

## Catégorisation comme néologisme et sentiment des locuteurs

Alizée Lombard & Richard Huyghe

Université de Fribourg

[alizee.lombard@unifr.ch](mailto:alizee.lombard@unifr.ch)

[richard.huyghe@unifr.ch](mailto:richard.huyghe@unifr.ch)

### Résumé

Certaines définitions mentionnent la capacité à susciter un sentiment de nouveauté lexicale comme une propriété essentielle des néologismes. Dans cet article, nous évaluons les corrélats d'une telle conception en nous intéressant aux variations du sentiment néologique selon deux propriétés linguistiques : la nouveauté formelle et la régularité des procédés de création lexicale. Nous menons une expérience d'identification des néologismes, dont les résultats montrent que l'intuition varie non seulement en fonction des individus, mais également des deux propriétés linguistiques examinées.

Mots-clés : sentiment néologique, néologisme, polysémie, construction morphologique

### Abstract

Neologisms are sometimes defined with respect to their ability to trigger neological intuition, i.e. the impression of lexical novelty. In this paper, we investigate the correlates of such a definition by exploring the variation of neological intuition according to two linguistic properties: formal novelty and lexical regularity. Through a behavioral experiment with native French speakers, we demonstrate that neological intuition depends not only on individuals, but also on the two linguistic factors we study.

Keywords: neological intuition, neologism, polysemy, word formation

### 1. Introduction

La catégorisation des néologismes comme tels peut s'avérer difficile, y compris pour les linguistes eux-mêmes. Les expériences menées par le passé sur le « sentiment néologique », c'est-à-dire sur l'intuition qu'ont les locuteurs du caractère nouveau de certains mots ou expressions, ont toujours indiqué un accord relativement faible entre annotateurs (Gardin & *al.* 1974, Sablayrolles 2003, Ben Hariz Ouenniche 2009). Identifier les néologismes est une opération de catégorisation subjective,

dépendante des propriétés et des pratiques langagières des locuteurs, qui de ce fait implique une importante variation interindividuelle.

On a pourtant cru pouvoir s'appuyer sur le sentiment néologique pour clarifier les contours d'une catégorie dont l'extension pose problème. Si, comme le souligne Corbin (1987), le recours au sentiment néologique ne permet pas de fonder une théorie du lexique valide, la question de savoir s'il faut ou non inclure dans la définition de la catégorie néologique la capacité de susciter un sentiment de nouveauté linguistique reste posée. D'une part, l'idée que les mots nouveaux ne soient pas perçus comme tels par les locuteurs peut sembler contradictoire avec l'idée même de nouveauté. D'autre part, pouvoir se fonder sur l'intuition linguistique pour circonscrire la néologie permettrait de résoudre certaines difficultés notoires, comme celle de savoir à partir de quand un néologisme employé dans la langue courante ne doit plus être considéré comme une expression nouvelle.

À y regarder de plus près, la variation du sentiment néologique dépend de différents paramètres individuels, dont les caractéristiques sociologiques des locuteurs (âge, catégorie sociale, lieu de vie, etc.). Cependant, on peut faire l'hypothèse qu'il existe aussi des facteurs linguistiques qui déterminent l'identification des expressions nouvelles. On peut penser par exemple que, indépendamment de la subjectivité et des spécificités des individus, les néologismes dans (1), qui relèvent de types néologiques différents, ne seront pas uniformément identifiés comme tels par des locuteurs et locutrices du français :

- (1) a. On n'en dira pas plus pour ne pas **divulgâcher** l'intrigue.
  - b. Ce raccourci n'est d'ailleurs pas innocent tant le réalisateur a participé à l'**iconisation** de la figure du gangster italo-américain au cinéma.
  - c. Une personne **toxique** a le don de se faire passer pour la victime.
- (web)

Dans cet article, nous nous proposons d'étudier la manière dont des locuteurs non experts des questions de linguistique repèrent la nouveauté lexicale. Nous nous interrogeons plus précisément sur les facteurs linguistiques qui déterminent la catégorisation des mots nouveaux comme tels. Notre objectif est d'évaluer le degré de saillance des différents néologismes de manière expérimentale. Plus précisément, nous étudierons la variation du sentiment néologique selon deux propriétés linguistiques : (i) la (non-)nouveauté formelle et (ii) la (non-)régularité des procédés de construction lexicale.

L'article est organisé comme suit. Dans la section 2, nous présenterons la typologie des néologismes sur laquelle notre étude se fonde, ainsi que les deux propriétés linguistiques spécifiquement analysées. Dans la section 3, nous décrirons l'ensemble de notre méthodologie expérimentale. La section 4 présentera et discutera les résultats de l'expérience menée.

## 2. Propriétés des types néologiques

Dans cette partie, après avoir brièvement présenté la typologie sur laquelle se base notre étude, nous analyserons les différents types néologiques selon deux propriétés transcatégorielles, la nouveauté formelle et la régularité lexicale, qui pourraient avoir une influence sur la saillance néologique.

### 2.1. Typologie des néologismes

Il existe de nombreuses typologies des néologismes (Sablayrolles 2000, Cabré Castellví 2006, Gérard & al. 2017, Cartier 2018, Poix 2018, Lavale Ortiz 2019, *inter alia*). Notre objectif ici n'est pas de comparer ni de discuter extensivement ces typologies, mais d'en extraire quelques types élémentaires qui suffiront à définir le cadre de notre étude. Nous distinguerons très schématiquement six types de néologismes, illustrés dans chacun des exemples ci-dessous :

- (2)
  - a. Sur les crédençes, au salon vide : nul **ptyx** (Mallarmé, *Sonnet en X*)
  - b. Certaines célébrités vont à contre-courant, en se faisant **détatouer**.
  - c. Aujourd'hui, 40 entreprises européennes sont des **licornes**.
  - d. **Cliquez la photo** pour l'agrandir.
  - e. Toute l'équipe **est au taquet** et se réjouit de vous retrouver dès ce samedi.
  - f. Monsanto est suspecté d'être un adepte du **ghostwriting**.  
(web)

La néologie *ex nihilo*, telle que l'exemplifie le nom *ptyx* dans (2a), donne lieu à des unités lexicales créées sans base lexicale préexistante. La néologie morphologique produit des nouvelles unités lexicales au moyen d'une opération sur la forme et généralement le sens d'une base lexicale, comme pour *détatouer* à partir de *tatouer* dans (2b). La néologie sémantique consiste en l'assignation d'un nouveau sens à une forme déjà existante (Sablayrolles 2010, Reutenauer 2012, Gérard & Kabatek 2012), comme c'est le cas de *licorne* dans (2c), employé au sens de 'start-up d'une valeur boursière de plus d'un million de dollars'. La néologie syntaxique consiste à créer une nouvelle construction syntaxique pour une unité lexicale dont la forme et le sens sont déjà existants. La transitivation du verbe *cliquer* dans (2d), construit avec un objet direct plutôt qu'un oblique (*cliquer sur la photo*), en est un exemple. La néologie phraséologique donne lieu à des expressions polylexicales nouvelles, issues de l'emploi conventionnalisé d'un syntagme (Baldwin & Kim 2010, Polguère 2015, Constant & al. 2017), comme c'est le cas de l'expression *être au taquet* dans (2e), qui est une locution verbale récemment lexicalisée. Enfin, la néologie par emprunt consiste à importer et à adapter dans une langue cible des unités lexicales ou expressions étrangères, à l'instar du nom anglais *ghostwriting* dans (2f) (Haspelmath & Tadmor 2009, Winter-Froemel 2009, Sablayrolles & al. 2011).

## 2.2. Analyse des types et saillance néologique

Les types de néologismes présentés ci-dessus varient selon différentes propriétés linguistiques, dont (i) la nouveauté formelle et (ii) la régularité lexicale. La première distingue les néologismes selon qu'ils constituent ou non un nouveau signifiant. Par exemple, *ptyx* dans (2a) est une forme complètement nouvelle, alors que *licorne* dans (2c) est une forme déjà existante. La seconde propriété distingue les néologismes d'après la (non-)régularité du procédé qui sous-tend leur création, c'est-à-dire d'après leur (non-)prédictibilité formelle, syntaxique ou sémantique. Par exemple, le verbe *détatouer* est formé en français de manière régulière, car il résulte d'une opération morphosémantique de préfixation dont les spécifications sont les mêmes pour une partie au moins des mots qu'elle produit. À l'inverse, l'expression *être au taquet* est formée de manière irrégulière, car on ne peut pas déterminer son sens par compositionnalité, ni prédire son figement syntaxique.

Notre étude se focalise sur deux cas : les néologismes morphologiques et les néologismes sémantiques. L'analyse de ces deux types de néologismes permet, d'une part, de distinguer clairement entre nouveauté formelle et non formelle et, d'autre part, de polariser la régularité, en vertu du fait que chacun de ces deux types forme des classes distinctes de néologismes irréguliers ou réguliers. Ce second point appelle quelques précisions.

La régularité des néologismes morphologiques peut être définie selon l'opposition entre morphologie grammaticale et extragrammaticale (Dressler 2000, Fradin & al. 2009, Mattiello 2013). Contrairement aux procédés grammaticaux, les procédés extragrammaticaux ne sont pas à proprement parler productifs, car ils forment des lexèmes dont la forme et/ou le sens ne sont pas prédictibles. Ainsi, le sens et la forme du mot *distriation*, construit à partir de *restauration* et de *distribution*, et dénotant une activité commerciale combinant la distribution de biens et l'offre de restauration, ne peuvent être déduits *a priori* des éléments qui composent le néologisme. Ni le segment commun autour duquel s'opère la fusion des lexèmes de base, ni la construction sémantique exacte ne peuvent être déterminés par avance, de sorte que l'opacité du mot construit requiert une désambiguïsation en contexte. La régularité des néologismes sémantiques dépend des propriétés de transfert polysémique permettant de générer des sens nouveaux à partir de sens existants. Selon J. Apresjan (1974), une relation de polysémie est régulière si elle est instanciée par au moins deux lexèmes non synonymes. Il est ainsi possible d'identifier et d'analyser sémantiquement des patrons de polysémie régulière (Briscoe & Copestake 1991, Ostler & Atkins 1991, Pustejovsky 1995, Barque 2008, Srinivasan & Rabagliati 2015). Par exemple, le patron PARTIE DU CORPS > PARTIE D'OBJET permet par métaphore d'étendre le sens de noms comme *tête*, *pied*, *nez*, *bras*, *bouche*, *aile*, *dent*, *bec*, etc. (Svorou 1994, Bertin 2018). À l'inverse, une relation polysémique est irrégulière lorsqu'elle est instanciée par une forme lexicale seulement. Tel est le cas du nom *licorne* dans le sens de 'start-up à forte valorisation boursière', créé par analogie avec l'animal en vertu de son caractère mythique. Les différents sens des polysèmes irréguliers entretiennent une similarité, mais ne sont pas dérivés l'un de

l'autre selon une règle sémantique (Brocher & al. 2018). En l'occurrence, la relation polysémique est imprédictible et idiosyncratique.

On peut supposer que l'intuition des locuteurs sur le caractère néologique des unités lexicales dépend des paramètres (i) et (ii) décrits ci-dessus, ce qui conduit à deux hypothèses. D'une part, une forme nouvelle serait plus saillante néologiquement qu'une forme existante. C'est ce que suggèrent les résultats des études de J.-F. Sablayrolles (2003) et S. Ben Hariz Ouenniche (2009) sur la variation du sentiment néologique. D'autre part, une forme irrégulière serait plus saillante néologiquement qu'une forme régulière. H.-J. Schmid (2008) suggère dans ce sens qu'il existe des différences dans l'implantation cognitive ("entrenchment") des mots construits, selon que ceux-ci sont formés de manière régulière ou non. En extrapolant, nous pouvons supposer que ces disparités existent également pour les néologismes, ce qui se traduit par une variation du sentiment néologique en fonction de la régularité du procédé de création lexicale. Notons que cette hypothèse a été formulée explicitement pour les néologismes sémantiques (Carston & Wearing 2014, Barque & al. 2018), mais n'a jamais été mise à l'épreuve. Pour tester les deux hypothèses évoquées, nous proposons d'effectuer une expérience sur les jugements métalinguistiques des locutrices et locuteurs natifs du français.

### 3. Méthodologie expérimentale

Dans cette section, nous décrivons la méthodologie adoptée dans notre expérience sur le sentiment néologique. Sont successivement présentés le principe de l'expérience, les participant-es, le matériel linguistique utilisé et les hypothèses générales, reformulées en fonction des variables expérimentales.

#### 3.1. Principe de l'expérience

L'expérience envisagée consiste à présenter à des locutrices et locuteurs natifs du français des phrases inventées, contenant ou non des néologismes de différents types. Elle prend la forme d'un sondage d'une durée d'une vingtaine de minutes<sup>1</sup>, se subdivisant en deux tâches successives. La première tâche consiste à indiquer pour chaque phrase, le plus rapidement possible, si elle contient un néologisme ou non<sup>2</sup>, et la seconde à identifier le néologisme en question dans les items jugés positifs. Ces tâches sont complétées en ligne sur l'ordinateur personnel des participant-es, en présence des linguistes menant l'expérience.

Deux éléments sont observés : la (non-)détection des néologismes et le temps de réponse (à la première tâche). Ces éléments nous donnent, respectivement, des informations explicites sur les jugements métalinguistiques des participant-es, et des

---

<sup>1</sup> L'expérience a pris en moyenne 22 minutes (var. 37,324) aux 96 participant-es qui ont effectué le sondage dans son intégralité.

<sup>2</sup> La consigne exacte est la suivante : « Indiquez si la phrase suivante contient un mot nouveau ou un mot existant employé dans un sens nouveau ». Le terme *néologisme* est intentionnellement omis afin d'éviter de possibles malentendus liés à cette notion.

informations implicites sur leur temps de réflexion, leurs hésitations ou leurs difficultés de compréhension des phrases stimuli.

### 3.2. Participant-es

Les personnes ayant participé à l'expérience sont des étudiant-es de bachelor en sciences humaines de l'Université de Fribourg (Suisse). Elles sont de langue maternelle française et appartiennent à la même tranche d'âge (voir tableau 1). Notons que parmi les 96 participant-es, 62 ont pu être rétribué-es en crédits académiques dans le cadre d'un module de formation expérimentale, tandis que les 34 autres étaient bénévoles.

Tableau 1 : Caractéristiques des participant-es

Nombre	96
Âge minimal	18
Âge moyen	21,05
Âge maximal	29

### 3.3. Matériel expérimental

L'expérience met en jeu quatre conditions néologiques (C1-C4) de 17 items chacune et une condition non néologique (C0) de 40 items, instanciées respectivement par des phrases avec et sans néologisme. Nous décrivons dans cette sous-section les conditions de sélection des néologismes, puis les enjeux liés à la création des phrases relevant de chaque condition.

Pour garantir que les participant-es ne soient pas influencé-es dans leurs décisions par l'éventuelle lexicalisation de certains néologismes attestés, nous avons créé de toutes pièces la plupart des néologismes constituant notre matériel expérimental. Nous nous sommes assurés que ceux-ci n'avaient pas d'entrée dans deux dictionnaires de référence (*Trésor de la Langue Française informatisé* et *Petit Robert électronique*<sup>3</sup>), et dans le cas des néologismes morphologiques, pas d'occurrence dans deux volumineux corpus du français (*frWaC* (Baroni & al. 2009) et *frTenTen12* (Jacubiček & al. 2013)<sup>4</sup>). L'existence des néologismes sémantiques testés est plus

---

<sup>3</sup> Le nom *licorne*, qui dans son sens de 'start-up d'une valeur boursière de plus d'un million de dollars' est entré dans le *Petit Robert* en 2018, constitue une exception à cette contrainte. Nous avons néanmoins conservé ce terme dans notre matériel expérimental, en estimant soit que ce nouveau sens serait encore inconnu des participant-es, soit qu'il serait connu mais toujours perçu comme nouveau et identifié comme tel.

<sup>4</sup> Les corpus *frWaC* et *frTenTen12* comptent respectivement de 1,3 et 5,7 millions de mots collectés sur Internet en 2010 et 2012, dans l'ensemble du domaine.fr. Ces corpus présentent l'avantage de regrouper des données linguistiques non nécessairement contraintes par la norme. Le corpus *frWaC* peut être librement consulté à l'adresse suivante :

difficile à évaluer en corpus. Nous avons effectué des sondages aléatoires dans les corpus consultés et dans *Google* pour vérifier que les candidats ciblés n'étaient pas fréquemment employés dans le sens visé dans l'expérience.

Parmi les néologismes retenus, nous avons sélectionné 44 noms et 25 verbes, répartis également entre les différentes conditions expérimentales, de manière à nous assurer que les participant-es ne se concentrent pas par entraînement sur une même catégorie grammaticale. Les néologismes morphologiques irréguliers (C1) sont des amalgames ou des mots-valises, c'est-à-dire des mots formés par troncation de deux bases lexicales et fusion avec segment commun (Fradin 2015). Ils sont choisis ou créés de manière à ce que les bases mobilisées dans leur formation soient plus ou moins apparentes, comme dans *chocolattirance* 'goût pour le chocolat' vs *distriation* 'distribution et restauration'. Les néologismes morphologiques réguliers (C2) comprennent 9 suffixations, 6 préfixations et 2 compositions. Par définition, ils instancient des patrons constructionnels attestés dans le lexique. Par exemple, le nom *aide-pianiste* et le verbe *refoudroyer* sont créés respectivement par composition de *aide* avec un nom d'humain pour dénoter une personne qui en assiste une autre (ex. *aide-soignant*, *aide-ménagère*), et par préfixation en *re-* pour décrire la réitération d'une action (ex. *refaire*, *rediscuter*). La plupart des néologismes sémantiques irréguliers sélectionnés (C3) sont métaphoriques, en vertu d'une nette tendance de la métaphore à être plus irrégulière que la métonymie. Figurent parmi C3 des créations néosémiques comme *chaton* dans le sens de 'mot doux', *gouttière* dans le sens de 'partie du visage située entre la bouche et le nez', et *dénoyauter* dans le sens de 'priver quelqu'un de sa confiance en soi'. Les néologismes sémantiques réguliers (C4), quant à eux, instancient différents patrons de polysémie régulière, dont trois patrons métonymiques s'appliquant à des noms et six patrons métaphoriques s'appliquant à des noms ou des verbes. Par exemple, *vase* 'contenu du vase' et *argile* 'plat en argile' sont respectivement créés selon les patrons CONTENANT > CONTENU (ex. *verre*, *bol*) et MATIÈRE > OBJET FABRIQUÉ DANS LA MATIÈRE (ex. *verre*, *bronze*), tandis que *typhon* 'grande quantité' et *barrir* 'X (non animé) produire un son sonore et aigu' sont créés selon les patrons PHÉNOMÈNE MÉTÉOROLOGIQUE > QUANTITÉ (ex. *avalanche*, *tempête*) et X (ANIMÉ) PRODUIRE UN SON > X (NON ANIMÉ) PRODUIRE UN SON (ex. *rugir*, *ronfler*).

Les néologismes sélectionnés sont insérés dans des phrases telles que celles présentées sous (3) :

- (3) a. Les vendeurs jouent sur la chocolattirance des passants. (C1)
- b. Le filtre rouge permet un assombrissage subtil des photos. (C2)
- c. Cet homme écrit des chatons adressés à sa compagne. (C3)
- d. L'employé est accablé d'un typhon de soucis. (C4)
- e. Le maçon s'est planté une grosse écharde dans le doigt. (C0)

---

[https://www.clarin.si/noske/run.cgi/corp\\_info?corpname=frwac&struct\\_attr\\_stats=1&subcorpora=1](https://www.clarin.si/noske/run.cgi/corp_info?corpname=frwac&struct_attr_stats=1&subcorpora=1)

Les phrases en question répondent à un certain nombre de contraintes. Elles doivent être simples, d'un niveau de langue ordinaire, dénuées de variation diaphasique ou diastratique, et d'une longueur similaire entre conditions néologiques et non néologiques (cf. tableau 2). Les néologismes sélectionnés figurent à différentes places syntaxiques, mais jamais en position saillante (c'est-à-dire en tout début ou en toute fin de phrase). Dans le cas des néologismes sémantiques, une contrainte additionnelle était nécessaire : le contexte de la phrase doit bloquer l'interprétation du lexème cible dans son sens originel. Ainsi les phrases (3c) et (3d) sont-elles formées de manière à ce que *chaton* et *typhon* ne puissent être compris comme dénotant respectivement un animal et un phénomène météorologique. Il s'ensuit que les phrases employées pour tester les néologismes sémantiques sont, dans leur interprétation littérale, nécessairement absurdes, et qu'il peut être difficile de prédire la manière dont le sentiment d'incongruité sémantique qui en résulte sera appréhendé par les participant-es au moment d'effectuer l'expérience.

Tableau 2 : Longueur des phrases selon les conditions expérimentales

	C1 : Morphologique irrégulier	C2 : Morphologique régulier	C3 : Sémantique irrégulier	C4 : Sémantique régulier	C0 : Sans néologisme
Moyenne (caractères)	53,24	52,71	55,24	54,65	54,78
Écart-type (caractères)	6,52	6,49	6,99	7,04	6,69
Moyenne (syllabes)	14,59	14,24	15,06	14,53	14,33
Écart-type (syllabes)	2,24	2,05	1,92	2,15	2,03

Ajoutons que, du fait de la simplicité de leur structure et des éléments lexicaux qu'elles comprennent, il est possible que les phrases construites pour l'expérience rendent saillants les néologismes ciblés, générant ainsi un effet plafond dans leur détection. Pour réduire cet effet, nous avons choisi de déséquilibrer le nombre d'items néologiques et non néologiques, en postulant que les participants tendraient à maintenir un équilibre entre réponses positives et négatives. C'est pourquoi l'expérience comporte 68 phrases contenant un néologisme contre 40 sans néologisme, pour un total de 108 stimuli.

### 3.4. Hypothèses

De manière à adapter nos hypothèses à la méthodologie expérimentale choisie et aux variables observées, nous postulons qu'un sentiment néologique fort se traduit par un taux de détection élevé et un temps de réponse faible. Nos hypothèses



conduisent ainsi à quatre prédictions. D'une part, les taux de détection devraient être plus élevés pour les conditions morphologiques (C1 et C2) que pour les conditions sémantiques (C3 et C4), et plus élevés pour les conditions irrégulières (C1 et C3) que pour les conditions régulières (C2 et C4). D'autre part, les temps de réponse devraient être plus faibles pour les conditions morphologiques (C1 et C2) que pour les conditions sémantiques (C3 et C4), et plus faibles pour les conditions irrégulières (C1 et C3) que pour les conditions régulières (C2 et C4).

#### **4. Résultats et discussion**

Cette section présente et discute les résultats de notre expérience. Nous examinons pour chacune des quatre conditions expérimentales le taux de détection des néologismes (section 4.1.) et le temps de réponse des participant-es (section 4.2.), avant d'analyser plus en détail les résultats obtenus item par item (section 4.3.). Précisons que les données finales ont été analysées au moyen de modèles mixtes, dans l'environnement *R* (R Core Team 2015), à l'aide du package *lme4* (Bates & al. 2014). Les modèles mixtes permettent de produire des régressions linéaires ou logistiques en y ajoutant des paramètres aléatoires tenant compte des variations individuelles. Le calcul de la significativité de chaque effet dans les régressions est réalisé par comparaison du modèle avec et du modèle sans l'effet en question, au moyen d'un test du ratio de vraisemblance.

##### **4.1. Détection des néologismes**

Les données de détection<sup>5</sup> ont été analysées par un modèle de régression logistique mixte avec comme effet fixe l'interaction entre la nouveauté formelle et la régularité lexicale. Comme effets aléatoires, nous avons inclus les ordonnées à l'origine par personne et par item, et les pentes par personne pour les deux prédicteurs. L'effet d'interaction ainsi que les effets simples des deux prédicteurs, calculés à partir d'un modèle sans interaction, sont globalement significatifs (voir tableau 3).

---

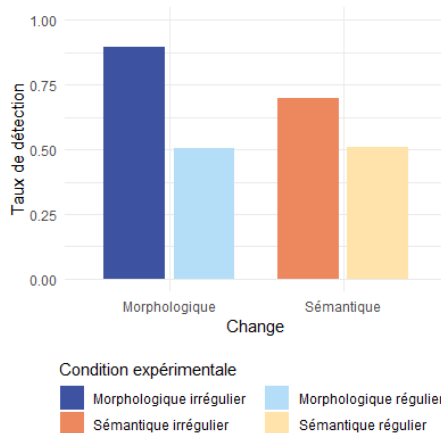
<sup>5</sup> Précisons que certains résultats ont été éliminés de l'analyse. Tel est le cas de l'ensemble des données fournies par une personne, lorsque les réponses de celle-ci aux phrases sans néologisme présentent en moyenne un taux de détection supérieur à la moyenne des réponses fournies par la totalité des participant-es augmentée de 2,5 fois l'écart-type – dans les faits, une participation a été ainsi écartée (1,031% des données totales). De même, les temps de réponse inférieurs à 200 ms ne sont pas pris en compte dans les analyses, cette durée étant clairement insuffisante pour lire les phrases présentées – 0,444% des données conservées après la première étape de filtrage est concerné.

Tableau 3 : Résumé des résultats de la régression logistique avec comme ordonnée à l'origine les néologismes morphologiques irréguliers

	Estimateur	Écart-type	z-valeur	deg. lib.	p-valeur
Ordonnée à l'origine	2,7172	0,2402	11,314		
Irrégulier à régulier	-2,6743	0,2911	-9,188	1	4,938e-13
Morphologique à sémantique	-1,5196	0,3114	-4,880	1	0,002939
Interaction	1,4627	0,3888	3,762	1	0,0003315

Les taux de détection des différents types de néologismes sont présentés dans la figure 1. L'effet de la régularité lexicale semble particulièrement fort, avec une augmentation entre régulier et irrégulier de 38,95% pour les néologismes morphologiques, et de 18,91% pour les néologismes sémantiques. Il apparaît, par ailleurs, que l'effet de la nouveauté formelle dépend des modalités de la régularité lexicale. Les taux de détection entre néologismes sémantiques et morphologiques augmentent de 19,64% pour les conditions irrégulières alors qu'il n'y a presque aucune différence (-0,4%) pour les conditions régulières. Ces effets de changement de modalité ont été analysés plus en détail dans R au moyen du package *lsmeans* (Lenth 2016). Selon ces analyses supplémentaires, l'effet au sein des conditions irrégulières est significatif ( $p < 0,0001$ ) tandis que celui au sein des conditions régulières ne l'est pas ( $p = 0,8484$ ). Les néologismes morphologiques ont significativement plus de chance d'être détectés que les néologismes sémantiques lorsqu'ils sont irréguliers, mais il n'existe aucune différence significative lorsqu'ils sont réguliers.

Figure 1 : Taux de détection des néologismes par condition



Nos hypothèses concernant la détection des néologismes sont donc vérifiées par les données recueillies, sauf pour l'effet de la nouveauté formelle, qui n'est que partiellement en accord avec nos prévisions, puisque son influence sur le sentiment néologique n'est pas présente dans le cas des néologismes réguliers. Au contraire, l'effet de la régularité lexicale est significatif sur chaque modalité du prédicteur lié à la nouveauté formelle, les néologismes irréguliers ayant dans tous les cas plus de chances d'être détectés que les néologismes réguliers. L'effet de cette propriété est, de plus, particulièrement important. Lorsqu'il s'agit d'identifier les nouvelles unités lexicales, l'aspect formel ne semble pas autant frapper l'esprit que l'originalité du procédé de création, que celui-ci soit de nature sémantique ou morphologique.

#### 4.2. Temps de réponse

Les temps de réponse<sup>6</sup> ont été analysés par un modèle de régression linéaire mixte avec comme effets principaux la nouveauté formelle et la régularité lexicale, sans effet d'interaction, ainsi qu'un prédicteur mesurant la longueur des items en nombre de caractères. Comme effets aléatoires, nous avons inclus les ordonnées à l'origine par personne et par item, et les pentes par personne pour la nouveauté formelle. Les effets retenus dans le modèle sont significatifs (voir tableau 4).

Tableau 4 : Résumé des résultats du modèle de régression linéaire avec comme ordonnée à l'origine les néologismes morphologiques irréguliers

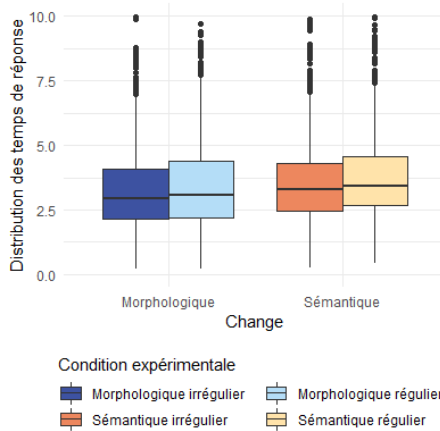
	Estimateur	Écart-type	t-valeur	deg. lib.	p-valeur
Ordonnée à l'origine	0,677412	0,104334	6,493		
Irrégulier à régulier	0,092908	0,02892	3,213	1	0,0008086
Morphologique à sémantique	0,081608	0,02399	3,402	1	0,001446
Longueur (caractères)	0,007804	0,00183	4,264	1	4,158e-05

La figure 2 représente la distribution des temps de réponse selon les quatre conditions expérimentales. Il apparaît globalement que les néologismes morphologiques sont repérés plus rapidement que les néologismes sémantiques, et

<sup>6</sup> Les temps de réponse ont nécessité un filtrage supplémentaire par rapport à la détection. Pour simplifier le modèle, seuls les néologismes détectés sont pris en compte dans l'analyse (65,06% des réponses). Les temps de réponse supérieurs à la moyenne par condition et par personne augmentés de 2,5 fois l'écart-type sont également écartés (2,03% des données restantes), afin de ne pas tenir compte des valeurs extrêmes. Enfin, les données conservées ont été log-transformées de manière à diminuer l'asymétrie de leur distribution.

que les néologismes irréguliers sont repérés plus rapidement que les néologismes réguliers. Après analyse des estimateurs, la nouveauté formelle semble avoir un effet légèrement plus prononcé que la régularité lexicale. Il semble difficile à ce stade de déterminer avec précision pourquoi les participant·es prennent plus de temps pour identifier les néologismes sémantiques que les néologismes morphologiques. Les premiers pourraient notamment entraîner plus de difficultés de compréhension, augmentant ainsi le temps de traitement des énoncés. Il semble plus facile d'imputer la différence de temps de réponse entre néologismes réguliers et irréguliers à de plus grandes hésitations à juger nouveau un lexème lorsqu'il est régulier et donc semblable à d'autres lexèmes déjà existants. Quoiqu'il en soit, nos hypothèses concernant le temps de réponse sont vérifiées par les données recueillies.

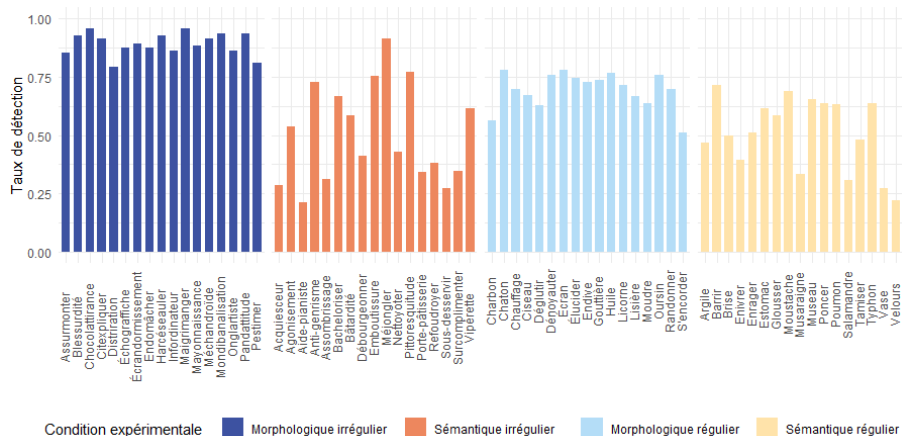
Figure 2 : Distribution des temps de réponse par condition



#### 4.3. Variation par néologisme

Après avoir analysé les différences de saillance néologique globale entre nos quatre conditions expérimentales, nous examinons à présent les différences de (non-)détection entre lexèmes néologiques. Comme on peut le voir dans la figure 3, il existe une différence marquée entre les mots relevant des conditions irrégulières et ceux relevant des conditions régulières. La variation entre items paraît en effet plus importante dans le second cas que dans le premier cas. On peut donc s'interroger sur les facteurs susceptibles d'expliquer cette différence.

Figure 3 : Taux de détection par néologisme



Un premier facteur pourrait tenir au choix des procédés de création néologique. L'homogénéité des procédés irréguliers, à savoir la formation d'amalgames dans le cas des néologismes morphologiques, et la création métaphorique dans le cas des néologismes sémantiques, pourrait être responsable de l'homogénéité des taux de détection. Les procédés de construction sélectionnés pour former les néologismes réguliers sont plus hétérogènes : ils consistent, dans le cas des néologismes morphologiques, en divers types d'affixation et de composition et, dans le cas des néologismes sémantiques, en différents transferts par métaphore ou par métonymie. De cette diversité pourrait découler l'hétérogénéité des taux de détection observés. Pour être étayée, une telle hypothèse devrait néanmoins être mise à l'épreuve. Il conviendrait, dans une expérience future, de varier les modes de création des néologismes irréguliers et d'observer si cette variation entraîne de plus grandes disparités dans les résultats.

Un second facteur explicatif des différences de saillance entre néologismes réguliers (qui n'exclut pas le facteur précédent) pourrait être la différence de productivité entre les procédés employés. En effet, puisque la productivité joue un rôle dans la représentation des mots construits dans le lexique mental (Dal Maso & Giraudou 2018), elle pourrait également affecter le traitement des néologismes morphologiques. De fait, on trouve parmi les néologismes morphologiques les plus détectés dans notre expérience des mots comme *méjongler*, *pittoresquitude*, *emboutissure*, et parmi les moins détectés des mots comme *sous-desservir*, *acquiesceur*, *assombrissage* — or les affixes *mé-*, *-ure* et *-itude* paraissent nettement moins productifs en français contemporain que *sous-*, *-eur* et *-age*. Aucune étude n'a été menée au sujet des néologismes sémantiques, mais une hypothèse semblable pourrait être formulée concernant la corrélation entre productivité des patrons de polysémie et saillance néologique. Par exemple, on observe dans notre expérience

que les acceptions métaphoriques créées sur le patron CAUSER UN ÉTAT CHEZ X (ANIMÉ) > CAUSER UN ÉTAT CHEZ X (NON ANIMÉ) (ex. *Je trouve que la scène finale enivre agréablement le film*) sont moins détectées que celles construites sur le patron X (ANIMÉ) PRODUIRE UN SON > X (NON ANIMÉ) PRODUIRE UN SON (ex. *La tondeuse à gazon barrit dans la montée*) — or on peut penser que le second patron est moins régulièrement instancié en français<sup>7</sup>. Qu'elles soient de nature morphologique ou sémantique, les différences de productivité pourraient déterminer différents degrés de régularité des procédés, induisant des différences de saillance entre les néologismes qu'ils permettent de former. Il serait ici aussi nécessaire de conduire une expérience spécifique permettant de tester cette hypothèse.

## 5. Conclusion

Comme l'indiquent D. Capin et ses collègues (ce numéro), la catégorisation subjective peut s'appliquer à tout type d'entités, ce qui inclut les objets linguistiques. Dans cet article, nous nous sommes intéressés au sentiment néologique de locuteurs natifs du français (non linguistes), en sondant expérimentalement leurs intuitions linguistiques sur le caractère nouveau de certains mots. Identifier les néologismes est une activité éminemment subjective ; cependant, par-delà la variation interindividuelle, il existe des principes universels qui influencent la catégorisation comme expression nouvelle. Quelles que soient les prédispositions des individus à détecter les néologismes, des facteurs stables et partagés déterminent leurs jugements métalinguistiques. Notre étude a ainsi montré que la catégorisation comme « mot nouveau ou doté d'un sens nouveau » était fluctuante, et que cette fluctuation dépendait non seulement des individus, mais également des propriétés linguistiques des items néologiques. Il est apparu notamment que les néologismes morphologiques avaient plus de chances d'être reconnus comme expressions nouvelles que les néologismes sémantiques, et que les néologismes irréguliers avaient plus de chances d'être reconnus comme expressions nouvelles que les néologismes réguliers. Les résultats de l'expérience menée ont notamment fait apparaître l'importance du second facteur, c'est-à-dire de la (non-)régularité des procédés de formation lexicale, en tant que prédicteur des jugements métalinguistiques sur le caractère de nouveauté des mots ou des significations.

Ces résultats permettent d'analyser l'extension variable et la délimitation de la catégorie néologique, selon les différentes options définitionnelles proposées dans la littérature. Si les néologismes sont définis comme des mots qui suscitent un sentiment néologique (Sablayrolles 2009), il faut soit admettre que la catégorie ait des frontières instables, certaines innovations lexicales étant possiblement considérées ou non comme des néologismes, soit inclure la possibilité d'une

---

<sup>7</sup> Il n'existe pas dans la littérature de définition explicite de la productivité polysémique, ni de ressources permettant de déterminer avec certitude le nombre de formes instanciant un patron de polysémie donné en français. Toute expérience visant à évaluer la corrélation entre saillance néologique et productivité polysémique nécessiterait que soient préalablement établis des critères permettant d'estimer la régularité des différents patrons de polysémie.

gradation de la néologie (selon des degrés de néologicit ), certaines innovations lexicales  tant consid r es comme plus fortement n ologiques que d'autres. Si,   l'inverse, les n ologismes sont d finis ind pendamment de leur capacit    susciter un sentiment n ologique (Cartier 2018), il devient plus facile d' tablir une distinction claire entre items n ologiques et non n ologiques et de d fendre une conception plus cat gorique de la classification comme n ologisme. Dans ce second cas, les n ologismes ne sont pas n cessairement saillants en tant qu'expressions nouvelles, la variation du sentiment n ologique t moignant simplement de leur h t rog nit  de construction linguistique, et de variations d'ordre sociolinguistique. Pour le dire autrement, la saillance n ologique ne peut pas constituer un crit re d'identification g n rale de la cat gorie n ologique si l'on souhaite disposer d'une cat gorie m talinguistique claire (vs approximative).

## R f rences bibliographiques

- APRESJAN J. (1974), "Regular Polysemy", *Linguisticae* 42, 5-32.
- BALDWIN T. & KIM S. N. (2010), "Multiword Expressions", *Handbook of Natural Language Processing* sous la dir. de Indurkha N. et Damerau, F. J., CRC Press, Boca Raton, USA, 267-292.
- BARONI M., BERNARDINI S., FERRARESI A. & ZANCHETTA E. (2009), "The wacky wide web: a collection of very large linguistically processed web-crawled corpora", *Language Resources and Evaluation* 43, 209-226.
- BARQUE L. (2008), *Description et formalisation de la polys mie r guli re du fran ais*, Th se de l'Universit  Paris-Diderot.
- BARQUE L., HAAS P. & HUYGHE R. (2018), « Polys mie r guli re et n ologie s mantique : constitution d'une ressource pour l' tude des sens nouveaux », *Neologica* 12, 91-108.
- BATES D., MAECHLER M., BOLKER B. & WALKER S. (2014), "lme4: Linear mixed-effects models using eigen and s4. r package version 1.1-6", <http://CRAN.R-project.org/package=lme4>.
- BEN HARIZ OUENNICHE S. (2009), « Diminuer les fluctuations du sentiment n ologique », *Neologica* 3, 37-51.
- BERTIN T. (2018), *La polys mie des noms de parties du corps humain en fran ais : analyse s mantique de art re, bouche, coeur  paule et pied*, Th se de l'Universit  de Rouen.
- BRISCOE T. & COPESTAKE A. (1991), "Sense extensions as Lexical Rules", *Proceedings of the IJCAE, Workshop on Computational Approaches to Non-Literal Language*, Sydney, Australia, 12-20.
- BROCHER A., KOENIG J.P., MAUNER G. & FORAKER S. (2018), "About sharing and commitment: the retrieval of biased and balanced irregular polysemes", *Language, Cognition and Neuroscience* 33, 443-466.
- CABR  CASTELLV  M. T. (2006), "La clasificaci n de neologismos : una tarea compleja", *Alfa : Revista de Ling stica* 50, 229-250.
- CARSTON R. & WEARING C. (2014), "Metaphor, hyperbole and simile: A pragmatic approach" *Language and Cognition* 3, 283-312.
- CARTIER E. (2018), *Dynamique lexicale des langues :  l ments th oriques, m thodes automatiques, exp rimentations en fran ais contemporain*, M moire d'Habilitation   diriger des Recherches, Universit  Paris 13 - Villetaneuse.

- CONSTANT M., ERYIGIT G., MONTIY J., VAN DER PLASZ L., RAMISCH C., ROSNERZ M. & TODIRASCUK A. (2017), "Multiword expression processing: A survey", *Computational Linguistics* 43, 837-892.
- CORBIN D. (1987), *Morphologie dérivationnelle et structuration du lexique*, Tübingen, Niemeyer.
- DAL MASO S. & GIRAUDO H. (2019), "On the interplay between family and series effects in morphological masked priming", *Morphology* 29, 293-315.
- DRESSLER W. U. (2000). "Extragrammatical vs. marginal morphology". *Marginal and extragrammatical morphology*, sous la dir. de Doleschal U. et Thornton A., München: Lincom Europa.
- FRADIN B., MONTERMINI F. & PLÉNAT M. (2009), "Morphologie grammaticale et extragrammaticale", *Aperçus de morphologie du français*, sous la dir. de Fradin B., Kerleroux F. et Plénat M., Presses universitaires de Vincennes, 21-45.
- FRADIN B. (2015), "21. Blending", *Handbook of Word-Formation. An International Handbook of the Languages of Europe*, Mouton de Gruyter, 386-413.
- GÉRARD C., BRUNEAU L., FALK I., BERNHARD D. & ROSIO A.L. (2017), « Le logoscope: observatoire des innovations lexicales en français contemporain », *La neología en las lenguas románicas: recursos, estrategias y nuevas orientaciones*, <https://hal.archives-ouvertes.fr/hal-01388255>.
- GÉRARD C. & KABATEK J. (2012), "Introduction : la néologie sémantique en questions", *Cahiers de lexicologie* 100, 11-36.
- GARDIN B., LEFÈVRE G., MARCELLES C. & MORTUREUX M.-F. (1974), « À propos du « sentiment néologique » », *Langages* 36, 45-52.
- HASPELMATH M. & TADMOR U. (2009), *Loanwords in the world's language: A Comparative Handbook*, Mouton de Gruyter.
- JAKUBÍČEK M., KILGARRIFF A., KOVÁŘ V., RYCHLÝ P. & SUCHOMEL V. (2013), "The tenten corpus family", *7th International Corpus Linguistics Conference CL6802013*, Lancaster, 125-127, <http://ucrel.lancs.ac.uk/cl2013/>.
- LAVALLE ORTIZ R. M. (2019), "Bases para la fundamentación teórica de la neología y el neologismo: la memoria, la atención y la categorización", *Círculo de Lingüística Aplicada a la Comunicación* 80, 201-226.
- LENTH R.V. (2016), "Least-squares means: The R package lsmeans", *Journal of Statistical Software* 69, 1-33.
- MATTIELLO E. (2013). *Extra-grammatical Morphology in English Abbreviations, Blends, Reduplicatives, and Related Phenomena*, Mouton de Gruyter.
- OSTLER N. & ATKINS B. (1991), "Predictable Meaning Shift: Some Linguistic Properties of Lexical Implication Rules", *Lexical Semantics and Knowledge Representation: First SIGLEX Workshop Proceedings*, sous la dir. de Pustejovsky J. et Bergler S., Springer-Verlag, Berlin.
- POIX C. (2018), "Neology in children's literature: A typology of occasionalisms", *Lexis* 12, <https://journals.openedition.org/lexis/2111>.
- POLGUÈRE A. (2015), « Non-compositionnalité : ce sont toujours les locutions faibles qui trinquent », *Verbum* XXXVII (2), 257-280.
- PUSTEJOVSKY J. (1995), *The Generative Lexicon*, MIT Press, Cambridge.
- R CORE TEAM (2015), "R: A language and environment for statistical computing", <http://www.R-project.org/>.



- REUTENAUER C. (2012), *Vers un traitement automatique de la néosémie : approche textuelle et statistique*, Thèse de l'Université de Lorraine.
- SABLAYROLLES J.-F. (2000), *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes*, Paris, Honoré Champion.
- SABLAYROLLES J.-F. (2003), « Le Sentiment néologique », *L'Innovation lexicale*, sous la dir. de J.-F. Sablayrolles, Paris, Champion, 279-295.
- SABLAYROLLES J.-F. (2010), « Néologisme homonymique, néologisme polysémique et évolution de sens : pour une restriction de la néologie sémantique », *Neologia e neologismos em diferentes perspectivas*, sous la dir. de Alvès I. M., Paulistana, CNPQ, 83-100.
- SABLAYROLLES J.-F., JACQUET-PFAU C. & HUMBLEY J. (2011). « Emprunts, créations "sous influence" et équivalents ». *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité*, Oct 2009, Lisbonne, Portugal, 325-339.
- SCHMID H.-J. (2008), "New words in the mind: Concept-formation and entrenchment of neologisms", *Anglia - Zeitschrift für englische Philologie* 126, 1-36.
- SRINIVASAN M. & RABAGLIATI H. (2015), "How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy", *Lingua* 157, 124-152.
- SVOROU S. (1994), *The Grammar of Space*, John Benjamins.
- WINTER-FROEMEL E. (2009), « Les emprunts linguistiques : enjeux théoriques et perspectives nouvelles », *Neologica* 3, 79-122.