HAL open science

Traitement et suivi de la diffusion de néologismes repérés par des extracteurs

Jean-François Sablayrolles

To cite this version:

Jean-François Sablayrolles. Traitement et suivi de la diffusion de néologismes repérés par des extracteurs. Neologica : revue internationale de la néologie, 2019, La néologie à l'ère de l'informatique et de la révolution numérique, 13, pp.55-69. hal-03086047

HAL Id: hal-03086047

https://hal.science/hal-03086047v1

Submitted on 22 Dec 2020

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.

HAL JFS 79 2019c/
Sablayrolles Jean-François (USPC et laboratoire HTL UMR 7597)
« Traitement et suivi de la diffusion de néologismes repérés par des extracteurs », Journées d'études organisées par Silvia Domenica Zollo, Naples 23 mai 2017, *Neologica* 13, 2019, p. 55-69

Résumé
Malgré ses avantages sur la collecte manuelle, la veille néologique automatisée sur grands corpus reste imparfaite, du fait de lacunes des corpus d'exclusion, de problèmes d'homonymie, de nouveaux emplois de mots existants…. Elle peut néanmoins être améliorée. L'analyse des néologismes identifiés comme tels, hapax révélateurs de la créativité individuelle ou néologismes qui se diffusent largement, pose aussi de nombreux problèmes que seuls des humains compétents peuvent résoudre.

Mots-clés : néologismes, veille néologique, corpus d'exclusion, hapax, homonymie

Abstract
In spite of its advantages over manual extraction, automatic monitoring of neologisms is less than satisfactory, due to incomplete exclusion corpora, problems of homonymy, new uses of existing words… It can however be improved. The analysis of neologisms identified as such, hapax revealing individual creativity or wide-spread neologisms pose many problems that only a competent human can deal with.

Keywords: neologisms ; monitoring neologisms ; exclusion corpus: hapax: homonymy

Traitement et suivi de la diffusion de néologismes repérés par des extracteurs

Introduction

L'objectif de cette contribution n'est pas mince puisqu'il ne s'agit pas moins que de passer en revue des méthodes de veille néologique, leurs faiblesses[1], des voies pour tenter d'y remédier, et cela en fonction des objectifs qu'on s'est fixés. Pour faire simple, soit on s'intéresse à l'évolution du lexique de la langue, en ne prenant en compte que les néologismes qui se diffusent et perdent, à plus ou moins brève échéance, ce statut, soit on s'intéresse à la créativité lexicale des locuteurs et on ne fait pas de sélection en fonction de la diffusion postérieure à la création.

Nous nous concentrerons sur l'étude de la diffusion dans la presse généraliste française de néologismes repérés par des extracteurs (Néoveille, Logoscope, Pompamo…)[2] en vue de leur analyse dans des bases comme Neologia ou Néoveille. La première, qui se contentait de l'analyse des néologismes qu'on y intégrait manuellement, est désormais en partie obsolète puisque la seconde entièrement conçue par Emmanuel Cartier se charge de la veille néologique à partir de données de presse disponibles (des fils RSS) en proposant des candidats néologismes (mais avec encore beaucoup de bruit) et a retenu une grande partie des

[1] Leurs indéniables avantages sont bien connus et sont rappelés brièvement au début de la première partie : elles permettent de traiter des données incomparablement plus nombreuses, ne sont pas sujettes à des erreurs dues à l'inattention ni à des fluctuations du sentiment néologique, par exemple.

[2] Pour tous ces outils, nous renvoyons aux présentations faites par leurs créateurs et indiquées dans la bibliographie : Emmanuel Cartier (seul ou avec des collaborateurs pour Neologia et Néoveille) ; Christophe Gérard pour le Logoscope, et l'ATILF pour Pompamo. Pour l'extraction automatique, nous renvoyons aussi aux travaux de Maria-Teresa Cabré, d'Antoinette Renouf et de Patrick Drouin pour les termes. Ces travaux sont cités dans la bibliographie.

champs d'analyse que l'auteur de cet article avait demandé au même Emmanuel Cartier d'intégrer dans la construction de la base Neologia.

1. Les difficultés de l'extraction automatique des néologismes et des pistes

L'extraction manuelle a comme défaut de ne pouvoir traiter que des corpus de petite taille et d'être sujette aux intermittences de l'attention humaine. Le recours aux outils informatiques d'extraction sur de grands corpus est désormais incontournable. Mais il présente aussi des écueils qu'il faut connaître pour mieux s'en prémunir.

1.1. Les néologismes formels

La *doxa* qui veut qu'un néologisme formel soit facilement identifiable par rapport à des innovations sémantiques est souvent prise en défaut, par excès ou par omission.

1.1.1. Par excès : de « mauvais » candidats

Le premier point à prendre en compte est l'incomplétude des dictionnaires pris comme corpus d'exclusion. Cela fait déjà longtemps par exemple qu'André Dugas (1990 et 1992) a pointé les difficultés posées au traitement automatique des documents écrits par des lacunes systématiques des dictionnaires, et il proposait, pour y remédier, d'introduire systématiquement dans les dictionnaires électroniques des formes possibles absentes des dictionnaires traditionnels tels des verbes préfixés en *re-* ou des lexies préfixées par *auto-*. *Rehurler*, *redormir* ou *autoflagellation* seraient proposés comme candidats néologismes, ce qu'ils ne sont sans doute pas. Cette solution pose le problème de la productivité des règles et de la nouveauté[3]. Par ailleurs certains préfixés par *re-* sembleraient bien devoir être considérés comme néologismes si on les trouvait attestés, comme *reradicalisation* d'un *déradicalisé* par exemple. Une réflexion sur ce point des lacunes systématiques des dictionnaires conventionnels reste à entreprendre. Si *maurassien* n'est assurément pas néologique[4], qu'en est-il de *pohérisation* ou de *s'autogifler*, etc. ?

Un deuxième point est la réintroduction, volontaire souvent mais pas nécessairement, de mots qui ont existé dans la langue mais qui en ont disparu (nommés *paléologismes* par Sablayrolles 2000 et 2010). Comme ils ont disparu des dictionnaires du français contemporain pris traditionnellement comme corpus d'exclusion, ils seraient proposés par l'extracteur comme candidats néologismes, mais ils sont présents dans le *Littré*, le *Larousse du XXe s. en 6 volumes*, l'*Encyclopédie Quillet*, le *TLF*. C'est le cas de *frétillon* employé par Richard Jorif dans *Le burelain*, dont les lecteurs pourraient penser que c'est une création de l'auteur par suffixation du verbe *frétiller*. Cependant l'abondance de cas de ce type dans l'œuvre met en garde le lecteur attentif qui repère ces paléologismes comme *sonica*, *la pelle se moque du fourgon*[5], etc.

Si ces cas de présentation de « mauvais » candidats néologismes ne sont pas très graves, puisque l'expert humain, indispensable, peut corriger l'erreur à condition d'être très vigilant,

---

[3] Voir les actes de la Journée ConSciLa dans *Neologica. Revue internationale de néologie*, 12, 2018.

[4] C'est pourtant ce que pensait un informaticien qui s'occupait de construire un système de veille néologique au Laboratoire de Linguistique Informatique de Paris 13 (LLI), il y a plus de vingt ans, et qui était contredit par ses collègues linguistes.

[5] Les contextes de ces trois exemples sont : « Julien fut aussitôt courtisé. Deux fillettes l'accaparèrent, un peu trop frétillones au goût de Frédéric » (p. 209) où il y a néanmoins une innovation flexionnelle puisque *frétillon* est normalement épicène ; « Il travaillait pour élever son fils et payait ses impôts sonica », p. 207, c'est-à-dire 'à point nommé, justement' ; « Mamitate fit cependant provende, sans l'avouer, d'un petit nombre de moralités tranchantes […] 'la pelle se moque du fourgon' qui avait la caution de Mme de Sévigné » et qui se dit d'une personne ridicule qui se moque d'une autre qui ne l'est pas moins. C'est synonyme de *C'est l'hôpital qui se moque de la charité*.

les cas d'oublis de la machine sont moins connus et plus graves, car moins facilement rattrapables.

1.1.2. Par défaut : des candidats oubliés

De véritables néologismes peuvent en effet être homonymes de lexies attestées dans les corpus d'exclusion et échapper ainsi au signalement. L'autre erreur serait de considérer, si on les repère, qu'il s'agit alors nécessairement de néologismes sémantiques. Or, dans de nombreux cas, ce sont de véritables néologismes formels, dont il existe plusieurs types.

Le cas le plus évident et le moins contestable est celui où les lexies homonymes sont construites sur des bases différentes. Les deux verbes *auditer* viennent l'un de la conversion du nom *audit* (il est conventionnellement attesté et présent dans les dictionnaires) et l'autre de la dérivation inverse d'*auditeur* dans *un auditeur sachant auditer*, fabriqué sur le modèle du virelangue *un chasseur sachant chasser*.

Parfois le néologisme est un « mot complexe construit » – dans la terminologie de Danielle Corbin (1987, 1988), qui a largement fait école – homonyme d'un mot complexe non construit conventionnel. Le verbe néologique *dé-nommer* « ôter d'une liste de promouvables » analysé par Michel Arrivé ([2005] 2010) est un préfixé à sens négatif (*dé-*) sur *nommer* « inscrire quelqu'un sur une liste de promotion » qui se trouve être homonyme du verbe conventionnel *dénommer* « donner un nom à quelqu'un ou quelque chose ».

Dans une troisième situation, la base peut être identique mais la structure dérivationnelle différente. C'est ainsi que Danielle Corbin (1990) montre qu'il existe au moins quatorze mots *antialcoolique* différents, ayant chacun sa structure morphologique propre avec le sens associé à chacune de ces structures. Sur ces homonymes, elle montre que quatre sont attestés[6], même s'il n'y a au mieux qu'une entrée dans les dictionnaires traditionnels.

L'homonymie peut résulter aussi, et c'est plus délicat à identifier, de l'application à une même base d'emplois différents d'un même mot. Le néologisme *filmer* « recouvrir d'un film protecteur » un bol de mayonnaise (Arrivé [2005] 2010) est la conversion du mot *film* « pellicule protectrice » en verbe de la même manière que l'est le plus ancien et conventionnel *filmer* « enregistrer des images sur une pellicule de cinéma » à partir de *film* « pellicule photosensible susceptible d'enregistrer des images ». Aucun lien sémantique direct ne permet de passer du sens attesté conventionnellement au sens du néologisme : ce sont deux lexies homonymes « enregistrer des images » et « protéger ».

Encore plus délicats à trancher – et nous sommes revenu sur les premières analyses, de type sémantique, que nous avions longtemps faites nous rendant compte qu'elles étaient fautives, ou du moins peu plausibles, et leur avons substitué une origine formelle, par dérivation – sont les cas comme « *la prise de train* » qui constitue une entorse à l'emploi du nom *prise* qui ne peut avoir comme compléments que des toponymes dont on s'est emparé du référent (*la prise de Troie*), des médicaments (*la prise régulière d'antibiotiques est grosse de risque*), etc. mais qui excluent la nourriture (**la prise d'un sandwich calme la faim*), les voies de circulation (**la prise de l'autoroute fait gagner du temps*), les moyens de transport (**la prise de l'avion est préférable à *la prise d'un bateau*). Il est en effet peu probable qu'il y ait là une métaphore affirmant l'équivalence du fait de prendre le train à la prise d'un médicament, même dans ce contexte d'une publicité de la SNCF où il est question de santé : « la prise de train bénéficie à la santé de votre voiture ». Il s'agit beaucoup plus probablement

---

[6] En voici la structure morphologique et le sens :
i) 'qui combat l'usage de l'alcool' : [(anti)af [alcool]N]Adj +ique (intégrateur paradigmatique)
ii) 'qui combat l'alcoolisme' : [(anti)af [[[alcool]N(ique)af]Adj ((isme)af]N]A + ique (intégrateur paradigmatique)
iii) 'personne qui combat l'alcoolisme' : [[(anti)af[[[alcool]N(ique)af]A(isme)af]]N]A +ique]N
iv) 'qui a pour objet la lutte contre l'alcoolisme : [[(anti)af[[[alcool]N(ique)af]A(isme)af]N]A+ique(isme)af]N(ique)af]A

de la nominalisation de *prendre le train*, qui est une construction parfaitement régulière. Nous venons de lire dans un mémoire de Master : « la soumission d'un questionnaire ». Or le sens de *soumission* attesté conventionnellement est celui de « dépendance », qui ne convient pas à un questionnaire. Il s'agit là aussi de la très probable nominalisation de *soumettre*, régulièrement employé pour *soumettre un questionnaire*. Les noms d'action *prise* et *soumission* sont en effet les noms correspondant aux prédicats verbaux *prendre* et *soumettre*, même si ce ne sont pas à proprement parler des dérivés morphologiques.

Notons encore le problème des recréations, étalées dans le temps ou quasiment simultanées. Un locuteur peut très bien recréer une lexie non conventionnelle dont il ignore qu'elle a déjà été créée par quelqu'un d'autre avant lui. Le critère chronologique fait comme si tous les emplois d'un néologisme étaient des réemplois de la première occurrence. Dans un certain nombre de cas, c'est sans doute faux et des créations parallèles sont souvent probables, comme l'attestent la surprise de celui qui croyait avoir inventé *comater* lorsqu'il découvre que ce verbe circule depuis longtemps, ou la triple création de *examination* par trois élèves dans un devoir en temps limité, sans que les circonstances matérielles de l'examen puissent faire imputer cette similitude au copiage des élèves les uns sur les autres : c'est simplement la nominalisation de « elle examinait la lettre » qui figurait dans le texte à commenter, extrait de *Belle du Seigneur* d'Albert Cohen.

1.2. Les néologismes dits sémantiques et les néologismes d'emploi

Les néologismes sémantiques sont traditionnellement réputés plus difficiles à identifier, ce qui est probablement faux puisque les vrais néologismes sémantiques sont sans doute très repérables, nettement plus que certains néologismes formels complètement réguliers : *une voiture incontinente*[7] ne passe pas inaperçu à la différence des suffixés *fienteux* ou *feuilleteur* créés par Laforgue[8]. Mais ceci est vrai pour un humain, pas pour une machine qui se fonde uniquement sur la présence ou l'absence d'une unité graphique dans un corpus d'exclusion.

Des recherches récentes d'extraction automatique de candidats néologismes d'emploi (Gérard & Kabatek 2012) se fondent sur la présence inattendue d'une lexie dans le cotexte large où elle se trouve car elle relève d'un autre domaine. Cette intrusion d'une lexie dans un domaine qui n'est pas traditionnellement le sien constitue un indice de nouveauté. Ainsi un terme culinaire comme *caviar* dans des articles de sport se révèle une innovation d'emploi quand il est question d'une bonne passe dans une partie de football (Rastier & Valette 2009). Mais pour cela, il faut que les domaines d'emploi de la lexie soient préalablement définis. Or il n'y a pas à notre connaissance de telles bases qui aient été systématiquement constituées. Le recours à l'extraction des mots lexicaux utilisés dans les articles de dictionnaires est une solution qui a été mise en œuvre (Reutenauer 2012), mais il faut sans doute une vérification humaine. Sans faire de 'dictionnaire bashing', ceux-ci présentent des lacunes et des erreurs, et surtout ils n'ont pas été conçus pour ce type d'utilisation. S'ils l'étaient, les informations entrées seraient probablement différentes.

Sans doute le repérage d'une infraction à la distribution d'une lexie est-il plus probant. Il s'agit d'infraction au cotexte étroit, mais si ce repérage peut fonctionner sur des phrases courtes, dans des domaines spécialisés, il semble impossible à mettre en œuvre pour des phrases longues, des textes littéraires qui posent toute une série de problèmes. C'est précisément l'intérêt du modèle des classes d'objets de Gaston Gross pour les études

---

[7] Expression employée pour décliner l'offre de remiser une voiture dans un box privé du fait des fuites d'huile du véhicule.

[8] C'est le correcteur orthographique qui a attiré notre attention sur ces créations qui avaient échappé à notre vigilance dans la préparation d'un cours pourtant consacré au lexique des *Complaintes* de Laforgue. *Fienteux* se trouve dans *La complainte de l'orgue de barbarie* : « Des berceaux fienteux / Aux bières de même / Bons couples sans gêne, / Tournez deux à deux. » *Feuilleteur* se trouve dans *À Paul Bourget* : « Puissent les feuilleteurs du quai / En rentrant se r'intoxiquer / De vos AVEUX, ô pur poète ! / C'est la grâce que je m'souhaite. »

néologiques qui nous avait conduit à rejoindre le Laboratoire de Linguistique Informatique (LLI) de Paris 13. Mais les domaines décrits (voies de communications, humains, coups…) et les ressources disponibles sont assez restreints, trop en tout cas pour servir de base à l'extraction automatique des néologismes d'emploi. Les dictionnaires explicatifs et combinatoires imaginés par Mel'čuk seraient également efficaces mais ils sont loin de couvrir toute la langue.

Avant de passer à la présentation sinon de solutions du moins de voies dans lesquelles aller pour améliorer le système d'extraction, il faut encore prendre en compte l'existence de faux néologismes sémantiques. Ne revenons pas sur les cas examinés précédemment de néologismes homonymiques ou de nominalisations comme *la prise de train*. Il reste que beaucoup d'emplois différents d'emplois plus anciennement attestés relèvent d'évolutions lentes et insensibles du sens et pas de néologie sémantique à proprement parler.

Le départ entre des néologismes sémantiques et des évolutions non néologiques n'est pas toujours fait. Il l'est même rarement, avec la tendance à tout reverser indistinctement dans l'une ou l'autre catégorie. Ullmann ([1952] 1969) explique tout par des tropes – et c'est ce que fait aussi la tradition (ainsi que la linguistique cognitive à la suite de Lakoff et Johnson ([1980] 1986)) – qui relèvent de la néologie sémantique alors que la néosémie de Rastier et Valette (2009) traite toutes les évolutions sémantiques sans distinguer de néologie sémantique. Les deux mécanismes explicatifs concurrents ne sont pas exclusifs l'un de l'autre, mais ils s'opposent généralement par le fait que les évolutions de sens sont nombreuses, inconscientes, perceptibles seulement a posteriori, durables et largement répandues (Nyckees 1998) alors que les néologismes sémantiques sont rares, volontaires, perçus comme tels, au moment de leur création ou de leur réemploi, avec un poids stylistique net et demeurent souvent des hapax (Sablayrolles 2018).

1.3. Des solutions

Une solution à l'extraction automatique des néologismes par corpus d'exclusion réside dans l'actualisation permanente de celui-ci, fondée sur la compétence humaine, comme c'est possible sur Néoveille. D'une part un certain nombre de fautes sont mises dans un filtre pour qu'elles ne soient plus proposées comme candidats néologismes, d'autre part, le dictionnaire est complété par des lexies non néologiques qui en étaient absentes. Petit à petit le fonctionnement du système s'améliore et le nombre de candidats néologismes indus diminue. Cette manière de procéder prendra du temps mais c'est sans doute le plus expédient à l'heure actuelle, et l'efficacité ira en augmentant.

Une autre manière de procéder est de comparer deux corpus homogènes d'époques différentes : les lexies qui apparaissent seulement dans le second sont les candidats néologismes. Mais il faut se méfier de l'incomplétude et des lacunes « accidentelles » toujours possibles du premier corpus, si grand soit-il. Par ailleurs l'homogénéité complète est toujours sujette à caution. En outre, un hapax n'est pas nécessairement un néologisme. Enfin la néologicité d'une lexie peut parfois durer longtemps si elle ne connaît pas une large et rapide diffusion. Cela peut aller jusqu'à plusieurs dizaines d'années (les moyennes d'âge des mots entrés dans les dictionnaires attestent ce caractère lent de la perte de la néologicité et aussi, il faut le reconnaître, une certaine frilosité – souvent justifiée – des lexicographes). Il peut donc se faire qu'une lexie présente dans le premier corpus, pris comme corpus d'exclusion, soit encore néologique dans le corpus récent. L'acception « collaboration » de *synergie* a ainsi mis près d'un siècle à devenir conventionnelle entre ses premières attestations, fin XIXe, et l'explosion de sa diffusion fin XXe siècle. Et le mot existait en biologie depuis le XVIIIe siècle. Il n'est donc pas sûr que ce système soit supérieur au corpus d'exclusion traditionnel, surtout si celui-ci est mis à jour régulièrement. Ceci vaut

essentiellement pour la néologie formelle, avec néanmoins le maintien de certains problèmes qui ne peuvent être traités que de manière manuelle, par un humain.

Mais la mise en évidence d'incongruités distributionnelles avec le co-texte proche et le contexte de l'énoncé (le domaine du texte par rapport au sémantisme de la lexie) permettront sans doute de pointer aussi bien des néologismes homonymiques, que des extensions d'emploi ou de véritables néologismes sémantiques. Ce sont des pistes à approfondir, comme est en train de le faire le travail en cours sur la plateforme Néoveille créée par Emmanuel Cartier.

2. Les difficultés d'une analyse fine des néologismes

Une fois repérés, les néologismes doivent être analysés et c'est toute une nouvelle série de problèmes qui se posent. Les grilles d'analyse diffèrent selon les groupes qui pratiquent la veille néologique. Certaines solutions ou simplifications adoptées ici et là sont sujettes à caution ou au moins à discussion, et l'éventuelle conjonction de plusieurs niveaux d'innovation doit être prise en compte.

2.1. Quels champs d'analyse retenir ?

La longueur des fiches d'analyse et le nombre des informations pour chaque néologisme varient grandement selon les équipes. Il peut y avoir cinq / six champs ou plusieurs dizaines (il y en a plus d'une trentaine dans la base Neologia, construite par Emmanuel Cartier à partir du cahier des charges précis établi par l'auteur du présent article), mais leur nombre a été légèrement restreint dans Néoveille ; on ne cesse cependant d'en réintroduire, du fait de leur nécessité ou de leur intérêt. Ces différences viennent, comme nous l'avons indiqué en introduction, de la différence des objectifs, pas toujours bien explicités : analyse de la créativité lexicale ou évolution du lexique de la langue, avec parfois le sous-objectif de confectionner des dictionnaires. Dans ce cas les informations requises, de type lexicographique, sont moindres que celles exigées par une étude de la créativité lexicale sous tous ses aspects. L'examen sous toutes ses coutures du néologisme prend d'autant plus de temps que la tâche n'est guère automatisable et que certains champs exigent une réflexion parfois approfondie avant d'être remplis, même quand on est aguerri, à plus forte raison quand on ne l'est pas. On manque de bras et de têtes bien formés à la tâche. Celle-ci excède ce qu'un individu seul peut faire et l'homogénéisation des résultats d'analyse des membres d'une équipe n'est pas facile à obtenir, comme le montrent les enquêtes que nous avons conduites à ce sujet (Sablayrolles 2003 ; Ben Hariz 2009). Mais ce n'est pas inaccessible.

2.2. Simplicité apparente et véritable complexité : éviter des confusions

La naïveté d'informaticiens qui pensaient identifier automatiquement la matrice du néologisme en même temps que son extraction repose sur une méconnaissance de la complexité des faits et la confusion entre l'apparence morphologique de surface et le ou les procédés à l'œuvre dans l'émergence du néologisme. Il n'est que de rappeler les analyses de *antialcoolique* de Danielle Corbin (1990) pour comprendre qu'un même signifiant apparent correspond à une multiplicité de schémas morphologiques différents assortis chacun d'un sens différent (voir supra la note 6 avec la structure et le sens des quatre homonymes attestés sur quatorze lexies possibles). Le sens des exemples que nous citons souvent de *désagrémenter*[9]

[9] Ce verbe figure dans l'incipit du roman *Le burelain* de Richard Jorif :
« – Quelle affaire te désagrémente ? se demanda, tout haut, Frédéric.

et *asexualisantes*[10] conduit à analyser le premier comme construit par conversion du nom *désagrément* et non par préfixation du verbe *agrémenter* et le second comme construit par suffixation sur *asexualiser* et non par préfixation sur *sexualisant*. Construit par préfixation *désagrémenter* signifierait « ôter un embellissement », alors que son sens en contexte est « provoquer une sensation désagréable ». De même la construction par préfixation de *asexualisant* signifierait « qui ne sexualise pas » et impliquerait que les êtres dont on parle sont naturellement asexués. Ce qui n'est pas le cas puisqu'il s'agit d'une troupe de danseurs et danseuses dont le vêtement gomme le genre.

La solution rapide qui consiste à identifier dans des cas de ce type des bases avec un préfixe et un suffixe à côté des lexies créées par préfixation et de celles créées par suffixation mêle la configuration morphologique apparente et l'identification de la matrice morphologique[11]. Tous les pourcentages établis par types de procédés de formation ensuite pâtissent de cette confusion initiale.

D'autres confusions ont aussi des conséquences sur les conclusions présentées à propos de la néologie de telle ou telle langue à telle ou telle époque. Il en va ainsi, par exemple, de l'identification comme affixes des formants savants (*phil*(*e*), *anthrop*, *graph*(*e*), *log*(*ue*), etc., dont beaucoup d'arguments[12] conduisent à penser qu'il vaut mieux les considérer comme des éléments de composition. Du coup, dans certains tableaux établissant la répartition des néologismes par matrice, la composition n'apparaît pas. Ce qui paraît pour le moins contre-intuitif, et mériterait, de ce fait, d'être réexaminé.

La répartition par domaines pâtit également d'une confusion entre le domaine sémantique du néologisme et celui de l'énoncé dans lequel on le trouve, les deux ne coïncidant pas nécessairement. En cuisine, un nouvel instrument ménager, comme le *coupe-friture* inventé par Boris Vian (et Paul Goraguer) dans *La complainte du progrès (les arts ménagers*)[13], n'est pas un aliment, comestible. Tout ce qui relève du domaine de la cuisine ne se mange pas. Les indices pris en compte pour identifier le domaine des énoncés peuvent diverger et aboutir à des résultats différents. Alors que nous relevons beaucoup de néologismes sociétaux (en particulier dans les pages 'société' des quotidiens et hebdomadaires), ils sont minoritaires dans le Logoscope.

2.3. Les trois niveaux d'innovation des néologismes

Le problème du nombre des matrices responsables de l'émergence d'un néologisme est un vrai problème : n'y en a-t-il qu'une ou peut-il y en avoir plusieurs ? Et, dans ce cas, quelles sont les éventuelles restrictions, quantitatives et qualitatives ? La tendance à considérer simultanées plusieurs opérations successives et ou à confondre la structure morphologique et la matrice à l'œuvre (*UMPiste* est créé par suffixation sur un sigle et pas à la fois par siglaison et suffixation, *aneigissage* est fait par suffixation en *-age* sur la base possible non attestée *aneigir* et pas par préfixation et suffixation simultanées sur *neige*) nous avaient conduit à

Il lui arrivait parfois, dans l'éparpillement du réveil, surtout quand un point d'angoisse l'avait un peu travaillé, de commettre quelque méchant néologisme, dont il se repentait presque aussitôt »

[10] L'adjectif qualifie les tuniques portées par les danseurs et danseuses pour une chorégraphie, dans un article d'un critique de *Télérama*.

[11] Par ailleurs, un doute existe sur la création de lexies par ajout simultané d'un préfixe et d'un suffixe, procédé – appelé parasynthétique par A. Darmesteter à la fin du XIXe siècle – dont l'existence a été contestée par D. Corbin, avec une argumentation convaincante.

[12] Beaucoup d'entre eux peuvent figurer en première ou en dernière position ; ils appartiennent au stock sinon illimité du moins très vaste de tous les mots grecs et latins ; de ce fait leur sémantisme est de type lexical, autant de points qui les écartent du fonctionnement des affixes. Ils entrent comme formants de composés savants (ou parfois de composés hybrides) au fonctionnement un peu différent des autres types de composition.

[13] Il figure parmi les ustensiles que reprend le personnage de la chanson, après les avoir offerts à sa femme, au moment où il la répudie : la tourniquette / à faire la vinaigrette / le ratatine-ordure, etc.

restreindre à une seule le nombre des matrices à l'œuvre. Mais des cas récalcitrants et la lecture des travaux de Peter Koch (avec ou sans Paul Gévaudan, 2010 et 2000 respectivement) nous ont convaincu que plusieurs matrices pouvaient co-exister, mais relevant de groupes différents. Une matrice formelle peut co-agir avec une matrice sémantique (*pélicanesque* est créé par suffixation et métaphore[14]) et une dimension stratique (entre autres des influences étrangères) peut également être présente, comme dans *escorteuse* qui combine les trois aspects : suffixé sur le verbe *escorter*, c'est un (faux ?) euphémisme[15] et l'existence de l'anglais *escort girl* doit jouer un rôle. Mais un seul ou deux de ces trois aspects peuvent être présents, l'autre ou les deux autres n'étant pas pertinents. Tout cela a une incidence sur le nombre des matrices identifiées et sur les calculs des pourcentages ensuite.

Conclusion

L'intérêt de la plateforme Néoveille créée par Emmanuel Cartier et auquel l'auteur du présent article collabore est de permettre aussi bien l'étude de la créativité lexicale (tout ce qui est relevé et reconnu néologisme par le linguiste), que celle de l'évolution de la langue en prenant en compte la diffusion réelle dans le temps et la diversité des supports où les néologismes apparaissent d'une part et d'autre part en pointant des évolutions sémantiques non tropiques de lexies existantes. Le suivi de la diffusion des néologismes est en effet un atout de cet outil très performant.

Il reste que l'analyse de la néologie est une tâche chronophage qui demande du personnel bien formé linguistiquement dans le domaine de la lexicologie en général et de la néologie en particulier et exige des discussions entre véritables experts pour prendre des décisions, le plus argumentées possible linguistiquement dans les cas épineux qui ne manquent pas d'apparaître. Mais cela est stimulant intellectuellement et fait avancer la réflexion et la connaissance.

Jean-François SABLAYROLLES
Université Paris 13 Sorbonne-Paris-Cité
Laboratoire d'Histoire des théories linguistiques (HTL), UMR 7597 CNRS

Références bibliographiques

ARRIVE Michel ([2005] 2010), *Verbes sages et verbes fous*, Limoges, 1ère éd., Limoges, Lambert-Lucas, puis 2nde éd. Belin.

ATILF, www.cnrtl.fr/outils/pompamo/.

BEN HARIZ Soundous (2009), « Diminuer les fluctuations du sentiment néologique », *Neologica. Revue internationale de néologie*, 3, Paris, Classiques Garnier, p. 37-51.

CABRE Maria-Teresa et YZAGUIRRE Lluis de (1995), « Stratégie pour la détection semi-automatique des néologismes de presse », *TTR : Traduction, Terminologie, Rédaction : Études Sur le Texte et ses transformations*, Montréal, PQ, Canada. TTR 1995, 8/2, p. 89-100.

CARTIER Emmanuel (2016), « Néoveille, système de repérage et de suivi des néologismes en sept langues », *Neologica. Revue internationale de néologie*, 10, Paris, Classiques Garnier, p. 101-131.

CARTIER Emmanuel (2017), « Néoveille, a Web Platform for Neologism Tracking », *Proceedings of the EACL 2017 Software Demonstrations*, Valencia, Spain, April 3-7 2017, p. 95-98.

[14] Du pélican du poème romantique qui se sacrifie pour ses enfants ; le sens est donc « pathétique ».
[15] Il ne s'agit pas d'accompagnatrice ni de dame de compagnie, mais d'un type de prostituée.

CARTIER Emmanuel (dans ce volume), « La créativité lexicale dans le français contemporain. Troncations, autres mécanismes néologiques, traitement informatique des données ».

CARTIER Emmanuel et VIAUX Julie (2018), « Étude de la pénétration des anglicismes de type N ou ADJ(-)*Ving* à partir d'un corpus contemporain journalistique : les exemples de *bashing* et *shaming* en français contemporain », *Emprunts néologiques et équivalents autochtones. Études interlangues*, Christine Jacquet-Pfau, Andrzej Napieralski et Jean-François Sablayrolles éd., Łódź, Presses de l'Université de Łódź, p. 11-34.

CARTIER Emmanuel et SABLAYROLLES Jean-François (2010), « 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), María Teresa Cabré, Ona Domènech, Rosa Estopà, Judit Freixa et Mercè Lorente éd., Barcelone, Université Pompeu Fabra, IULA, sèrie activitats 22, p. 757-766 (759-767 en fait).

CARTIER Emmanuel, Jean-François SABLAYROLLES, Najet BOUTMGHARINE, HUMBLEY John, BERTOCCHI Massimo, TALLARICO Giovanni, JACQUET-PFAU Christine et KÜBLER Natalie (2018), « Détection automatique, description linguistique et suivi des néologismes en corpus : point d'étape sur les tendances du français contemporain », 6ème Congrès Mondial du Linguistique Française CMLF, Mons, 9-13 juillet 2018, URL : https://doi.org/10.1051/shsconf/20184608002.

CORBIN Danielle (1987), *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Max Niemeyer ; diffusion française, Villeneuve d'Ascq, P.U.L.

CORBIN Danielle (1988), « Pour un composant lexical associatif et stratifié », *D.R.L.A.V.*, 38, Paris, Université Paris VIII, p. 63-92.

CORBIN Danielle (1990), « Homonymie structurelle et définition des mots construits, vers un dictionnaire dérivationnel », *La définition*, Jacques Chaurand et Francine Mazière éd., Paris, Larousse, p. 175-192.

DROUIN Patrick (2002), *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*, Thèse de doctorat, Montréal, Université de Montréal.

DROUIN Patrick, (s.d.), « Guide de l'utilisateur », *TermoStat*, URL : http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html.

DUGAS André (1990), « La création lexicale et les dictionnaires électroniques », *Langue française*, 87, p. 23-29.

DUGAS André (1992), « Le préfixe *auto-* », *Langue française*, 96, p. 20-29.

GERARD Christophe et KABATEK Johannes (2012), « Introduction, la néologie sémantique en questions », *Cahiers de lexicologie*, 100, Paris, Classiques Garnier, p. 11-36.

GERARD Christophe, FALK Ingrid et BERNHARD Delphine (2014), « Traitement automatisé de la néologie : pourquoi et comment intégrer l'analyse thématique ? », *Actes du 4ème Congrès Mondial de Linguistique Française* (CMLF), Berlin, Web of Conferences 2014, http://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf4_01208.pdf.

GÉRARD Christophe, BRUNEAU Lauren, FALK Ingrid, BERNHARD Delphine et ROSIO Ann-Lise (2017), « Le Logoscope : observatoire des innovations lexicales en français contemporain », *La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones*, Joaquín García Palacios, Goedele De Sterck, Daniel Linder, Nava Maroto García, Miguel Sánchez Ibáñez et Jesús Torre Del Rey éd., *La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones*, Peter Lang, p. 339-355.

GEVAUDAN Paul et KOCH Peter (2010), *Sémantique cognitive et changement sémantique. Grandes voies et chemins de traverse de la sémantique cognitive*, Mémoire de la Société de linguistique de Paris, XVIII, p. 103-145.

KOCH Peter (2000), « Pour une approche cognitive du changement sémantique lexical : aspect onomasiologique », *Théories contemporaines du changement sémantique*, (Mémoires de

la Société de Linguistique de Paris N.S., 9), François Jacques éd., Louvain, Peeters, p. 75-95.

LAKOFF George et JOHNSON Mark (1980), *Metaphors We Live By*, Chicago et London, The University of Chicago Press, traduction française par Michel de Formel (1986), *Les métaphores de la vie quotidienne*, Paris, Éditions de Minuit, coll. « Propositions ».

NYCKEES Vincent (1998), *La sémantique*, Paris, Belin, coll. « Sujets ».

RASTIER François et VALETTE Mathieu (2009), « De la polysémie à la néosémie », *Le français moderne*, « La problématique du mot », 77, p. 97-116.

RENOUF Antoinette (2010), « Identification automatique de la néologie lexicologique et sémantique : questions soulevées par notre méthode », *Actes del I Congrés Internacional de neologia de les Llengües Romàniques* (Barcelone (7-10 mai 2008), María Teresa Cabré, Ona Domènech, Rosa Estopà, Judit Freixa et Mercè Lorente éd., Barcelone, Université Pompeu Fabra, IULA, sèrie activitats 22, p. 129-141.

RENOUF Antoinette (2016), « Big data and its consequences for Neology », *Neologica. Revue internationale de néologie*, 10, Paris, Classiques Garnier, p. 15-37.

RENOUF Antoinette (2017), « Some corpus-based observations on determinologisation », *Neologica. Revue internationale de néologie*, 11, Paris, Classiques Garnier, p. 21-48.

REUTENAUER Coralie (2012), « Propositions pour la détection automatique de la néologie sémantique », *Cahiers de lexicologie*, 100, Paris, Classiques Garnier, p. 139-162.

SABLAYROLLES Jean-François (2000), *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes*, Paris, Honoré Champion, coll. « Lexica. Mots et dictionnaires ».

SABLAYROLLES Jean-François (2002), « Fondements théoriques des difficultés pratiques du traitement des néologismes », *Revue française de linguistique appliquée*, vol. VII-1, Lexique : recherches actuelles, p. 97-111.

SABLAYROLLES Jean-François (2003), « Le sentiment néologique », *L'innovation lexicale*, Jean-François Sablayrolles éd., Paris, Honoré Champion, coll. « Lexica – Mots et Dictionnaires », p. 279-295.

SABLAYROLLES Jean-François (2008), « Néologie et dictionnaire(s) comme corpus d'exclusion », *Néologie et terminologie dans les dictionnaires*, Jean-François Sablayrolles éd., Paris, Honoré Champion, coll. « Lexica - Mots et Dictionnaires », p. 19-36.

SABLAYROLLES Jean-François (2010), « Néologie, classes d'objets et extraction automatique », *Actes del I Congrés Internacional de neologia de les Llengües Romàniques* (Barcelone, 7-10 mai 2008), María Teresa Cabré, Ona Domènech, Rosa Estopà, Judit Freixa et Mercè Lorente éd., Barcelone, Université Pompeu Fabra, IULA, sèrie activitats 22, p. 143-149.

SABLAYROLLES Jean-François (2010), « Archaïsme : un concept mal défini et des utilisations littéraires contrastées », Colloque *Stylistique de l'archaïsme*, Cerisy, 26-30 septembre 2007, Laure Himy-Pieri et Stéphane Macé éd., Cerisy, France, Presses Universitaires de Bordeaux, coll. « Poétique et stylistique », p. 43-65.

SABLAYROLLES Jean-François (2018), « Néologie et/ou évolution du lexique ? Le cas des innovations sémantiques et celui des archaïsmes », *Actes du colloque « NEOLEX : la néologie lexicale à travers les âges », Études de Linguistique et d'Analyse des Discours / Studies in Linguistics and Discourse Analysis*, Denis Jamet éd., Lyon, Presses de l'Université de Lyon 3, URL : https://revues.univ-lyon3.fr/elad-silda/.

ULLMANN Stephen ([1952] 1969), *Précis de sémantique française*, Berne, A. Francke S.A.