



**HAL**  
open science

# Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille

Emmanuel Cartier

► **To cite this version:**

Emmanuel Cartier. Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille. L'emprunt en question(s) : conceptions, réceptions, traitements lexicographiques, 2019. hal-02537344

**HAL Id: hal-02537344**

**<https://hal.science/hal-02537344>**

Submitted on 8 Apr 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.

# Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille

Emmanuel Cartier  
Université Paris 13 SPC, LIPN-RCLN UMR 7030 CNRS, Labex EFL,  
emmanuel.cartier@lipn.univ-paris13.fr

## Introduction

Parmi l'ensemble des procédés néologiques, l'emprunt occupe une place à part puisque le matériau provient d'un autre système linguistique. Il occupe également une place de choix dans les études néologiques, car il s'agit d'un phénomène intrinsèque aux langues, la très grande majorité des langues se développant non pas dans un environnement clos et étanche à toute influence externe, mais au contraire dans un écosystème dont elles ne sont qu'un composant. Cela est encore plus vrai maintenant que les moyens de communication permettent des échanges au niveau mondial, chacun étant exposé, via la communication électronique, de façon quasi instantanée, à des messages linguistiques en provenance de la quasi-totalité des autres régions du monde, dans d'autres langues, et pouvant en subir les influences.

Cet article portera sur les emprunts en français contemporain, à partir des néologismes repérés et validés dans la plateforme Néoveille (Cartier 2016, 2018). Nous commencerons par exposer quelques notions préliminaires permettant de délimiter et de préciser ce que nous entendrons par *emprunt*. Ensuite, quelques principes méthodologiques seront présentés pour effectuer l'étude applicative. Enfin, nous ferons une synthèse sur les particularités des emprunts en français contemporain.

## 1. Jalons théoriques

### 1.1. Définitions et périmètre de l'emprunt linguistique

La première caractéristique de l'emprunt linguistique est de constituer un groupe à part parmi les mécanismes néologiques, puisqu'il provient d'un système linguistique externe à la langue dans laquelle il se produit (Sablayrolles 2016). Il s'agit d'un des effets du contact et des échanges entre les langues.

Ces contacts aboutissent à plusieurs phénomènes que nous pouvons analyser selon au moins trois points de vue : point de vue des mécanismes formateurs d'emprunts ; point de vue du cycle de vie des emprunts ; point de vue des motivations de l'emprunt.

Distinguons tout d'abord l'emprunt d'autres formes de contact qui dépendent de la relation entre les deux langues :

- [ l'*interférence linguistique (code mixing)* consiste à effectuer un mélange de deux ou plusieurs langues : ce premier phénomène est lié à des situations sociolinguistiques spécifiques de bi- ou multilinguisme, il peut être limité à des groupes sociolinguistiques spécifiques et aboutit parfois à des créoles ; pour ce qui concerne le français, de nos jours, des zones métropolitaines frontalières (notamment en Alsace), un certain nombre de pays francophones non métropolitains connaissant la cohabitation de plusieurs communautés linguistiques (voir Reutner 2017 pour une synthèse sur ce point) et des communautés linguistiques très restreintes, sont concernés par ce phénomène, qui ne sera pas notre objet ici ;
- [ l'*alternance codique (code switching)* consiste, dans un texte dans une langue L2, à changer de code linguistique, pour émettre une séquence de discours (pouvant aller de la lexie au texte) dans la langue L1 ; ce phénomène ne fait pas partie de l'emprunt linguistique,

- puisqu'il s'agit d'un phénomène dans lequel l'énonciateur assume le passage d'une langue à l'autre, et ne cherche aucunement à intégrer le passage en langue L1 dans la langue L2 ;
- [ *l'emprunt*, dans cette perspective, se limite à une situation où le contact de langue laisse intactes les langues L1 et L2, et où le matériau de la langue 2 se trouve circonscrit à des propriétés linguistiques identifiables. Cependant, on voit bien que le phénomène d'emprunt est en continuité avec l'alternance codique, d'autant que cette dernière se manifeste également dans les textes.

Du point de vue des **formes linguistiques de l'emprunt**, de nombreuses typologies ont été proposées<sup>1</sup>. Dans le cadre de ce travail, la typologie proposée par (Sablayrolles, 2016) est utilisée. Celui-ci distingue trois familles de phénomènes induits par les contacts de langue : les emprunts lexicaux véritables, les créations d'équivalents (généralement proposés par les organismes institutionnels de régulation linguistique) et d'autres néologismes subissant une influence étrangère mais où l'une des matrices internes prédomine.

Les *emprunts lexicaux* proprement dit comprennent les lexies dont le signifiant (avec ou non adaptation phonique et/ou graphique : *staff, lobby, mildiou* < *mildew, paquebot* < *packet boat*) et/ou le signifié (*réaliser* dans le sens de l'anglais *to realize*, 'comprendre') sont directement importés dans la langue réceptrice. Ces vrais emprunts s'opposent à toute une série de cas où une matrice interne prédomine, avec une influence étrangère pouvant prendre différentes formes :

- la *traduction*, lorsque un formant existant dans la langue réceptrice est utilisé en place du formant étranger, l'influence ressortissant alors au sens, qui est généralement intégralement repris (par exemple *souris* < *mouse*, où c'est l'apparition d'une nouvelle acception de *souris*, par la matrice sémantique par métaphore, qui prédomine) ;
- le *calque morphologique* : lorsque la traduction opère sur un mot complexe, on parle alors de calque morphologique (exemples : *gratte-ciel* < *sky-scraeper, cheval de troie* < *Trojan Horse*) ;
- l'allogénisme ou faux emprunt : « Sous les dénominations faux emprunts et allogénismes sont rangées des lexies qui n'existent pas dans la langue censée être la source, mais qui sont fabriquées dans la langue cible avec des formants issus d'une autre langue (*tennisman* par exemple), en particulier les hybrides : *serial menteur* ou les influences de structure : *royale attitude*. » (Sablayrolles, 2016 : 5) ;
- la *synthèse néologique* (ou *trou comblé*) : il s'agit des cas où la création lexicale ne peut être expliquée que par l'influence d'un sens provenant d'une autre langue, sans que l'on retrouve de lien entre les formes (exemple de la synapsie *lanceur d'alerte* < *whistle-blower*).

Nous renvoyons à (Sablayrolles, 2016) pour une présentation détaillée de cette typologie.

Du point de vue du **cycle de vie des emprunts**, plusieurs notions ont été proposées, tentant de couvrir les différents phases saillantes de ce type d'innovation lexicale :

- le *xénisme* s'applique « à un terme étranger qui désigne une réalité inconnue ou très particulière et dont l'emploi s'accompagne, nécessairement, d'une marque métalinguistique qui peut être soit une paraphrase descriptive, soit une note explicative en bas de page quand il s'agit d'un texte écrit » (Guilbert, 1975:92). Il est très proche de l'alternance codique, puisqu'il consiste à citer une réalité propre à une langue L2 dans la langue L1. Restreint aux éléments lexicaux, la mention de la lexie est assumée comme appartenant à la mémoire linguistique de L1 : quelques lexies dorénavant ressenties comme empruntées (ou plus ressenties comme telles) sont passées par l'état de xénisme : *sushi, glasnost, jazz*, etc. Cependant, généralement, le xénisme désigne une réalité culturelle propre à L1, ce qui bloque son intégration plus avant dans la langue L2. Par exemple, un très grand nombre de plats *locaux* sont désignés par leur nom d'origine, adapté ou non. Le *feiao tropeiro*, plat

<sup>1</sup> Dans le monde anglo-saxon et germanique, la typologie de (Haugen 1950) reste la base des typologies (voir Backus et al. 2009, pour une revue, et Winter-Froemel 2009).

- typique brésilien à base de haricot, de lardons et de saucisses est un exemple de terme qui sans doute restera à l'état de xénisme, pour des raisons à la fois culturelles et linguistiques ;
- le *pérégrinisme* désigne l'état d'un xénisme qui diffuse dans plusieurs couches sociales sans toutefois se fixer; xénismes et pérégrinismes constitueraient dans ce cadre des *états* du phénomène d'emprunt<sup>2</sup> ; Nous renvoyons à (Sablayrolles, ce volume) pour une étude détaillée de cette notion, aboutissant à la rejeter ;
  - *l'emprunt* proprement dit, dans ce cadre, désignerait un état d'intégration encore plus avancé : « Il y a emprunt linguistique quand un parler A utilise et finit par intégrer une unité ou un trait linguistique qui existait précédemment dans un parler B (dit langue source) et que A ne possédait pas. L'unité ou le trait emprunté sont eux-mêmes appelés emprunts » (Dubois *et al.* 2007 : 177)<sup>3</sup> . On notera à ce propos que les anglo-saxons utilisent deux termes distincts pour désigner le processus (*borrowing*) et le résultat « lexical » (*loanword*). On remarque ici que la définition inclut un critère d'*intégration avancée* de la lexie en L2, ce qui fournit un nouveau critère de distinction entre emprunt et xénisme ;
  - le cycle de vie d'une lexie empruntée ne s'arrête pas au moment de son émergence, ce qui nous fait diverger de la définition précédente, qui limite l'emprunt aux lexies arrivées au stade de l'intégration. Or il faut distinguer les phases d'émergence, de diffusion et de lexicalisation. D'autres paramètres permettent d'identifier les phases d'intégration des emprunts et notamment, d'un point de vue exclusivement linguistique<sup>4</sup> : intégration phonologique (*riding coat* > *redingote*, *mais baby-sitting* > *baby-sitting*), morphologique, intégration dans le système de morphologie productive (*facebook* > *facebookisation*, *facebookeur*, *se défacebookeur*...).

Enfin, du **point de vue des motivations** de l'emprunt, les chercheurs en ont identifié quatre principales, qui selon les cas peuvent se cumuler :

- l'emprunt est lié à un besoin dénomiatif et la langue réceptrice ne dispose pas d'une lexie correspondante : on recourt alors à un terme en usage dans une autre langue, qui, de nos jours, de par la mondialisation, est généralement la *lingua franca* anglo-américaine. Cette situation se présente notamment dans tous les domaines à composante technologique ;
- l'emprunt est une manière d'affirmer une identité culturelle ou idéologique : c'est généralement le cas lorsque les emprunts sont faits à d'autres langues que l'anglo-américain ;
- l'emprunt est lié au prestige ressenti d'une langue : c'est notamment le cas dans la presse féminine, ou encore dans un certain type de presse magazine parisienne (*Slate* par exemple), pour ce qui concerne l'anglo-américain.
- L'emprunt est motivé par le *jeu*.

<sup>2</sup> « At first, loans are 'xénismes' foreign words normally italicised or enclosed in quotes in a text, and generally translated. These may be nonce forms, or may enter a second stage of 'pérégrinisme', or true adoption, in which they begin to be used more widely, partly by non-bilinguals; at this stage, loans are still seen as foreign » (McMahon 1994 : 209). Voir aussi (Dubois *et al.* 2007 : 512) : « le pérégrinisme renvoie encore à la réalité étrangère [celle du xénisme] mais la connaissance de son sens est supposée partagée par l'interlocuteur. ». (Deroy 2013 [1956]:chap.IX) assimile xénisme et pérégrinisme.

Concernant le rapport entre le xénisme et le pérégrinisme, (Chadelat 2000) affirme que « Les pérégrinismes ne sont après tout que des mots voyageurs ou migrants considérés du point de vue linguistique, en fonction d'une place hypothétique au sein du système susceptible de les adopter, tandis que les xénismes sont ces mots étrangers considérés du point de vue des locuteurs en fonction de leur forme exotique ». Mais il conclut ensuite que la distinction est difficile à établir (voir Sablayrolles dans ce volume pour une analyse)

<sup>3</sup> On trouve une définition similaire dans (Thomason et Kaufman 1988:37) « Borrowing is the incorporation of foreign features into a group's native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features. »

<sup>4</sup> Nous nous limitons ici à des éléments linguistiques, mais il est nécessaire de prendre en compte les paramètres socio-pragmatiques de l'intégration (ou non) des emprunts - et des néologismes en général. Voir la section suivante.

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. Emprunts en question(s), La Lexicothèque, Limoges : Lambert-Lucas.

Enfin, un dernier point concerne l'éventuelle politique linguistique de la langue réceptrice, qui peut chercher à amoindrir une tendance à l'emprunt jugée trop invasive, notamment lorsque le processus est massif ou ressenti comme tel. Il s'agit d'un élément important dans le domaine français qui dispose depuis longtemps d'organismes régulateurs.

## 1.2. Cycle de vie des néologismes : perspectives linguistique, socio-pragmatique et cognitive

Les emprunts, tout en obéissant à des règles particulières, font partie de cette catégorie de lexies appelée *néologismes*, ou encore *innovations lexicales*. Les spécificités des emprunts proviennent essentiellement de ce que le matériau formel provient d'une autre langue (*matrice externe*), tandis que les autres procédés néologiques sont formés à partir du matériau interne disponible (*matrices internes*) (Sablayrolles 2017). Sablayrolles s'est, entre autres choses, intéressé à modéliser et décrire les mécanismes de formation des mots, en s'appuyant sur les travaux initiaux de (Tournier 1985). Il évoque (Sablayrolles, 2000, par exemple) l'importance des caractéristiques sociolinguistiques des énonciateurs et des circonstances énonciatives caractérisant les discours comprenant des néologismes.

D'autres courants linguistiques ont considéré que trois perspectives principales pouvaient être prises sur le phénomène néologique. (Schmid 2008 ; 2015) propose ainsi :

- une *perspective linguistique*, cherchant à décrire les propriétés phonologiques, morphologiques, syntaxiques et sémantiques des unités lexicales, et les mécanismes permettant d'expliquer les modifications de l'une ou de plusieurs de ces propriétés, aboutissant à l'innovation lexicale ; la *lexicalisation* (anglais : *lexicalization*) représente dans ce cadre le processus d'intégration ou non du néologisme dans le vocabulaire de la langue considérée ;
- une *perspective cognitive*, qui cherche à modéliser et décrire les mécanismes de formation et d'enracinement (*entrenchment*, voir plus loin) des unités lexicales dans l'esprit des locuteurs d'une communauté linguistique ; la notion de *formation de concept* (anglais : *concept-formation*) couvre dans cette perspective le processus d'intégration ou non du symbole linguistique dans l'esprit des individus (voir Schmid, 2016 pour un état de l'art);
- une *perspective socio-pragmatique*, qui cherche à modéliser les paramètres sociaux et pragmatiques permettant de décrire comment les innovations lexicales deviennent ou non graduellement intégrées à la mémoire collective. La notion d'*institutionnalisation* (anglais : *institutionalization*) subsume dans cette perspective le cycle de vie des néologismes pour la communauté linguistique.

(Schmid 2008 : 3) reprend alors à son compte le terme d'*establishment* (Bauer 2001 : 46) pour subsumer les trois termes de lexicalisation, formation de concept et institutionnalisation. Il reprend par ailleurs les trois phases classiques pour décrire le cycle de vie des innovations lexicales : *creation*, *consolidation* et *establishing*, qui se déclinent selon les trois perspectives (voir figure 1).

**Figure 1.** Trois perspectives et trois phases du cycle de vie des innovations lexicales (Schmid 2008 : 3)



Perspectives: Stages:	Structural perspective	Socio-pragmatic perspective	Cognitive perspective
creation	(product of) nonce-formation	(process of) nonce-formation	pseudo-concept
consolidation	stabilization	spreading	(process of) hypostatization
establishing	lexicalized lexeme	institutionalized lexeme	hypostatized concept

Ces trois phases saillantes (que l'on retrouve aussi, par exemple, dans (Bybee 2010 ; Bybee 2016; Traugott and Trousdale 2013)) doivent donc être modélisées selon les trois perspectives.

**D'un point de vue cognitif**, la notion d'*entrenchment*<sup>5</sup> est centrale, liée au degré d'automatisme de l'accès à la lexie (ou plutôt au symbole linguistique, l'association forme-sens) : l'émergence correspond alors à l'apparition d'une construction nouvelle (nouvelle forme ou nouvel usage), avec une très faible fréquence et une très faible exposition, nécessitant pour l'interprétation une analyse des symboles constitutifs de l'innovation et du contexte au sens large ; la *routinisation* de l'accès intervient avec une plus grande répétition aboutissant finalement à une *automatisme de l'accès*. Nous renvoyons à (Schmid 2016) pour une présentation des approches cognitives.

**D'un point de vue linguistique**, la description des matrices lexicogéniques (Sablayrolles 2017) donne une base solide à la description, ainsi que la description des propriétés phonologiques, morphologiques, sémantiques et contextuelles proposées dans (Sablayrolles 2010 ; Cartier et Sablayrolles 2009). Les descriptions statistiques et probabilistes issues des travaux du distributionnalisme permettent également d'affiner les descriptions linguistiques, en décrivant les profils combinatoires et distributionnels des innovations lexicales (voir section 2.3 et Cartier et al. 2018).

**Enfin, d'un point de vue socio-pragmatique**, la linguistique variationniste a mis au jour l'évidence des variations d'emploi et l'idéalisation fautive que constitue la notion de langue unique. Nous devons à (Coseriu 1964) les premières intuitions sur les paramètres permettant d'identifier ces strates de savoirs linguistiques : variation diatopique (français des régions, français des pays francophones, etc.), variation diastratique (liée à l'existence de sous-groupes sociologiques dont les usages linguistiques sont différenciés), variation diaphasique (ou registres de langue). (Koch & Oesterreicher 1990) ont plus tard détaillé la variation diamesique (oral / écrit) pour élaborer un modèle basé sur la distance communicative et explicitant un continuum entre une communication « intime » et une communication formelle et normée. Il est nécessaire de prendre en compte les conditions socio-pragmatiques d'apparition des innovations : il est en effet évident qu'un changement linguistique est d'abord une variation linguistique qui se diffusera ou non. De ce point

<sup>5</sup> Cette notion, introduite par (Langacker 1987 : 59 ; voir Schmid 2015 pour une revue), et qu'on traduit généralement par *enracinement*, fonde le processus d'ancrage socio-cognitif des signes linguistiques (définis comme l'association d'une forme et d'un sens), notamment en relation avec le degré d'automatisme de l'accès aux lexies, à la fois en production et en compréhension. Le plus ou moins fort enracinement cognitif est généralement lié à la plus ou moins grande exposition à des occurrences de la lexie, c'est-à-dire lié à la *répétition des occurrences*. Cet enracinement, de par son lien avec les conditions pragmatico-sociales, est instable dans le temps, et variable selon les locuteurs. En dehors de l'intérêt explicatif de cette notion, cette approche fonde une étude statistique-distributionnelle en corpus : les répétitions de séquences linguistiques reflètent l'enracinement (ou corrélativement le déracinement) socio-cognitif des signes linguistiques. Le phénomène de répétition doit par ailleurs être replacé dans son contexte global : c'est un processus continu et évolutif, instable, qui nécessite de prendre en compte la variabilité des corpus selon l'axe diachronique, et selon l'axe socio-pragmatique.

de vue, quelques travaux ont appliqué et opérationnalisé cette notion de diastratie, soit en considérant qu'il était possible d'identifier les variétés linguistiques en identifiant les propriétés de groupes sociaux, par la classe sociale, l'âge, le genre, etc. (voir par exemple Labov 2001), soit en modélisant la communauté linguistique comme un ensemble de groupes sociaux, et en modélisant la diffusion de l'innovation en distinguant les groupes innovateurs, les groupes adopteurs précoces et tardifs, et les retardataires (Rogers 2003 [1983]). Ce modèle s'accompagne d'une théorie décrivant l'innovation réussie, au moyen de la courbe sigmoïde, liant l'adoption de l'innovation à son emploi progressif par ces différents groupes sociaux. (voir Nevalainen 2015 et Feltgen et al. 2017 pour une application au changement lexical).

Avec ces différents jalons théoriques, nous abordons maintenant la méthodologie que nous avons suivie pour mener notre étude.

## 2. Méthodologie

Dans cette partie, nous évoquons les points essentiels de la méthodologie suivie, qui est liée à l'utilisation de la plateforme Néoveille et aux hypothèses théoriques adoptées. Nous renvoyons à (Cartier *et al.* 2018) pour plus de détails sur la plateforme elle-même.

### 2.1. Corpus

Pour cette étude, nous avons utilisé le corpus disponible dans Néoveille. Le corpus dynamique du français contemporain est composé de 245 sources de presse récupérées deux fois par jour via les fils RSS publiés par les éditeurs, depuis 2015. Chaque source d'information comprend des informations diatopiques (pays, région pour l'espace hexagonal) et diastratiques (domaine). Un aperçu est présenté dans le tableau 1.

**Tableau 1.** Extraits sur la répartition diatopique et diastratique (pour la presse hexagonale)

Pays	Nbre titre de presse	Nbre d'articles	Domaine	Nbre d'articles
France	156	912 504	Général	730 386
Algérie	53	75 291	Sport	99 802
Canada	4	72 693	Presse féminine	27 108
Belgique	4	39 503	Politique	13 945
Sénégal	25	39 296	Informatique	11 451
Total		1 139 287	Industrie	9 527
			Sciences	4 516

Les différences dans la distribution par pays sont liées au nombre de titres de presse, d'une part, et au démarrage de la récupération, plus tardive (début 2016) pour les pays francophones hors métropole. Concernant les domaines, la liste est liée, en la simplifiant, à l'ontologie des thèmes proposées par l'IPTC<sup>6</sup>.

### 2.2. Validation des emprunts

Dans Néoveille, les néologismes formels sont d'abord détectés automatiquement, puis les experts humains doivent valider ou non les propositions du système. Ces validations ont été effectuées en suivant le protocole suivant : chaque membre du groupe de travail pour le français annote

<sup>6</sup>Cette association professionnelle internationale regroupe les principaux organes de presse. Voir : <https://iptc.org/standards/media-topics/>

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. Emprunts en question(s), La Lexicothèque, Limoges : Lambert-Lucas.

individuellement sur la plateforme une partie des néologismes candidats, sur la base d'une fiche d'instructions détaillant les catégories de néologismes et de non-néologismes. Puis, lors de réunions collectives mensuelles, une validation est effectuée, les cas litigieux étant tranchés sur la base d'un vote majoritaire. Ce processus de validation a permis de valider, sur deux ans et six mois (juin 2015-2017), un peu plus de 21 000 néologismes et 1 429 emprunts à composante formelle. Il convient de noter qu'étant donné que le système initial ne détecte que les néologismes ayant une composante formelle, un certain nombre d'emprunts ne sont pas représentés dans notre corpus (calques syntaxiques et sémantiques).

### 2.3. Description linguistique des emprunts

Une fois validés, les néologismes sont ensuite décrits linguistiquement, en utilisant une version simplifiée de la micro-structure développée dans (Sablayrolles 2010). Cela aboutit à une microstructure comprenant les champs détaillés dans le tableau 2, les deux derniers champs étant propres aux emprunts.

**Tableau 2.** Microstructure utilisée dans Néoveille

Informations	Définition succincte
Partie du discours	Catégorie morphosyntaxique parmi : nom, verbe, adjectif, etc.
Classe sémantique	Classe sémantique générique. Inspirée de (Le Pesant et Mathieu-Colas 1998)
Définition	
Procédé(s) néologique(s) impliqués	Le ou les mécanismes néologiques impliqués dans l'innovation lexicale, en partant de la typologie de (Sablayrolles 2017)
Configuration syllabique	Description générique et détaillée de la configuration syllabique de la lexie, au moyen des notions de syllabes ouverte et fermée.
Configuration morphologique	Décomposition morphologique de l'innovation, au moyen des notions de radical, d'affixe et de formant.
Lexie base	Identification de la ou des lexies ayant servi de base au néologisme
Partie du discours lexie base	Identification de la partie du discours de la lexie base, ou de la racine.
Influence langue	Indication de la langue origine de l'emprunt ou des influences présentes dans la création d'un néologisme, formé par une / des matrices internes
Mode influence	Type de l'influence, parmi l'une des catégories suivantes : traduction (ou calque sémantique), calque morphologique, formant emprunté, sens emprunté, structure empruntée, synthèse néologique <sup>7</sup>

À ces informations, il faut ajouter trois informations linguistiques qui sont disponibles de manière automatique sur la plateforme depuis 2018 : la *famille morphologique* associée à l'innovation étudiée ; le *profil combinatoire* des occurrences dans le corpus, permettant de détecter les collocations (Firth 1957), les collostructions (Stefanovitsch et Gries 2003) et les constructions lexico-syntaxiques les plus fréquentes ; le *profil distributionnel*<sup>8</sup>, permettant d'accéder aux lexies sémantiquement similaires (et donc notamment (quasi-)synonymes, hyperonymes et hyponymes).

<sup>7</sup> voir Sablayrolles 2016 et section 1.1.

<sup>8</sup> La notion de profil distributionnel provient des intuitions du distributionnalisme. Harris évoque très tôt cette façon d'exploiter la distribution des lexies :

« ...if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution. » (Harris, 1954:43)

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. Emprunts en question(s), La Lexicothèque, Limoges : Lambert-Lucas.



Nous illustrons les deux premières informations dans le tableau 3, la troisième nécessitant un corpus plus étendu pour obtenir des résultats fiables<sup>9</sup>.

**Tableau 3.** Informations linguistiques automatiquement détectées dans Néoveille, pour *Food* (voir aussi Hildenbrand et Jacquet-Pfau, 2016)

Type d'information	Description sommaire	Exemples pour <i>food</i>
Famille morphologique	Ensemble des lexies formées sur la même base (y compris mot composé à trait d'union)	<i>foodies, fooding, foods, food-biz, food-market(s), food-truck(s), food-deco, foodeur(s), foodflock, foodista(s)...</i> liste complémentaire (noms propres) : <i>Food4Good, FoodChéri, FoodOrganic, FoodStocks, FoodTech, FoodTemple, FoodWatch, Foodora ...</i>
Profil combinatoire	Ensemble des collocations, des collostructions et des constructions lexico-syntaxiques représentatives <sup>10</sup>	<u>Collocations</u> : <i>fast food</i> (16), <i>slow food</i> (16), <i>street food</i> (11), <i>raw food</i> (9), <i>junk food</i> (7), <i>food market</i> (7) <u>Collostructions</u> : <i>tendance food</i> (10) => ADJ <i>phénomène food</i> (9) => ADJ <i>projet food</i> (5) => ADJ <i>Det (masc) food</i> (10) => NOM <u>Constructions lexico-syntaxiques</u> : <i>food + verbe</i> : <i>aller, débarquer, arriver, consister, cartonner...</i>

## 2.4. Description socio-pragmatique des emprunts

La description interne – linguistique – doit être complétée par une description socio-pragmatique, qui permet de caractériser les propriétés des contextes d'emploi : caractéristiques des énonciateurs, des supports matériels de diffusion, des types de discours, domaines, contextes d'emploi. Le courant de l'analyse du discours fournit dans ce cadre plusieurs pistes pour caractériser les discours dans lesquels s'insèrent les innovations lexicales (voir Charaudeau 1995 et Paveau 2017 par exemple). Les métadonnées proposées dans (Siepmann *et al.* 2016) pour caractériser les textes du Corpus de référence du Français Contemporain peuvent également servir de guide. Dans le cadre de Néoveille, actuellement, seulement trois propriétés sont disponibles : le domaine, lié à l'information thématique fournie par les producteurs eux-mêmes ; le pays ou la région d'origine des journaux, le

---

À partir de cette intuition, un nouveau champ, la *sémantique distributionnelle* (Baroni et Lenci, 2010) verra le jour, ainsi que des applications pratiques (Mikolov *et al.* 2013), permettant, à partir de larges corpus, d'identifier les lexies en relation de similarité sémantique, et donc, en diachronie, d'étudier l'évolution de cette signature sémantique.

<sup>9</sup> Pour illustrer le profil distributionnel, citons par exemple *tsunami*, qui, des années 1900 à nos jours, a pour lexies sémantiquement similaires « raz-de-marée, phénomène violent, etc ». À partir des années 1970, on voit apparaître « phénomène social, tendance, etc » montrant une extension de sens (*tsunami politique, social*, etc). Enfin, depuis 2004, apparaissent de nouveaux synonymes « grand nombre de, plusieurs, etc. », liés à l'apparition de l'emploi comme déterminant complexe au sens de « grand nombre d'éléments, avec connotations de grande puissance et de soudaineté » (*un tsunami d'applaudissements*). (Hamilton, 2016) proposera une première application de ce principe, sur le corpus Google Ngrams. Il montre par exemple comment l'adjectif *gay* est sémantiquement similaire à *daft, tasteful, sweet, pleasant*, dans les années 1900, puis sémantiquement similaire à *bisexual, homosexual, lesbian* à partir des années 1970 (à noter que le *dictionnaire historique de la langue française* date les premières apparitions en anglais d'Amérique du Nord de ce sens au début des années 50 comme adjectif et en 1971 comme nom). Cette technique, *modulo* les limitations dues au corpus source, est l'une des voies les plus prometteuses pour détecter des changements sémantiques.

<sup>10</sup> Nous indiquons ici la fréquence constatée dans le corpus Néoveille.

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. *Emprunts en question(s)*, La Lexicothèque, Limoges : Lambert-Lucas.

type de discours, qui est *a priori* toujours l'article de presse. Ces informations sont évidemment largement insuffisantes.

## 2.5. Suivi du cycle de vie

L'ensemble des propriétés linguistiques et socio-pragmatiques précédentes permettent de se faire une idée précise des innovations lexicales à un moment *t*. Cependant, l'évolution diachronique des innovations lexicales est tout aussi importante pour les caractériser. De ce point de vue, plusieurs notions importantes sont détaillées ci-après.

### 2.5.1. Évolution fréquentielle

La fréquence est l'élément le plus évident à la fois pour mesurer une émergence, mais également une diffusion et une institutionnalisation. En effet, il s'agit d'un facteur rendant directement compte de l'exposition cognitive à l'innovation, et donc de son enracinement ou non (*entrenchment*). Trois phases-clés doivent être identifiées : l'émergence, la diffusion et l'institutionnalisation.

Concernant l'émergence, nous avons utilisé la fréquence brute plutôt que l'une des mesures disponibles pour calculer une fréquence relative<sup>11</sup>. Ce choix tient à la définition de *néologisme* que nous avons adoptée dans cette étude et dans le projet Néoveille : un néologisme est une unité linguistique nouvelle, dès sa première apparition, et quel que soit son destin ultérieur. Cette conception va à l'encontre de nombreuses modélisations, notamment anglo-saxonnes<sup>12</sup> qui considèrent qu'il faut distinguer entre *hapax legomenon* (littéralement, qui ne se produit qu'une seule fois), *nonce-word* (mot ad-hoc, ou occasionnalisme, créé pour une circonstance déterminée, et donc ayant peu de chances de se répéter) et *néologisme*, défini alors comme une formation ayant déjà passé le stade de l'émergence initiale. Pour notre part, nous considérons que tous ces termes renvoient à un même phénomène de créativité lexicale, et qu'il est nécessaire d'étudier le phénomène dès son apparition. Nous verrons que ce choix théorique permet notamment d'intéressantes conclusions sur la légitimité de la notion d'hapax en linguistique de corpus. Il faut enfin noter que la détection de formes nouvelles dans le corpus Néoveille ne préjuge pas de l'existence de cette lexie antérieurement. C'est pourquoi, afin de réduire au minimum le recours à la notion de *sentiment néologique*, nous avons pris comme critère d'émergence d'une innovation lexicale l'absence d'attestation, dans le sens repéré, avant 2010, en vérifiant dans le corpus de Google Books, ainsi que dans le corpus *Europresse*.

En dehors de la saisie de l'émergence d'une innovation, la fréquence peut être exploitée pour mesurer la diffusion voire la lexicalisation d'une lexie nouvelle. Mais à quel seuil de fréquence pouvons-nous fixer qu'une diffusion, puis une institutionnalisation, se produisent ? Deux techniques sont envisageables : d'une part, exploiter un corpus de référence pour obtenir les fréquences constatées des lexies dont nous savons qu'elles sont institutionnalisées et s'en servir comme *fréquence de référence*. Cette approche n'est pas sans difficulté, puisque, en dehors de la notion de *corpus de référence*, qui n'est toujours qu'une approximation des usages réels, nous savons que la distribution fréquentielle des lexies obéit au principe de Zipf, et que l'immense majorité des lexies ont une fréquence relative extrêmement basse. Cependant, il est toujours possible d'obtenir une moyenne fréquentielle pour des tranches de lexique. Il serait également possible d'utiliser un corpus de référence diachronique pour établir des modèles d'évolution lexicale, du point de vue fréquentiel.

D'autre part, l'évolution fréquentielle est toujours différentielle, et une phase de diffusion sera toujours marquée par un accroissement des occurrences. (Rogers 2003 [1983]) fournit un modèle d'évolution fréquentielle encore plus général (courbe en S) qui peut servir de référence, mais qui

---

<sup>11</sup> La linguistique de corpus a proposé de nombreuses mesures de fréquence. Nous renvoyons à (Dal et Namer 2015) pour une présentation accessible.

<sup>12</sup> Nous renvoyons par exemple à (Crystal, 2000 : 219) qui renvoie aux définitions du *Oxford English Dictionary*.

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. *Emprunts en question(s)*, La Lexicothèque, Limoges : Lambert-Lucas.

nécessite de croiser la fréquence avec les paramètres diastratiques : une diffusion se produit lorsque l'innovation apparaît dans des groupes sociaux autres que celui de l'innovateur. Deux études récentes (Nevallainen, 2015 ; Feltgert *et al.*, 2017) ont montré que cette approche était prometteuse pour le suivi des évolutions lexicales, même si les données de référence concernent des évolutions en diachronie longue. Dans Néoveille, à l'heure actuelle, nous utilisons la simple fréquence brute.

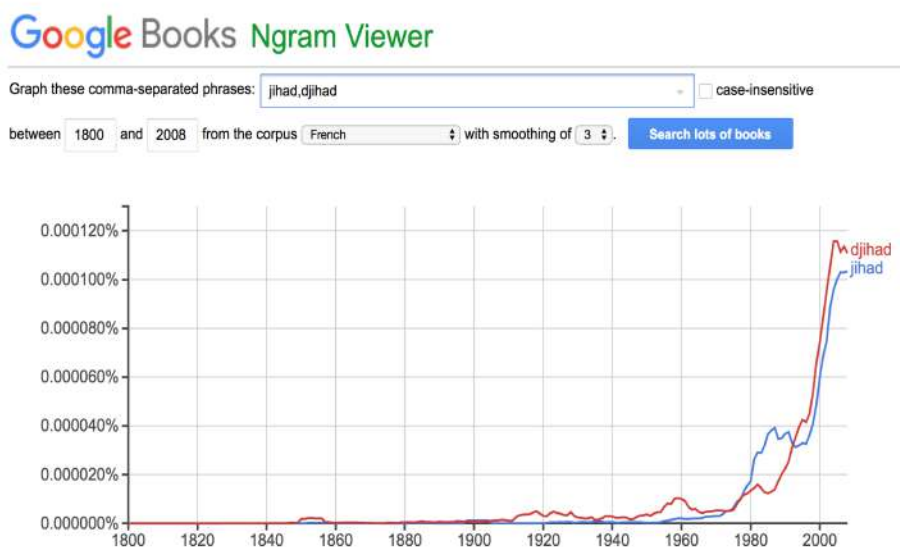
### 2.5.2. Évolution des paramètres diastratiques ou socio-pragmatiques

Un autre paramètre fondamental pour mesurer la diffusion d'une innovation consiste à observer les contextes socio-pragmatiques d'apparition des occurrences. Une innovation intervient toujours dans un contexte donné, déterminé par ses propriétés socio-pragmatiques. Lorsque les contextes socio-pragmatiques évoluent, il s'agit alors certainement du signe d'une diffusion. Nous illustrerons ce point dans l'étude descriptive.

### 2.5.3. Intégration phonologique et orthographique

L'intégration phonologique des unités empruntées est une étape importante. Du point de vue des formes scripturales, deux exemples récents illustrent le processus, *twitter* (et ses dérivés), emprunté à l'anglo-américain international, et *djihad* (et ses dérivés), emprunté à l'arabe. Pour le premier, la création déonomastique a dès le départ modifié la base *tweeter*, qui est son origine. En français, les deux orthographes sont encore en concurrence, mais la forme de la lexie initiale connaît tout de même une fréquence bien plus importante (*tweeter*, *tweet*, *retweet*, *tweetos*, *etc.*). Pour *djihad*, la forme *jihad*, si elle est largement moins employée, reste tout de même vivace, selon le corpus Néoveille (2015-2018). En remontant à l'émergence en français de la notion (vers 1850, selon Google Ngrams) on constate une concurrence plus vive jusqu'aux années 2010 (voir figure 1). On voit bien dans ces cas l'intérêt d'une étude sur corpus, puisqu'alors il est possible de suivre les évolutions d'usage, et les tendances qui en découlent. Dans ce cadre, dans le cadre du projet Néoveille, nous utilisons, outre le corpus récupéré quotidiennement depuis mi-2015, les données fournies par Google Ngrams, qui couvre un empan temporel de 1800 à 2008, et le corpus *Europresse*, qui dispose d'un corpus de presse exhaustif de 1945 à nos jours, avec des fonctionnalités de recherche plus limitées.

**Figure 2 – évolution fréquentielle comparative entre les formes *djihad* et *jihad* (Google Ngrams)**





#### 2.5.4. Intégration morphosyntaxique

Un aspect complémentaire au précédent concerne l'intégration morphosyntaxique des termes empruntés. En effet, l'usage (par opposition à l'emploi métalinguistique) nécessite que les lexies adoptent le comportement morphosyntaxique de l'une ou l'autre des parties du discours existantes dans la langue d'accueil. De ce point de vue, un suivi automatique de la famille morphologique de la lexie considérée, ainsi que du comportement lexico-syntaxique (profil combinatoire), permet de se faire une idée précise de l'intégration ou non d'une lexie. Les emplois métalinguistiques, par complément, sont le signe d'une émergence, et d'une faible diffusion.

#### 2.5.5. Intégration à la morphologie productive

Un autre paramètre permettant d'évaluer l'intégration d'une lexie est son intégration ou non à la morphologie productive de la langue considérée. Les langues disposent en effet d'une série de formants (affixes et morphèmes grammaticaux) extrêmement productifs et disponibles, c'est-à-dire dont la règle d'application peut librement s'appliquer à une ou plusieurs parties du discours, sans ou avec très peu de contraintes sémantiques. Citons<sup>13</sup>, pour le français, les préfixes *anti-*, *ex-*, *non-*, *mini-*, *ultra-*, *mi-*, *post-*, *hyper-*, *auto-*, *demi-*, *sous-*, *semi-*, *quasi-*, *pro-*, *multi-* etc., les morphèmes *-er* et *-ment* et les suffixes : *-iser*, *-is-ation*, *is-ateur/trice*, *-isme*, *-ien(ne)(s)*, *-iste*, *-eur(euse)(s)*, *-itude*, *-esque*, *-able*, *-erie*, *-issime*, *-ité*, *-ique*, etc. Lorsqu'une innovation commence à créer des dérivés de la sorte, son intégration est en bon chemin. Parmi les emprunts, plusieurs lexies liées aux réseaux sociaux ont rapidement franchi cette étape : *twitt/tweet*, *facebook*, *instagram*.

#### 2.5.6. Productivité réalisée, en expansion et potentielle

Un autre concept important pour le suivi du cycle de vie des innovations lexicales, dans un état d'intégration avancé, concerne leur *productivité*. Ce concept a pour but de rendre compte de la propriété de certaines lexies ou de certains formants à permettre la création d'autres unités par préfixation, suffixation (voir 2.4.3) mais aussi par composition (les schèmes *-clé*, ou *-phare*) et par fracto-composition (*e-*, *cyber*, *eco-*). Prenons *e-* et *cyber-*. Ils ne font pas *classiquement* partie des affixes, mais leur comportement s'en approche de par leur capacité à permettre la génération d'un grand nombre de formations lexicales : *e(-)* est le résultat de la troncation de *electronic*, et a été importé d'abord dans *email*, puis, depuis quelques années, se rencontre dans un nombre grandissant de lexies (*e-gouvernance*, *e-addictif*, *e-recruter*, *e-scooter*, *e-manif*, etc.), au point de devenir, en français comme dans d'autres langues, un *formant emprunté productif*. Pour mesurer cette productivité, nous nous appuyons sur les travaux de (Baayen 2009) qui distingue trois acceptions de la notion : la *productivité réalisée* (*realized productivity*) désigne les réalisations déjà constatées dans le passé à partir du formant ou de la lexie, et qui peuvent être comptabilisées dans un dictionnaire ; la *productivité en expansion* (*expanded productivity*) quantifie le nombre d'hapax nouveaux (de néologismes) créés par la catégorie dans un corpus dynamique. La *productivité potentielle* (*potential productivity*), enfin, mesure l'étendue maximale possible de cette productivité, difficilement mesurable mais dont on peut décrire la règle (et ses contraintes). Par exemple, *non-* a une productivité potentielle plus grande que *ex-* car il s'applique à des noms et des adjectifs alors que *ex-* s'adjoint uniquement aux noms. (Baayen 2009) propose des mesures pour chacune de ces productivités, mais elles ne sont pas réellement opérationnalisables (voir Dal et Namer 2015 pour une revue critique). Dans Néoveille, nous parlerons généralement de la *productivité en expansion*, tandis que concernant la productivité potentielle, si elle n'est pas mesurable, elle peut être décrite par les contraintes d'application de la règle.

#### 2.5.7. Evolution du profil combinatoire

<sup>13</sup> Pour une liste exhaustive en français contemporain, voir (Cartier *et al.* 2018).



On peut distinguer deux phases dans l'évolution du profil combinatoire : généralement, le profil combinatoire initial, étant donné la nouveauté du concept, nécessite un attirail métalinguistique permettant d'explicitier le sens de l'innovation. Ainsi, pour le xénisme *cutlet*, une glose sera insérée dans une parenthèse : « Les apprentis cuistots apprendront à réaliser ...des cutlets (galettes panées)... » (*Le Parisien*, 12/04/2016). Les différents procédés de glose ont été décrits depuis bien longtemps (Rey-Debove 1997 ; Steuckardt et Niklas-Salminen 2003, 2005). Ils sont sans doute l'un des premiers critères d'identification des innovations lexicales, même si on les rencontre également dans les discours de vulgarisation scientifique et dans les discours didactiques pour des termes déjà implantés. On constate ensuite, lorsque l'énonciateur considère que la lexie a suffisamment diffusé, le passage à des emplois non-autonymiques, montrant une diffusion accrue du terme, par exemple : « son look off-duty fait partie des plus épiés de la planète mode. » (*Elle*, 27 mars 2018). Quoi qu'il en soit, une fois le passage à l'usage effectué, le profil combinatoire permet d'approcher à la fois le comportement morphosyntaxique et, par induction, le sens de la lexie (ici, par exemple, *look*, *allure*, *style*, *dimension* sont les collocations les plus fréquentes). C'est également le point d'entrée pour l'étude des innovations sémantiques.

### **2.5.8. Mise en place et évolution du profil distributionnel**

Le profil distributionnel est en quelque sorte le complément de l'approche précédente, et signale une diffusion accrue : pour qu'une stabilisation dans une communauté linguistique se produise, il est nécessaire qu'un nombre minimal d'expositions en contextes ait eu lieu. Ce n'est qu'à cette condition qu'une stabilisation du sens, dans le réseau sémantique global et dans le champ sémantique spécifique de la lexie, peut avoir lieu. Le profil distributionnel ne peut enfin pas être calculé sans un minimum d'occurrences (pour des exemples d'exploitation du profil distributionnel, voir note 9).

## **3. Les emprunts en français contemporain (2015-2017)**

Dans cette section, nous analysons les résultats du travail mené sur la détection, la description et le suivi des emprunts à composante formelle du français détectés et validés sur la plateforme Néoveille, de 2015 à fin 2017. Le travail de validation des néologismes candidats, ainsi que la description linguistique des néologismes validés a été effectué par le groupe de travail pour le français<sup>14</sup>. Nous présentons tout d'abord les tendances générales, puis détaillons les spécificités des emprunts, en suivant notre plan méthodologique.

### **3.1. Synthèse sur les emprunts repérés et validés dans Néoveille**

À partir d'environ 250 sources de fils de presse nationale et régionale récupérées automatiquement de juillet 2015 à fin 2017, un corpus de 1 143 912 articles pour un total de plus de 92 millions de mots (1 037 876 formes différentes) a été constitué. Parmi environ 35 000 néologismes formels candidats repérés automatiquement, 22 475 néologismes ont été validés, correspondant à 726 222 occurrences. Les néologismes représentent donc 2,16 % des formes rencontrées, et, au niveau du nombre d'occurrences, 0,78 %.

### **Répartition par mécanisme**

---

<sup>14</sup> Ont participé à ce travail : Najet Boutmgharine, Emmanuel Cartier, Massimo Bertocci. Par ailleurs, ont participé à la validation collective des analyses : Jean-François Sablayrolles, John Humbley, Christine Jacquet-Pfau, Natalie Kübler, Giovanni Tallarico. Voir (Cartier *et al.*, 2018) pour une synthèse sur les néologismes en français contemporain.

A paraître dans Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (2018), éd. *Emprunts en question(s)*, La Lexicothèque, Limoges : Lambert-Lucas.

Dans le tableau 4, qui donne la répartition par matrices, les colonnes 2 et 3 indiquent le nombre de néologismes différents, les colonnes 4 et 5 le nombre d'occurrences, et la colonne 6 le nombre moyen d'occurrences par matrice.

**Tableau 4.** Synthèse sur les mécanismes néologiques.

Mécanisme néologique principal	Nombre de néologismes (formes uniques)		Nombre d'occurrences de néologismes		Moyenne d'occ. par forme néologique
	Nb	%	Nb	%	
préfixation	17 051	<b>75,87%</b>	485 566	<b>66,86%</b>	28
composition	1 646	<b>7,32%</b>	31 173	4,29%	19
emprunt	1 429	6,36%	132 104	<b>18,19%</b>	<b>92</b>
suffixation	1 245	5,54%	65 262	8,99%	<b>52</b>
fracto-composition	791	3,52%	7 039	0,97%	9
onomatopée	92	0,41%	665	0,09%	7
troncation	73	0,32%	2 678	0,37%	<b>37</b>
composition savante	68	0,30%	479	0,07%	7
compocation	47	0,21%	1 043	0,14%	22
composition hybride	33	0,15%	213	0,03%	6
mot-valise	9	0,04%	100	0,01%	11
Totaux	22 475	100,00%	726 222	100,00%	

On constate que le procédé le plus utilisé est la préfixation (75 % des formes néologiques). Les emprunts représentent 6,36 % du contingent, soit 1 429 formes uniques. Le nombre d'occurrences révèle une spécificité des emprunts, puisque de ce point de vue, ils représentent 18 %, (soit une moyenne de 92 occurrences en moyenne) du total. Cela est dû à un certain nombre d'emprunts dont nous souhaitons analyser l'implantation et non simplement l'émergence (notamment toutes les lexies des réseaux sociaux, et leurs dérivés : *blog, facebook, twitter, instagram*, etc.). De plus, de nombreux emprunts, déjà implantés ou en voie de l'être, ne sont pas comptabilisés ici, car ils ressortissent à d'autres matrices néologiques, notamment les affixations (cela est encore vrai pour les *buzzwords* des réseaux sociaux : *facebook, instagram, twitter*, empruntés il y a quelques années, qui ont très rapidement été intégrés à la morphologie productive du français : *facebooker, facebookage, instagrammeur, twitos, tweeker, twitterisation*, etc.). Il faut également ajouter aux 1 430 emprunts environ un millier de xénismes<sup>15</sup>.

Pour les emprunts, la langue source la plus représentée est l'anglais<sup>16</sup> à environ 91%, suivi de l'espagnol (4,5%), de l'arabe (3%) et de l'italien (2%). Les xénismes ont des langues sources beaucoup plus diversifiées. On retrouve ici une différence fondamentale entre les emprunts anglais et les emprunts à d'autres langues : tandis que les seconds dénotent pour une très grande majorité des concepts culturels spécifiques, les premiers renvoient à des aires culturelles variées (Chesley et Baayen 2010). On peut également y voir une différence dans la perception collective de l'anglais international, perçu comme une langue de prestige, tandis que les autres langues sont perçues comme des marqueurs d'identités. Ces résultats consolident les chiffres proposés par (Martinez 2009) sur les provenances des emprunts enregistrés dans *Le Petit Robert* de 1997 à 2011.

La forte imprégnation du vocabulaire anglo-saxon n'est pas une particularité française ni un processus récent (Pulcini *et al.* 2012 : 2-3). Dans l'aire francophone, comme ailleurs, plusieurs facteurs permettent de l'expliquer : d'une part, depuis la fin de la Seconde Guerre mondiale, la puissance économique et politique américaine, qui s'est manifestée notamment par de nombreuses migrations culturelles et technologiques vers les États-Unis, une politique économique

<sup>15</sup> Dans Néoveille, étant donné le statut des xénismes, pouvant en cas de forte croissance d'emploi devenir des emprunts à proprement parler, nous les conservons dans une base spécifique.

<sup>16</sup> Ou plus exactement la lingua franca anglo-américaine (voir par exemple Christiansen 2015)

d'exportation massive, la stabilisation de l'anglais comme *lingua franca* dans de nombreux domaines et le nombre croissant d'anglophones non-natifs, a évidemment favorisé les emprunts à l'anglo-américain. Le développement de la radio, de la télévision et du cinéma a été après les années 70 un nouveau facteur d'essor. Enfin, l'avènement de la communication électronique, à partir des années 1995-2000, puis le développement des réseaux sociaux, depuis 2010, n'ont fait qu'accroître ce phénomène, qui est global. Ces raisons expliquent un taux important d'emprunts dans certains domaines (informatique, industrie, monde du travail en général, pratiques sociales). Par ailleurs, il existe à l'évidence un prestige de l'anglais américain dans certaines couches sociales francophones : presse féminine, presse internationale disposant d'une version locale (*Slate*, *Huffington Post*). On trouve dans notre corpus des traces à la fois de la diffusion dans toute la communauté linguistique d'emprunts liés aux réseaux sociaux, mais également dans le domaine des pratiques professionnelles et de loisirs (voir tableau 9, colonne autres).

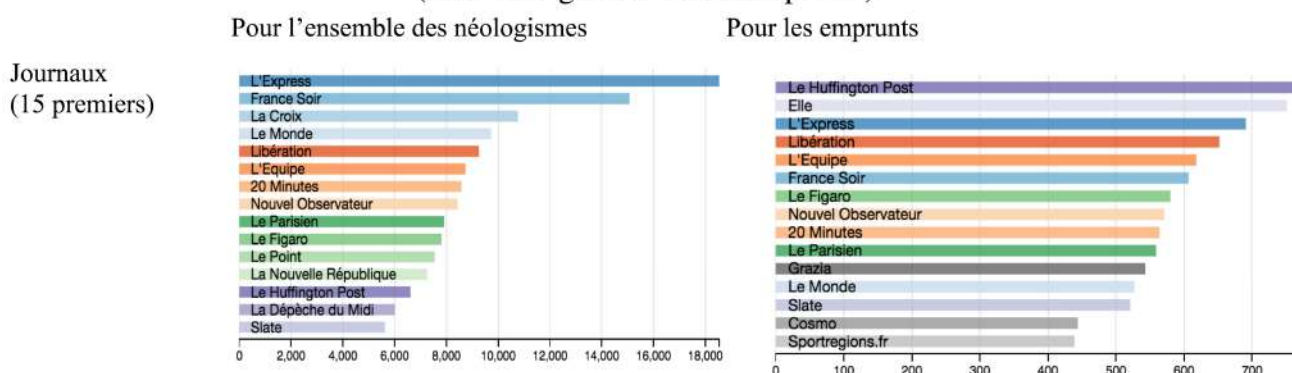
### Répartition par parties du discours

La distribution par parties du discours est la suivante, pour l'ensemble des néologismes : Nom (79,61 %), Adjectifs (9,76 %), Verbe (8,34 %) et Adverbe (2,29 %). Une comparaison avec la distribution constatée sur le corpus complet<sup>17</sup> montre que la néologie s'applique principalement sur les noms et les adjectifs, les verbes et les adverbes étant sous-représentés. Pour les emprunts, nous obtenons la répartition suivante : 83,8 % de noms, 9,7 % d'adjectifs et 6,5 % de verbes. À noter cependant que de très nombreux emprunts nominaux sont également attestés en tant que verbes via le morphème flexionnel *-er*, et de très nombreuses affixations appliquées sur une base empruntée (voir section 3.1.4.).

### Répartition par journaux, domaine, pays

En ramenant les chiffres globaux précédents aux paramètres diastratiques et diatopiques disponibles sur le plateforme, on obtient les résultats suivants :

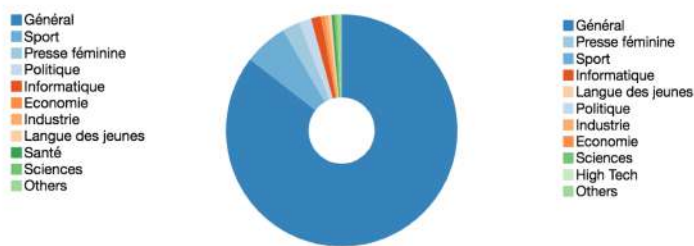
**Figure 2.** Comparatif des répartitions par domaines et par journaux (tous néologismes versus emprunts)



<sup>17</sup> La distribution constatée, restreinte aux mêmes catégories, est la suivante : 53,19 % de noms, 26,50 % de verbes, 11,33 % d'adjectifs, 8,98 % d'adverbes.



## Domaines



Les journaux les plus productifs sont relativement différenciés : pour l'ensemble des néologismes, *L'Express* (46 889 occurrences, 6,45 %), *Libération* (28 551, 3,93 %), *France Soir* (27 828), *Le Huffington Post* (26 237) et *Le Monde* (25 701). Par contre, pour les emprunts, la presse la plus prolifique est représentée par le *Huffington Post*, *Elle*, *L'Express*, *Libération* puis *L'Équipe* et *France Soir*. On peut voir dans cette modification de la distribution une trace des types de presse où les emprunts apparaissent et/ou se rencontrent le plus souvent : la presse internationale disposant d'une version française (*Huffington Post* et *Slate*), la presse féminine (*Elle*, *Grazia* parmi les quinze premiers), la presse populaire (*France Soir*, *Le Parisien*), la presse généraliste (*Libération*, *L'Express*), la presse sportive (*L'Équipe*) et dans une moindre mesure la presse informatique et les journaux économiques.

Cette interprétation est confirmée par la répartition en domaines : pour l'ensemble des néologismes, trois domaines sont prédominants<sup>18</sup>, sport (10 %), presse féminine (10 %) et informatique (5 %); les tendances sont accentuées pour les emprunts : presse féminine (15 %), sport (12 %) et informatique (9 %). Ces domaines ont des occurrences pour près de 50 % des emprunts relevés. Si nous croisons emprunts et presse féminine, les journaux les plus productifs sont : *Grazia*, *Elle*, *Styles*, *Madame Figaro* et *Cosmo*. Si nous focalisons sur *Grazia*, on constate que près de 70 % des néologismes qui s'y trouvent sont des emprunts ! La presse féminine, certains titres en particulier, ont donc cette caractéristique de recourir massivement aux emprunts (et plus particulièrement aux anglicismes).

Cette caractéristique est encore confirmée lorsque nous étudions les seules premières occurrences des emprunts : près de 50 % d'entre elles se trouvent dans la presse féminine, ce qui appuie l'idée que non seulement ce type de presse recourt massivement aux anglicismes, mais constitue également un lieu privilégié d'émergence de ce type de néologismes, jouant le rôle d'innovateur. Nous présentons dans le tableau 6 des exemples d'emprunts selon les domaines.

**Tableau 5.** Exemples d'emprunts dans les domaines de la presse féminine, de l'informatique et du sport

Presse féminine	Informatique	Sport	Autres
Styling, contouring, fashion-week, shopper, coming-out, e-shop, lifestyle, street-food, casual, hype, pop-streetware, lipstock, blur, slow-food...	Geek, boost, hashtag, playlist, emoji, tag, datacenter, ransomware, snapchat, big-data, instagrammer, scrapbooker, pop-up, blockchain, deep learning, chatbot, hoax,	Wild-card, snowboard, running, aquabike, snorkeling, futsal, kiteboarding, wakesurfing...	Think-tank, fake-news, talk, prime-time, call, push, reviewer, subprime, flat-tax, coworking, guest-star, fact-checking, pop-up store, debrief, car-jacking, bashing, start-up(eur), bingewatcheur, data-journalisme, ghostwriting, sugardaddy, ...

<sup>18</sup> Nous excluons le domaine « général », qui représente plus de 75 % des néologismes, mais les articles correspondants mériteraient sans doute une caractérisation plus fine.



	hotspot, live-streaming, animoji, podcaster, web-réputation...		
--	--	--	--

### 3.2. Cycles de vie des emprunts

#### 3.2.1. Émergence et hapax : remise en cause de la pertinence de la notion d'hapax et proposition de critères pour caractériser l'émergence

Comme nous l'avons vu, la moyenne d'occurrences est relativement faible par néologisme, d'autant plus si nous limitons le comptage aux occurrences survenant dans des documents différents. La déviation standard<sup>19</sup> est cependant importante (111 par forme néologique, 237 par nombre total d'occurrences), montrant qu'il existe un petit nombre de néologismes qui sont employés de façon massive dès leur apparition (notamment toutes les innovations liées à une actualité : *loi-travail*, *nuit-debout*, *penelopegate*, *cyberattaquant*, *street(-)wear*, *street(-)art*, etc.). Si nous utilisons la médiane, le nombre d'occurrences tombe à 4 : une très large majorité de néologismes sont donc principalement des hapax ou des quasi-hapax. Les figures 3 et 4 présentent les données de manière plus détaillée : pour les six types de néologismes principaux (en ordonnée), on identifie la distribution des néologismes par nombre d'occurrences (en abscisse).

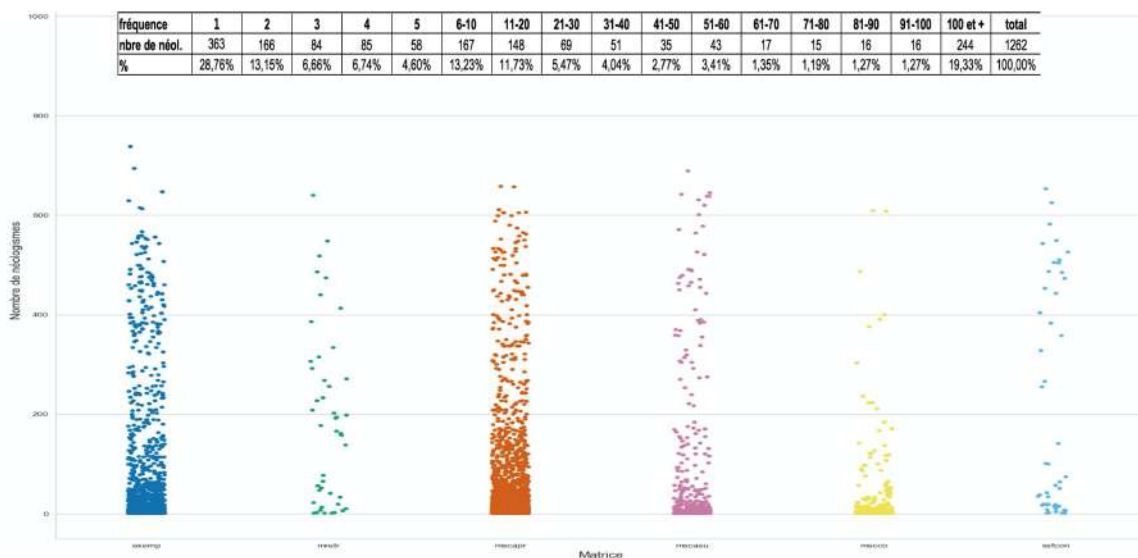


Figure 3. Distribution des néologismes par fréquence (en abscisse), pour six types de néologismes (en ordonnée). Le tableau détaille la répartition des fréquences pour les emprunts.

<sup>19</sup>En statistique, on appelle *déviation standard* ou *écart-type* (*standard deviation*), la distance entre la valeur minimale et la valeur maximale d'une série. Cette mesure permet d'approcher la dispersion d'une distribution. La médiane rend compte de la valeur moyenne, en additionnant toutes les valeurs individuelles. Elle rend donc mieux compte de la tendance générale d'une série.

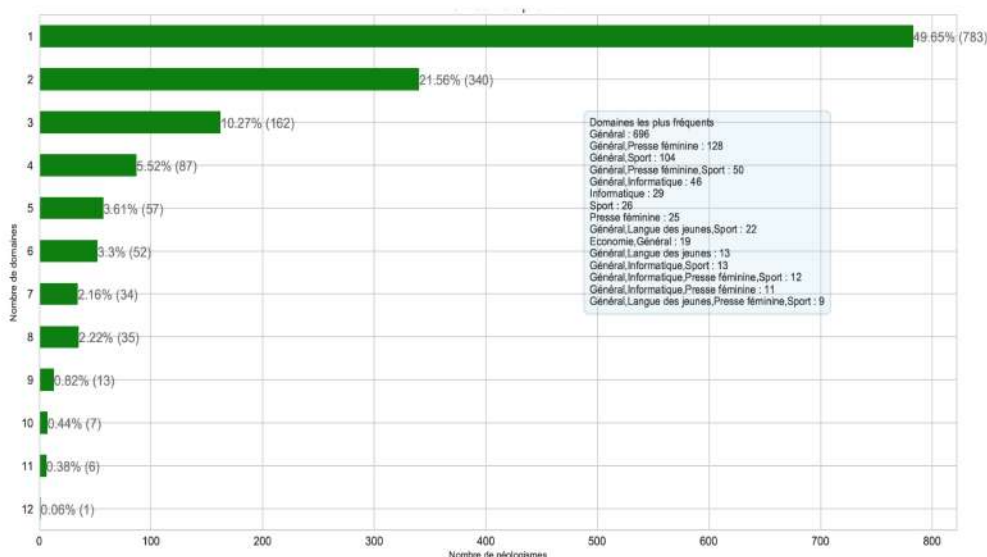


Figure 4. Distribution des emprunts par nombre de domaines représentés. Le tableau interne détaille les combinaisons de domaine les plus fréquentes.

On constate que les hapax proprement dits (une seule occurrence) représentent environ 25% du total des néologismes, mais il existe un continuum entre les fréquences très basses et les fréquences plus hautes, sans qu'il soit possible de tracer une frontière *linguistique* pertinente entre les hapax et les autres néologismes (notamment ceux ayant une fréquence « faible »). Si nous étendons le moment d'émergence à deux semaines à partir de la première apparition, nous constatons que plus de 70 % des néologismes sont représentés, ce qui tend à montrer que la non-diffusion serait mieux définie comme une répétition « faible » sur une période courte, plutôt que via la notion d'hapax. On pourrait même affiner cette définition : *l'émergence est une période courte (de l'ordre de quelques jours ou quelques semaines) durant laquelle une innovation lexicale apparaît et peut se répéter, très généralement dans le même domaine que celui de la première apparition.* Par exemple, *zebracake* et *tigercake*, malgré un *buzz* (7 et 9 occurrences) en avril 2016, sont restés cantonnés à la presse féminine dans les rubriques culinaires. De même pour *street-girl*, qui est resté cantonné à la presse féminine. La figure 3 illustre la validité de ce critère, en montrant la répartition des emprunts selon le nombre de domaines représentés : on constate que près de 50 % des innovations sont cantonnées à un seul domaine (principalement le domaine général : 696, l'informatique : 29, le sport : 26 et la presse féminine : 25). Par contre, dès qu'une innovation est représentée dans plus d'un domaine, il s'agit d'un signe de diffusion (par exemple, ici la combinaison domaine général-presse féminine comprend 128 lexies, la combinaison sport, général 104).

Ces critères additionnels à la notion d'hapax pour caractériser l'émergence sont liés à la structuration contemporaine de la communication : d'une part, la communication via internet accélère la connaissance des autres informations publiées, favorisant les *mêmes* et en tout cas la diffusion ; d'autre part, de nombreux groupes de presse détiennent plusieurs titres, et il est fréquent de voir des répétitions survenant quasiment au même moment, dans des journaux différents, souvent avec reprise intégrale de la phrase complète (voire de l'article...). Le texte suivant, par exemple, se retrouve, à quelques heures d'intervalle, dans trois titres différents, le 18 mars 2016 (nous soulignons les innovations, partiellement basées sur trois emprunts anciens (*cookie*, *doughnut* et *brownie*, début du 19<sup>e</sup>), sur le mode de la compocotation):

Les croisements sont à la mode. Ils ont déjà accouché du crookie (croissant mélangé avec un cookie Oreo), le duffin (mariage entre doughnut et muffin anglais), le bronut (brioche feuilletée sucrée), et surtout le cronut, union entre le



croissant et le doughnut, du chef français Dominique Ansel, basé à New York.  
(18 mars 2016, *L'Express*, *Libération*, *Le Parisien*)

Enfin, un dernier critère est particulièrement utile pour détecter une phase d'émergence : la mise en exergue de la lexie, d'une part, et l'existence d'une glose dans l'environnement immédiat du néologisme. Cette glose est évidemment requise pour la très grande majorité des innovations, dans un souci de compréhension par les destinataires : « ... nombre d'applications identifiées par Symantec comme étant des " malwares ". A la croisée des malwares et des adwares, cette génération d'apps... » (*Le Monde Informatique*, 13 avril 2016). Cependant, là encore, ce critère n'est pas suffisant, notamment lorsque l'énonciateur considère que le sens est inférable par compositionnalité : « ...Réponse jeudi où se tiendra juste après le défilé, une grande after-party ... » (*Elle*, 21/02/2017). Nous renvoyons à (Jacquet-Pfau 2018) pour une étude détaillée de ces marqueurs de glose pour les emprunts.

Une dernière caractéristique des quasi-hapax est liée au périmètre sémantique des unités lexicales : elles désignent dans leur très grande majorité, soit des réalités locales – à la limite du xénisme – (*dibbuk*, *wapeningen*, *escrache*), soit des concepts circonscrits à un domaine spécifique (*pika-don*, *nanotrading*, *cutlet*, *fadeaway*) ou à des pratiques sociales confidentielles (*selfie-whore-stick*, *denki-buro*, *nightswapping*).

### 3.2.2. Diffusion des emprunts : exemples et enseignements

Parmi les emprunts repérés, plus de 85 % sont des hapax ou des quasi-hapax. Mais qu'en est-il des 15 % restants ? Suivons les chemins de leur diffusion.

#### 3.2.2.1. Adaptations phonologique, orthographique et morphosyntaxique

On peut faire l'hypothèse que les emprunts obéissent aux différentes phases communes d'intégration. Il se produit également, étant donné que le matériau emprunté provient d'un autre système, une adaptation au système du français aux niveaux phonologique, orthographique et morphosyntaxique. Concernant l'adaptation orthographique, le corpus Néoveille ne présente pas de particularité concernant les emprunts à des langues alphabétiques, dont les lexèmes sont généralement rendus tels quels<sup>20</sup>. Pour l'arabe, la translittération donne parfois lieu à des hésitations (*jihadiste*, *djihadiste*, avec cependant une prédominance depuis environ 2010 de la version *dj-*).

Concernant l'adaptation phonologique, elle est plus difficile à déterminer étant donné le corpus écrit. Notons cependant le cas de *check*, qui dans tous les cas conserve la prononciation anglaise (/ʃ/).

L'adaptation morphologique, pour les anglicismes, ne présente pas, généralement, de difficultés pour les noms (sans ajout de morphème) et les verbes (par ajout du morphème *-er*), mais une particularité déjà étudiée par (Saugera 2012 ; 2017 : 123-138) concerne les adjectifs empruntés à l'anglais, qui, au pluriel, s'accordent ou non, selon le cas. Dans notre corpus, on retrouve ainsi un grand nombre d'adjectifs en *-y*, invariables (*arty*, *sketchy*, *glowy*, *skinny*, *girly*, *creepy*, *healthy*, *edgy*, *catchy*, *flashy*, *bluesy*, etc.), au point qu'il est probable que ce formant soit devenu un formant suffixal productif en français.

#### 3.2.2.2. Intégration à la morphologie productive

Un autre signe de diffusion concerne l'intégration à la morphologie productive. La grande majorité des emprunts relevés dans notre corpus subissent ce « passage » à la dérivation. Les exemples les plus typiques sont liés aux bases *nom propre* issus des réseaux sociaux (tableau 6). Dans ce tableau, nous distinguons les dérivations par morphèmes grammaticaux (nom et verbe) des affixations

<sup>20</sup> Notons toutefois le cas de *twitter* / *tweet*, qui, pour l'emprunt simple, est plus souvent non-adapté, mais qui l'est la plupart du temps dans ses versions dérivées (*twittos*, *twictée*, *twitonaute*, *twitcam*, etc.). Un autre exemple concerne les dérivés à base *instagram*, dont les dérivés connaissent deux graphies concurrentes : *instagrammeur*, *instagrameur*. Enfin, le pluriel des noms peut donner lieu à des variantes (*smartwatches* ou *smartwatchs*).

proprement dites. Nous constatons que l'étendue des dérivations est liée à la popularité du réseau, *Twitter* étant largement en tête. Toutes ces dérivations appliquent les procédés affixaux les plus productifs du français contemporain (Cartier *et al.* 2018). On notera la particularité de *twitto(s)*, pour désigner le ou les émetteur(s) d'un *tweet*, directement emprunté de l'anglais et sans concurrence du dérivé français pourtant très bien implanté en *-eur(euse)* et utilisé pour les autres termes. De même, on notera que seuls *twitter* et *snapchat* ont une lexie (directement empruntée) pour désigner le type de message (*tweet* et *snapchat*, parfois tronqué en *snap*). Cependant, *Facebook* a aussi permis l'emprunt de *like* (n) et *liker* (v).

**Tableau 6.** échantillon de dérivations attestées pour cinq bases nom propre issues des réseaux sociaux

	facebook	twitter	instagram	snapchat	youtube
Intégration morphologique (morphème flexionnel)	facebooker (v)	twit(s) (n) tweeter (v) twitto(s) (n)	instagram(m)er	snapchater (v) snapchat(s) (n) > snap(s)	youtuber (v)
Intégration morphologie productive (affixes, fracto-lexèmes et formants savants)	facebookeur(e use) facebookien(n e) facebooking anti-facebook facebookisme	twitteur(euse) tweeteur (euse) re(-)tweeter tweeterisation tweeting anti-tweet demi-tweet non-tweet auto-tweet pseudo-tweet tweetesque tweetable tweetonade	instagram(m)eur(euse)	snapchat(t)eur(euse) snapchat(t)ien(ne)	youtubeur(euse) youtubing

Si l'on s'intéresse aux emprunts à base *nom commun* (plus ou moins récents), on pourra faire les mêmes constats pour les emprunts les plus populaires (*blog, food, hashtag, check, shop, geek, market, game*, etc.). Par exemple, l'histoire de la pratique journalistique du *fact-checking* (terme dont les premières attestations en français datent de 1998, mais dont l'emploi connaîtra un premier pic avec les élections américaines de 2012 et un second, plus intense encore, avec celles de 2016 et en France en 2017) est éloquent : jusqu'aux élections américaines de 2012, les seuls emplois attestés sont *fact-checking*, avec recours quasi-systématique à la glose-traduction (par exemple : « Au printemps, le site va aussi s'allier avec d'autres médias afin de développer le "fact checking", une méthode répandue dans les pays anglo-saxons qui consiste à vérifier les chiffres et les affirmations des hommes politiques. », AFP, 18/02/2011). Puis des emplois non métalinguistiques apparaissent (« le fact-checking fait sa rentrée sur les ondes radio », *Libération*, 18/09/2012). Mais c'est seulement avec les élections américaines fin 2016 et les élections françaises en 2017 qu'une cohorte de dérivés fait son apparition (*fact-checker, factcheckeur, fact-checkings*) ainsi que, plus récemment encore, des composés sémantiquement liés (notamment le *fast-fact-checking* devenu *fast-checking*). Cependant, l'emploi verbal reste limité à l'infinitif, sauf timides exceptions (« les médias qui fact-checkeront les articles litigieux », *Libération*, 11/01/2017).

Enfin, dans le même ordre d'idée, la popularité d'une base lexicale empruntée se traduit également par la génération de composés et de fractocomposés. Toujours sur l'exemple *twitter*, nous relevons : *tweet-boomerang, tweet-choc, tweetosphère, tweet-série, feu-tweet, tweeteur-en-chef* ainsi que plusieurs autres formations directement empruntées : *tweetwall, acrostweet, live-tweet, tweetdeck, tweetstorm, fake-tweet, tweetbot, commander-in-tweet*, etc.



### 3.2.2.3. Processus de mise en place d'un profil combinatoire

La mise en place d'un *usage* des néologismes passe par l'abandon progressif (sauf visée didactique spécifique) des marques métalinguistiques, qui accompagnent l'émergence des lexies. De ce point de vue, les données permettent d'observer le passage de l'un à l'autre. Prenons l'exemple de *ghosting*, cette pratique consistant, dans une relation amoureuse, à disparaître brusquement, sans plus répondre aux sollicitations du/de la partenaire. Le terme apparaît dans la presse américaine en 2014<sup>21</sup>. Très rapidement le terme se répand aussi en français, avec dérivation morphologique (*ghoster qn*, *ghosteur (euse)*, *ghosté(e)*), et intégration à la morphologie productive (*anti-ghosting*, *néo-ghosting*). Si l'on scrute les emplois de *ghosting*, dans le corpus Néoveille, sur 17 attestations (voir extraits tableau 8), on constate que les emplois métalinguistiques (guillemets, glose) tendent à disparaître, en tout cas dans la presse féminine et la presse magazine parisienne. Cette tendance est encore plus forte avec le verbe *ghoster*, apparu dans un second temps, et dont l'emploi transitif, également emprunté (to ghost someone > *ghoster* + *Nom* ; également emploi passif *être ghosté par Nom*, et factitif : *se faire ghoster par Nom*) et les contraintes argumentales sur l'objet (*personne, mec, type, Pauline, etc.*) dessinent très rapidement l'usage syntactico-sémantique.

**Tableau 7.** échantillon de contextes avec *ghosting* depuis début 2016

Date	Journal	Domaine	Extrait
24/05/18	<i>Elle</i>	Presse féminine	...Et dans le registre -ing de nos comportements amoureux, le <b><i>ghosting</i></b> demeure le plus célèbre : on disparaît sans un mot ...
30/03/18	<i>Slate</i>	Général	...La pratique du <b><i>ghosting</i></b> – la rupture sans explication – en est le signe...
14/03/18	<i>Slate</i>	Général	...Le no-show, équivalent du <b><i>ghosting</i></b> mais version Guide Michelin C'est tellement simple que la...
26/12/17	<i>Nouvel Observateur</i>	Général	...Elle avait choisi à un moment le " <b><i>ghosting</i></b> ", c'est-à-dire de disparaître totalement dans une forme de...
25/10/17	<i>Le Progrès</i>	Général	...Inutile de préciser que ces cas de <b><i>ghosting</i></b> se produisent en grande partie suite à des relations...
15/10/17	<i>Nouvel Observateur</i>	Général	...me dissimule pas, je ne fais pas du ' <b><i>ghosting</i></b> ' (l'art de disparaître en pleine séduction)...
05/06/17	<i>Nouvel Observateur</i>	Général	Faut-il encore expliquer ce qu'est le <b><i>ghosting</i></b> ...
07/02/17	<i>Libération</i>	Général	...On avait déjà recensé le <i>ghosting</i> , qui consiste à disparaître sans donner de nouvelles...
13/10/16	<i>Cosmo</i>	Presse féminine	...Le <b><i>ghosting</i></b> est plus violent qu'une rupture amoureuse normale...
13/10/16	<i>Cosmo</i>	Presse féminine	...% des filles ont déjà vécu l'expérience charmante du <b><i>ghosting</i></b> ...
13/10/16	<i>Cosmo</i>	Presse féminine	...Dans le cas du <b><i>ghosting</i></b> , le drapeau blanc persiste à flotter au vent...
01/09/16	<i>Cosmo</i>	Presse féminine	...Le <i>benching</i> , pourquoi est -ce pire que le <b><i>ghosting</i></b> ...

Nous renvoyons également au tableau 3, qui détaille les collocations, collostructions et constructions syntaxiques que nous avons pu constater à partir de la lexie *food*.

<sup>21</sup> Nous prenons appui sur l'analyse faite sur Wikipedia et reprise par le Collins, qui a introduit le terme en 2015, qui date l'émergence « écrite » du terme de l'article ci-après : <https://jezebel.com/charlize-theron-broke-up-with-sean-penn-by-ghosting-him-1712760688>

### 3.2.2.4. Évolution des contextes socio-pragmatiques des innovations

L'exemple précédent illustre un autre phénomène : l'importance de l'inscription socio-pragmatique des innovations lexicales : *ghosting*, *ghoster* ne sont plus glosés dans la presse féminine, mais le sont encore dans la presse généraliste, montrant qu'ils restent, pour le grand public, en phase d'émergence. L'attestation d'un néologisme hors de son domaine d'émergence est donc un autre signe de sa diffusion. Dans Néoveille, les paramètres pour décrire ces propriétés sont encore trop grossiers (domaine, journal, pays de la source d'information), mais permettent néanmoins de constater des diffusions différenciées. Dans les figures 5 et 6, à titre d'illustration, nous pouvons visualiser la distribution temporelle (2016-2018) des domaines pour deux innovations lexicales, l'une datant des années 2000 (*smartphone*) et une autre bien plus récente (*smartwatch*) : on constate que la première a maintenant pénétré beaucoup de domaines, montrant un emploi non limité à un groupe socio-économique, tandis que la seconde, dont on peut constater l'émergence dans les domaines *hightech* et *informatique*, se diffuse maintenant dans la presse généraliste.

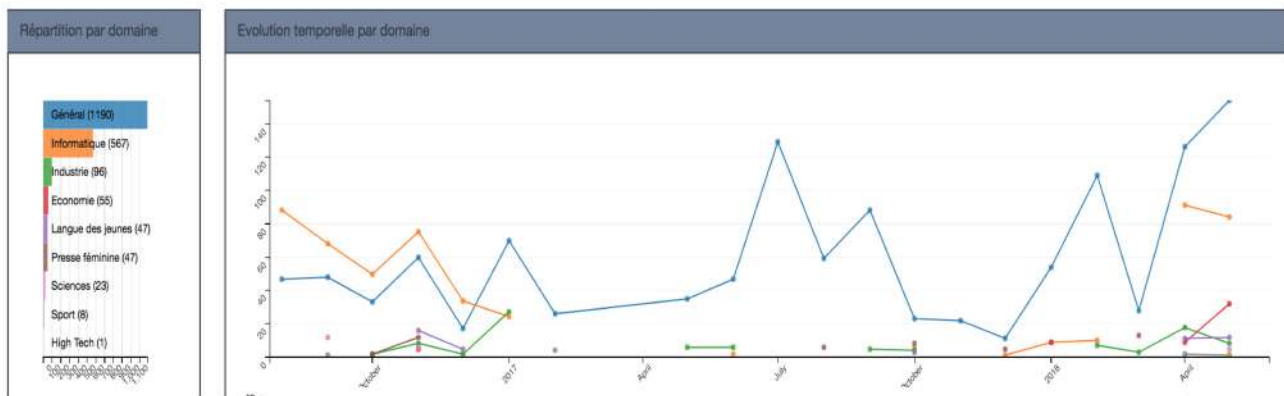


Figure 5 – distribution temporelle par domaine pour *smartphone*

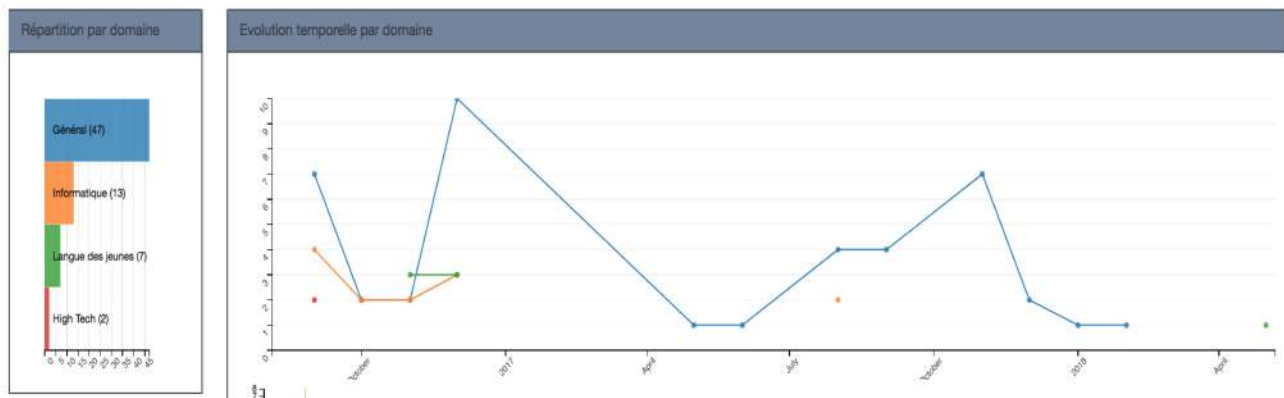


Figure 6 – distribution temporelle par domaine pour *smartwatch*

L'observation des modifications de domaine fournit au moins deux autres informations : d'une part, elle permet de connaître les restrictions éventuelles de domaine d'application de la lexie (par exemple *phablet(te)*, apparu dans le domaine informatique mais pénétrant peu les autres domaines ; les termes des réseaux sociaux n'ont pas de telles limitations). D'autre part, elle permet d'identifier, dans le cadre de la théorie de l'innovation (Rogers 2003 [1983]), les groupes sociaux *innovateurs* et *diffuseurs* (adopteurs), grâce au suivi temporel de la distribution domaniale.

### 3.3. Emprunts de patrons lexico-syntaxiques productifs.

Les emprunts à l'anglais ne se limitent pas au transfert de lexies. Du point de vue phonologique et orthographique, l'influence de l'anglo-américain est perceptible depuis longtemps (prononciation de *-ing*, *-ee-*, etc.). La pénétration est également visible par l'implantation de formants à fonctionnement affixal : pour ne citer que les plus productifs, citons *e-* (formateur de verbes et de noms, dans le sens « ayant une caractéristique électronique ou numérique »), *-y* (formateur d'adjectifs, dans le sens vague de « ayant la plupart des caractéristiques de la base ») et, moins récent, *-ing* (formateur de noms d'action ou d'événement). Dans le corpus Néoveille, ces formants ont une productivité importante : 86 lexies pour le premier (soit emprunts directs, moins de dix : *e-voting*, *e-shopping*, etc. soit hybrides : *e-défilé*, *e-vendeur*, *e-marché*, *e-citoyenneté*, etc.), 22 pour le second (uniquement par emprunts directs : *jazzy*, *buggy*, *girly*, *cosy*, *healthy*, *creepy*, *skinny*, *glowy*, *flashy*, *bluesy*, *catchy*, *crazy*, *bitchy*, *smoky*, *edgy*, *flexy*, *wavy*, etc.), 303 pour le dernier. Le morphème *-ing* est attesté depuis plus d'un siècle (*parking*, *camping*, *pressing*, *meeting*, *dancing*, etc.), formant essentiellement des noms de lieux où une action se déroule, par métonymie du sens anglais. Cependant, à partir des années 50, le morphème obtient le statut de quasi-suffixe, exprimant « une action, son résultat ou le lieu où se déroule cette action » (Dubois 1962 : 14). (Mudrochova 2017) étudie une trentaine de formes attestées dans *Le Petit Robert*, entre 1996 et 2002. La concurrence avec *-age* fait qu'il reste limité à l'expression de pratiques sportives (*running*, *beatboxing*, *snorkeling*, *cardiotraining*, etc.) professionnelles (*networking*, *packaging*, *branding*, *fact-checking*, *coworking*, *crowdfunding*...) ou socio-culturelles (*bashing*, *ghosting*, *pet-sitting*) spécifiques sans équivalents synthétiques en français. C'est un des signes les plus flagrants d'un *code-switching* lié au prestige dans plusieurs journaux (de façon conséquente dans le *Huffington Post* et la presse féminine).

Une autre caractéristique des emprunts à l'anglais concerne l'émergence de patrons lexico-syntaxiques productifs. Notamment : les formations en *-gate* (62 occurrences : *dieselgate*, *couscousgate*, *penaltygate*, *penelopegate*, etc.), *-friendly* (41 lexies : *vidéo-friendly*, *nudistes-friendly*, *lobby-friendly*, *éco-friendly*, etc.), *street-* (26 lexies : *streetstyle*, *street-artiste*, etc.), *food-* (29 lexies : *food-truck*, *foodosphère*, *foodocratie*, *foodivores*, *street-fooders*, etc.), *-bashing* (11 lexies : *agribashing*, *sucre-bashing*, *macronbashing*, etc.), *-shaming* (14 lexies : *fatshaming*, *name-shaming*, *skillshaming*), *it-* (8 lexies : *it-jean*, *it-bag*, etc.), *serial-* (2 lexies : *serial-buteur*, *serial-cendrillonneur*). Nous relevons également dans notre corpus 144 occurrences du patron N/ADJ-Ving (*car-jacking*, *home-staging*, *speed-dating*, *speed-watching*, *binge-viewing*, *ride-sharing*). Ce dernier cas semble montrer l'émergence d'une nouvelle structure syntaxique en N/ADJ N, clairement empruntée à l'anglo-américain (voir Cartier et Viaux 2018).

La diffusion, l'implantation de ces formants et schémas lexico-syntaxiques nous semble un phénomène plus profond que les simples emprunts. En reprenant l'hypothèse de l'*impulsion synthétique* (*synthetic imperative*) (Picone 1991 ; 1996), le français, langue analytique, tendrait maintenant à plus de formations synthétiques, ce qui se manifeste clairement avec la productivité des schémas empruntés à l'anglais N/ADJ-N, ainsi que l'implantation du suffixe *-ing*.

### 3.4. Emprunts et politique linguistique

Les lexies empruntées peuvent entrer en concurrence avec des lexies déjà implantées, ou avec les recommandations des organismes associés aux politiques linguistiques de la langue réceptrice. Dans le cadre francophone, la politique linguistique est particulièrement active, à la fois pour la métropole et dans la zone québécoise. De nombreuses études ont déjà été faites sur le sujet, récentes (Saugera 2017 ; Humbley 2010 ; Steuckardt 2008) et moins récentes (Picone 1996). Beaucoup démontrent que les propositions, dans le cadre métropolitain, ne donnent pas les résultats escomptés, mais les outils de suivi sont encore très impressionnistes : il serait urgent de développer des outils de suivi comparatif des nouveaux emprunts et de leurs équivalents forgés ou non. Dans le cadre de Néoveille, un tel suivi fréquentiel comparatif sera prochainement disponible, et un autre



projet permettra sous peu de suivre temporellement les implantations comparées des emprunts et des propositions des collègues d'experts de la DGLFLF<sup>22</sup>.

## Conclusion et Perspectives

Nous avons présenté dans ce travail à la fois quelques jalons théoriques, une méthodologie pour l'analyse des différents aspects de l'objet « néologisme », et les résultats d'analyse concernant les emprunts en français contemporain (2015-2018). Les travaux ont été menés dans le cadre de la plateforme Néoveille, qui permet de détecter semi-automatiquement, de décrire linguistiquement et de suivre l'évolution temporelle des néologismes sous trois points de vue complémentaires : évolution temporelle, évolution linguistique et socio-pragmatique.

La présentation théorique a insisté sur l'importance de considérer trois moments saillants de la vie des néologismes : l'émergence, l'éventuelle diffusion et l'éventuelle lexicalisation. Pour décrire et caractériser ces moments, nous avons insisté sur l'importance de combiner trois points de vue ; un point de vue linguistique, permettant de décrire les mécanismes de formation, mais également les profils combinatoires et distributionnels ; un point de vue socio-pragmatique, permettant d'identifier les lieux d'occurrences des néologismes. Un point de vue cognitif, qui permet de faire reposer l'évolution des innovations sur le principe de l'*entrenchment*, qui se manifeste par la répétition. À l'aide de ces trois points de vue, nous pouvons modéliser un système capable de détecter semi-automatiquement l'émergence et de suivre l'évolution des innovations sur de grands corpus dynamiques. C'est là l'objet de la plateforme Néoveille, développée depuis 2015, et qui continue d'évoluer au gré des travaux descriptifs menés par les groupes de travail dans les différentes langues du projet (allemand, chinois, espagnol, français, grec, italien, néerlandais, polonais, portugais, russe, tchèque).

Notre étude des emprunts a suivi les principes décrits dans la seconde partie, et plusieurs enseignements en découlent : tout d'abord, les emprunts en français contemporain représentent environ 6 % (soit 1 429 lexies) des formations nouvelles en français. À ce chiffre, il faut ajouter environ un millier de xénismes, ainsi qu'un nombre conséquent d'autres formations sur une base empruntée. Plus de 90 % de ces emprunts proviennent de la *lingua franca* anglo-américaine, tandis que les xénismes sont bien plus diversifiés. Il existe par ailleurs des domaines privilégiés de recours aux emprunts : la mode, le sport, les domaines technologiques et économiques. Les journaux les plus prolifiques appartiennent à la presse féminine, à la presse magazine parisienne, ainsi qu'à la presse populaire. Les emprunts, enfin, comme l'ensemble des néologismes, sont à plus de 75 % des hapax ou des quasi-hapax, ou plus exactement des néologismes à faible ou très faible diffusion. Il est difficile de dire si les emprunts à l'anglo-américain sont liés à des nécessités dénominatives, ou bien motivés par l'expression d'une identité liée au prestige de l'anglais. En effet, depuis plusieurs dizaines d'années, et de manière plus prégnante encore depuis l'avènement des communications numériques, l'anglo-américain est quasiment une seconde langue intégrée à laquelle nous sommes confrontés régulièrement, à la fois au niveau personnel et professionnel. Cette situation est vraie pour la majorité des langues actuelles. Cela aboutit à faire de cette *lingua franca globale* une ressource immédiatement disponible pour l'innovation lexicale, et qui fournit également des moules morphosyntaxiques productifs spécifiques, dotés d'un trait *synthétique* de plus en plus implanté en français.

L'étude des néologismes à plus forte fréquence montre que, si les domaines des réseaux sociaux et des pratiques sociales spécifiques connaissent une grande popularité, les emprunts sont présents dans de nombreux domaines. Plusieurs paramètres nous ont permis de présenter quelques

---

<sup>22</sup> Le projet Néonaute, financé par la DGLFLF en 2017, regroupe le LIPN (Université Paris 13, coordinateur), le LILPA (Université de Strasbourg) et la Bibliothèque Nationale de France (département des archives du web), et permettra de suivre l'évolution diachronique et diastratique des néologismes repérés par Néoveille et le Logoscope, ainsi qu'un échantillon des propositions de la DGLFLF en remplacement d'anglicismes, et ce sur un corpus de presse allant de 2011 à 2016, le corpus Néoveille permettant de suivre les néologismes depuis lors.

cas de diffusion, au travers d'une diffusion non pas seulement liée à une répétition stabilisée, mais également à plusieurs phénomènes linguistiques, notamment l'intégration à la morphologie grammaticale et productive du français. Les paramètres socio-pragmatiques sont également importants, même si de ce point de vue, il conviendrait de préciser les paramètres pour approcher plus finement les cheminements des innovations.

## Références bibliographiques

- BACKUS Add, DORLEIJN Margreet, BULLOCK B.E. et TORIBIO, A. J. (2009). « Loan translations versus code-switching ». *The Cambridge handbook of linguistic code-switching*, 75-94.
- BAAYEN R. Harald (2001). *Word Frequency Distributions*. Dordrecht: Kluwer.
- BARONI Marco et LENCI Alessandro (2010). « Distributional memory: A general framework for corpus-based semantics ». *Computational Linguistics*, 36(4):673–721.
- BAUER Laurie (2001). *Morphological Productivity*. Cambridge: Cambridge UP
- BYBEE Joan (2010). *Language, usage and cognition*. Cambridge University Press.
- BYBEE Joan (2016). *Language Change*. Cambridge University Press.
- CARTIER Emmanuel et SABLAYROLLES Jean-François (2009). « Néologismes, dictionnaires et informatique ». *Cahiers de lexicologie*, 2008(93), 175-192.
- CARTIER Emmanuel (2016). « Neoveille, système de repérage et de suivi des néologismes en sept langues », *Neologica* 10, p. 101-131.
- 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 ». *Folia Litteraria Romanica*. p.11-34.
- CARTIER Emmanuel (2017), « Neoveille, a Web Platform for Neologism Tracking ». *Proceedings of the EACL 2017 Software Demonstrations*, Valencia, Spain, April 3-7 2017.
- CARTIER Emmanuel, SABLAYROLLES Jean-François, BOUTMGHARINE Najet, HUMBLEY John, BERTOCCI Massimo, JACQUET-PFAU Christine, KÜBLER Natalie et TALLARICO Giovanni (2018). « Détection automatique, description linguistique et suivi des néologismes en corpus : point d'étape sur les tendances du français contemporain ». *Actes du Congrès Mondial de Linguistique française 2018*, Mons 9-13 juillet 2018.
- CHADELAT Jean-Marc (2000). *Valeur et fonctions des mots français en anglais à l'époque contemporaine*. Editions L'Harmattan.
- CHARAUDEAU Patrick (1995). « Une analyse sémiolinguistique du discours ». *Langages*, 96-111.
- CHESLEY Paula et BAAYEN R. Harald (2010). « Predicting new words from newer words: Lexical borrowings in French ». *Linguistics*, 48(6), pp. 1343-1374.
- CHRISTIANSEN Thomas Wulstan (2015). « The rise of English as the global lingua franca. Is the world heading towards greater monolingualism or new forms of plurilingualism? » *Lingue e Linguaggi* (15), 129-154.
- CORBIN Danielle (1987) *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Max Niemeyer Verlag.
- COSERIU Eugenio (1964). *Pour une sémantique diachronique structurale*. Klincksieck, 186 p.
- DAL Georgette et NAMER Fiammetta (2015). « La fréquence en morphologie : pour quels usages ? » *Langages*, (1), 47-68.
- DEROY Louis (2003)[ 1956] *L'emprunt linguistique*. Presses Universitaires de Liège.
- DUBOIS Jacqueline (1962) *Étude sur la dérivation suffixale en français moderne et contemporain*. Librairie Larousse.
- DUBOIS Jacqueline (2007). *Linguistique & sciences du langage*, Larousse.
- FELTGEN Quentin, FAGARD Benjamin et NADAL Jean-Pierre (2017). « Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change ». *Royal Society open science*, 4(11).
- FIRTH John Rupert (1957). *Papers in Linguistics 1934–1951*. Oxford University Press.

- GÉVAUDAN Paul et KOCH Peter (2010). « Sémantique cognitive et changement lexical. Grandes voies et chemins de traverse de la sémantique cognitive. » *Journée scientifique de la Société de Linguistique de Paris*, p. 103-146.
- GOTTLIEB Harald (2005). « Anglicisms and translation. In and out of English: For better, for worse », dans G. Anderman & M. Rigers (eds), Clevedon : *Multilingual Matters*, 161-184.
- GRIES Stephan (2010). Behavioral profiles: « A fine-grained and quantitative approach in corpus-based lexical semantics ». *The Mental Lexicon*, 5(3):323–346.
- GUILBERT Louis (1975). *La créativité lexicale*. Paris : Larousse.
- HAMILTON, W. L., Leskovec, J., and Jurafsky, D. (2016). Cultural shift or linguistic drift ? Comparing two computational measures of semantic change, ACL 2016.
- HARRIS Zellig Sabbataï (1954). « Distributional structure ». *Word*, 10(2-3):146–162.
- HAUGEN, E. (1950), « The analysis of linguistic borrowing », *Language* 26, 210-31.
- HILDENBRAND Zuzana et Jacquet-Pfau Christine (2016). « Les pratiques alimentaires, un domaine très ouvert aux emprunts. Analyse en français et en tchèque », dans Z. Hildenbrand, A. Kacprzak et J.-F. Sablayrolles (éd.), *Emprunts néologiques et équivalents autochtones en français, en polonais et en tchèque*, Limoges, Lambert-Lucas, coll. « La Lexicothèque », 2016, p. 176-225.
- HUMBLEY John (2010). « Peut-on encore parler d’anglicisme? » Dans *Lexique, normalisation, transgression*. Mes Mots Edition. Limay, 2010. p. 21-45.
- JACQUET-PFAU Christine (2018). « À propos des emprunts néologiques dans le discours journalistique : marquage et commentaires », dans D. Bernard, M. Boisseau, C. Gérard, T. Grass et E. Todirascu, *Créations lexicales et variations contextuelles*, Actes du colloque LTT, Strasbourg 2015, Limoges, Lambert-Lucas [sous presse].
- KOCH Peter (2000). « Pour une approche cognitive du changement sémantique lexical : aspect onomasiologique ». *Mémoires de la société linguistique de Paris*, pages 10–75.
- LABOV William (2001). *Principles of linguistic change, tome 2 : social factors*, Oxford, Blackwell.
- LOUBIER Chistiane (2011). *De l’usage de l’emprunt linguistique*. Office québécois de la langue française.
- LANGACKER Ronald (1987). *Foundations of cognitive grammar. Volume 1. Theoretical prerequisites*. Stanford University Press Stanford.
- LANGACKER Ronald (1991). *Foundations of cognitive grammar: Volume II :Descriptive Application*. Stanford university press.
- LE PESANT Denis, Mathieu-Colas Michel (1998). Introduction aux classes d’objets. *Langages*, 6-33.
- MARTINEZ Camille (2009). *L’évolution de l’orthographe dans les Petit Larousse et les Petit Robert 1997-2008 : une approche généalogique du texte lexicographique*. Thèse de doctorat soutenue à l’université Cergy-Pontoise.
- McMAHON April (1994). *Understanding language change*. Cambridge : Cambridge University Press.
- MEILLET Antoine (1904). « Comment les mots changent de sens ». *L’Année sociologique* (1896/1897-1924/1925), 9:1–38.
- MIKOLOV Tomas, YIH Wen-tau et ZWEIG Geoffrey (2013). « Linguistic regularities in continuous space word representations ». In *Proceedings of HLT-NAACL*, volume 13, pages 746–751.
- MILROY James et MILROY Lesley. (1985). « Linguistic change, social network and speaker innovation ». *Journal of linguistics*, 21(2), 339-384.
- MUDROCHOVÁ Radka (2017). « À propos des mots en-ing d’origine anglaise issus du dictionnaire le Petit Robert ». *Linguistica Pragensia*, 27(1), 7-19.
- NEVALAINEN Tertu (2015). « Descriptive adequacy of the S-curve model in diachronic studies of language change ». In *Varieng, Studies in Variation, Contacts and Change in English, vol.16: Can We Predict Linguistic Change?* Helsinki.
- PAVEAU Marie-Anne (2017). *L’analyse du discours numérique. Dictionnaire des formes et des pratiques*. Hermann (Editions).
- PULCINI Virginia, Furiassi Christiano et González Félix (Eds.) (2012). *The Anglicization of European Lexis*. John Benjamins Publishing.
- REUTNER Ursula (ed.) (2017). *Manuel des francophonies*. Walter de Gruyter GmbH & Co KG.



- ROGERS Everett (2003 [1983]) *The Diffusion of Innovations*. Fifth Edition. The Free Press, New York.
- REY-DEBOVE Josette (1997). *Le métalangage*, Paris : Armand Colin
- SABLAYROLLES Jean-François (2010). « Neologia une base de données pour la gestion des néologismes ». *Actes del I congrès internacional de neologia de les llengües romàniques. Barcelone: UPF/IULA*, (22), 757-766.
- 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 (2010). « Néologie, classes d'objets et extraction automatique », actes du Congrès Cineo (néologie des langues romanes) de Barcelone (7-10 mai 2008), 2010, p. 143-149.
- SABLAYROLLES Jean-François (2017). *Les néologismes. Créer des mots français aujourd'hui*, Petits guides de la langue française, n° 29, Paris, Garnier et *Le Monde*.
- SABLAYROLLES Jean-François et PRUVOST Jean (2016). *Les néologismes*. Presses Universitaires de France-PUF, collection Que sais-je ?
- SAUGERA Valérie (2012). « The inflectional behavior of English-origin adjectives in French ». *Journal of French Language Studies*, 22(2), 225-250.
- SAUGERA Valérie (2017). *Remade in France: Anglicisms in the Lexicon and Morphology of French*. Oxford University Press.
- SCHMID Hans-Jörg (2008), « New words in the mind: Concept-formation and entrenchment of neologisms. » *Anglia* 126 (1), 1-36.
- SCHMID Hans-Jörg (2015). « A blueprint of the entrenchment-and-conventionalization model ». *Yearbook of the German Cognitive Linguistics Association*, 3(1):1-27.
- SCHMID, Hans-Jörg (Ed.). (2016). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Walter de Gruyter GmbH & Co KG.
- SIEPMANN Dirk, BÜRCEL Christoph et DIWERSY Sascha (2016). « Le Corpus de référence du français contemporain (CRFC), un corpus massif du français largement diversifié par genres » *Actes de CMLF 2016, SHS Web of Conferences*, 27 11002.
- STEFANOWITSCH Anatol et GRIES Stephan Th. (2003). « Collostructions: Investigating the interaction of words and constructions ». *International journal of corpus linguistics*, 8(2):209-243.
- STERN Gustaf (1931). *Meaning and change of meaning ; with special reference to the english language*.
- THOMASON Sarah-Grey et KAUFMAN Terence (1988). *Language contact, creolization, and genetic linguistics*. Berkeley : University of California Press.
- STEUCKARDT Agnès et Niklas-Salminen Aïno (éds.) (2003). *Le mot et sa glose*, Coll. Langues et langage, n° 9, Aix-en-Provence, Publications de l'Université de Provence.
- STEUCKARDT Agnès et Niklas-Salminen Aïno (éds.) (2005). *Les marqueurs de glose*, Aix-en-Provence, Publications de l'Université de Provence.
- STEUCKARDT Agnès (2008). « L'emprunt, un événement linguistique ». *Neologica: revue internationale de la néologie*, no 2, p. 9-17.
- TOURNIER Jean (1985). *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Champion Books.
- TRAUGOTT Elisabeth Closs et TROUSDALE Graeme (2013). *Constructionalization and Constructional Changes*. Oxford University Press, Oxford Studies in Diachronic and Historical Linguistics.
- WINTER FROEMEL Esme, 2009, « Les emprunts linguistiques : enjeux théoriques et perspectives nouvelles », *Neologica* 3, p. 79-122.