire », cinq onglets comprenant divers groupes d'informations. Le dernier, consacré au contexte et à la situation de production, peut se démultiplier à l'infini dans une même fiche. C'est-à-dire que l'on peut ajouter autant de contextes que l'on veut pour un néologisme. Il n'est bien sûr pas question d'entrer tous les contextes où l'on retrouve un néologisme déjà entré. Ce serait une perte de temps inutile. Trois situations ont justifié l'aménagement de l'état initial de la base qui ne pouvait comporter qu'un seul contexte, pour en accueillir plusieurs dans une même fiche. Tout d'abord on a voulu se donner la possibilité de consigner des emplois antérieurs à celui qui a fait entrer le néologisme dans la base. On peut ainsi remonter dans le temps, éventuellement jusqu'à la première occurrence, le moment de la création où le néologisme est encore un hapax. Ensuite, on peut entrer des contextes particulièrement éclairants, en particulier quand des gloses, explicatives ou évaluatives, accompagnent l'émission du néologisme. Enfin, on peut consigner les emplois de néologismes qui ont encore ce statut dans une nouvelle année civile. De ce point de vue, nous nous sommes alignés sur la pratique des autres équipes du réseau Néologie des langues romanes, NEOROM, qui basculent sur la plateforme commune tous les éléments considérés comme néologismes, année civile par année civile. L'ajout d'un

nouveau contexte dans la base Neologia permet alors de sélectionner le néologisme pour l'année en cours en vue de son transfert sur la base commune.

### 1.3. La ou les pages contextes

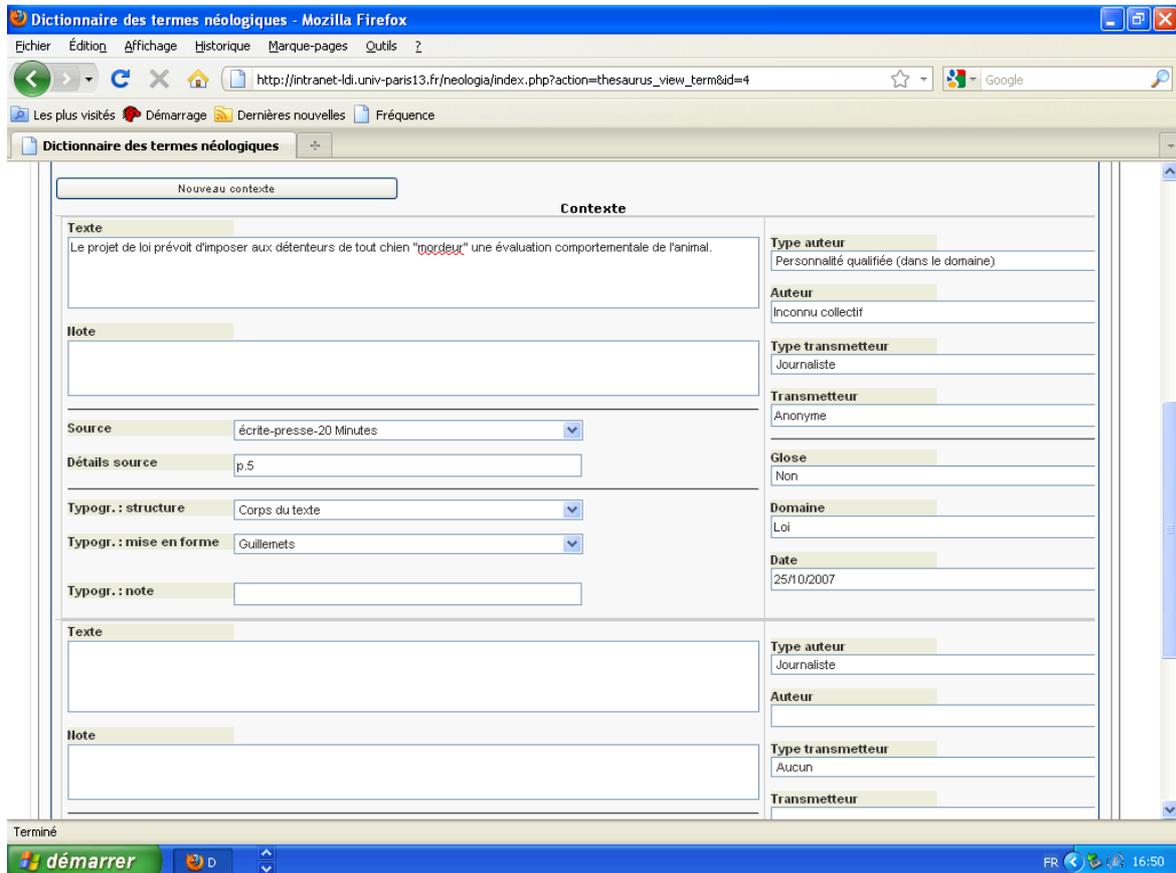


fig 3

Dans cette partie sont consignées toutes les informations permettant d'identifier les conditions d'énonciation : la source, orale ou écrite, précise avec la date d'émission, le contexte où figure le néologisme, c'est-à-dire la phrase ou le paragraphe où il apparaît, le contexte d'énonciation éventuellement (circonstances locales, temporelles...), l'énonciateur qui peut se confondre ou non avec le créateur du néologisme et nous distinguons les deux quand c'est nécessaire, avec des informations sur leur statut respectif (journaliste, homme politique, spécialiste du domaine, sans statut spécial...). À côté de ces informations d'état civil pourrait-on dire, sont également relevées des informations sur la présence d'une glose, et, pour les néologismes écrits, sur leur place dans le texte (corps du texte, titre, sous-titre, chapeau, légende, etc.) et sur leurs attributs graphiques (guillemets, italiques, gras,

soulignement, etc.). Tous ces champs peuvent être interrogés, avec l'obtention de la liste de tous les items qui correspondent à la requête.

À la différence du domaine couvert par ce cinquième onglet réservé à un ou plusieurs contextes, le contenu de tous les autres ne se dédouble pas mais est unique car y sont consignées des informations concernant le néologisme même, communes à tous les contextes où il apparaît.

#### 1.4. L'onglet morpho-syntaxe

The screenshot shows a web browser window titled "Dictionnaire des termes néologiques - Mozilla Firefox". The address bar shows the URL: [http://intranet-ldi.univ-paris13.fr/neologia/index.php?action=thesaurus\\_view\\_term&id=4](http://intranet-ldi.univ-paris13.fr/neologia/index.php?action=thesaurus_view_term&id=4). The page content is titled "Edition d'un terme". It features several input fields and dropdown menus: "Statut" (A valider), "Auteur" (Array), "Terme" (mordeur), "Langue" (Français), "Définition" (Qui peut mordre et blesser, à propos de certains types de chiens.), and "Commentaire" (Projet de contrôle de chiens susceptibles d'être dangereux, après plusieurs faits divers de chiens ayant causé de graves blessures voire la mort d'enfants (ou d'adultes). On parle de chiens catégorisés. Cet adjectif a été intégré (depuis?) dans le Hachette (au moins depuis le millésime 2009, mais il est encore absent du PLI 2008 et du PR 2010.)). Below the form is a tabbed interface with "Morpho-syntaxe" selected, showing "Catégorie" (Adjectif), "Sous-cat." (Qualificatif subjectif), and "Modèle morphologique" (En cours d'implémentation). At the bottom are "Sauvegarder" and "Annuler" buttons. The footer reads "A propos du dictionnaire terminologique NEOLOGIA".

fig 4

Le premier domaine, traditionnel et attendu, est consacré à l'analyse en parties du discours (nom, verbe, adjectif, adverbe essentiellement, mais pas exclusivement car on peut trouver aussi des actualisateurs qui sont des déterminants) et catégories grammaticales afférentes à chacune d'elles. Notons que le nombre, et le genre pour les adjectifs, ne sont pas ceux du contexte mais ceux du néologisme (sg/pl/indif et m/f/indif).

## 1.5. L'onglet sémantique

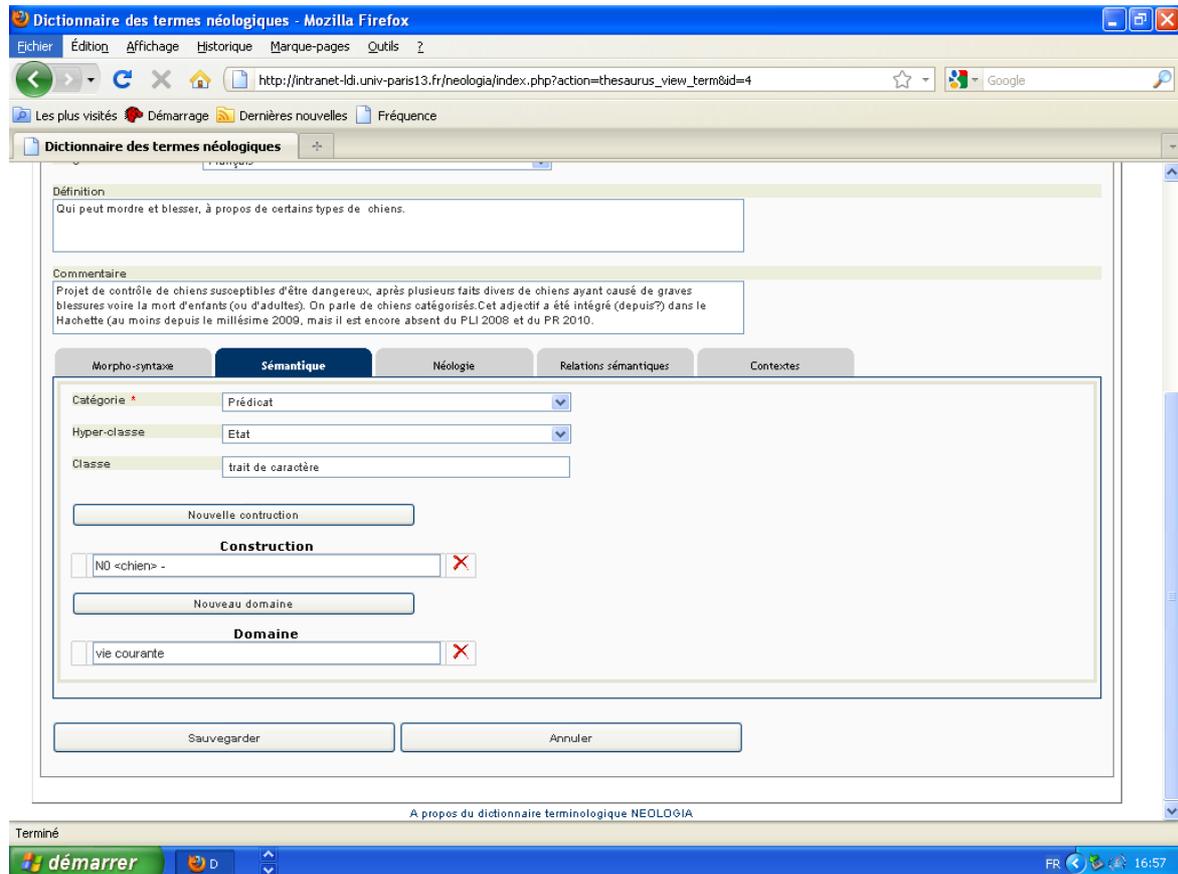


fig 5

Le second domaine est plus original et propre au laboratoire LDI dans la mesure où il consigne les propriétés syntactico-sémantiques des néologismes en termes de classes d'objets. Un néologisme est soit un prédicat soit un argument soit un actualisateur et il relève d'une des huit hyperclasses humain, animal, végétal, inanimé concret, inanimé abstrait, locatif, temps, événement. Suivent, pour les arguments, leur appartenance à une classe d'objets précise, et pour les prédicats, leur schéma argumental. Toute cette partie de l'analyse est encore en cours d'élaboration dans la base et devra se développer en concertation avec d'autres membres du laboratoire pour plus d'efficacité et d'interaction avec les autres dictionnaires de prédicats et d'arguments élaborés au LDI. L'adjectif *mordeur* est un prédicat relevant de l'hyperclasse des états ; il exprime un trait de caractère permanent et a comme seul argument l'animal qu'est le chien.

## 1.6. L'onglet néologie

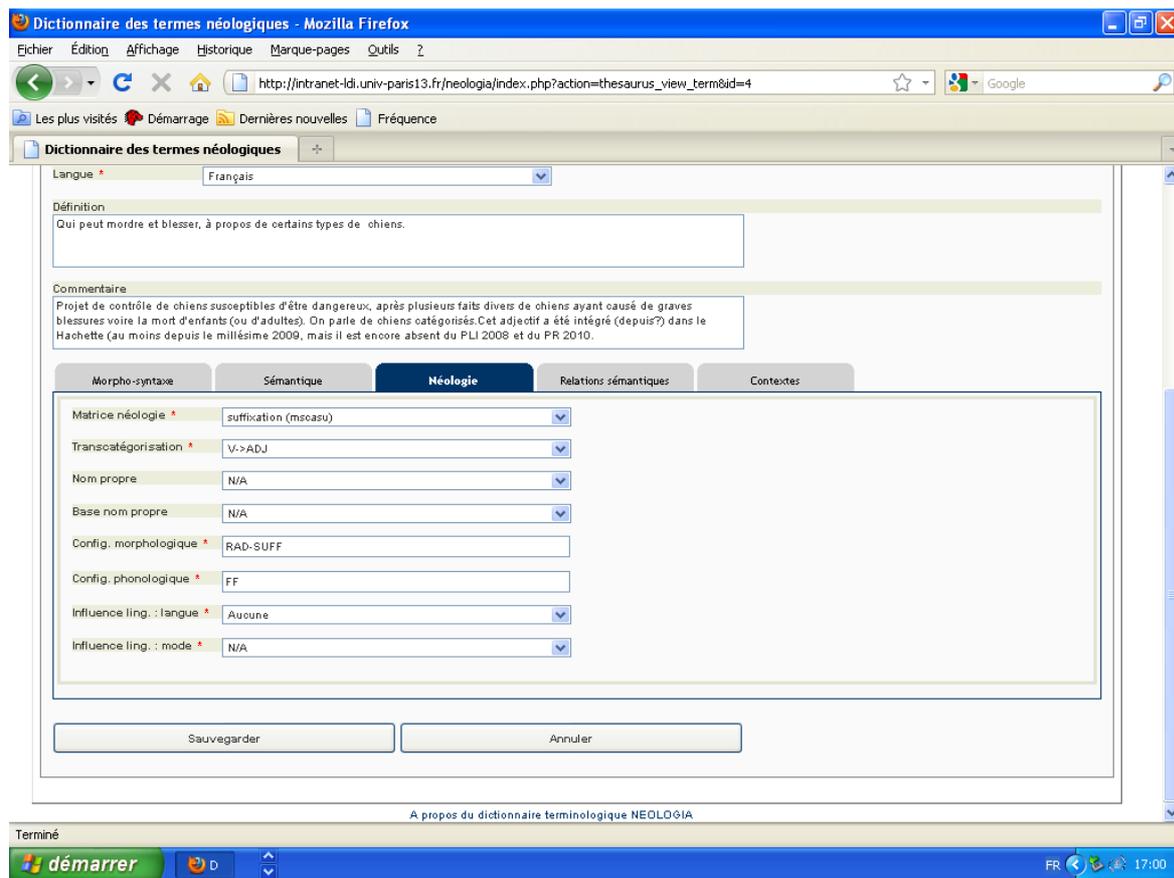


fig 6

Le troisième domaine est en revanche complètement analysé : il s'agit d'informations proprement néologiques : l'identification de la matrice lexicogénique, l'indication d'une transcatégorisation<sup>3</sup> éventuelle, la présence d'un nom propre comme base, la configuration morphologique puis la configuration syllabique et l'éventuelle influence d'une langue étrangère dans l'apparition de néologismes autres que les emprunts proprement dits<sup>4</sup>.

### 1.7. L'onglet relations sémantiques

Le quatrième domaine est consacré aux relations sémantiques de synonymie, antonymie, hyperonymie, etc. et aussi de traduction. Il n'est pas systématiquement renseigné

<sup>3</sup> Sous cette dénomination nous rangeons tous les cas où apparaît un changement de parties du discours entre le mot de base et le néologisme fait à partir de lui, quelle que soit la matrice qui opère ce changement de catégories grammaticales : préfixation (verbe *se dénouiller* sur l'adjectif *nouille* « bête »), suffixation (adjectif *tarantinesque* sur le nom *Tarantino*), dérivation inverse (verbe *prester* à partir des noms *prestation* ou *prestataire*), conversion (nom *la glisse* à partir du verbe *glisser* dont le radical est *gliss-*), etc.

<sup>4</sup> Ainsi *souris* (d'ordinateur) est créé en français par métaphore à partir du nom du petit rongeur, sous l'influence de l'anglais *mouse*. Les modes d'influence repérés jusqu'à ce jour sont : traduction (*souris / mouse*), calque (*gratte-ciel / sky scraper*), utilisation d'un formant d'origine étrangère (dans un composé hybride : *serial menteur*), utilisation d'une structure étrangère (en particulier l'ordre régressif des composés déterminant-déterminé : *royale attitude*), trou à combler (*hameçonnage* pour *phishing* par exemple).

pour l'instant, car ce n'est pas une priorité de l'équipe néologie, mais il pourra l'être quand on le voudra, là encore pour se trouver en synergie avec d'autres travaux du laboratoire.

## **2. L'exploitation du dictionnaire néologique comme base de données et de calculs**

L'exploitation du dictionnaire néologique Neologia comme base de données n'en est encore qu'à ses balbutiements, faute de temps et de bras pour tout faire, mais nous avons néanmoins procédé à quelques tests et plusieurs projets sont en cours, notamment dans le cadre de travaux d'étudiants de master et de thèse.

### **2.1. Exemple des adjectifs**

On peut ainsi, par exemple, si on travaille sur les adjectifs, obtenir, à un moment donné, le nombre d'adjectifs et calculer la proportion qu'ils représentent par rapport à l'ensemble des néologismes collectés. On en compte actuellement 307 sur 1466, ce qui représente approximativement 21 %, avec 272 adjectifs prédicatifs, qui représentent 18,5% de l'ensemble et 88,6% des adjectifs (les autres étant des adjectifs relationnels). On peut ensuite savoir immédiatement combien, parmi les adjectifs prédicatifs, sont obtenus par « transcatégorisation » à partir d'un mot de base appartenant à une catégorie grammaticale différente (combien sont formés à partir de noms : il y en a actuellement 98 sur les 120 adjectifs dénominaux, les autres étant des adjectifs relationnels ; de verbes : 10 actuellement dont *mordeur*, etc.), combien d'entre eux sont formés par préfixation, suffixation, conversion, etc. Le tableau ci-dessous indique ainsi, comme illustration, la liste des huit adjectifs prédicatifs avec une transcatégorisation verbe -> adjectif, créés par suffixation. En quelques clics, on peut ainsi radiographier l'état de la néologie adjectivale en français. C'est un progrès énorme pour les études néologiques. Un perfectionnement en cours d'élaboration consistera dans l'indication automatique du nombre d'occurrences des néologismes dans certaines sources (non encore arrêtées), ce qui permettra de mesurer leur circulation et leur diffusion.

The screenshot shows a web browser window titled 'Dictionnaire des termes néologiques - Mozilla Firefox'. The address bar shows the URL 'http://intranet-ldi.univ-paris13.fr/neologia/index.php'. The page content is divided into several sections:

- Filters:** A series of dropdown menus for various criteria: Définition, Note, Matrice (set to 'Suffixation'), Transcat. (set to 'V->ADJ'), Conf. morph., Nom propre, Base IIP, Conf. phon., Infl. ling. (set to 'Tous les éléments'), Mode infl. (set to 'Tous les éléments'), Statut (set to 'Tous les éléments'), and Auteur de la fiche (set to 'Tous les éléments').
- Search and Display:** A 'Rechercher' button and a display setting 'Affichage/page: 20'.
- Results Table:** A table with 8 rows and 7 columns. The columns are: Langue, Terme, Cat. Synt., Cat. Sem., Matrice, Statut, and Auteur de la fiche. Each row has a small icon on the left.

Langue	Terme	Cat. Synt.	Cat. Sem.	Matrice	Statut	Auteur de la fiche
FR	mordeur	Adjectif	Prédicat	mscasu	A valider	Inconnu collectif
FR	mélangiste	Adjectif	Prédicat	mscasu	A valider	Catherine Firmin-Didot
FR	Boboisant	Adjectif	Prédicat	mscasu	A valider	Anonyme
FR	Césarisable	Adjectif	Prédicat	mscasu	A valider	Inconnu
FR	Oscarisable	Adjectif	Prédicat	mscasu	A valider	Inconnu
FR	Viabiliseur	Adjectif	Prédicat	mscasu	A valider	Nicolas Schmaltz
FR	Différencialiste	Adjectif	Prédicat	mscasu	A valider	Rokhaya Diallo
FR	respirant	Adjectif	Prédicat	mscasu	A valider	

At the bottom of the page, there is a footer: 'A propos du dictionnaire terminologique NEOLOGIA' and a system tray showing 'Terminé', 'démarrer', and the time '17:07'.

fig 7 (tout ne tenant pas sur un écran, le haut du tableau est absent)

## 2.2. Une petite difficulté, surmontable

Une petite difficulté pratique vient du fait que, la base étant par nature évolutive, avec des ajouts permanents de nouveaux néologismes, les résultats des calculs peuvent évoluer au fil des jours. Mais il ne s'agit pas d'une difficulté rédhibitoire ou insurmontable. D'une part, si les chiffres bruts augmentent avec le développement de la base, les proportions ne varient sans doute que dans de faibles proportions, mais seules plusieurs études étalées dans le temps permettront de le confirmer. D'autre part, il faut prendre garde à opérer tous les calculs dans un laps de temps court afin que tous soient faits sur des données identiques. L'exploitation des résultats obtenus peut, elle, être effectuée sur une longue période sans dommage. Enfin on peut sélectionner un ensemble clos avec la sélection de dates (une ou des années civiles achevées).

## 2.3. Des choix à opérer dans les requêtes...

Mais pour cela il faut préalablement résoudre le problème, bien réel, qui réside dans le nombre sinon infini du moins dépassant de loin les capacités humaines d'un individu et même d'une équipe de procéder à toutes les requêtes multicritères qui s'offrent : il y en a des centaines de milliers. Il faut donc opérer des choix, et là ce n'est que l'esprit de finesse des chercheurs qui peut les guider dans les requêtes susceptibles de révéler au mieux la néologie d'une langue, et qui sont donc à opérer prioritairement, lors de la première phase de l'étude, afin qu'elles s'appliquent toutes aux mêmes données.

#### ***2.4. ... en fonction d'objectifs préétablis***

Et l'établissement de ces requêtes multicritères dépend fondamentalement de l'objectif que l'on se fixe au départ, avec essentiellement le choix entre l'étude de la créativité lexicale des locuteurs natifs à un moment donné ou celle de l'évolution du lexique de cette même langue au même moment. Dans le premier, cas tous les items de la base sont exploités. Dans le second, seuls ceux qui montrent une certaine diffusion doivent être retenus. Dans la mesure où nous sommes lexicologues et non lexicographes, ce sont les créations lexicales et les moyens mis en œuvre pour cela qui nous intéressent, plus que les seules lexies qui se diffusent. C'est d'ailleurs aussi le choix opéré dans le groupe Neorom initié et dirigé par Teresa Cabré. Avouons aussi que cette solution est plus facile à mettre en œuvre dans la mesure où la sélection des lexies qui se diffusent est épineuse, malgré les moyens disponibles désormais de mesure de la diffusion dans de grands corpus, mais ces informations, longues à mettre en place, ne fournissent que des indices et les décisions sont à prendre au cas par cas, par l'examen du nombre des attestations mais aussi du nombre et de la qualité des supports différents où ils apparaissent.

#### **Conclusion**

Au terme de cette présentation, on peut dire que c'est une petite révolution dans la création et l'exploitation de dictionnaires de néologismes qui a été rendue possible par le développement des nouvelles technologies. L'avantage réside autant dans l'automatisation ou la rapidité d'exécution de certaines tâches que dans la nécessité de préciser les objectifs et les méthodes de travail afin que l'outil informatique puisse fonctionner. C'est un outil stimulant qui permet de concrétiser des rêves d'analyse formulés il y a plusieurs décennies. Mais la

machine ne travaille pas seule, et, malgré le désir et l'enthousiasme, les forces et le temps manquent souvent pour tirer tout le parti de ces infinies possibilités.

### **Indications bibliographiques**

Cartier Emmanuel et Sablayrolles Jean-François (2008), « Néologismes, dictionnaires et informatique », *Cahiers de lexicologie* n° 93, 2008-2, p. 175-192.

Cartier Emmanuel et Sablayrolles Jean-François (2010) « Neologia, une base de données pour la gestion des néologismes », actes du Congrès Cineo (néologie des langues romanes) de Barcelone (7-10 mai 2008), 2010, p. 757-766 (759-767 en fait).

Cartier Emmanuel et Sablayrolles Jean-François (à paraître), « Nouvelles technologies, nouveaux modèles linguistiques et néologie », Journée de l'Association des sciences du langage (ASL) décembre 2009

Gril (Gabriel Merle, Robert Perret, Jennifer Vince, Claudie Juilliard) (1990, 1992, 1994) *Néologie Lexicale*, n° 2, 4, 6, UFR EILA, Université Paris-7 Denis Diderot et (1989) *Les mots nouveaux apparus depuis 1985*, Belfond.

Sablayrolles Jean-François (2000), *La néologie en français contemporain*, Champion.