

Université de Montréal

**Analyse comparative de la terminologie  
des médias sociaux :  
contribution des domaines de la communication et de  
l'informatique à la néologie**

par Julien-Claude Charlebois

Département de linguistique et traduction  
Faculté des arts et des sciences

Mémoire présenté  
en vue de l'obtention du grade de Maître en traduction  
option recherche

août 2019

© Julien-Claude Charlebois, 2019

*Ce mémoire intitulé*

**Analyse comparative de la terminologie des médias sociaux :  
contribution des domaines de la communication et de l'informatique à la néologie**

*Présenté par*  
**Julien-Claude Charlebois**

*A été évalué par un jury composé des personnes suivantes*

**Patrick Drouin**  
Président-rapporteur

**Marie-Claude L'Homme**  
Directrice de recherche

**Monique Cormier**  
Membre du jury

## Résumé

L'objectif de cette étude est de repérer des néologismes à partir de corpus de textes français au moyen d'une méthode semi-automatique. Plus précisément, nous extrayons les néologismes de corpus associés à deux domaines différents, mais traitant du même thème, nous examinons leur répartition et nous les classons selon leur type.

L'étude s'appuie sur l'analyse de corpus traitant des médias sociaux. Le premier aborde les médias sociaux du point de vue de la communication, l'autre le fait du point de vue de l'informatique. Ces points de vue ont été privilégiés, car la communication considère ce qui a trait l'utilisation des médias sociaux et l'informatique aborde leur cartographie. La méthode fait appel à l'extracteur de termes TermoStat pour recenser la terminologie des médias sociaux pour chaque point de vue. Ensuite, nous soumettons les 150 termes les plus spécifiques de chaque point de vue à une méthode de validation divisée en trois tests destinés à valider leur statut néologique : des dictionnaires spécialisés, des dictionnaires de langue générale et un outil de visualisation de n-grammes. Finalement, nous étiquetons les néologismes selon la typologie de Dubuc (2002). L'analyse des résultats de la communication et de l'informatique est comparative.

La comparaison des deux corpus révèle les contributions respectives de la communication et de l'informatique à la terminologie des médias sociaux en plus de montrer les termes communs aux deux disciplines. L'étude a également permis de repérer 60 néologismes, dont 28 sont exclusifs au corpus de la communication, 28 exclusifs à celui de l'informatique et 4 communs aux deux corpus. La recherche révèle également que les composés par subordination sont les types de néologismes les plus présents dans nos résultats.

**Mots-clés** : terminologie, néologie, médias sociaux, communication, informatique, extraction semi-automatique de termes, corpus d'exclusion

## Abstract

The objective of this study is to identify the neologisms within corpora of French texts by means of a semi-automatic method. More precisely, we will extract the neologisms from corpora associated to two different areas; however dealing with the same topic, we examine their distribution and we classify them according to their type.

This study is based on an analysis of two corpora within social media. The first one approaches social media from the point of view of communication, and the other approaches it from the point of view of computer science. We prioritize these two points of view being that communication is used as the main source of social media's utilization and that computer science allows us to understand what is involved to allow for social media to be functional. For this method, we use the TermoStat term extractor in order to take census of terminology for each point of view. We then submit 150 of the most specific terms related to each point of view by way of an exclusion corpus from which we divide into three different tests meant to validate their neological status: specialized dictionaries, general language dictionaries, and a visualization tool for n-grams. Lastly, we label the neologisms according to Dubuc's (2002) typology. The analysis of the results obtained for communication and computer science uses a comparative method.

The comparison of the two corpora reveals the respective contributions from communication and computer science with respect to the terminology of social medias, as well it demonstrates common terms found within the two disciplines. This examination also allowed for the identification of 60 neologisms; of which 28 are exclusive to the corpus of communication, another 28 are exclusive to that of computer science, and four were found to be common to both corpora. This research also reveals that subordinate compounds are the most present types of neologisms according to our results.

**Keywords** : terminology, neology, social medias, communication, computer science, semi-automatic term extraction, exclusion corpus

## Table des matières

<b>RÉSUMÉ</b> .....	<b>I</b>
<b>ABSTRACT</b> .....	<b>II</b>
<b>TABLE DES MATIÈRES</b> .....	<b>III</b>
<b>LISTE DES TABLEAUX</b> .....	<b>VII</b>
<b>LISTE DES FIGURES</b> .....	<b>IX</b>
<b>LISTE DES SIGLES</b> .....	<b>X</b>
<b>REMERCIEMENTS</b> .....	<b>XI</b>
<b>INTRODUCTION</b> .....	<b>1</b>
<b>CHAPITRE 1 : NÉOLOGIE ET NOTIONS CONNEXES</b> .....	<b>4</b>
1.1    LA NÉOLOGIE ET LE NÉOLOGISME.....	4
1.1.1 <i>La durée de vie des néologismes</i> .....	5
1.1.2 <i>La néologie en terminologie</i> .....	6
1.2    LA TYPOLOGIE DES NÉOLOGISMES.....	7
1.2.1 <i>La typologie des néologismes de Guilbert (1973)</i> .....	7
1.2.2 <i>La typologie des néologismes de Sablayrolles (2000)</i> .....	9
1.2.3 <i>La typologie des néologismes de Dubuc (2002)</i> .....	13
1.2.4 <i>Comparaison des typologies de Guilbert (1973), Sablayrolles (2000) et Dubuc (2002)</i> .....	17
<b>CHAPITRE 2 : MÉTHODES DE REPÉRAGE ET DE VALIDATION DES NÉOLOGISMES</b> .....	<b>18</b>
2.1    LA COLLECTE MANUELLE.....	18
2.1.1 <i>L'identification de CN</i> .....	18
2.1.2 <i>La validation des néologismes</i> .....	19
2.2    LA COLLECTE ASSISTÉE PAR ORDINATEUR.....	19
2.2.1 <i>L'identification de CN</i> .....	20
2.2.1.1    Le montage de corpus de textes.....	20
2.2.1.2    Le repérage de CN.....	20
2.2.1.2.1    La veille néologique.....	20
2.2.1.2.2    Les extracteurs de termes.....	21
2.2.2 <i>La validation des néologismes</i> .....	22
2.3    ÉTUDES ANTÉRIEURES.....	24
2.3.1 <i>Cabré et Yzaguirre (1995)</i> .....	24

2.3.2 Paquin (2007).....	25
2.3.3 Barbu (2011) .....	26
2.3.4 Falk et al. (2014) .....	26
2.3.5 Cartier (2016).....	27
2.3.6 Lafrance (2016).....	27
2.3 LE CHOIX DES MÉTHODES POUR LA RÉALISATION DE CETTE ÉTUDE .....	28
<b>CHAPITRE 3 : MÉTHODOLOGIE .....</b>	<b>29</b>
3.1 LE DOMAINE SPÉCIALISÉ.....	29
3.1.1 <i>Le choix du domaine des médias sociaux</i> .....	29
3.1.2 <i>Les deux points de vue</i> .....	30
3.1.2.1 <i>Le choix des deux points de vue</i> .....	31
3.1.2.2 <i>Le point de vue de la communication</i> .....	32
3.1.2.3 <i>Le point de vue de l'informatique</i> .....	32
3.2 LE MONTAGE DES CORPUS.....	32
3.2.1 <i>Les thèmes des textes</i> .....	33
3.2.2 <i>La comparabilité des corpus</i> .....	34
3.2.3 <i>La période cible</i> .....	35
3.2.4 <i>Le genre des textes</i> .....	35
3.2.5 <i>Les critères linguistiques</i> .....	37
3.2.6 <i>La taille des corpus et des textes</i> .....	38
3.2.7 <i>Les difficultés rencontrées</i> .....	41
3.3 L'IDENTIFICATION DES CT PAR TERMOSTAT.....	43
3.3.1 <i>Le formatage des textes</i> .....	43
3.3.2 <i>La soumission de COMMU et INFOR à TermoStat</i> .....	44
3.3.3 <i>Les candidats-termes</i> .....	45
3.3.3.1 <i>Les 150 premiers termes</i> .....	47
3.3.3.2 <i>L'élagage</i> .....	48
3.3.3.2.1 <i>Les formes multiples</i> .....	48
3.3.3.2.2 <i>Les erreurs de repérage</i> .....	49
3.3.3.2.3 <i>Les unités non terminologiques</i> .....	49
3.4 LES MÉTHODES DE VALIDATION DES NÉOLOGISMES.....	50
3.4.1 <i>Les dictionnaires de la communication</i> .....	52
3.4.2 <i>Les dictionnaires de l'informatique</i> .....	52
3.4.3 <i>Les dictionnaires des médias sociaux</i> .....	53
3.4.4 <i>L'outil de visualisation de n-grammes</i> .....	53

3.4.5 Les dictionnaires de langue générale .....	54
3.5 LA VALIDATION DES NÉOLOGISMES.....	55
3.6 L'ÉTIQUETAGE DES NÉOLOGISMES.....	56
<b>CHAPITRE 4 : ANALYSE DES RÉSULTATS .....</b>	<b>57</b>
4.1 LA COMPARAISON ENTRE COMMU ET INFOR.....	57
4.1.1 Les termes récurrents dans les deux corpus.....	61
4.1.2 Les termes présents dans un seul des deux corpus .....	63
4.1.3 Les termes concurrents .....	64
4.1.3.1 Les termes concurrents dans le même corpus .....	65
4.1.3.2 Les termes concurrents dans des corpus différents .....	67
4.1.4 La répartition des parties du discours et des syntagmes nominaux des termes de COMMU et INFOR.....	67
4.1.5 La répartition des termes simples et complexes de COMMU et INFOR.....	70
4.1.6 L'influence des codes de recherches sur les résultats.....	71
4.2 LA VALIDATION DES CN PAR LES MÉTHODES DE FILTRAGE .....	72
4.2.1 La soumission des CN aux dictionnaires spécialisés .....	72
4.2.1.1 La répartition des CN et des exclusions de COMMU et INFOR après la soumission aux dictionnaires spécialisés .....	79
4.2.1.2 La répartition des exclusions selon le type de dictionnaire .....	81
4.2.2 La recherche des CN dans l'outil de visualisation de n-grammes.....	82
4.2.3 La soumission des CN aux dictionnaires de langue générale .....	86
4.2.3.1 La soumission des CN au TLFi.....	87
4.2.3.2 La soumission des CN à Usito.....	89
4.3 LES NÉOLOGISMES DES MÉDIAS SOCIAUX.....	90
4.3.1 La répartition des néologismes et des exclusions de COMMU et INFOR.....	92
4.3.2 La répartition des parties du discours et des syntagmes nominaux des néologismes de COMMU et INFOR .....	93
4.3.3 La répartition des termes simples et complexes néologiques de COMMU et INFOR .....	95
4.4 LES PROCÉDÉS NÉOLOGIQUES UTILISÉS POUR LA CRÉATION DES NÉOLOGISMES DE COMMU ET INFOR.....	97
4.4.1 Les néologismes étiquetés.....	97
4.4.1.1 Les néologismes de COMMU étiquetés .....	97
4.4.1.1.1 La répartition des néologismes de COMMU.....	100
4.4.1.1.2 Les néologismes répertoriés dans INFOR.....	102
4.4.1.1.2.1 La répartition des néologismes de INFOR.....	104
4.4.1.2 La comparaison de la répartition des néologismes entre COMMU et INFOR.....	105
4.4.2 La comparaison de la répartition des néologismes entre COMMU et INFOR.....	105
4.4.3 Les procédés de néologisation de Dubuc (2002) absents de COMMU et INFOR.....	107

<b>CONCLUSION .....</b>	<b>110</b>
<b>RÉFÉRENCES.....</b>	<b>112</b>
<b>ANNEXE 1 : LE CORPUS DE LA COMMUNICATION .....</b>	<b>XIII</b>
<b>ANNEXE 2 : LE CORPUS DE L'INFORMATIQUE .....</b>	<b>XXXVI</b>
<b>ANNEXE 3 : LES THÈMES DE RECHERCHE DU DÉPARTEMENT DE COMMUNICATION DE L'UNIVERSITÉ DE MONTRÉAL .....</b>	<b>L</b>
<b>ANNEXE 4 : LES THÈMES DE RECHERCHE DU DÉPARTEMENT D'INFORMATIQUE DE L'UNIVERSITÉ DE MONTRÉAL .....</b>	<b>LII</b>
<b>ANNEXE 5 : LES CT DE COMMU ET INFOR REJETÉS DE L'ANALYSE .....</b>	<b>LIV</b>



## Liste des tableaux

Tableau 1 <i>La typologie des néologismes de Sablayrolles (2000, pp. 212-237)</i> .....	9
Tableau 2 <i>La typologie des néologismes de Dubuc (2002)</i> .....	14
Tableau 3 <i>Les thèmes de recherche attribués aux textes sélectionnés dans COMMU</i> .....	34
Tableau 4 <i>Les thèmes de recherche attribués aux textes sélectionnés dans INFOR</i> .....	34
Tableau 5 <i>La répartition des textes en termes de tailles : COMMU</i> .....	38
Tableau 6 <i>La répartition des textes en termes de tailles: INFOR</i> .....	40
Tableau 7 <i>La liste des 150 termes les plus spécifiques aux médias sociaux de COMMU et INFOR après l'élagage</i> .....	58
Tableau 8 <i>Les termes récurrents dans COMMU et INFOR</i> .....	62
Tableau 9 <i>Les termes quasi synonymiques de COMMU</i> .....	65
Tableau 10 <i>Les termes quasi synonymiques de COMMU et INFOR</i> .....	67
Tableau 11 <i>Les termes de COMMU utilisés dans les codes de recherche</i> .....	71
Tableau 12 <i>Les termes de INFOR utilisés dans les codes de recherche</i> .....	71
Tableau 13 <i>La soumission des CN de COMMU aux dictionnaires spécialisés</i> .....	73
Tableau 14 <i>La soumission des CN de INFOR aux dictionnaires spécialisés</i> .....	76
Tableau 15 <i>Liste des CN de COMMU après l'interrogation dans Google Books Ngram Viewer</i> .....	84
Tableau 16 <i>Liste des CN de INFOR après l'interrogation dans Google Books Ngram Viewer</i> .....	85
Tableau 17 <i>Liste des CN de COMMU à vérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer</i> .....	88
Tableau 18 <i>Liste des CN de INFOR à vérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer</i> .....	88
Tableau 19 <i>Liste des CN de COMMU et INFOR à vérifier dans Usito</i> .....	89
Tableau 20 <i>Les néologismes des médias sociaux : COMMU</i> .....	90
Tableau 21 <i>Les néologismes des médias sociaux : INFOR</i> .....	91
Tableau 22 <i>Les néologismes étiquetés : COMMU</i> .....	98
Tableau 23 <i>Les néologismes étiquetés : INFOR</i> .....	102
Tableau 24 <i>La répartition des procédés néologiques dans COMMU et INFOR</i> .....	106

Tableau 25 *Types de procédés néologiques absents des résultats par catégorie* ..... 108

## Liste des figures

Figure 1 <i>La répartition des genres des textes : COMMU</i> .....	36
Figure 2 <i>La répartition des genres des textes : INFOR</i> .....	37
Figure 3 <i>L'interface de TermoStat</i> .....	45
Figure 4 <i>Capture d'écran des premiers CT de COMMU proposés par TermoStat</i> .....	46
Figure 5 <i>Capture d'écran des premiers CT de INFOR proposés par TermoStat</i> .....	46
Figure 6 <i>Capture d'écran de Google Books Ngram Viewer</i> .....	54
Figure 7 <i>La répartition des parties du discours et des syntagmes nominaux dans COMMU</i> ..	68
Figure 8 <i>La répartition des parties du discours et des syntagmes nominaux dans INFOR</i> .....	68
Figure 9 <i>La répartition des termes simples et complexes : COMMU</i> .....	70
Figure 10 <i>La répartition des termes simples et complexes : INFOR</i> .....	70
Figure 11 <i>La répartition des CN et des exclusions de COMMU après la soumission aux dictionnaires spécialisés</i> .....	80
Figure 12 <i>La répartition des CN et des exclusions de INFOR après la soumission aux dictionnaires spécialisés</i> .....	80
Figure 13 <i>La répartition des exclusions de COMMU selon le dictionnaire</i> .....	81
Figure 14 <i>La répartition des exclusions de INFOR selon le dictionnaire</i> .....	82
Figure 15 <i>Répartition des néologismes et des exclusions : COMMU</i> .....	92
Figure 16 <i>Répartition des néologismes et des exclusions : INFOR</i> .....	93
Figure 17 <i>Répartition des parties du discours et des syntagmes nominaux des néologismes : COMMU</i> .....	94
Figure 18 <i>Répartition des parties du discours et des syntagmes nominaux des néologismes : INFOR</i> .....	94
Figure 19 <i>Répartition des termes simples et complexes des néologismes : COMMU</i> .....	96
Figure 20 <i>Répartition des termes simples et complexes des néologismes : INFOR</i> .....	96
Figure 21 <i>La répartition des types de néologismes : COMMU</i> .....	101
Figure 22 <i>La répartition des types de néologismes : INFOR</i> .....	105

## Liste des sigles

Adj : adjectif

Adv : adverbe

COMMU : Corpus de la communication

CN : candidat-néologisme

CT : candidat-terme

DII : Dictionnaire de l'informatique et d'internet

DIN : Dictionnaire informatique & numérique

DJM : Dictionnaire du journalisme et des médias

INFOR : corpus de l'informatique

L : Lexicom

LMS : Lexique des médias sociaux

N : nom

PDD : partie du discours

PPAdj : participe passé adjectival

Prep : préposition

SN : syntagme nominal

TLFi : Trésor de la langue Française informatisé

V : verbe

VRS : Vocabulaire des réseaux sociaux

## Remerciements

Ce mémoire n'aurait sans doute jamais pu être réalisé sans le soutien extraordinaire de mon réseau.

Merci, Marie-Claude. Sur le plan de la recherche, j'ai eu la chance de travailler avec la meilleure. Sur le plan humain, j'ai rencontré une personne inspirante, passionnée et extrêmement généreuse. Je te dis mille fois merci pour tes conseils, ton soutien, ton engagement dans mon projet et ta bonne humeur impérissable. Je suis fier du travail que nous avons accompli ensemble et ça serait un véritable plaisir de travailler avec toi à nouveau.

Merci, papa et maman. Vous m'avez enseigné à être persévérant et travaillant, ce qui m'a permis de me rendre jusqu'ici. Vous avez toujours laissé toute la place à mes intérêts et vous m'avez donné tous les outils pour entretenir mes passions. Je suis très fier d'être le fils de parents aussi aimants. Merci, ma très chère sœur qui croit toujours en moi. Je t'aime. Tu peux toujours compter sur moi.

Merci, Marc-Olivier. Ce parcours, on l'a fait ensemble et je n'aurais pas voulu le faire sans toi. Tu es le MEILLEUR. Allons célébrer immédiatement, puis passons au prochain chapitre.

Merci, Nadjat. Chaque fois que je penserai à mon passage au deuxième cycle, j'aurai une bouffée d'amour pour toi.

Merci, mes amours de collègues du premier cycle. Andréanne, Béatrice et Patrick, je ne saurais pas vous traduire l'amour que j'ai pour vous. Jessica, si je suis le verbe, t'as vraiment été mon complément.

Merci, Benoit Robichaud. Merci, Pascale Bellemare. Merci, Clémence Bourget. Merci, Patrick Drouin. Vous m'avez tellement facilité la vie.

Merci, Mélanie. Tu ne doutais tellement pas de moi que ç'en était fatiguant... mais ça m'a donné toute la confiance dont j'avais besoin. Merci, Kenny. Tu as littéralement été parfait. Grâce à toi, je ne me suis jamais senti tout seul. Merci, Tessa. J'ai vraiment pu compter sur toi quand j'avais besoin de garder les deux pieds sur terre. Merci, Gabrielle. C'est fou à quel point tu m'inspires. Merci, Chloé. Tes paroles auront guidé mon chemin du début à la fin. Merci, Maude. Quand je te cherche, je te trouve dans mon cœur. Merci, Vanessa. Continuons de viser plus haut ensemble.

Merci, mes tantes et mes oncles. Vos encouragements ont toujours eu un plus gros impact que vous le pensez sur mon ambition et sur ma confiance.

Merci, les Activités culturelles de l'Université de Montréal. Merci, le TUM. Merci, Dominic. Merci, Amélie. Merci, Chloée. La vie étudiante à l'Université de Montréal, sans vous, c'est pas une vie.

Merci, l'Espace Thèsez-Vous et tous ses bénévoles. Merci, Esther.

Merci au Département de linguistique et Traduction de l'Université de Montréal. Je garde un souvenir très spécial de mon passage dans cette maison.

Finalement, merci à tous mes amis et tous mes proches. J'ai la chance d'avoir le meilleur réseau. Chaque encouragement m'a permis de me rendre à la fin. Je vous en suis reconnaissant à jamais.

## Introduction

Au tournant du XXI<sup>e</sup>, les médias sociaux ont pris d'assaut le Web et se sont petit à petit implantés dans les mœurs. Que ce soit YouTube, Facebook, Instagram ou d'autres médias sociaux, ils sont désormais des plateformes incontournables pour la majorité des utilisateurs d'Internet et des téléphones intelligents. Le nombre d'abonnés ne fait que grimper année après année. En 2019, nous faisons face à une nouvelle réalité : les métiers des médias sociaux. Des influenceurs, des entreprises, des gouvernements se sont investis dans ces nouveaux médias pour redéfinir leur image et propager leur message. Les utilisateurs sont maintenant des clients, le contenu une marchandise. Les médias sociaux ne servent plus seulement à connecter des internautes. Ils servent de plateforme de vente pour les entrepreneurs, de catalogue d'employés potentiels pour les entreprises, de boussole électorale pour les gouvernements, de microsociétés pour les sociologues, de place publique pour les activistes, de canal de communication pour les journalistes et les artistes, etc. Le domaine des médias sociaux enchaîne les innovations.

Une des conséquences de l'innovation technologique, peu importe le domaine, est l'arrivée de nouveaux concepts qui doivent être nommés par la langue. L'innovation dans un domaine spécialisé implique irrémédiablement l'innovation *lexicale*, car comment parler de quelque chose qui n'existait pas avant ? Les études en néologie s'intéressent à ces nouvelles unités lexicales. La terminologie côtoie la néologie, parce qu'elle se manifeste abondamment dans les domaines spécialisés, ce qui a particulièrement attiré l'attention des chercheurs et des institutions depuis la fin du XX<sup>e</sup> siècle (Humbley, 2018).

Plusieurs chercheurs se sont penchés sur la néologie en français, dont Guilbert (1973), Sablayrolles (2000) et Dubuc (2002). Ce dernier s'est consacré à l'étude de la terminologie du point de vue terminologique, et comme Humbley (2018), il s'est intéressé aux besoins de la langue de spécialité en matière d'innovations terminologiques.

Dubuc (2002, p. 117-131) met l'accent sur le fait que la néologie se doit de répondre à un *besoin*. Il définit différents types de besoins en les classant en quatre catégories de néologie.

La première est la *néologie stylistique*. Celle-ci a pour but de remplacer des mots dont le sens a subi un glissement péjoratif : par exemple, *aveugle* → *malvoyant*, *gestion* → *management*, *concierge* → *salubriste*.

La deuxième s'appelle la *néologie technologique*. Elle se manifeste lorsqu'il y a innovation ou découverte dans un domaine.

La troisième catégorie est la *néologie sociale*. Elle s'impose lorsqu'il y a eu changement de mœurs dans une société. Par exemple, l'arrivée des femmes sur le marché du travail a nécessité la féminisation des titres professionnels (ex. : *policière*).

La dernière est la *néologie fonctionnelle*. Celle-ci sert à créer des mots plus pratiques. Par exemple, un terme long comme *acier inoxydable* devient *inox*, une périphrase comme *l'épreuve du feu* devient *anti-feu*, les termes plats ou abstraits deviennent concrets comme les *voies d'accès* des autoroutes qui deviennent des *bretelles*, et des termes sont empruntés à d'autres disciplines, comme *cadre* emprunté au domaine militaire par le domaine de la gestion.

Plusieurs études, notamment en diachronie, ont été réalisées dans l'optique de repérer la néologie d'un domaine en particulier : Paquin (2007) s'est intéressée au terrorisme ; Lafrance (2016) à l'informatique. Ces études ont permis de montrer que l'utilisation de certaines méthodes semi-automatiques se révèlent efficaces pour l'identification d'unités potentiellement néologiques. Nous présentons dans ce mémoire les méthodes qui ont aidé ces chercheuses à trouver des néologismes. Ces méthodes ont aussi servi à Cabré et de Yzaguirre (1995), Barbu (2011) et Falk, Bernhard et Gérard (2014), qui, eux, se sont plutôt concentrés sur les néologismes dans la presse.

L'objectif de cette étude est double. Premièrement, nous voulons faire une comparaison de la terminologie des médias sociaux en français en compilant des corpus qui empruntent le point de vue de deux disciplines : la communication et l'informatique. Deuxièmement, nous voulons repérer les néologismes présents dans ces corpus et déterminer quels sont ces types de néologismes.

Pour cette recherche, nous avons compilé deux corpus de textes qui traitent des médias sociaux. Le premier corpus aborde le domaine du point de vue de la communication, alors que le deuxième le fait du point de vue de l'informatique. Nous avons soumis ces corpus à



l'extracteur de termes TermoStat (Drouin, 2003) pour établir les terminologies des médias sociaux de chaque point de vue, que nous avons comparées. Après un nettoyage des résultats, nous avons sélectionné les 150 termes les plus spécifiques de chacun des corpus à l'aide de TermoStat et, à partir de ces termes, les néologismes ont été validés au moyen d'un filtrage. Trois tests ont été utilisés pour valider les néologismes : la soumission à des dictionnaires spécialisés, le repérage dans un outil de visualisation de n-grammes et la soumission à des dictionnaires de langue générale. Finalement, les néologismes ont été étiquetés par types, selon la typologie de Dubuc (2002). Nous avons en dernier lieu analysé leur répartition.

Ce mémoire est divisé en quatre chapitres. Le premier chapitre s'attarde sur des notions importantes en néologie. Le deuxième chapitre décrit les méthodes utilisées par les experts pour repérer des néologismes. Il présente également des études qui s'apparentent à la nôtre. Le troisième chapitre montre la méthodologie qui a été appliquée pour la réalisation de ce projet. Le quatrième chapitre présente divers résultats que notre méthode a permis d'obtenir : la comparaison de la terminologie de deux domaines abordant un même thème, ainsi que les néologismes et les procédés néologiques observés dans chacun des domaines. La conclusion résume les apports principaux de notre étude et évoque quelques pistes pouvant être explorées par la suite.

## Chapitre 1 : Néologie et notions connexes

Ce premier chapitre s'attarde sur les définitions proposées par différents auteurs. Nous présentons les notions reliées à la néologie qui seront utiles pour la réalisation de cette recherche. Nous résumons également différentes typologies qui décrivent les *néologismes*, l'objet d'étude de la néologie.

### 1.1 La néologie et le néologisme

Les termes *néologie* et *néologisme* sont répertoriés dans des encyclopédies et des dictionnaires écrits au début du XVIII<sup>e</sup> siècle (Sablayrolles, 2000). Le terme *néologie* veut littéralement dire « nouveau » (néo) et « parole », « discours » (logos) (Pruvost et Sablayrolles, 2016, p. 3). Deux premières acceptions du terme *néologie* sont : « l'étude des processus de formation des termes nouveaux et des termes ainsi formés » (Office québécois de la langue française, 1985). Le terme *néologie* est polysémique. Il est aussi utilisé pour décrire le « processus de formation de nouvelles unités lexicales » ("néologie," 1972). Les *néologismes* sont ce sur quoi la néologie se concentre. Ce sont des « unités[s] lexicale[s] nouvelle[s] » (Office québécois de la langue française, 2008).

Si les premières traces des termes *néologie* et *néologisme* datent du XVIII<sup>e</sup> siècle, les langues ont toujours eu recours à la néologie (Sablayrolles, 2000). Le « [...] mot nouveau est toujours formé *relativement* à une situation sociohistorique particulière, pour y remplir un rôle communicationnel précis [...] » (Gérard, 2018, p. 9). « C'est pour remédier à [une] carence qu'il faut créer des termes nouveaux, des néologismes » (Dubuc, 2002, p. 115).

Par exemple, durant le dernier siècle, le domaine technique s'est étendu et la recherche scientifique s'est développé grâce à la découverte d'une abondance de nouvelles connaissances. Nous sommes à une époque où la néologie en langues de spécialité bat son plein. Gérard (2018, p. 11) décrit les domaines scientifiques et techniques comme « néologènes », c'est-à-dire qu'ils « favorisent la formation de néologismes ».

La néologie ne se produit pas seulement dans les domaines techniques. Elle se manifeste lorsqu'un nouveau concept qui n'a jamais été nommé apparaît dans une société. Toutefois, ce concept peut très bien avoir existé depuis des centaines d'années dans une autre société et avoir été nommé maintes fois à l'étranger. À l'ère de la mondialisation, nous avons vu disparaître les frontières entre les cultures. Les échanges interculturels entraînent le partage de concepts. Mais comment nommer ces nouveaux concepts étrangers dans la langue d'accueil ? Parfois ces concepts gardent leur signifiant dans la langue emprunteuse, parfois pas. Dans ce dernier cas, il importe de trouver des termes pour désigner ces concepts dans la langue d'accueil.

### 1.1.1 La durée de vie des néologismes

Les néologismes sont des *mots nouveaux*. On peut toutefois se demander à partir de quand un néologisme *n'est plus* un néologisme, ou en d'autres termes, quand perd-il son caractère de nouveauté ? S'il est inclus dans un dictionnaire, est-il toujours un néologisme ?

Les chercheurs ne s'entendent pas sur les réponses à donner à ces questions. Selon Sablayrolles (2000, p. 167), on peut émettre l'hypothèse selon laquelle ce n'est que lors de son premier emploi qu'un mot est considéré néologique. Il ajoute, par ailleurs, que « son sort est incertain ». En effet, la création d'un mot n'implique pas sa réutilisation par la suite. Même des mots qui étaient bien établis dans la langue sont devenus de plus en plus rares dans l'usage, pour finalement disparaître complètement. Gérard (2018, p. 9) énonce que « trois quarts des néologismes produits [...] ne s'affranchissent jamais du texte où ils voient le jour ».

Au contraire, peut-être qu'un néologisme se répandra rapidement dans le discours, parfois parce qu'il en est question dans les médias ou parce qu'il s'agit d'un tout nouveau concept qui s'installe dans les mœurs. Lorsqu'un tel néologisme entre dans l'usage, il est difficile de poser une date d'expiration précise à son statut néologique. Toutefois, « [o]n peut considérer qu'il y a néologisme tant que, pour l'ensemble des usagers concernés par l'utilisation du mot, l'effet d'insolite produit par l'innovation lexicale n'a pas disparu » (Dubuc, 2002, p. 115).

### 1.1.2 La néologie en terminologie

Les différentes instances linguistiques, comme l'Office québécois de la langue française ou l'Académie française, ainsi que les chercheurs en terminologie et en linguistique se penchent sur la néologie. Sablayrolles (2000) présente quatre exemples de domaines de recherche qui explorent la néologie : la morphologie, la lexicologie, l'interlocution et la terminologie. Notre étude se situe dans ce dernier domaine de recherche.

Humbley (2018, p. 33-34) s'intéresse particulièrement à la néologie dans le champ d'études de la terminologie qu'il renomme « néologie terminologique ». Il décrit la néologie terminologique comme un besoin spécifique d'une langue de spécialité qui peut apparaître autant dans le domaine industriel qu'en aménagement linguistique, lors de la mise à jour de dictionnaires et bases terminologiques, ou encore dans l'analyse de l'évolution de la langue dans le temps. La néologie peut se manifester aussi bien par l'arrivée d'une nouvelle unité lexicale dans la langue que par l'extension de sens d'un mot (ou d'un terme) déjà présent dans la langue. Ces deux formes de manifestations portent les noms de *néologie de forme* et la *néologie de sens*. Les néologismes de sens sont plus difficiles à repérer et ne peuvent être identifiés qu'en contextes (dans le texte).

Gérard (2018, p. 10) mentionne que, au cours des dernières années, la majorité des recherches se sont concentrées sur la néologie du point de vue de l'évolution de la langue dans le temps ou sur la néologie observée dans des domaines en particulier. Toutefois, il note que les recherches qui s'intéressent plutôt à la néologie dans un genre discursif particulier, chez un groupe social en particulier ou sur un *thème* en particulier, sont beaucoup moins nombreuses. Pour ces raisons, nous avons décidé de travailler sur un thème en particulier dans cette étude par souci d'originalité.

Selon Humbley (2006, p. 91-93), la néologie pose deux problèmes. Le premier est l'incohérence entre la théorie et la pratique. Celle-ci se manifeste lors du repérage des néologismes. Dans la théorie, si un terme est relevé dans un texte comme un néologisme potentiel par un processus de repérage, il faut ensuite vérifier s'il est absent de la nomenclature d'un dictionnaire (cela peut se faire au moyen d'un corpus d'exclusion). S'il est absent de la nomenclature, il est considéré comme un néologisme (Humbley, 2006). Pourtant, cette pratique

laisse planer une incertitude sur le véritable statut de ce terme, car ce corpus d'exclusion ne pourra jamais être représentatif de l'intégralité de la langue actuelle. Le deuxième problème évoqué par l'auteur est que, pour de nombreux terminologues, l'aspect de nouveauté des termes qu'ils incluent dans leurs terminologies n'a que peu d'importance. En effet, ils recherchent la terminologie d'un domaine à un moment donné, et l'aspect chronologique revêt une importance moindre dans leur sélection.

## 1.2 La typologie des néologismes

Les auteurs utilisent plusieurs typologies pour décrire les néologismes, pour les classer et pour montrer le procédé de *néologisation* qui a mené à la création d'un néologisme. Cette section présente quelques-unes de ces typologies.

### 1.2.1 La typologie des néologismes de Guilbert (1973)

Dans son étude « Théorie du néologisme », Guilbert (1973, p. 19-24) établit une typologie selon laquelle la néologie s'opère de cinq façons dans la langue.

La première est par la *néologie phonologique*. Le néologisme phonologique viendrait perturber le système des voyelles et des consonnes établi dans la langue (dans ce cas-ci, la langue française). L'auteur donne comme exemple de ce type de néologisme les créations onomatopéiques, notamment celles de la bande dessinée (*glop*), le substantif *gaz* (adaptation phonétique du grec *Khaos*), l'abréviation, comme dans *manif* (*manifestation*), la « transposition phonologique d'expressions condensées graphiquement » (Guilbert, 1973, p. 19), comme dans *ZUP* (*zone à urbaniser en priorité*) et l'adaptation d'emprunts, comme dans *gadget*, qui servent de base à une dérivation (*gadgétiser*).

Vient ensuite la *néologie syntaxique*. Les néologismes syntaxiques sont créés à partir d'unités qui existent déjà dans la langue : *base + affixe* (Guilbert, 1973, p. 20). Les néologismes de cette catégorie peuvent avoir subi une forme d'affixation. Celles-ci sont la suffixation (*godilleur*), la préfixation (*pré-inscription*) et la formation parasynthétique (*conurbation*). Ces néologismes peuvent aussi avoir été créés par composition. Ils peuvent être des composés savants (*orthophonie*), mais aussi comporter une partie savante fusionnée à une partie française

(*pico-seconde*). Les premiers sont composés de mots non autonomes, alors que les seconds sont composés d'un mot non autonome et d'un mot autonome. La troisième forme de composition se fait à partir de deux mots autonomes qui peuvent se manifester sous les formes N + N (*pneu-neige*), V + N (*protège-nuque*) ou Adj + N (*jeune-loup*). La dernière forme de néologie syntaxique est celle de la « transposition directe de la séquence syntagmatique » (Guilbert, 1973, p. 21). Ces types de composés ne présentent pas de forme de jonction (*sécurité sociale*, *rectangle blanc*). Cela inclut les locutions verbales (*prendre l'air*), les locutions adverbiales (*coup par coup*) et les locutions prépositionnelles (*en prise directe sur*). La forme non abrégée des sigles entre aussi dans cette catégorie comme dans *zone à urbaniser en priorité (ZUP)*.

La *néologie sémantique* est le troisième type de procédé néologique identifié par l'auteur. Celle-ci concerne l'ajout de sens à un lexème déjà présent dans la langue. Elle se manifeste sous plusieurs formes. D'abord, par « le changement du groupement des sèmes afférents à un lexème » comme dans *la catastrophe* → *en catastrophe* (Guilbert, 1973, p. 22). Ce type de néologie fait parfois appel à des figures de style, comme la synecdoque, la métaphore, la comparaison ou la métonymie. Ensuite, on trouve la « néologie syntagmatique » (*pirate* → *pirate de l'air*). Il y a également les cas où les lexèmes changent de catégorie grammaticale comme l'adjectif *belle* qui devient un nom. Le dernier type de néologie de cette catégorie transfère un terme spécialisé vers la langue courante ou l'inverse. C'est un changement de statut.

La quatrième est l'*emprunt*. Les emprunts sont des mots d'une langue étrangère qui sont intégrés dans une langue d'accueil. Ils peuvent se présenter sous plusieurs formes. La première est sous la forme d'une adaptation graphique ou phonétique (*zoom* → *zoum*) ; la deuxième opère une transformation de la sémantique du mot (*planning* employé dans le sens de *plan*). Cette forme garde le sens de la langue d'origine, mais elle subit une adaptation morphologique dans la langue d'accueil (*realize* → *réaliser*). Finalement, les *xénismes* sont des mots qui font référence à des réalités qui sont étrangères dans la culture d'accueil.

Pour finir, le cinquième type est la *néologie graphique*. Celle-ci est influencée par l'oral. Le discours publicitaire peut réunir plusieurs mots en un seul sans espace (dans un but stylistique), la prononciation peut changer l'orthographe d'un mot et la graphie d'un mot peut changer sa prononciation (*promptitude*).

### 1.2.2 La typologie des néologismes de Sablayrolles (2000)

Sablayrolles (2000) divise les procédés néologiques d'une manière qui diffère de la classification de Dubuc (2002). Il les classe dans cinq grandes catégories : *phonétique/graphie, morphologie, domaine sémantique et figures de rhétorique, domaine syntaxique et emprunt*. Chacune de ces grandes catégories comprend des subdivisions et des sous-sections supplémentaires.

Selon l'auteur, il est difficile de classer certains néologismes à cause de leur double appartenance. En effet, certains néologismes sont formés à partir de plus d'un procédé, par exemple *réaliser (comprendre)*, qui nous vient de l'anglais. C'est à la fois une nouveauté sémantique et un emprunt (Sablayrolles, 2000, p. 208). Le Tableau 1 résume le classement de Sablayrolles (2002, p. 212-237).

Tableau 1 *La typologie des néologismes de Sablayrolles (2000, pp. 212-237)*

1. Phonétique/graphie		
1.1 Créations absolues		
1.1.1 Création <i>ex nihilo</i>	<i>Kodak</i>	Création par la combinaison arbitraire de sons.
1.1.2 Onomatopée	<i>vroum</i>	Reproduction en langue d'un bruit.
1.1.3 Idéophone, phonestème, puissance suggestive des sons	<i>glance</i> <i>gleam</i>	Association d'un signifié à des sons. En anglais, « éclat lumineux » à <i>gl-</i> .
1.2 Plusieurs cas de réduction		
1.2.1 Aphérèse ou troncation antérieure	<i>pitaine (capitaine)</i>	Suppression du début du signifiant.
1.2.2 Syncope, troncation médiane (modification interne)	<i>cap'taine</i>	Suppression d'un élément central du signifiant.
1.2.3 Apocope, troncation extérieure, abréviation	<i>sympa</i>	Suppression de la fin d'un signifiant.
1.3 Deux types d'ajouts		
1.3.1 Réduplication, redoublement	<i>pipi</i>	Répétition d'une partie de la lexie.
1.3.2 Insertion	<i>peinctresse</i>	Ajout à l'intérieur d'une lexie d'un élément supplémentaire.
1.4 Altérations de divers types		

1.4.1 Mauvaise articulation, paronomase, approximation	<i>infractus</i>	Déformation, volontaire ou non, d'une lexie.
1.4.2 Fausse coupe	<i>alcool</i>	Les frontières entre les morphèmes ne sont pas celles qui correspondent à celles qui étaient originelles. Ex. : Les mots d'origine arabe.
1.4.3 Verlan, javanais, loucherbem	<i>verlan</i>	Déformation par inversion, ajout, modification de phonèmes.
1.4.4 Graphie	<i>œuf course</i>	Modification de l'orthographe d'un mot qui n'entraîne pas nécessairement de modification phonétique, mais qui suggère parfois d'autres sens.
<b>2. Morphologie</b>		
<b>2.1 Réduction de forme</b>		
2.1.1 Sigle	<i>SDF</i>	Lexie composée d'initiales de plusieurs lexies.
2.1.2 Acronyme	<i>radar</i>	Lexie qui est un sigle, mais qui est prononcée comme un mot et non lettre par lettre.
2.1.3 Fractomètre, irradiation, représentation	<i>Irangate (de Watergate)</i>	Transfert de sens qui opère du signifiant global d'une lexie vers une partie de celui-ci.
<b>2.2 Affixation</b>		
2.2.1 Préfixation	<i>re-lire</i>	Ajout d'un affixe devant la base.
2.2.2 Dérivation/suffixation	<i>opérationnel</i>	Ajout d'un affixe à la droite de la base.
2.2.3 Dérivation régressive, dérivation inverse	<i>galop</i>	Suppression d'un élément existant dans la lexie pour en créer une nouvelle.
<b>2.3 Composition</b>		
2.3.1 Composé régulier	<i>compte rendu</i> <i>Artésiano-flamand</i> <i>vidéosurveillance</i> <i>gratte-ciel</i>	Rapprochement ou fusion de deux ou plusieurs lexies indépendantes.
2.3.8 Composé savant, recomposé, composé avec des quasimorphèmes, des paléomorphèmes ou des pseudomorphèmes	<i>latinophobe</i>	Composés français avec des formants anciens. Ce ne sont pas des emprunts, parce qu'ils n'existent pas dans la langue ancienne.
2.3.9 Composé par particule, synapsie, syntème	<i>pomme de terre</i>	Plusieurs lexies autonomes jointes par des prépositions.



2.3.10 Composé hybride	<i>télévision</i> (gréco-latin)	Les deux éléments constitutifs n'appartiennent pas à la même langue
2.4 Lexie, autre qu'un composé, combinant plusieurs lexies		
2.4.1 Mot-valise, amalgame, mot portemanteau, mot-centaure, croisement, télescopage, mot-tiroir, mot-gigogne, emboîtement	<i>Paristoric</i>	La combinaison de deux ou plusieurs lexies plus ou moins altérées qui combine les signifiés des diverses lexies présentes.
2.4.2 Association privilégiée, cliché, stéréotype, mots tandems, mots blocs	<i>ignorance crasse</i> <i>attention soutenue</i>	L'attrance de certains mots pour d'autres et la forte probabilité de leur emploi ensemble.
2.4.3 Syntagme en voie de figement	<i>nouvelle vague</i> (cinéma français des années 70)	Fréquence d'emploi élevée, privilégiée par le locuteur.
2.4.4 Changement d'un élément dans une locution figée.	<i>Un Marceau long sachant compter</i>	Allusion avec détournement à des éléments linguistiques existant.
3. Domaine sémantique et figures de rhétorique		
3.1 Extension de sens, appauvrissement de sens	<i>arriver (rive)</i>	La lexie dénomme un ensemble plus large que ce qu'elle dénommait avant.
3.2 Restriction de sens, enrichissement de sens	<i>pondre</i> (avant était synonyme de <i>poser</i> )	La lexie dénomme un sous-ensemble par rapport à ce qu'elle dénommait avant.
3.3 Remotivation, réactivation, réactualisation	<i>intraitable</i> (dans le sens <i>qui ne peut pas être traité</i> et non dans le sens <i>impitoyable</i> ).	Redonner un sens compositionnel à une lexie qui s'était démotivée.
3.4 Étymologie populaire, fausse étymologie	<i>remède de bonne femme</i> ( <i>remède de bonne fame</i> )	Remotivation inconsciente dû à l'ignorance de l'étymologie d'origine.
3.5 Métaphore	<i>taupe</i> (engin de travaux publics qui creuse des tunnels)	Une lexie utilisée pour dénommer un nouveau référent qui présente des similitudes avec l'autre.
3.6 Métonymie	<i>verre</i> ( <i>boisson</i> )	Un rapport de contiguïté entre le signifié d'origine et l'autre (matière-objet, contenant-contenu, etc.)
3.7 Synecdoque	<i>voile</i> (un <i>bateau à voile</i> )	Emploi de la partie pour le tout.
3.8 Antonomase	<i>le divin marquis</i> ( <i>Sade</i> )	Utilisation d'un nom propre au lieu d'un nom commun, ou inversement, pour les qualités qu'il possède à un haut degré
3.9 Litote	<i>Décéder</i> ( <i>mourir</i> )	Expressions amoindries pour désigner des termes tabous.
3.10 Antiphrase	<i>une respectueuse</i> ( <i>une prostituée</i> )	Dire le contraire de ce qu'on pense.

3.11 Oxymore	<i>chaud-froid</i>	Association de deux termes qui normalement s'excluent l'un l'autre.
3.12 Hypocristique : terme caressant	<i>pyrowoman</i>	Création de mots spéciales pour des êtres ou des objets par pulsions affectives.
3.13 Calembour, jeu phonique/graphique	<i>œuf-reusement</i>	Décalage entre plusieurs signifiés possibles. La graphie peut parfois être responsable d'une deuxième lecture.
3.14 Amphibologie	<i>Le grand chantier (travaux ordonnés ou désordre)</i>	Ambiguïté, car deux sens sont possibles.
3.15 Paradoxe	<i>il faut colériser le débat politique</i> (pourtant, le but est d'apaiser la colère)	Proposition inattendue qui va à l'encontre de l'opinion établie.
<b>4. Domaine syntaxique</b>		
4.1 Conversion, transfert de classe, modification des traits inhérents, dérivation impropre, recatégorisation, transcatégorisation (pour certains)	<i>ça m'esclave (V) sévère (Adv)</i>	Changement catégoriel d'une lexie.
4.2 Modification du type de construction	<i>ça craint</i>	Changement de l'environnement syntactico-sémantique.
4.3 Modification des traits de sélection (nature du sujet ou du complément du type : animé vs. Inanimé, comptable vs. Non-comptable, abstrait vs. Concret, etc.)	<i>illuminer des énigmes</i> (Complément d'objet direct abstrait au lieu du concret attendu)	Modification des traits de sélection (nature du sujet ou du complément du type : animé vs. Inanimé, comptable vs. Non-comptable, abstrait vs. Concret, etc.)
<b>5. Emprunt</b>		
5.1 Emprunt aux langues vivantes	<i>tennisplayer</i> (anglais)	Lexie empruntée à une langue étrangère vivante.
5.2 Emprunts aux langues anciennes	<i>mélancolie</i> (latin)	Lexie empruntée à une langue morte.
5.3 Emprunt à une langue véhiculaire (religion, culture...)	<i>amen</i>	Lexie empruntée à une langue véhiculaire.
5.4 Xénisme, emprunt tel quel, alloglotte	<i>Big bang</i>	Lexie avec un aura d'étrangeté Pas encore intégrée.
5.5 Pérégrinisme	<i>paquebot</i> ( <i>packet boat</i> de l'anglais)	Emprunt à une langue étrangère francisé.
5.6 Calque morphologique, emprunt adapté, francisation	<i>gratte-ciel</i>	Francisation d'une lexie étrangère dont on garde la structure ou l'image.
5.7 Calque sémantique, équivalence	<i>réaliser</i> (pour <i>comprendre</i> emprunté à l'anglais <i>realize</i> )	Emprunt d'une acception nouvelle pour une lexie qui existe déjà dans la langue.

5.8 Emprunt à un dialecte, régionalisme	pas d'exemple	Emprunt d'une lexie présente seulement dans une région géographique
5.9 Emprunt à un sociolecte, argot, etc.	<i>loufoque</i>	Emprunt d'une lexie à une classe sociale définie.
5.10 Emprunt à une forme ancienne de la langue	<i>synergie</i>	Lexie considérée comme un archaïsme ou un paléologisme
5.11 Emprunt à un technolecte (par la langue commune)	<i>atomique</i>	Emprunt de la langue commune à la langue de spécialité.
5.12 Emprunt d'un technolecte à un autre	<i>panne</i> (emprunté à la navigation par la mécanique)	Emprunt d'une langue de spécialité à une autre langue de spécialité.
5.13 Emprunt d'un technolecte à la langue commune	<i>souris</i> (emprunté par l'informatique)	Emprunt de la langue de spécialité à la langue commune
5.14 Emprunt avec fausse analyse	<i>cache-flot</i> ( <i>cash flow</i> de l'anglais)	Emprunt par ignorance ou mauvaise interprétation.
5.15 Emprunt avec calembour	<i>œuf-course</i>	Mot-valise mettant en cause une lexie étrangère.
5.16 Emprunt à graphie incertaine	<i>skidoo / skidou</i>	Plusieurs synonymes pour une lexie empruntée.

Sablayrolles (2000, p. 237) mentionne que cette liste de 60 procédés aurait pu être beaucoup plus longue. Le regroupement de divers procédés sous le même (ex. : plusieurs types de composition) rend la liste moins précise et plus courte. Ces procédés s'appliquent à la langue générale, donc on ne sait pas si chacun de ceux-ci s'applique à la langue de spécialité.

### 1.2.3 La typologie des néologismes de Dubuc (2002)

Dubuc (2002, p. 118) affirme qu'il a deux procédés de formation des néologismes en français : *la formation indirecte* et *la formation directe*. D'une part, la formation indirecte se manifeste de trois façons, soit l'*extension sémantique*, le *changement grammatical* et l'*emprunt*. D'autre part, la formation directe est la « création d'une nouvelle lexicale, simple ou complexe (Dubuc, 2002, p. 123) ». Elle se fait au moyen de la composition et de la dérivation. Nous avons réuni les types de néologismes proposés par Dubuc (2002) dans le Tableau 2.

Tableau 2 *La typologie des néologismes de Dubuc (2002)*

1. Formation indirecte		
1.1 L'extension sémantique	Étirement du sens d'un terme, mais en conservant une parenté sémantique.	
1.1.1 Glissement de la relation logique	Déplacement du point de vue par rapport à la notion que représente le terme.	
1.1.1.1 Du concret à l'abstrait/de l'abstrait au concret	<i>tribune</i>	Le lieu physique devient un genre d'émission.
1.1.1.2 De la matière à l'objet	<i>acétate</i>	Une pellicule en acétate devient un <i>acétate</i> .
1.1.1.3 Du signe au signifié	<i>courbe</i>	En statistiques, elle était d'abord la représentation graphique d'une distribution. Elle devient la distribution elle-même.
1.1.1.4 De la partie au tout	<i>tourne-disque</i>	Une partie de l'objet dénomme le tout.
1.1.1.5 De la cause à l'effet	<i>lavage</i>	L'action de laver le minerai devient la quantité de minerai lavé.
1.1.1.6 Du contenant au contenu	<i>verre</i>	Le contenu du verre est appelé par le contenant.
1.1.2 Analogie	Une nouvelle acception est donnée à un mot à cause de sa ressemblance avec une autre réalité.	
1.1.2.1 Analogie de forme	<i>grue</i>	Un nom d'animal qui ressemble à la forme de la machine.
1.1.2.2 Analogie de fonction	<i>diagnostic</i>	En évaluation d'entreprise, il sert à évaluer l'état d'une entreprise. Ce terme vient de la médecine où le médecin établit un <i>diagnostic</i> de maladie après avoir examiné un malade.
1.1.3 Assimilation (Seule de sa catégorie)	Le transfert du signifiant d'un objet à un nouvel objet, ce qui modifie la sémantique du terme. Par exemple <i>ferret</i> , dénomme le morceau de fer au bout d'un lacet, qui a été remplacé par un morceau de plastique au fil du temps, mais qui a gardé le même signifiant.	
1.2 Les changements grammaticaux	Utilisation des mécanismes grammaticaux pour créer des néologismes.	
1.2.1 Changements de catégorie	<i>informatique</i>	Le N devient un Adj.
1.2.2 Changements de genre	<i>dépanneuse</i>	Le terme <i>dépanneur</i> qui désigne un mécanicien devient <i>dépanneuse</i> pour désigner la machine.
1.2.3 Changements de nombre	<i>actifs</i>	Le terme <i>actif</i> désigne l'ensemble des éléments du patrimoine d'une personne, alors que les <i>actifs</i> désignent les éléments individuellement.

1.2.4 Changement de statut	<i>watt</i>	Le nom propre peut devenir un nom commun.
1.3 Les emprunts	Le transfert d'une unité lexicale d'un système ou d'un sous-système linguistique à un autre.	
1.3.1 Emprunts extérieurs	L'emprunt à un système linguistique étranger.	
1.3.1.1 Emprunt intégral	<i>starter</i>	L'unité est transposée dans le système linguistique emprunteur sans modification formelle et désigne une réalité connue des locuteurs de son nouveau système. C'est un emprunt intégral même si le sens de l'unité a été un peu modifié lors du transfert de système.
1.3.1.2 Xénisme	<i>bortsh</i>	Il est comme l'emprunt intégral, à la différence qu'il désigne une réalité qui est étrangère aux locuteurs de la langue emprunteuse.
1.3.1.3 Calque	<i>chien chaud</i>	Une transposition littérale dans la langue emprunteuse de la forme linguistique étrangère, comme ici pour <i>hot dog</i> .
1.3.1.4 Emprunt naturalisé	<i>conteneur</i>	Une modification de la forme linguistique étrangère pour qu'elle s'intègre mieux au système linguistique emprunteur, comme dans le cas de <i>container</i> .
1.3.1.5 Faux emprunt	<i>badwill</i>	Un mot formé de toute pièce à partir d'éléments morphologiques de la langue prêteuse, mais qui n'existe pas tel quel dans cette langue, comme pour l'antonyme de <i>goodwill</i> qui se manifeste en anglais sous la forme de <i>negative goodwill</i> .
1.3.2 Emprunts intérieurs (Seuls de leur catégorie)	Le transfert d'une unité lexicale d'un sous-système à un autre à l'intérieur du même système linguistique. Par exemple, le domaine de l'audiovisuel emprunte <i>platine</i> au domaine de l'horlogerie.	
<b>2. Formation directe</b>		
2.1 La dérivation	L'ajout d'un suffixe à un radical ou la suppression d'un élément terminal d'un mot.	
2.1.1 La dérivation propre ou normale	<i>styliste</i>	L'ajout du suffixe <i>-iste</i> , qui signifie <i>spécialiste</i> , au radical <i>style</i> .
2.1.2 La dérivation régressive ou impropre	<i>conserve</i>	C'est une dérivation à rebours. On a enlevé la finale du V <i>conserver</i> pour créer un N.

2.2 La composition	La création de nouvelles unités lexicales à partir de mots ou d'éléments préexistants.	
2.2.1 La composition par préfixes	L'ajout d'un préfixe autonome ou non autonome à un autre mot pour former un mot nouveau.	
2.2.1.1 Les préfixes autonomes	<i>sous-chef</i>	Les prépositions et les adverbes entrent dans cette catégorie. Ils conservent leur sens usuel ou acquièrent une nouvelle charge.
2.2.1.2 Les préfixes non autonomes	<i>supervente</i>	Ils comprennent généralement les formants latins ou grecs.
2.2.2 Composition par mots à sens plein	Utilisation de mots ou d'éléments de mots qui ont déjà un sens précis.	
2.2.2.1 Composition par coordination	<i>jaune-vert</i>	L'union de deux mots de même nature avec ou sans trait d'union. Ils jouent sur le même plan.
2.2.2.2 Composition par subordination	<i>directeur général</i> <i>fondé au noir</i> <i>lève-tôt</i>	Les formants ne jouent pas sur le même plan. L'un est en dépendance de l'autre. Il peut y avoir un trait d'union ou pas.
2.2.2.3 Composition par racines empruntées aux langues anciennes (grecque ou latine)	<i>vélocipède</i>	La combinaison de mots ou de parties de mots à sens plein de langues anciennes respectant les transpositions orthographiques traditionnelles.
2.2.2.4 Composition par soudure d'éléments français	<i>autoroute</i>  <i>informatique</i> <i>infopub</i>  <i>pdg</i> <i>sida</i> <i>cégep</i>	Ce mode inclut plusieurs procédés. D'abord, l'application du modèle gréco-latin comme dans <i>autoroute</i> , où l'abréviation du mot <i>automobile</i> est soudée au mot <i>route</i> .  Ensuite, dans la contraction des mots-racines, comme dans <i>informatique</i> ou <i>infopub</i> qui prennent respectivement une partie des mots <i>information/automatique</i> et <i>information/publicité</i> . Le résultat est un mot-valise.  Finalement, dans la siglaison et l'acronymisation. Le premier crée une entité lexicale à partir de la première lettre de chaque élément principal du syntagme d'origine comme dans <i>pdg</i> ( <i>président-directeur général</i> ). Le second ressemble au premier, sauf qu'il y a un souci d'alterner les consonnes et les voyelles pour former un résultat prononçable, comme dans <i>sida</i> ( <i>syndrome d'immuno-déficiência acquise</i> ou <i>cégep</i> ( <i>collège d'enseignement général et professionnel</i> )).
2.2.2.5 Composition par onomatopée	<i>bip</i>	Ce mode cherche à reproduire graphiquement une caractéristique sonore de l'objet à désigner. Par exemple un <i>bip</i> comme synonyme de <i>téléavertisseur</i> .

Au total, Dubuc (2002) recense 28 procédés de formation néologique répartis dans cinq catégories principales.

#### **1.2.4 Comparaison des typologies de Guilbert (1973), Sablayrolles (2000) et Dubuc (2002)**

Nous avons présenté trois typologies des néologismes et de leurs procédés de formation. Deux des auteurs choisis (Gilbert et Sablayrolles) se sont intéressés à la néologie telle qu'elle se présente dans la langue générale ; le troisième (Dubuc) s'y est intéressé du point de vue de la terminologie. Ces typologies, qui, pourtant, cherchent toutes à rendre compte des néologismes, s'articulent de façons très différentes. Guilbert (1973) en présente moins de 20 procédés néologiques ; Sablayrolles (2000), quant à lui, en présente plus de 60 ; finalement, Dubuc (2002), 28 . Nous estimons que les typologies moins longues ne sont pas nécessairement moins riches.

Nous reprochons à la typologie de Guilbert (1973) de représenter les néologismes seulement en surface. Dans le cadre de cette recherche, il serait difficile d'étiqueter les néologismes au moyen de sa typologie, car chaque type de néologisme n'est pas décrit individuellement.

D'autre part, Sablayrolles (2000) présente dans sa typologie beaucoup trop de types de néologismes qui appartiennent à la langue orale. Nous rejetons donc sa typologie pour cette étude, car elle correspond moins à la langue de spécialité.

La typologie de Dubuc (2002) est celle que nous utilisons pour cette recherche, car elle a été spécifiquement conçue pour traiter de la néologie en terminologie, contrairement à Sablayrolles (2000). Par ailleurs, sa division des types de néologismes nous semble plus facile à manipuler dans le cadre de cette étude que celle de Guilbert (1973).

## Chapitre 2 : Méthodes de repérage et de validation des néologismes

Ce deuxième chapitre détaille les méthodes que les chercheurs utilisent pour le repérage de néologismes. Peu importe les méthodes, les étapes sont sensiblement les mêmes : l'identification de *candidats-néologismes* (CN) et la validation des néologismes. Nous pouvons diviser la collecte de néologismes en deux méthodes principales : la collecte de néologismes manuelle et la collecte de néologismes assistée par ordinateur.

### 2.1 La collecte manuelle

La première méthode de collecte des néologismes que nous présentons est la *collecte manuelle*. C'est la plus ancienne des méthodes et elle a été utilisée depuis bien avant la création des outils informatiques.

#### 2.1.1 L'identification de CN

Pour ce qui est de l'identification des néologismes, la méthode consiste à faire une lecture de textes et d'y repérer manuellement, donc à l'œil, les mots *susceptibles* d'être des néologismes.

Cette détection se fait en se fiant aux « compétences lexicales » du chercheur (Cabré et de Yzaguirre, 1995, p. 90). C'est un long processus qui exige de la part du chercheur une connaissance approfondie du domaine. Le chercheur doit se fier à son *intuition* lors du repérage. Les néologismes sont caractérisés par une aura de nouveauté ou d'étrangeté (Sablayrolles, 2000). C'est cette aura qui attire l'attention du chercheur vers les mots inconnus. Sablayrolles (2000) émet tout de même une mise en garde, car tous les mots inconnus ne sont pas forcément des néologismes.



### **2.1.2 La validation des néologismes**

Une fois l'identification complétée, le chercheur valide les CN afin de s'assurer qu'il s'agit bien de néologismes. Pour ce faire, il a recours à des outils de référence. Dans le cas de la collecte manuelle, ces outils sont des dictionnaires. Le chercheur parcourt les nomenclatures de ces ouvrages pour voir si les CN repérés s'y trouvent. Ces ouvrages servant à la validation font partie de ce qu'il est convenu d'appeler un corpus d'exclusion. Nous en parlerons davantage à la section 2.2.2. La validation peut aussi se faire auprès d'un expert du domaine étudié.

La collecte manuelle n'est pas nécessairement la plus efficace en termes de volume de CN répertoriés, surtout si elle est faite par un seul individu. « Une collecte individuelle manuelle aboutit à des résultats nécessairement limités, étant donné la faible proportion de productions qu'un individu seul peut effectivement dépouiller » (Sablayrolles, 2002, p. 98).

## **2.2 La collecte assistée par ordinateur**

La deuxième méthode de collecte de néologismes est la collecte assistée par ordinateur. La différence entre cette dernière et la collecte manuelle réside dans le fait que la collecte manuelle ne fait appel à aucun outil informatique, alors que la collecte assistée par ordinateur combine des tâches effectuées par des outils informatiques et des tâches réalisées manuellement par le chercheur. Les outils informatiques jouent un rôle facilitateur dans le repérage des néologismes. Cependant, Biskri, Meunier et Joyal (2004, p. 194) mentionnent que « l'intervention humaine est incontournable dès lors qu'il s'agit de traitement des langues naturelles pour des résultats de qualité ». Ils ajoutent que l'apport de l'aspect humain dans l'analyse des résultats de la machine permet « une analyse qui tien[t] compte de la perspective, de la subjectivité et des connaissances de l'utilisateur » (Biskri et al., 2004, p. 194), ce qui rend la recherche de néologismes plus flexible.

Les étapes de la collecte assistée par ordinateur sont les mêmes que celles de la collecte manuelle, soit l'identification de CN et la validation des néologismes. Dans la méthode assistée par ordinateur, l'identification des CN est réalisée différemment. Elle comporte plus d'étapes.

### **2.2.1 L'identification de CN**

La collecte de néologismes assistée par ordinateur nécessite le montage d'un corpus de textes (cette technique existe aussi dans la collecte manuelle) qui sera soumis à un logiciel pour l'extraction de CN.

#### ***2.2.1.1 Le montage de corpus de textes***

Le corpus de textes est un ensemble de fichiers qui est soumis à un logiciel pour l'identification de CN. Ce corpus peut être composé d'un seul texte, de plusieurs textes d'un seul auteur, ou de plusieurs textes produits par des auteurs différents. Le corpus peut ensuite être divisé en plusieurs sous-corpus qui seront mis en opposition par le chercheur dans l'analyse des résultats (Charaudeau, Maingueneau et Adam, 2002).

#### ***2.2.1.2 Le repérage de CN***

Le repérage des CN se fait à l'aide d'outils informatiques. Plusieurs ont été conçus pour automatiser la tâche du repérage. Nous présentons deux outils. Le premier est un logiciel de veille néologique et le second est un extracteur de termes.

##### ***2.2.1.2.1 La veille néologique***

La veille néologique est une méthode utilisée par les néologues pour l'extraction de CN selon des paramètres donnés. Neoveille est un de ces systèmes de veille néologique spécialisés dans le repérage et de suivi des néologismes (tout comme Logoscope : voir section 2.3.4). Cette plateforme est construite pour travailler sur sept langues différentes.

Le repérage automatique vise deux objectifs : le repérage des formes nouvelles et le repérage des « nouveaux usages/sens en corpus » (Cartier, 2016, p. 106). Ce sont ces deux types de néologismes (de forme et de sens) que nous avons vus précédemment (1.1.2).

Pour repérer les néologismes de forme, le logiciel de veille néologique recherche les mots inconnus. Neoveille est assez efficace pour être capable de repérer les erreurs

typographiques et de distinguer les mots contenant des erreurs des véritables CN. Le repérage des unités lexicales inconnues se fait également au moyen d'un analyseur morphosyntaxique comme TreeTagger (Schmid, 1995).

Pour le repérage des néologismes de sens, Cartier (2016, p. 108-109) mentionne l'utilisation de logiciels capables d'établir un « profil combinatoire » qui décrit dans quel genre de contextes une lexie se retrouve. Le profil combinatoire décrit un sens spécifique. Si les contextes montrent un profil combinatoire différent, ceux-ci peuvent révéler des cas de néologie. Neoveille utilise, lui aussi, la méthode du profil combinatoire (Cartier, 2016).

#### 2.2.1.2.2 *Les extracteurs de termes*

Un autre outil utilisé pour repérer des CN est l'extracteur de termes. TermoStat (Drouin, 2003) est l'un de ces outils qui ont été utilisés à cette fin.

TermoStat est un logiciel « d'acquisition automatique de termes qui exploite une méthode de mise en opposition de corpus spécialisés et non spécialisés en vue de l'identification de termes » (Drouin, 2010). Ce logiciel utilise une « technologie hybride qui combine outils statistiques et linguistiques en vue de l'extraction de candidats termes simples et complexes » (Bourassa et Drouin, 2005, p. 5).

Le logiciel traite le corpus en deux temps. En premier lieu, il utilise le logiciel d'étiquetage TreeTagger (Schmid, 1995) qui détermine à quelle partie du discours appartient chaque occurrence du corpus. À partir de TreeTagger (Schmid, 1995), TermoStat produit les matrices syntaxiques du corpus à l'aide d'un modèle de référence (ex. : *Nom + Préposition + Nom*) (Drouin, 2010). Pour ces matrices trouvées, il calcule la fréquence d'occurrence des candidats-termes à l'intérieur du corpus soumis (Bourassa et Drouin, 2005). En deuxième lieu, il compare les fréquences des unités dans le corpus soumis aux fréquences de ces mêmes unités dans le corpus de référence. En français, le corpus de référence contient 28 500 000 mots et est bâti à partir des articles de journaux spécialisés de 2002 du journal *Le Monde* (Drouin, 2010). Cette comparaison est faite à l'aide du calcul de spécificité de (Lafon, 1980) :

«  $\log P(X=b) = \log (a+b)! + \log (N-(a+b))! + \log (b+d)! + \log (N-(b+d))! - \log N! - \log a! - \log b! - \log (N-(a+b+d))!$  » (Drouin, 2010).

Ce calcul permet d'identifier « dans quelle mesure un candidat-terme (CT) est spécifique au corpus dépouillé » (Bourassa et Drouin, 2005, p. 5). C'est la différence significative entre les deux fréquences qui fait qu'un terme est identifié comme CT par TermoStat (Bourassa et Drouin, 2005). La comparaison de ces fréquences résulte en un poids de spécificité pour chaque terme du corpus soumis. Ce poids est attribué par le logiciel. Plus le poids de spécificité d'un terme est élevé, plus le logiciel évalue que le terme est spécifique à ce domaine spécialisé. Dans les fonctionnalités supplémentaires offertes par le logiciel, il y a la possibilité pour le chercheur de « consulter tous les contextes d'occurrence d'un candidat [...] » (Bourassa et Drouin, 2005, p. 6). Cela permet au chercheur d'évaluer en contexte si le candidat est véritablement un terme du domaine ou voir l'environnement contextuel afin de déterminer si c'est un néologisme de sens.

Toutes ces étapes sont automatisées. Comme les extracteurs de termes se basent sur la spécificité et la fréquence en comparant le corpus soumis au corpus de référence, leur application en néologie peut s'avérer utile, car nous pouvons supposer que les néologismes d'un corpus ont des fréquences atypiques et une spécificité élevée lors de la comparaison avec le corpus de référence.

### **2.2.2 La validation des néologismes**

Cette section s'attarde sur les méthodes utilisées pour la validation des néologismes.

« Il est bien connu que la nouveauté n'existe pas en elle-même, sinon par référence à autre chose ; une recherche sur la néologie impose donc de définir le cadre dans lequel une lexie est nouvelle ainsi que le code avec lequel elle est en relation » (Estopà, 2009, p. 553) . Il s'agit en fait de comparer du neuf avec du vieux (Boulanger, 2010). Le vieux est représenté par tous les autres mots déjà existants qui seraient contenus dans un gigantesque dictionnaire (si un tel dictionnaire existait). C'est donc ici qu'entrent en jeu les corpus d'exclusion qui ne représentent que partiellement ce dictionnaire hypothétique (Boulanger, 2010). Le corpus d'exclusion est

cette « référence » mentionnée par (Estopà, 2009) un peu plus haut et il agit en tant que guide pour les chercheurs en néologie.

Les corpus d'exclusions entrent dans la méthodologie de la recherche en néologie comme méthode de filtrage en 1979 grâce aux travaux de Boulanger. L'auteur évalue les ouvrages qui formeront le corpus d'exclusion d'une étude à l'aide de six critères (Boulanger, 1979, p. 43) :

- a. La nouveauté de l'ouvrage ;
- b. la qualité du contenu de l'ouvrage ;
- c. la diffusion de l'ouvrage ;
- d. la disponibilité de l'ouvrage ;
- e. l'adéquation du contenu avec l'état de la langue décrite ;
- f. la réputation de l'auteur ou de l'organisme auteur.

Ces critères de sélection facilitent l'évaluation de l'intérêt que présente un ouvrage dans l'optique de la constitution d'un corpus d'exclusion.

Le corpus d'exclusion peut être composé d'ouvrages de référence lexicographiques ou terminologiques. Ce sont des lexiques ou des dictionnaires numériques ou papiers (Sablayrolles, 2002). Sablayrolles (2002) recommande de ne pas limiter un corpus d'exclusion à un seul dictionnaire, car les ouvrages sont tous imparfaits et présentent invariablement des lacunes. Il le montre avec l'exemple de *bug* qui se trouve dans le dictionnaire *Le Petit Robert*, mais qui est absent dans le *Lexis* (Sablayrolles, 2002, p. 101). Il est donc important de bien choisir les ouvrages qui feront partie du corpus d'exclusion, car ce sont ceux-ci qui guideront la validation des néologismes. Il est préférable que les ouvrages qui constituent le corpus d'exclusion soient contemporains ou mis à jour.

Une fois le corpus d'exclusion établi, les CN sont recherchés dans celui-ci. Si le corpus d'exclusion comporte quatre ouvrages, le CN doit être absent des quatre ouvrages pour être considéré comme un néologisme.

Le chercheur doit établir ce qu'il entend par *néologisme*. Il pourrait décider de ne pas considérer un terme comme néologisme s'il est présent dans la nomenclature d'un des ouvrages du corpus d'exclusion (Estopà, 2009). Toutefois, cela dépend de sa définition du caractère néologique d'un terme. Si un terme vient d'entrer dans une nomenclature, s'agit-il toujours d'un néologisme ? Pendant combien de temps un terme garde-t-il son caractère néologique ? Est-ce que c'est un néologisme seulement lors de sa première apparition dans l'hapax, ou bien garde-t-il son statut pendant cinq ans ? Dix ans ? (Sablayrolles, 2006). Nous pouvons nous entendre sur le fait que « [l]'étiquette de néologisme est accrochée au mot qui a résisté à l'épreuve du corpus d'exclusion (Boulangier, 2010, p. 41) ».

Certains logiciels qui servent à l'identification de CN sont dotés d'un corpus d'exclusion interne, ce qui permet d'automatiser cette étape. Dans le cas de Neoville, cette étape est indissociable à l'étape d'identification, car le logiciel identifie les CN au moyen d'un corpus textuel de référence (Cartier, 2016).

Finalement, la validation des néologismes peut aussi se faire en faisant appel à un expert du domaine étudié. Les experts sont au courant des nouveaux concepts de leurs domaines et peuvent éclairer le chercheur dans sa validation des néologismes grâce à leur expertise.

## **2.3 Études antérieures**

Cette section présente des études qui font appel aux méthodes que nous avons décrites jusqu'ici dans des contextes qui peuvent se comparer au contexte de notre étude.

### **2.3.1 Cabré et Yzaguirre (1995)**

Dans le cadre d'une étude pour l'Observatori de Neologia, Cabré et de Yzaguirre (1995) ont cherché des néologismes dans la presse. Ils ont privilégié ce genre de textes, car ils estiment que le passage dans des textes généraux fait en sorte que ces néologismes sont plus établis.

En ce qui concerne l'identification de néologismes, les chercheurs ont utilisé la collecte manuelle en se fiant aux compétences de leurs collaborateurs. Pour valider ces néologismes, ils

ont eu recours à la méthode du corpus d'exclusion. En fait, leur corpus d'exclusion était divisé en deux parties distinctes, car leur recherche en néologie portait sur deux langues (le catalan et le castillan). Pour le catalan, ils ont utilisé trois éditions de la *Enciclopèdia catalana* (1982, 1990 et 1993) ainsi qu'un dictionnaire de langue générale intitulé *Diccionari general de la llengua catalana* (1932). Pour le castillan, ils ont sélectionné trois dictionnaires, dont deux éditions de la *Real Academia Española* (1989 et 1992) et une édition de *VOX-Biblograf* (1987). Cette étude traitait de néologismes de presse, ce qui explique l'utilisation exclusive de dictionnaires de langue générale. Il est intéressant de constater qu'ils ont opté pour plusieurs éditions du même dictionnaire, car les mots qu'ils contiennent ne font pas que s'ajouter : certains sont ajoutés, d'autres sont retirés.

Après la création de fiches néologiques, les néologismes étaient discutés en comité pour valider leur statut. Dans les conclusions de leur étude, ils déplorent que le projet soit onéreux et qu'il nécessite beaucoup de ressources humaines, et ils prévoient se tourner vers la collecte assistée par ordinateur.

### **2.3.2 Paquin (2007)**

Dans sa recherche, Paquin (2007) a utilisé une méthode lexicométrique lorsqu'elle a comparé deux corpus du terrorisme. La chercheuse s'est penchée sur l'évolution de la terminologie du terrorisme en diachronie. Les deux corpus comparables se focalisaient sur des périodes d'approximativement cinq ans, avant et après le 11 septembre 2001. Elle a utilisé le logiciel TermoStat pour sonder ces corpus. L'étude emprunte aussi une approche comparative. Grâce à l'extracteur de termes, elle a réalisé plusieurs expérimentations qui lui permettaient de comparer des tranches temporelles de ces corpus de textes.

Dans son étude, elle a montré une partie de la néologie qui s'est opérée dans la terminologie du terrorisme après les attentats sur le World Trade Center à New York. En effet, elle a repéré des termes qui ne seraient apparus qu'après les attentats du 11 septembre 2001 : *jihad défensiste, arme de terreur de masse, hyperterrorisme, etc.*

### 2.3.3 Barbu (2011)

Encore une fois dans le domaine du journalisme, Barbu (2011) décrit le projet NEOROM (qui s'intéresse aux néologismes dans les langues romanes). Cette étude cherchait des néologismes de presse dans la langue roumaine. Le corpus est bâti à partir d'articles de journaux en ligne et compte environ 14 millions de mots.

Pour l'identification des CN, le système utilise le concordancier LUCON. Après cette identification, les CN sont automatiquement recherchés dans le corpus d'exclusion.

Le corpus d'exclusion de Barbu (2011) est composé de la longue liste de dictionnaires électroniques roumains de la plateforme dexonline<sup>1</sup> (dexonline, 2019). L'utilisation de ces dictionnaires comme corpus d'exclusion a été privilégiée dans cette étude, car puisqu'ils sont accessibles électroniquement, la tâche est en partie automatisée. Barbu (2011) considère qu'un mot absent de ces dictionnaires est un néologisme. Ces mots sont inclus sous forme de fiche dans la base NEOROM.

### 2.3.4 Falk et al. (2014)

Falk et al. (2014) utilisent la méthode de la collecte assistée par ordinateur.

Les auteurs se sont intéressés à un corpus de textes journalistiques en français. Ils ont monté un corpus à partir des fils de presse RSS français. Après avoir appliqué un étiqueteur morphosyntaxique, ils ont extrait les unités inconnues au moyen du logiciel Logoscope<sup>2</sup> (lilpa, 2018). Ils ont ensuite soumis les unités inconnues extraites à leur corpus d'exclusion après avoir répertorié 692 formes inconnues dans un corpus de 51 000 mots.

---

<sup>1</sup> La plateforme dexonline est repérable à l'adresse <https://dexonline.ro>. Nous ne savons pas exactement combien de dictionnaires contenait cette liste en 2011, car ce n'est pas détaillé dans l'étude de Barbu (2011). En revanche, l'accès est toujours possible en ligne, et en 2019 la base compte 67 dictionnaires (ou lexiques) électroniques.

<sup>2</sup> Le logiciel Logoscope est un outil de veille néologique repérable à l'adresse <http://logoscope.unistra.fr>.



Le corpus d'exclusion utilisé par les auteurs est basé principalement sur le lexique en ligne Morphalou<sup>3</sup> (Centre national de ressources textuelles et lexicales, 2012)

Après avoir repéré ces CN grâce au corpus d'exclusion, ils ont analysé les CN manuellement pour en extraire les véritables néologismes.

Les auteurs concluent de leur analyse que leur méthode pour gérer les formes inconnues s'est révélée efficace. Bien qu'ils aient rejeté plusieurs de ces formes, ils ont pu valider 81 néologismes.

### **2.3.5 Cartier (2016)**

Cartier (2016) présente la plateforme conçue par l'Université Paris 13 : Neoveille.

Le corpus soumis à Neoveille est en fait un fil RSS sélectionné par le chercheur. Le logiciel peut parcourir des fils de journaux de langue générale ou de langue de spécialité. Pour l'identification de CN, Neoveille segmente les mots, puis fait une analyse morphosyntaxique et syntaxique. Pour repérer les néologismes de forme, le logiciel de veille néologique recherche les mots inconnus et les identifie au moyen d'un corpus lexicographique de référence (Cartier, 2016, p. 106) qui agit à titre de corpus d'exclusion. Quant au repérage des néologismes de sens, Cartier (2016, p. 108-109) mentionne l'utilisation d'outils capables d'établir un « profil combinatoire ».

Neoveille a la particularité d'automatiser un maximum d'étapes. La superposition de plusieurs filtres fait en sorte que les résultats nécessitent moins d'intervention humaine. La plateforme est toujours en développement.

### **2.3.6 Lafrance (2016)**

Lafrance (2016) a elle aussi travaillé en diachronie avec un corpus en anglais de l'informatique dans le but de faire ressortir la néologie et la nécrologie (termes en désuétude)

---

<sup>3</sup> L'outil Morphalou est un lexique en ligne repérable à l'adresse <https://www.cnrtl.fr/lexiques/morphalou/>.

en comparant les deux corpus qu'elle a montés sur deux périodes. Son premier corpus couvrait des textes publiés entre 2001 et 2005 ; le deuxième réunissait des textes publiés entre 2006 et 2010. Tout comme Paquin (2007), la chercheuse a utilisé l'extracteur de termes TermoStat pour le repérage des unités terminologiques. Dans son analyse, elle a observé la fréquence, le rang et la spécificité des termes.

Ses résultats ont montré que sa méthodologie permettait effectivement de recenser la nouveauté et l'obsolescence dans la terminologie d'un corpus spécialisé.

### **2.3 Le choix des méthodes pour la réalisation de cette étude**

Dans cette section, nous vous présentons les différentes méthodes que nous avons choisies pour la réalisation de notre étude.

Nous avons choisi d'utiliser la méthode de la collecte assistée par ordinateur. Notre choix s'est arrêté sur celle-ci, car elle nous semble plus efficace que la collecte manuelle. Cependant, nous ne pourrions pas automatiser toutes les étapes de notre recherche, faute de ressources. Certaines étapes seront donc manuelles et d'autres assistées par ordinateur.

Pour l'identification de candidats-néologismes, nous utiliserons l'extracteur de termes TermoStat. Cela implique que nous devons monter un corpus de textes électroniques que nous soumettrons au logiciel. Cette étape de l'identification des CN sera donc en partie automatisée.

Pour la validation des néologismes, nous utiliserons un corpus d'exclusion que nous devons monter spécialement pour le thème que nous abordons. Nous comptons inclure des ressources électroniques et des ressources papier.

## **Chapitre 3 : Méthodologie**

Pour réaliser cette étude, nous appliquons la méthode de la collecte des néologismes assistée par ordinateur.

Nous utilisons le logiciel TermoStat (Drouin, 2003) que nous avons présenté à la section 2.2.1.2.2 pour l'identification des candidats-néologismes (CN) et des outils de référence pour la validation des néologismes. Nous expliquons l'application de ces notions, de ces méthodes et de ce logiciel dans ce troisième chapitre. Nous avons également exploité des méthodes qui sont propres à la réalisation de ce projet en particulier. En ce qui concerne l'étiquetage des néologismes, nous utilisons la typologie des néologismes de Dubuc (2002) présentée à la section 1.2.2.

Dans cet ordre, nous décrivons le domaine spécialisé choisi et la manière dont nous le délimitons (3.1), le montage des corpus de textes (3.2), l'identification des CN (3.3), le montage du corpus d'exclusion (3.4), la validation des néologismes (3.5), puis la validation des néologismes (3.6).

### **3.1 Le domaine spécialisé**

Dans cette section sont décrites les raisons qui nous ont mené à choisir les médias sociaux comme domaine spécialisé, notre définition de ce domaine ainsi que ce qui nous a fait choisir la communication et l'informatique comme différents points de vue pour aborder les médias sociaux.

#### **3.1.1 Le choix du domaine des médias sociaux**

La néologie est prolifique dans les domaines où il y a une abondance de nouveaux concepts. Il est donc sage de sélectionner un domaine en expansion plutôt qu'un domaine bien établi. Il faut également écarter les domaines dans lesquels il n'y a eu aucun changement majeur ou grande découverte dans les dernières années qui pourraient avoir modifié le paysage

terminologique de ceux-ci. Ce domaine doit donc être actuel pour que nous puissions avoir accès à un nombre suffisant de documents numériques qui traitent du sujet.

Les domaines des nouvelles technologies de l'information semblent respecter ces critères, puisque celles-ci sont omniprésentes au XXI<sup>e</sup> siècle et ne cessent d'évoluer.

Nous avons choisi le domaine des médias sociaux. Ce domaine nous semble être un bon candidat pour cette étude, car il est très récent. Maintenant que les médias sociaux sont intégrés dans la vie quotidienne, de nouveaux métiers ont vu le jour et forcément, les nouveaux experts ont écrit sur de nouvelles réalités. Les textes écrits par ces experts sont susceptibles de faire appel aux termes clés du domaine. Mieux encore, comme ce sont des médias accessibles à partir du Web, il y a de fortes chances qu'il y ait une abondance de textes sur le sujet en format numérique, ce qui est indispensable dans une méthode de collecte assistée par ordinateur.

Toutes ces raisons font en sorte que nous pensons que le domaine des médias sociaux est, comme le dit Gérard (2018, p. 11), « néologène », et qu'il constitue un bon choix de domaine pour cette étude.

### **3.1.2 Les deux points de vue**

Cette section présente le processus du choix des disciplines de la communication et de l'informatique qui incarnent les points de vue à partir desquels nous analyserons la néologie des médias sociaux.

Une partie de cette étude consiste à observer la terminologie d'un domaine de spécialité en créant deux corpus de textes issus de deux disciplines distinctes qui traitent des médias sociaux. (Pour éviter toute confusion, nous parlerons de *domaine* lorsque nous parlons des médias sociaux et de *discipline* pour parler des deux points de vue par lesquels nous analyserons le domaine.)

Bien sûr, cet exercice peut être fait sans problèmes à partir de plusieurs disciplines ; toutefois, il convient de choisir deux disciplines qui soient en mesure de fournir assez de textes récents et qui se concentrent sur les médias sociaux.

### ***3.1.2.1 Le choix des deux points de vue***

Nous avons décidé de nous concentrer sur des disciplines universitaires, car cela nous permet non seulement de bien délimiter les points de vue, mais aussi d'avoir accès au genre de texte scientifique qui est susceptible d'être très riche en terminologie.

En ce qui concerne le choix de ces deux disciplines, il est intéressant de prendre des disciplines assez éloignées pour faire ressortir des néologismes concurrents ou propres à chacune d'elles. À l'opposé, il est possible que certains termes soient présents dans les corpus de ces deux disciplines. Il va de soi qu'il faut choisir des disciplines qui ont un lien direct avec le domaine des médias sociaux pour réussir à monter un corpus suffisamment volumineux pour faire notre analyse.

Les disciplines de la communication et de l'informatique abordent les médias sociaux. Ces deux disciplines, pourtant très éloignées (la communication relève des sciences sociales et l'informatique des sciences exactes), peuvent être représentées comme le *contenu* et le *contenant* des médias sociaux. La communication peut être considérée comme le *contenu*, car l'utilité première des médias sociaux est de permettre aux utilisateurs de *communiquer* entre eux. Les experts en communication utilisent les médias sociaux comme outils de travail, mais cherchent aussi à en définir les usages et à trouver de nouvelles techniques d'exploitation de ces plateformes. Pour sa part, l'informatique peut être envisagée comme le *contenant*, car elle représente le support de ce type de médias et elle traite de leur fonctionnement en tant que tel. Ce sont les rouages de l'informatique qui permettent aux utilisateurs d'avoir accès aux médias sociaux. Sans un support numérique, les médias sociaux n'existent pas. L'intérêt de cette distinction *contenu/contenant* pour notre étude réside dans le fait que l'informaticien et l'expert en communication qui travaillent sur les médias sociaux utilisent ces outils de manières très différentes. Cette distance entre ces deux disciplines nous laisse croire qu'elles peuvent nous fournir deux terminologies des médias sociaux distinctes, et potentiellement, des néologismes qui leur sont exclusifs.

### ***3.1.2.2 Le point de vue de la communication***

La communication nous semble être une discipline incontournable pour notre analyse de la terminologie des médias sociaux. Dans les dernières années, des cours complets sont consacrés à ces nouveaux réseaux dans les programmes de communication de certaines universités. Par exemple, l'Université de Montréal offre le cours *Communautés virtuelles, réseaux sociaux* à ses étudiants du programme de science de la communication.

### ***3.1.2.3 Le point de vue de l'informatique***

L'informatique est aussi une discipline importante au XXI<sup>e</sup> siècle et longue est la liste de revues, articles, thèses, mémoires, manuels et ressources sur le sujet. L'abondance de thèmes différents abordés par l'informatique pourrait permettre de relever des termes différents dans des contextes différents.

## **3.2 Le montage des corpus**

Pour le montage des corpus de la communication (COMMU) et de l'informatique (INFOR), nous nous sommes inspiré de la méthode de Dury (2004). Nous avons apporté quelques modifications à sa méthode, puisque, dans sa recherche, l'autrice analysait la terminologie en diachronie alors que, de notre côté, nous comparons deux corpus de deux disciplines. Nous nous sommes également inspiré de la méthode de Lafrance (2016) et de celle de Paquin (2007). Même si elles travaillent sur des corpus en diachronie et, nous, sur des corpus de disciplines différentes, elles ont monté leurs corpus dans le même but que nous : repérer des néologismes.

Nous commençons la description du montage de notre corpus avec la sélection des thèmes des textes (3.2.1), puis, dans l'ordre, nous abordons la comparabilité entre le corpus de la communication et le corpus de l'informatique (3.2.2), la période cible de la publication des textes sur laquelle nous nous sommes arrêté (3.2.3), le genre des textes sélectionnés (3.2.4), les critères linguistiques qui ont guidé notre choix de textes (3.2.5), la taille des corpus et des textes (3.2.6) et, finalement, les difficultés rencontrées lors du montage des corpus (3.2.7).

Pour chaque texte qui fait partie de nos corpus, nous avons rempli une fiche signalétique qui sert d'outil de référence, mais aussi de descripteur du contenu des textes. Ces fiches signalétiques sont inspirées de la méthode de Marshman (2003). Les fiches signalétiques de COMMU sont données à l'Annexe 1 et celle de INFOR dans l'Annexe 2.

### 3.2.1 Les thèmes des textes

En abordant les médias sociaux selon les points de vue des disciplines de la communication et de l'informatique, nous devons nous assurer que les textes que nous choisissons traitent non seulement du domaine des médias sociaux, mais également des thèmes de ces deux disciplines. Pour nous en assurer, nous nous sommes servi des thèmes de recherche abordés par ces disciplines. Nous avons consulté les onglets *communication* et *informatique* dans le guide des disciplines de l'Université de Montréal (Département d'informatique et de recherche opérationnelle, 2019; Département de communication, 2019). Les thèmes de recherche de l'Université de Montréal ont été réunis sous la forme de tableaux. La liste complète des thèmes de recherche de la communication a été placée dans l'Annexe 3, tandis que celle de l'informatique a été placée dans l'Annexe 4.

Grâce à cette méthode, non seulement nous avons pu délimiter les thèmes que nous considérons *recevables* pour le montage de nos corpus, mais cela nous a également donné des pistes de recherche que nous n'avions pas envisagées. Un exemple de piste est celui de la *sécurité informatique*. Nous savions que c'était un thème de recherche qui existe en informatique, mais nous n'aurions peut-être pas pensé à ce thème pour notre recherche de textes si nous n'avions pas utilisé le guide des expertises.

Lors du montage des deux corpus, nous avons justifié chaque choix de textes en lui apposant ce que nous avons convenu de nommer un *thème* qui représente une thématique de recherche de cette discipline. De plus, nous avons pu utiliser certains de ces thèmes comme codes de recherche dans Google Scholar pour aborder des sujets encore plus précis, par exemple *médias sociaux + sécurité informatique*.

Chacun des textes sélectionnés doit aborder un des thèmes pour faire partie des corpus de la communication ou de l'informatique. Afin de justifier le choix de nos textes, nous ajoutons

dans la fiche descriptive de chacun des textes (Annexe 1 et Annexe 2) le ou les thèmes qui indique(nt) à quel(s) champ(s) d'expertise(s) des domaines de la communication et de l'informatique le texte appartient. Par exemple, pour justifier le choix d'un texte de la communication, celui-ci peut se voir attribuer un seul thème, soit celui de *culture numérique*, ou alors plusieurs thèmes, comme *communautés virtuelles* et *communication interprofessionnelle*.

Nous avons fait une sélection parmi les thèmes de recherche proposés afin de retenir ceux qui sont liés aux médias sociaux. Le Tableau 3 et le Tableau 4 présentent les thèmes qui ont été utilisés pour valider le choix de nos textes. On peut les retrouver dans la section *Thème* des fiches descriptives de nos textes (Annexe 1 et Annexe 2).

Tableau 3 *Les thèmes de recherche attribués aux textes sélectionnés dans COMMU*

Les thèmes de recherche attribués aux textes sélectionnés dans COMMU		
organisation sans but lucratif	communication politique	technologies interactives
controverses publiques	communication et santé	environnement d'apprentissage
réseaux sociaux	communications tactiques d'entreprises	interactions sociales
études des médias	époque contemporaine	études sur la surveillance
controverses sociotechniques	pratiques situées et institutionnalisées en milieu de travail	cyberculture
interaction humain-ordinateur	communication en situation de crise	communication organisationnelle
journalisme	identité occupationnelle	arts médiatiques
culture populaire	conflit organisationnel	médias
mouvements sociaux		

Tableau 4 *Les thèmes de recherche attribués aux textes sélectionnés dans INFOR*

Les thèmes de recherche attribués aux textes sélectionnés dans INFOR		
sécurité informatique	circuits booléens	sciences des données
réseaux sociaux	traitement automatique des langues naturelles (TALN)	algorithmique
médias sociaux	programmation	modèles statistiques
inférence bayésienne		

### 3.2.2 La comparabilité des corpus

Nous voulons monter deux corpus comparables pour répondre à l'un des objectifs de cette recherche qui est de faire ressortir les différences et les ressemblances entre les disciplines



de la communication et de l'informatique dans la terminologie du domaine des médias sociaux. De ce fait, nous établissons comme critères pour le montage des corpus que la répartition des genres de textes soit semblable, que les textes traitent de sujets qui sont inhérents à la discipline, que le thème des médias sociaux soit central dans ces textes, que les corpus soient de tailles équivalentes et que les textes soient rédigés durant la même période.

### **3.2.3 La période cible**

Lors de la constitution des corpus, nous avons délimité une période cible pour les textes qui les composent. Les premiers médias sociaux sont arrivés à la fin des années 1990. Toutefois, ce n'est vraiment qu'à partir de 2007, avec la sortie du premier téléphone intelligent de Apple, l'iPhone, que les médias sociaux se sont emparés de la vie quotidienne des utilisateurs de téléphones cellulaires. Nous aurions donc pu choisir de traiter des textes qui datent de 2007, mais pour circonscrire les résultats à une période de 10 ans, nous avons choisi de n'accueillir dans nos corpus que des textes publiés entre 2010 et 2019. Cela équivaut à un peu moins de 10 ans, car l'année 2019 n'est pas terminée à la date du montage des corpus. Cette période est aussi la plus féconde pour le domaine, car c'est autour des années 2010 que nombre de médias sociaux importants en ce moment ont vu le jour : Instagram en 2010, WhatsApp en 2009, Snapchat en 2011. Les deux corpus doivent répondre à ce critère, toujours pour assurer la comparabilité.

### **3.2.4 Le genre des textes**

La construction des corpus nécessite d'évaluer le genre de textes. Nous avons cherché principalement des textes scientifiques, car non seulement notre barème de choix des thèmes était la liste d'expertises des départements d'informatique et de communication de l'Université de Montréal, mais notre principal outil de recherche en ligne a été Google Scholar (nous avons aussi cherché quelques textes à l'aide de Google). Nous avons voulu inclure le plus de textes scientifiques possibles, mais si nous trouvions un texte qui n'était pas scientifique qui traitait tout de même de thèmes faisant partie de la liste des expertises, nous le conservions, peu importe son genre. Les textes qui n'appartiennent pas au genre scientifique représentent moins de 10 % de nos corpus.

Nous avons sélectionné les genres de textes suivants : *revues spécialisées, vulgarisation scientifique et journaux* (Ahmad, 1995, cité dans Marshman, 2003) . Nous avons ajouté quelques catégories que nous avons repérées dans nos textes : *rapports officiels, thèses et mémoires, actes de conférence et blogue*.

Pour chaque texte, nous avons indiqué un niveau de spécialisation (Marshman, 2003) :

1. *Spécialisés* : revues spécialisées, rapports officiels, actes de conférences, thèses et mémoires ;
2. *Semi-spécialisés* : vulgarisation scientifique ;
3. *Non spécialisés* : blogues, journaux.

Nous avons évalué le statut des textes en nous fiant au niveau d'expertise de l'auteur et le lectorat ciblé. Les deux graphiques présentés à la Figure 1 et la Figure 2 montrent la répartition des textes dans les corpus.

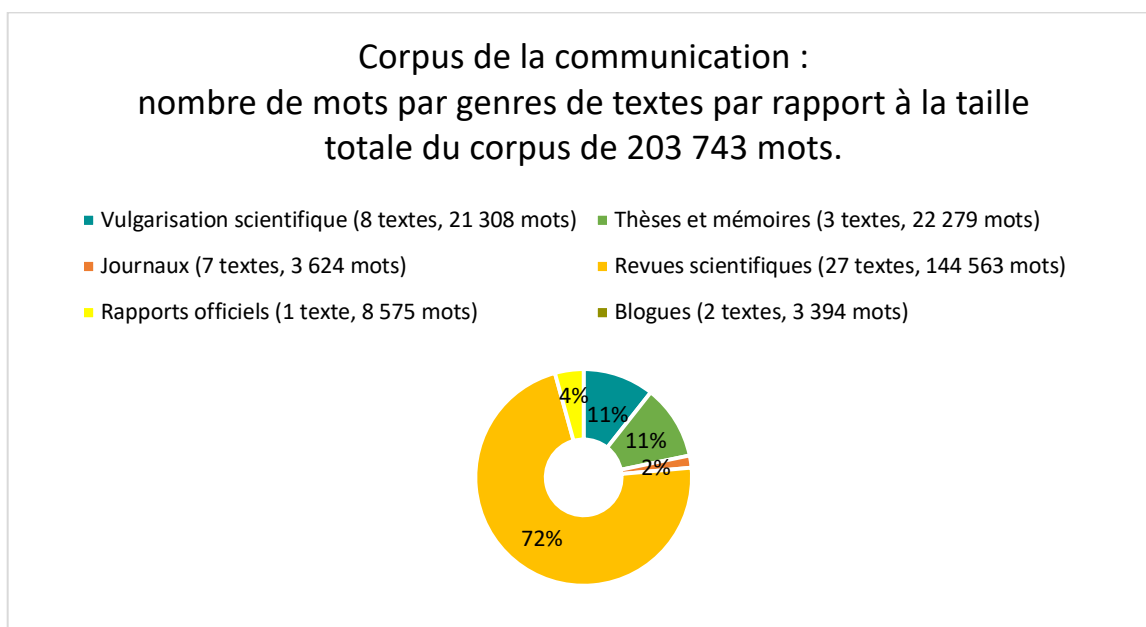


Figure 1 *La répartition des genres des textes : COMMU*

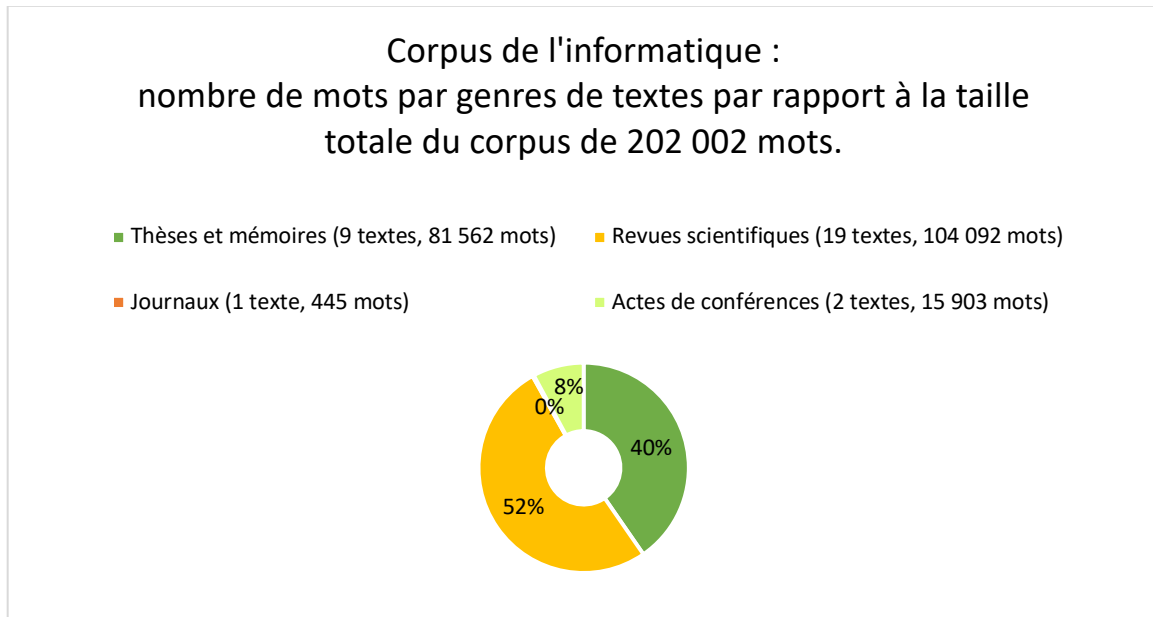


Figure 2 *La répartition des genres des textes : INFOR*

Comme nous pouvons le constater, les deux corpus ne sont pas construits avec la même répartition des genres de textes. Cependant, si nous ne nous fions pas au genre des textes, mais au type de spécialisation, les deux corpus sont formés à plus de 90 % de textes spécialisés, ce qui les rend comparables, à notre avis.

### 3.2.5 Les critères linguistiques

Pour monter deux corpus comparables, nous devons utiliser les mêmes critères linguistiques pour chacun.

La qualité de la rédaction des textes a été jugée primordiale. Pour chacun des textes, nous avons analysé quelques extraits pour voir s'ils contenaient des erreurs de langue. Nous nous fions à des extraits pour évaluer la qualité de la langue, car ils restent représentatifs de la qualité du texte. Les textes de mauvaise qualité sont repérables même dans un extrait. Nous savons toutefois que tous les textes que nous avons sélectionnés ne sont pas totalement exempts d'erreurs de langue.

Nous n'avons pas accepté dans nos corpus des textes qui n'ont pas été rédigés en français. Tous les textes traduits ont été rejetés de notre analyse, car nous cherchons à analyser

des textes idiomatiques et, dans le cas d'une traduction, le caractère idiomatique peut être plus faible.

Il n'était pas nécessaire selon nous de nous arrêter à une seule géolocalisation pour l'origine des textes. Nos textes viennent d'Amérique, d'Europe et d'Afrique. Bien que nous rédigeons ce travail dans un contexte nord-américain, la quantité de textes en français que nous pouvions trouver est trop petite pour la portée de l'analyse terminologique. De plus, comme il sera mentionné plus loin dans la section 3.2.7, il a été difficile de recueillir des textes pour le corpus de l'informatique. Plusieurs des textes de l'informatique choisis ont été rédigés par des universitaires magrébins. Sans l'expertise dans le domaine des médias sociaux des étudiants en informatique des pays africains, nous n'aurions pas pu avoir un corpus aussi riche en thèses et mémoires.

### 3.2.6 La taille des corpus et des textes

Pour ce qui est de la taille des corpus, nous avons dû réduire la taille que nous avions projetée au départ. Nous apporterons des précisions sur les raisons expliquant cette révision à la baisse dans la section 3.2.7. Nous avons donc monté deux corpus d'un peu plus de 200 000 mots. En termes de taille, COMMU s'élève à 203 743 mots alors que INFOR s'élève à 202 002 mots. Nous jugeons que l'écart de 1 741 mots aura peu d'impact sur les résultats. Le Tableau 5 et le Tableau 6 présentent les textes selon leur longueur. Ils sont regroupés par tranches de 2 000 mots. Cela permet de montrer la répartition de la taille des textes par rapport à la taille des corpus.

Tableau 5 *La répartition des textes en termes de tailles : COMMU*

La répartition en termes de tailles des textes : COMMU		
Moins de 2 000 mots	286	gangsternumerique
	351	militairesrusses
	434	antivaccins
	469	homosexuelmusulman
	475	sante
	534	rtrusse
	768	liguedulol
	782	talibans
	898	sixetapes
	1 107	batirstrategie

	1 109	multicanal
	1 193	fearofmissingout
	1 771	communicationdigitale
	1 984	oeilmoderne
<b>Sous-total</b>	<b>12 161</b>	<b>14 textes</b>
2 000 à 3 999 mots	2 064	marketingterritorial
	2 201	avantagesinconvenients
	2 536	wikipedia
	2 636	controlesocial
	3 374	communiquer
	3 505	cyberbullying
	3 619	biographisation
	3 781	transplateforme
	3 947	personnagesfictifs
3 969	saintamable	
<b>Sous-total</b>	<b>27 663</b>	<b>10 textes</b>
4 000 à 5 999 mots	4 081	professionnelscommunication
	4 156	netart
	4 382	hopitaux
	4 494	twittermariage
	4 698	youtubedebats
	4 775	affectsfacebook
	4 793	usagesprofessionnels
	4 809	cotedivoire
	5 058	vieprivee
	5 189	autoformation
	5 518	apprentissageangelangues
	5 543	recrutement
5 738	tweetstraces	
<b>Sous-total</b>	<b>63 234</b>	<b>13 textes</b>
6 000 à 7 999 mots	6 961	algerie
	7 046	publicitetrompeuse
	7 404	linkedincommunication
	7 430	youtubeobesite
	7 934	abcpourosbl
<b>Sous-total</b>	<b>36 775</b>	<b>5 textes</b>
8 000 à 9 999 mots	8 575	administrationquebecoise
	8 774	feuxdeforet
	9 080	erecrutement
<b>Sous-total</b>	<b>26 429</b>	<b>3 textes</b>
10 000 à 11 999 mots	10 000	Policing
	10 350	infomediation
<b>Sous-total</b>	<b>20 350</b>	<b>2 textes</b>
12 000 à 13 999 mots	13 166	revolutiontunisienne
<b>Sous-total</b>	<b>13 166</b>	<b>1 texte</b>
<b>Total</b>	<b>203 743</b>	<b>48 textes</b>

Tableau 6 *La répartition des textes en termes de tailles: INFOR*

La répartition en termes de tailles des textes : INFOR		
Moins de 2 000 mots	445	appevents
	953	introductiontwitter
	959	sentimentstwitter
	1 126	plateformesondy
	1 687	meltwater
<b>Sous-total</b>	<b>5 170</b>	<b>5 textes</b>
2 000 à 3 999 mots	2 789	mesusagemedicaments
	2 900	reinformation
	3 084	defistal
	3 157	ironietweets
	3 532	cartographie
	3 935	statistiquesinformation
<b>Sous-total</b>	<b>19 397</b>	<b>6 textes</b>
4 000 à 5 999 mots	4 565	emotionstweets
	5 226	opinionsblogosphere
	5 800	contextualisationtweets
<b>Sous-total</b>	<b>15 591</b>	<b>3 textes</b>
6 000 à 7 999 mots	6 132	profilutilisateur
	6 230	capturededonnees
	6 969	bigsocialdata
	7 274	securisationreseauxsociaux
	7 288	rechercheinformation
	7 421	systemesderecommandation
	7 839	donneesmanquantes
<b>Sous-total</b>	<b>49 153</b>	<b>7 textes</b>
8 000 à 9 999 mots	8 832	detectiontwitter
	8 977	propagationmemes
	9 333	integritecontextuelle
<b>Sous-total</b>	<b>27 142</b>	<b>3 textes</b>
10 000 à 11 999 mots	10 075	detectionrumeurs
	10 677	intelligenceartificielle
<b>Sous-total</b>	<b>20 752</b>	<b>2 textes</b>
12 000 à 13 999 mots	12 066	dangersreseauxsociaux
	12 856	memesreseauxsociaux
	12 909	diffusioninformation
	13 141	reseauxegocentriques
	13 825	detectionpointsdevue
<b>Sous-total</b>	<b>64 797</b>	<b>5 textes</b>
<b>Total</b>	<b>202 002</b>	<b>31 textes</b>

Comme nous pouvons le constater dans le Tableau 5 et le Tableau 6, COMMU est composé de 48 textes et INFOR, de 31 textes.

### 3.2.7 Les difficultés rencontrées

Pour chacun des points de l'étape du montage des corpus abordés précédemment, nous avons rencontré des difficultés.

Lorsque nous avons commencé à constituer les corpus, nous n'avions pas instauré de prime abord le système de thèmes. Nous ne sommes expert ni en communication ni en informatique. Comment savoir si les textes sélectionnés font partie des thèmes abordés dans ces disciplines ? La solution que nous avons trouvée à ce problème a été de consulter la liste des expertises des départements de communication et d'informatique de l'Université de Montréal. Une fois que nous avons réglé ce problème, nous étions plus confiant quant au choix des textes. En revanche, nous avons déjà commencé à monter notre corpus. Nous avons donc dû évaluer les textes que nous avons déjà sélectionnés pour vérifier qu'ils appartenaient bien à des thématiques présentées dans la liste des expertises. Certains textes se sont avérés hors sujets, ce qui nous a amené à les rejeter.

Les critères de comparabilité que nous avons fixés ont été mis à l'épreuve pour différentes raisons. D'abord, le montage des deux corpus s'est fait en deux temps. En effet, nous avons commencé le montage par COMMU et, après avoir amassé des textes cumulant environ 150 000 mots, nous nous sommes intéressé aux textes qui pourraient faire partie du corpus de l'informatique. Inévitablement nous avons cherché à comparer ce que nous avons trouvé pour l'informatique avec ce que nous avons déjà pour la communication. Cette étape de comparaison s'est avérée être un point tournant dans notre montage du corpus : le volume de texte de l'informatique en français est moindre que le volume de texte de la communication. Cela a entraîné un changement dans la taille des corpus, que nous aborderons un peu plus bas.

La période cible ne s'est pas avérée un si grand défi, puisque les médias sociaux sont un phénomène récent autant en communication qu'en informatique. La très grande majorité des textes repérés respectait la période cible. Cependant, à une vingtaine de reprises durant notre recherche de textes, nous avons déniché des textes intéressants pour notre corpus, mais qui dataient d'avant 2010, ce qui nous a amené à les rejeter.

Quant au type des textes, il apparaît dans nos corpus que certains types se retrouvent dans un seul des corpus. Dans COMMU, on retrouve les types *vulgarisation scientifique*,

*rapports officiels* et *blogues* qui sont absents du corpus de l'informatique. Dans INFOR, le genre *actes de conférences* y est présent, alors qu'il est absent de COMMU. Nous nous sommes demandé si nous devions inclure tous les mêmes genres de textes dans les deux corpus. Nous estimons que cela n'affecterait pas nos résultats de façon démesurée, car c'est surtout la valeur du niveau de spécialisation qui compte à nos yeux. Nous avons également fait l'exercice de relever le volume de chaque genre en termes de mots par rapport au corpus entier. Comme nous pouvons le constater dans les graphiques de la section 3.2.4 (Figure 1 et Figure 2), les genres de textes principaux dans les deux corpus sont *thèses et mémoires* et *revues scientifiques*. Cependant, leur répartition est différente dans chaque corpus. Là aussi, nous nous sommes demandé si cela aurait une incidence sur nos résultats. Nous supposons que les résultats ne pourront pas révéler si cela a eu une réelle incidence. Cependant, la difficulté que nous avons rencontrée à trouver des textes de l'informatique en français qui traitent des médias sociaux était telle que nous n'avons eu d'autre choix que de nous tourner principalement vers le genre *thèses et mémoires* dans le montage du corpus de cette discipline. Nous pensons qu'il s'agit d'une caractéristique représentative du domaine : les textes en français dans le domaine des médias sociaux en informatique sont des thèses et des mémoires en majorité. Nous avons également réorienté nos recherches de textes de l'informatique vers des études du traitement automatique des langues qui sont nombreuses à traiter des médias sociaux.

En ce qui concerne les critères linguistiques, nous nous sommes rendu compte qu'il était plutôt facile de trouver des textes en français pour le domaine de la communication, alors que la tâche était beaucoup plus ardue dans le domaine de l'informatique. Force est de constater que dans les sciences exactes, les auteurs sont plus portés à écrire en anglais. Il n'y a pas de revues scientifiques en informatique en français dans le répertoire des bibliothèques de l'Université de Montréal. Nous avons dû rejeter plusieurs textes traduits. Quant aux textes comportant des erreurs de langue, ils étaient plus fréquents dans le domaine de la communication ; nous avons remarqué que plusieurs auteurs semi-spécialisés ou non spécialisés écrivent sur le sujet des médias sociaux du point de vue de la communication, ce qui faisait en sorte que nous avons rencontré beaucoup plus de textes avec des erreurs de langue.

La taille des corpus a changé de manière importante. Avant de commencer la recherche, nous nous étions fixé comme objectif de créer deux corpus de 500 000 mots pour un total d'un



million de mots. En commençant à chercher des textes pour le domaine de l'informatique, nous nous sommes rapidement rendu compte que l'objectif de 500 000 mots était démesuré. Nous avons donc décidé de réduire nos deux corpus à 200 000 mots. L'objectif a été atteint, mais le montage du corpus de l'informatique a été trois fois plus long que celui de la communication et a nécessité beaucoup plus de codes de recherche. Nous avons utilisé seulement 6 codes de recherche pour trouver des textes intéressants pour COMMU (Tableau 3), alors que nous en avons utilisé 11 codes de recherche pour INFOR (Tableau 4).

Finalement, nous nous sommes efforcé de ne pas prendre trop de textes provenant de même auteur. Par exemple, nous avons dû rejeter des textes de Atefeh Farzindar malgré le fait que ses articles étaient tout à fait pertinents pour notre étude, car nous avions déjà quelques textes écrits par cette autrice.

### **3.3 L'identification des CT par TermoStat**

Pour l'identification les termes contenus dans les textes des médias sociaux, nous avons fait appel au logiciel TermoStat (Drouin, 2003). Cette partie de la recherche s'est déroulée en plusieurs étapes. Il a d'abord fallu faire le formatage des textes qui composent nos corpus avant de les soumettre au logiciel, nous avons ensuite soumis les deux corpus à TermoStat. Finalement, il a fallu faire un nettoyage manuel dans la liste des résultats.

#### **3.3.1 Le formatage des textes**

Les textes qui composent nos corpus sont tous issus de pages et documents Web. Pour les soumettre à TermoStat, nous avons dû effectuer un certain nombre de conversions. D'abord, nous avons copié et collé leur contenu dans un document Word.

À ce nouveau document, nous avons enlevé la plupart des sauts de paragraphes afin que le texte soit réuni en un seul bloc. Pour chacun des textes, des retraits ont été faits (bibliographie, index, liste des tableaux, etc.). La nature de ces retraits pour chacun des textes est expliquée dans les fiches signalétiques des textes qui se trouvent dans l'Annexe 1 et dans l'Annexe 2.

Une fois les modifications dans le texte terminées, nous passons à l'étape de l'enregistrement. Les textes ont été enregistrés en *UTF-8* sous le format *texte brut* (.txt).

Ensuite, nous avons renommé chacun des .txt par un titre révélateur de son contenu. Par exemple : memesreseauxsociaux.txt pour un texte traitant de la création d'un nouvel outil d'analyse et de visualisation de mèmes sur Internet.

Finalement, nous avons copié le contenu de chacun des documents .txt et nous les avons collés dans un nouveau document Word pour chaque corpus. Ces deux nouveaux documents contenant tous les textes sont nos deux corpus réunis. Ces corpus ont été renommés respectivement *COMMU* et *INFOR*, puis enregistrés en *UTF-8* et en .txt.

### **3.3.2 La soumission de COMMU et INFOR à TermoStat**

Une fois le formatage des deux corpus complété, ils peuvent être traités par le logiciel Termostat. Nous les avons versés à tour de rôle dans la boîte intitulée *choisir un fichier* (Figure 3 *L'interface de TermoStat*).

Nous avons sélectionné l'option *Français* pour la langue (Figure 3).

Puis, nous avons coché les boîtes *termes simples* et *termes complexes* (Figure 3). Nous avons choisi d'inclure les termes complexes dans l'analyse afin de nous assurer de repérer différents types de néologismes. Pour les termes simples, le logiciel propose un menu déroulant qui permet de sélectionner la partie du discours à laquelle doivent appartenir les termes analysés ; nous avons coché les quatre catégories à notre disposition (adjectifs, adverbes, noms et verbes), toujours dans le but de maximiser notre nombre de néologismes. La Figure 3 montre l'interface de TermoStat.

# TermoStat Web 3.0

## Nouveau corpus

Fichier  Aucun fichier choisi

Langue  ▾

Extraction  termes simples  termes complexes

Catégories (4) ▾

- adjectifs
- adverbes
- noms
- verbes

Figure 3 *L'interface de TermoStat*

### 3.3.3 Les candidats-termes

Après la soumission des deux corpus à TermoStat, une liste de candidats-termes classés en ordre de score de spécificité décroissant est générée par le logiciel.

La Figure 4 et la Figure 5 sont des captures d'écran de la première page de la liste pour COMMU et INFOR dans TermoStat.

termostat.ling.umontreal.ca/display\_results.php

Amazon.ca eBay Booking.com AliExpress Facebook Télé Fibe

Corpus >> commu2\_utf8 Julien2992 | Aide | Déconnexion

## Résultats

Liste des termes Nuage Statistiques Structuration Bigrammes

Candidat de regroupement	Score		Variantes orthographiques	Matrice
	Fréquence (Spécificité)			
médias sociaux	599	265.52	médias sociaux	Nom Adjectif
réseau social	479	236.39	réseau social réseaux sociaux	Nom Adjectif
médias	794	143.73	médias	Nom
vidéo	379	131.14	vidéo vidéos	Nom
corpus	207	124.96	corpus	Nom
social	1453	116.26	social sociale sociales	Adjectif
web	133	111.77	web	Nom
utilisateur	246	109.12	utilisateur utilisateurs utilisatrices	Nom
réseau	722	107.76	réseau réseaux	Nom
tweets	99	107.49	tweets	Nom
interaction	161	103.49	interaction interactions	Nom
plateformes	85	98.92	plateformes	Nom
communication	522	95.02	communication communications	Nom
usager	218	91.68	usager usagers usagère	Nom
twitter	68	88.87	twitter	Verbe
plateforme	85	86.07	plateforme	Nom
publicité rédactionnelle	63	85.49	publicité rédactionnelle publicités rédactionnelles	Nom Adjectif
socionumériques	63	85.49	socionumériques	Adjectif
contenu	301	85.18	contenu contenus	Nom
osbl	59	82.69	osbl	Nom

Figure 4 Capture d'écran des premiers CT de COMMU proposés par TermoStat

Corpus >> Infor\_utf8 Julien2992 | Aide | Déconnexion

## Résultats

Liste des termes Nuage Statistiques Structuration Bigrammes

Candidat de regroupement	Score		Variantes orthographiques	Matrice
	Fréquence (Spécificité)			
utilisateur	995	291.47	utilisateur utilisateurs utilisatrice	Nom
réseau social	725	287.85	réseau social réseaux sociaux	Nom Adjectif
al	538	239.47	al	Verbe
tweets	275	177.57	tweets	Nom
algorithme	297	177.41	algorithme algorithmes	Nom
graphe	272	174.65	graphe graphes	Nom
réseau	1190	170.7	réseau réseaux	Nom
médias sociaux	241	166.19	médias sociaux	Nom Adjectif
profil	493	150.57	profil profils	Nom
hashtags	187	146.3	hashtags	Nom
donnée	632	140.97	donnée données	Nom
utilisateur	200	138.2	utilisateur utilisateurs utilisatrice utilisatrices	Adjectif
n	239	135.42	n	Nom
social	1582	125.21	social sociale sociales	Adjectif
uds	130	121.84	uds	Nom
corpus	200	120.57	corpus	Nom
détection	215	120.23	détection	Nom
temporel	152	109.15	temporel temporels temporelle	Adjectif

Figure 5 Capture d'écran des premiers CT de INFOR proposés par TermoStat

Les termes ayant un score de spécificité élevé apparaissent plus haut dans la liste, même si ceux-ci ont parfois une fréquence d'occurrence moindre. Par exemple, le terme *web* a un score de spécificité de 111,77 dans COMMU (Figure 4). Il apparaît dans la liste avant *utilisateur* qui a un score de spécificité de 109,12, même si *utilisateur* a une fréquence d'occurrence de 246 contre 133 pour *web*.

Le logiciel a repéré dans le corpus COMMU 2 920 CT ; il en a repéré 3 834 dans le corpus INFOR. Dans le cadre de cette étude, nous n'analyserons pas tous les termes repérés. Nous nous concentrons sur un échantillon plus petit qui regroupe les termes avec les scores de spécificité les plus élevés après l'élagage des unités non terminologiques.

La liste ne comprend pas que des unités terminologiques. Le logiciel a repéré des unités non terminologiques et a fait des erreurs de repérage. Nous avons donc fait un tri manuellement en observant les contextes d'occurrence pour éliminer les candidats-termes à rejeter.

### ***3.3.3.1 Les 150 premiers termes***

Notre analyse prend en compte les 150 termes parmi les premiers CT proposés par TermoStat à partir de chacun de nos corpus (COMMU et INFOR). Ces termes ne sont pas nécessairement les 150 ayant obtenu les scores de spécificité les plus élevés, car nous avons fait un élagage des différentes erreurs de repérage de TermoStat jusqu'à ce que les listes de COMMU et de INFOR soient toutes deux d'une longueur de 150 termes. Du premier des 150 termes au dernier, la liste de COMMU compte des scores de spécificité de 265,52 à 25,27 ; pour INFOR, ils sont de 291,47 à 33,95. Ce ne sont toutefois pas 300 termes (l'addition des deux listes de 150 termes de COMMU et INFOR) que nous allons analyser, car certains termes sont présents à la fois dans la liste de la communication et dans celle de l'informatique. Nous traiterons de ces termes au chapitre 4.

Un plus grand nombre de CT sont rejetés plus nous avançons dans la liste (particulièrement après la première centaine de termes). Nous estimons qu'un échantillon de 150 termes peut donner une idée assez fidèle des contributions de chacune des disciplines et des types de leurs néologismes.

Ces deux listes de 150 termes seront chacune soumises au corpus d'exclusion pour le repérage des néologismes (3.5).

### **3.3.3.2 L'élagage**

Comme nous l'avons mentionné plus haut, ce ne sont pas tous les CT repérés par TermoStat qui sont analysés. L'élagage est une étape inévitable pour nous permettre de nous concentrer sur les termes du domaine des médias sociaux du point de vue de la communication et de l'informatique.

Cette méthode a également été utilisée par Lafrance (2016) (l'auteur parle de nettoyage) afin de faire le tri dans ses résultats après l'extraction. Elle a élagué trois types de CT, qu'elle catégorisait comme des « erreurs d'étiquetage », des « CT qui se répétaient » et des « CT n'ayant pas un statut terminologique » Lafrance (2016, p. 33-40). Nous élaguons trois types de CT qui peuvent être classés dans des catégories semblables à celles de Lafrance (2016) : les occurrences de formes multiples, les erreurs de repérage et les unités non terminologiques. Tous les CT élagués sont répertoriés dans l'Annexe 5. Les CT élagués en bleu sont les erreurs de repérage et les CT en brun sont les unités non terminologiques. Les formes multiples n'ont pas été mises dans ce tableau, parce que nous les avons regroupées sous une entrée unique dans les listes de 150 termes. Nous pouvons constater que les unités terminologiques sont celles qui occupent une plus grande place parmi les unités élaguées. En tout, 99 candidats-termes ont été élagués de COMMU (sans compter les formes multiples) et 140 de INFOR (sans compter les formes multiples).

#### *3.3.3.2.1 Les formes multiples*

Certains CT apparaissent sous plusieurs formes dans la liste de TermoStat. Par exemple, le terme complexe *média social* apparaît dans la liste de CT de COMMU sous la forme *média social* et *médias sociaux*. Lorsqu'un CT apparaissait sous plusieurs formes, nous conservions le rang de sa première occurrence pour la liste des 150 termes et nous le ramenions à sa forme au singulier. C'est le cas de *tweet* qui apparaissait en premier lieu au pluriel. Dans le cas d'un terme

apparaissant au féminin, nous le ramenons à la forme neutre du masculin. Par contre, pour le terme *vidéoblogueuses*, nous l'avons laissé au féminin, car la forme masculine n'apparaissait pas dans le corpus

#### 3.3.3.2.2 *Les erreurs de repérage*

Certains CT ne sont en fait que des erreurs de repérage par le logiciel. Par exemple, TermoStat a repéré dans le corpus INFOR les CT *ii*, *maxw* et *z* (Annexe 5). Ces erreurs de repérage peuvent être causées par le transfert de la page HTML vers le document *.txt*. Il peut également s'agir d'erreurs de retranscription (ex. : *z* au lieu de *2*) lors de la copie d'un document PDF vers le document Word qui sert à créer le *.txt*. Ces erreurs sont faciles à repérer dans la liste de CT, car elles ne ressemblent pas à des vrais mots.

#### 3.3.3.2.3 *Les unités non terminologiques*

Notre objectif est de réunir deux listes de 150 *termes* des médias sociaux, pour ensuite y trouver des néologismes. Le logiciel TermoStat compare un corpus spécialisé à un corpus de référence interne (2.2.1.2.2). C'est à partir de cette comparaison qu'il donne un score de spécificité à tous les *CT*, puis il les regroupe en ordre décroissant de spécificité. Toutefois, lors de la comparaison avec le corpus de référence, il se peut que certaines unités non terminologiques soient beaucoup plus présentes dans le corpus spécialisé que dans le corpus de référence : elles obtiennent, par conséquent, un score de spécificité plus élevé. Par exemple, dans la liste de CT de INFOR, le CT *corpus* apparaît avec un score de spécificité de 120,57. Pourtant, si nous nous fions aux contextes dans lesquels le CT *corpus* apparaît dans INFOR, nous nous rendons compte que c'est principalement dans des revues scientifiques et des thèses ou mémoire où le chercheur a utilisé des *corpus* de données. Comme notre corpus est composé de plusieurs textes de ces genres, cela explique pourquoi ce CT apparaît dans la liste de TermoStat avec un score élevé, et surtout une grande fréquence. Ces unités, comme que nous catégorisons comme *non terminologiques* sont rejetées, car elles ne font pas partie de la

terminologie comme telle des médias sociaux : elles déjouent le corpus de référence à cause de leur plus grande fréquence

L'Homme (2004, p. 3-4) donne des critères utiles pour sélectionner des termes à inclure dans un dictionnaire à partir d'un corpus spécialisé. Dans le cas de sa recherche, c'était pour la création de fiches terminologiques dans un dictionnaire de l'informatique. Selon la chercheuse, mises à part la fréquence et la répartition, ces quatre critères peuvent aider le chercheur à valider des termes :

- a. L'unité doit désigner une entité du domaine ;
- b. Les unités prédicatives doivent avoir pour actants des entités du domaine (si elles présentent le même sens lorsqu'elles sont combinées à des entités du domaine et avec des actants non spécialisés, elles sont exclues) ;
- c. Les dérivés morphologiques doivent être sémantiquement apparentés à des unités validées en a ou en b ;
- d. Finalement, les unités en relations paradigmatiques sont terminologiques si leur relation paradigmatique est avec une unité validée en a, b ou c.

### **3.4 Les méthodes de validation des néologismes**

En ce qui concerne le montage du corpus d'exclusion, nous nous sommes inspiré de l'étude de Cabré et de Yzaguirre (1995). En tout, ils ont utilisé cinq ouvrages de référence pour détecter les néologismes. Nous avons intégré deux dictionnaires de la communication et deux dictionnaires de l'informatique à notre corpus d'exclusion. Nous avons également sélectionné deux dictionnaires des médias sociaux, car nous émettons l'hypothèse qu'ils répertorient des termes qui sont indépendants des deux points de vue. Comme nous ne travaillons qu'avec une seule langue, nous pouvons soumettre nos deux corpus aux mêmes outils de validation. Il est important pour nous de consulter des ressources qui visent chacun des points de vue. Nous émettons l'hypothèse que certains termes relevés par l'extracteur appartiennent à un aspect plus large de la communication ou de l'informatique qui ne sauraient se retrouver dans des dictionnaires qui se concentrent exclusivement sur les médias sociaux. D'autre part, nous



émettons une autre hypothèse selon laquelle certains termes des médias sociaux spécifiques à la communication ou à l'informatique sont absents des dictionnaires de ces deux disciplines à cause de l'abondance des autres concepts à décrire qui ne concernent pas les médias sociaux. C'est pourquoi nous avons voulu inclure des dictionnaires qui traitent spécifiquement des médias sociaux. Les dates de publication des dictionnaires étaient importantes pour nous. Nous avons tenté de sélectionner des dictionnaires très récents. Nous nous sommes donné comme critère de sélectionner des dictionnaires publiés après 2010.

À la différence de Cabré et de Yzaguirre (1995) qui travaillaient avec deux langues différentes, le catalan et le castillan, notre étude ne porte que sur le français et se fait à partir de deux points de vue. Cependant, nous voyons une ressemblance entre notre recherche et celle de Cabré et de Yzaguirre (1995), car nous travaillons à partir de deux corpus comparables. Si au départ nous n'avions qu'un seul test (le corpus d'exclusion), nous avons finalement intégré l'outil de visualisation de n-grammes dans notre analyse *a posteriori*, car les dictionnaires ne suffisaient pas pour l'élagage des unités non néologiques.

Notre validation des néologismes est un enchaînement de méthodes. Ce ne sont pas tous les CN qui sont soumis à tous les tests. Le premier test est la soumission des CN aux dictionnaires spécialisés, le deuxième à l'outil de visualisation de n-grammes et le troisième aux dictionnaires de langue générale.

Nous avons donc soumis les listes de 150 termes des corpus de la communication et de l'informatique à six dictionnaires spécialisés (deux dictionnaires de la communication, deux dictionnaires de l'informatique et deux dictionnaires des médias sociaux). Si les CN sont toujours considérés comme des néologismes potentiels, ils passent au deuxième test, mais pas les autres. Après le premier test des dictionnaires spécialisés, il nous restait 69 CN pour COMMU et 91 CN pour INFOR. Pour le deuxième test, nous avons soumis ces 157 CN à l'outil de visualisation de n-grammes. Finalement, les dictionnaires de langue générale servent de dernier test après la soumission CN aux deux premiers tests.

Nous présentons d'abord les outils, puis nous expliquons leur application aux deux listes de 150 termes.

### 3.4.1 Les dictionnaires de la communication

Pour les dictionnaires de la communication, nous avons d'abord sélectionné le *Lexicom : Les 3 500 mots du marketing publicitaire, de la communication et des techniques de production* (Milon et Saint-Michel, 2015). Le fait que le dictionnaire contient 3 500 mots est un atout, car nous émettons l'hypothèse que si sa nomenclature liste autant de termes, le dictionnaire doit couvrir plusieurs thèmes de la communication en profondeur, notamment les médias sociaux. De plus, son année de publication (2015) se situe au centre de notre période de recherche des textes qui composent notre corpus (voir la section 3.2.3).

Le deuxième dictionnaire de cette catégorie que nous avons choisi est le *Dictionnaire du journalisme et des médias* (Le Bohec, 2010). Si le premier traitait des termes de la communication du point de vue du marketing, celui-ci les traite directement dans les médias et à partir d'un autre sous-domaine de la communication : le journalisme. C'est un ouvrage moins récent (2010), ce qui nous permet d'émettre l'hypothèse que les termes des médias sociaux listés dans le dictionnaire sont implantés dans la terminologie des médias sociaux en 2019, puisqu'ils auront fait l'objet d'une entrée de dictionnaire depuis 10 ans. D'autre part, nous nous attendons à ce que les avancées technologiques fassent en sorte que plusieurs termes plus récents ne soient pas répertoriés dans ce dictionnaire, ce qui nous permettra de faire l'analyse de ces néologismes.

### 3.4.2 Les dictionnaires de l'informatique

Pour les dictionnaires de l'informatique, nous avons choisi un ouvrage papier et un ouvrage numérique. Le premier est le *Dictionnaire informatique & numérique, 2<sup>e</sup> édition* (Lilen, 2014). Un peu comme le *Lexicom*, ce dictionnaire a été publié au milieu de notre période de recherche, donc, celui-ci devrait contenir des termes plus anciens ainsi que des termes qui sont apparus plus tard que certains textes qui composent notre corpus. Ce dictionnaire est composé de 1 400 entrées.

L'autre ressource choisie est le *Dictionnaire de l'informatique et d'internet* (Goguey, 2019), qui est une ressource en ligne. Cette dernière réunit 4 475 entrées terminologiques, ce

qui fait d'elle notre ressource qui répertorie le plus de termes. Les dernières mises à jour ont été faites en 2019, donc la ressource est à jour.

### **3.4.3 Les dictionnaires des médias sociaux**

Finalement, pour les ressources sur les médias sociaux, nous avons choisi les lexiques offerts par Termium plus et l'Office québécois de la langue française. La première est le *Lexique des médias sociaux* (Termium plus, 2018), une ressource en ligne qui compte 468 termes. La dernière mise à jour remonte à 2018, donc, c'est une ressource presque à jour.

L'autre est le *Vocabulaire des réseaux sociaux* (Office québécois de la langue française, 2018). Même s'il ne comporte que 68 notions, ces notions sont très importantes dans le domaine des médias sociaux. Il nous a permis d'exclure plusieurs termes qui n'étaient pas exclus par les 5 autres dictionnaires : *réseautage*, *Web social*, *gestionnaire de communauté* et *binette* (Tableau 13 et Tableau 14). Sa dernière mise à jour date également de 2018.

### **3.4.4 L'outil de visualisation de n-grammes**

Le deuxième test du filtrage des néologismes auquel nous avons soumis les CN qui n'ont pas été exclus par le test des dictionnaires spécialisés est celui de l'outil de visualisation de n-grammes.

Le domaine du traitement automatique des langues utilise des plateformes manipulant des *n-grammes*. « Un n-gramme est une séquence de *n* caractères consécutifs » (Jalam et Chauchat, 2002, p. 2). En d'autres termes, le repérage des n-grammes se fait par les chaînes de caractères.

Des recherches en terminologie ont été faites à partir de n-grammes, notamment Biskri et al. (2004), qui utilisaient les n-grammes pour repérer des termes complexes. C'est également le cas pour Korenchuk (2018), qui a utilisé les n-grammes pour repérer des familles morphologiques. Pour se faire, elle a utilisé les CT extraits d'un corpus pour faire des recherches

par chaîne de caractères. Elle mentionne aussi que ces outils sont utiles dans le traitement des mots composés et pour l'identification de termes à fréquence faible.

Une ressource gratuite en ligne qui propose une analyse en n-grammes à partir de larges corpus est Google Books Ngram Viewer (Google, 2013). Cet outil permet de visualiser les occurrences de syntagmes sur une période sélectionnée dans les Google Books (l'outil analyse les e-books parus qu'entre 1800 et 2008) . La Figure 6 montre une capture d'écran de Google Books Ngram Viewer.

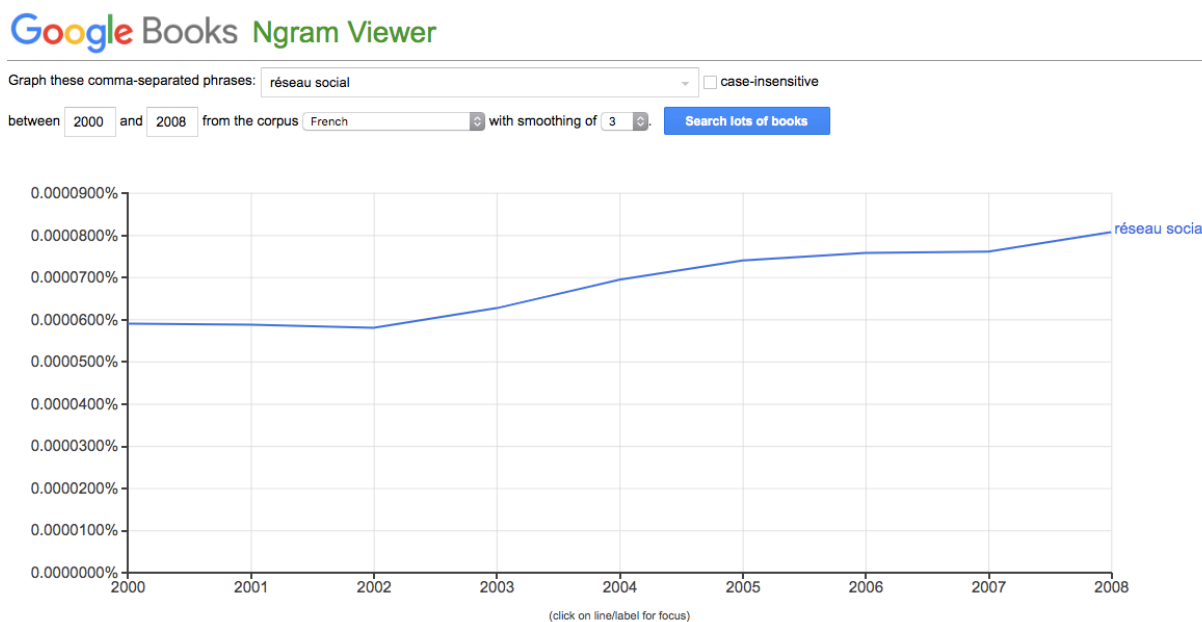


Figure 6 Capture d'écran de Google Books Ngram Viewer

Comme nous pouvons le voir sur la Figure 6, l'outil fournit une courbe de l'évolution dans le temps de la fréquence du syntagme qui fait l'objet de la recherche dans les Google Books sur la période sélectionnée. Si les n-grammes recherchés ne représentent pas une chaîne de caractères présente dans le corpus de Google Books, un message d'erreur apparaît. Cet outil peut être utile pour la recherche en diachronie ou pour le repérage d'unités néologiques.

### 3.4.5 Les dictionnaires de langue générale

Nous avons recours au dictionnaire le Trésor de la Langue Française informatisé (TLFi), car il permet de visualiser la date de la première d'apparition enregistrée. C'est ce facteur qui

nous a fait sélectionner ce dictionnaire, car pour les acceptions des CN faisant aussi partie de la langue générale, nous avons besoin de savoir à quelle date leur acception se matérialisant dans les contextes est apparue dans le dictionnaire afin de savoir si nous pouvons les considérer comme des néologismes.

Usito est aussi un dictionnaire en ligne. Il est alimenté chaque jour, ce qui le rend beaucoup plus à jour qu'une édition papier. Usito propose autant de contenu que 12 dictionnaires (dictionnaire orthographique, dictionnaire des québécismes, dictionnaire des anglicismes, etc.) et contient plus de 80 000 mots (Usito, 2017). Ces caractéristiques font de Usito un bon ouvrage pour notre corpus d'exclusion.

### **3.5 La validation des néologismes**

La méthode que nous avons utilisée pour valider les néologismes est en trois étapes.

D'abord, nous avons cherché chacune des entrées des deux listes de 150 termes dans les six dictionnaires spécialisés. Pour chaque terme, nous notons sa présence ou son absence dans chacun des dictionnaires (Tableau 13 et Tableau 14). Les CN présents dans ces dictionnaires sont automatiquement exclus.

Ensuite, tous les CN qui n'ont pas été exclus par les dictionnaires spécialisés ont été soumis à Google Books Ngram Viewer. Nous avons établi la période de recherche entre 2000 et 2008, comme l'outil de visualisation ne va pas plus loin que 2008 et que les tout premiers médias sociaux ont vu le jour au début des années 2000. Nous établissons que si le CN n'apparaît pas dans l'application, il s'agit d'un néologisme. Si le CN apparaît dans l'application et que sa courbe augmente d'un minimum de 50 % sur la période choisie, nous le considérons comme un néologisme. Les autres CN qui ne montrent pas une courbe saillante, nous les excluons. L'outil de visualisation de n-grammes est très utile pour repérer des termes complexes, mais il présente des limites : nous ne pouvons pas sélectionner la partie du discours des CN que nous interrogeons dans la plateforme et nous n'avons pas accès à des contextes. Ces lacunes nous ont amené à faire passer certains CN qui présentaient des résultats insatisfaisants au troisième test.

Nous nous sommes tourné vers les dictionnaires de langue générale. Plusieurs termes qui ressortaient toujours comme CN après les deux premières étapes nous semblaient avoir une acception dans la langue générale en plus de leur acception terminologique. Leur présence dans les domaines spécialisés est possible, puisque ces domaines ne contiennent pas nécessairement que de la terminologie très technique. Cela ne veut toutefois pas dire qu'il s'agit toujours de néologismes. Nous les avons répertoriés et nous les avons soumis aux dictionnaires de la langue générale. Nous les avons d'abord cherchés dans le TLFi. S'ils y étaient recensés avec la bonne acception, nous les avons exclus. Un problème avec le TLFi, c'est qu'il n'a pas été mis à jour depuis 1996. Pour remédier à ce problème, les CN restants ont été soumis à Usito. S'ils étaient recensés dans Usito, nous les avons exclus.

### **3.6 L'étiquetage des néologismes**

La dernière étape de cette étude consiste à étiqueter les néologismes obtenus à la suite du filtrage. Pour réaliser cette tâche, nous avons utilisé la typologie de Dubuc (2002) (1.2.2).

Nous avons d'abord répertorié les néologismes en indiquant leur rang de spécificité. Nous avons mis en évidence les néologismes qui sont présents dans les deux corpus.

Nous avons ensuite montré leur répartition dans les corpus par rapport aux CN exclus. Nous avons analysé leur répartition en fonction de leur partie du discours ou leur catégorie de syntagme nominal. Puis, nous avons réparti les néologismes en termes simples et en termes complexes.

Ensuite, nous avons procédé à l'étiquetage des néologismes, en fournissant à chacun un contexte d'utilisation issu des corpus. À partir des résultats de l'étiquetage, nous avons analysé la répartition des types de néologismes afin de déterminer lesquels sont les plus productifs et lesquels ne sont pas répertoriés dans les corpus de la communication et de l'informatique.

## Chapitre 4 : Analyse des résultats

Le chapitre 4 présente nos résultats et nos analyses. Il est divisé en quatre sections principales. Nous faisons la comparaison entre les listes de candidats-néologismes de COMMU et INFOR (4.1), nous expliquons les étapes du filtrage des CN (4.2), nous présentons les néologismes des médias sociaux identifiés par notre étude (4.3) et nous étiquetons ces néologismes par type (4.4).

### 4.1 La comparaison entre COMMU et INFOR

Cette section traite de la comparaison entre la terminologie des médias sociaux de la communication et de l'informatique. En posant un regard sur ces deux terminologies côte à côte, il est possible de repérer des ressemblances et des différences dans leur traitement des médias sociaux.

Ces ressemblances et différences peuvent avoir trait à la présence de termes récurrents dans les deux corpus, l'absence de termes de l'un des corpus, la présence de termes en concurrence dans les deux corpus (c.-à-d. de termes quasi-synonymes), la répartition des parties du discours, la répartition des termes simples et complexes, etc.

Les données sont présentées en ordre décroissant du score de spécificité (du score le plus élevé vers le moins élevé) dans le Tableau 7. C'est de cette façon que sont présentées les données dans TermoStat (Drouin, 2003). Le Tableau 7 est divisé en deux colonnes principales (COMMU et INFOR) et en trois colonnes secondaires qui indiquent les rangs des termes après l'élagage, les termes et les parties du discours auxquelles appartiennent les termes. Les termes présents dans les deux corpus sont en couleurs pour faciliter la lecture. Chaque case de couleur dans la colonne de gauche (COMMU) a son identique dans la colonne de droite (INFOR), ce qui permet de voir les différences de rang dans les deux terminologies. Après l'analyse, nous constatons que la convergence entre les rangs semble être faible, mais certains obtiennent tout de même presque le même rang dans les deux corpus, notamment *tweet* et *réseau social numérique*. Le seul terme récurrent dans les deux corpus qui occupe le même rang est *réseau social*.

Tableau 7 La liste des 150 termes les plus spécifiques aux médias sociaux de COMMU et INFOR après l'élagage

COMMU			INFOR		
Rang	Terme	PDD/SN	Rang	Terme	PDD/SN
1	média social	N+Adj	1	utilisateur	N
2	réseau social	N+Adj	2	réseau social	N+Adj
3	média	N	3	tweet	N
4	vidéo	N	4	algorithme	N
5	social	Adj	5	graphe	N
6	Web	N	6	réseau	N
7	utilisateur	N	7	média social	N+Adj
8	réseau	N	8	profil	N
9	tweet	N	9	hashtag	N
10	interaction	N	10	donnée	N
11	plateforme	N	11	utilisateur	Adj
12	communication	N	12	social	Adj
13	usager	N	13	détection	N
14	twitter	V	14	temporel	Adj
15	publicité rédactionnelle	N+Adj	15	similarité	N
16	socionumérique	Adj	16	jeu de données	N+Prep+N
17	contenu	N	17	information	N
18	internaute	N	18	twitter	V
19	Facebook	N	19	événement	N
20	ethos	N	20	profil utilisateur	N+Adj
21	outil	N	21	mème	N
22	usage	N	22	descripteur	N
23	réseautage	N	23	filtrage	N
24	fictionnel	Adj	24	plateforme	N
25	discursif	Adj	25	application	N
26	hashtag	N	26	polarité	N
27	numérique	Adj	27	contextuel	Adj
28	rédactionnel	Adj	28	contenu	N
29	blog	N	29	filtrage social	N+Adj
30	commentaire	N	30	système de recommandation	N+Prep+N
31	profil	N	31	Web	N
32	média socionumérique	N+Adj	32	interaction sociale	N+Adj
33	information	N	33	mésusage	N
34	usage professionnel	N+Adj	34	réseautage	N
35	réseau socionumérique	N+Adj	35	pertinent	N
36	réseau social numérique	N+Adj+Adj	36	catégorisation	N
37	identité professionnelle	N+Adj	37	annotation	N
38	Instagram	N	38	textuel	Adj
39	utilisation	N	39	itération	N
40	trace numérique	N+Adj	40	classification	N
41	interacteur	N	41	requête	N
42	blogueur	N	42	réseau social numérique	N+Adj+Adj
43	autoformation	N	43	centre d'intérêt	N+Prep+N
44	innovation	N	44	pertinence	N
45	informations personnelles	N+Adj	45	confidentialité	N
46	discours dominant	N+Adj	46	cluster	N



47	pratiques transplateformes	N+Adj	47	propagation	N
48	interagir	V	48	centralité	N
49	communicationnel	Adj	49	blog	N
50	privacy	N	50	tranche temporelle	N+Adj
51	podcast	N	51	détecter	V
52	ligne	N	52	déecté	PPAdj
53	interface	N	53	paramètre	N
54	visibilité	N	54	apprentissage automatique	N+Adj
55	Internet	N	55	clustering	N
56	URL	N	56	média	N
57	mode de réception	N+Prep+N	57	fonctionnalité	N
58	trace	N	58	détection de communauté	N+Prep+N
59	relationniste	N	59	réseau égocentrique	N+Adj
60	marque employeur	N+N	60	visualisation	N
61	informationnel	Adj	61	réinformation	N
62	podcaster	N	62	mot principal	N+Adj
63	conversationnel	Adj	63	annotateur	N
64	réseautage social	N+Adj	64	détection d'événements	N+Prep+N
65	identité professionnelle numérique	N+Adj+Adj	65	retweet	N
66	LinkedIn	N	66	collaboratif	Adj
67	relation	N	67	permission	N
68	youtuber	N	68	vecteur	N
69	cyberactiviste	N	69	besoin d'information	N+Prep+N
70	parrhésie	N	70	tâche	N
71	page Facebook	N+N	71	filtrage collaboratif	N+Adj
72	utilisateur	Adj	72	thématique	N
73	interaction sociale	N+Adj	73	bigramme	N
74	communication institutionnelle	N+Adj	74	collection de test	N+Prep+N
75	recrutement	N	75	pondération	N
76	micromessage	N	76	égocentrique	Adj
77	affect	N	77	topique	N
78	outil de communication	N+Prep+N	78	homophilie	N
79	fonctionnalité	N	79	F-mesure	N
80	Web social	N+Adj	80	analyse de sentiment	N+Prep+N
81	forumier	N	81	magnitude d'impact	N+Prep+N
82	site d'actualité	N+Prep+N	82	co-auteur	N
83	SNS	N	83	classifieur	N
84	stratégie de communication	N+Prep+N	84	Facebook	N
85	site	N	85	n-grammes	N
86	message	N	86	émoticône	N
87	réseau virtuel	N+Adj	87	URL	N
88	cyberbullying	N	88	interface	N
89	RSNP	N	89	profil social	N
90	plateforme de réseau	N+Prep+N	90	item	N
91	espace numérique	N+Adj	91	lien	N
92	relationnel	Adj	92	réseautage social	N+Adj
93	communiquer	V	93	résumé automatique	N+Adj
94	vie privée	N+Adj	94	graphe social	N+Adj
95	identité numérique	N+Adj	95	mot lié	N+Adj
96	nom de domaine	N+Prep+N	96	hyperlien	N
97	infomédiation	N	97	réplication	N
98	YouTube	N	98	dérivation	N

99	mise à jour	N+Prep+N	99	recommandation	N
100	bidirectionnelle	Adj	100	événement redondant	N+Adj
101	dévoilement	N	101	dimension utilisateur	N+N
102	publicité	N	102	information sociale	N+Adj
103	inclusif	Adj	103	site de réseautage	N+Prep+N
104	capture d'écran	N+Prep+N	104	TAL	N
105	lien	N	105	topique discursif	N+Adj
106	virtuel	Adj	106	discursif	Adj
107	interpersonnel	Adj	107	redondant	Adj
108	donnée	N	108	booléen	Adj
109	sujet éthique	N+Adj	109	conversationnel	Adj
110	politique d'utilisation	N+Prep+N	110	macro-précision	N
111	publication	N	111	flux de message	N+Prep+N
112	photo de profil	N+Prep+N	112	authentification	N
113	communauté Web	N+N	113	matrice	N
114	autodidaxie	N	114	média traditionnel	N+Adj
115	plateforme de réseau social	N+Prep+N+Adj	115	vectorel	Adj
116	gestionnaire de communauté	N+Prep+N	116	node	N
117	parrhésiste	N	117	lien faible	N+Adj
118	réseau social traditionnel	N+Adj+Adj	118	mesure de centralité	N+Prep+N
119	curation	N	119	hameçonnage	N
120	appropriation	N	120	moniteur positif	N+Adj
121	site d'information	N+Prep+N	121	prédiction	N
122	recherche	N	122	imitation	N
123	publicité trompeuse	N+Adj	123	prescription	N
124	partage	N	124	routage	N
125	réseau social professionnel	N+Adj+Adj	125	page web	N+N
126	relation virtuelle	N+Adj	126	Facebook	Adj
127	Twitter	N	127	mesure de similarité	N+Prep+N
128	site de réseautage	N+Prep+N	128	modifieur	N
129	follower	N	129	chaîne mémétique	N+Adj
130	outil de publication	N+Prep+N	130	mise à jour	N+Prep+N
131	networking	N	131	extraction	N
132	proxys	N	132	minimisation	N
133	sociabilité	N	133	application sociale	N+Adj
134	site Web	N+N	134	Twitter	N
135	connecté	PPAdj	135	pertinence système	N+N
136	échange	N	136	plate-forme sociale	N+Adj
137	sujet d'actualité	N+Prep+N	137	recommandation sociale	N+Adj
138	média traditionnel	N+Adj	138	poids de pertinence	N+Prep+N
139	référencement	N	139	sous-graphe	N
140	communication interpersonnelle	N+Adj	140	identification	N
141	transmédiatique	Adj	141	données personnelles	N+Adj
142	organisation émettrice	N+Adj	142	saillant	Adj
143	schéma publicitaire	N+Adj	143	fréquence quotidienne	N+Adj
144	compte utilisateur	N+Adj	144	framework	N
145	vidéoblogueuse	N	145	vecteur de support	N+Prep+N
146	outil de discussion	N+Prep+N	146	utilisateur actif	N+Adj
147	binette	N	147	énonciatif	Adj
148	contact	N	148	site de réseautage social	N+Prep+N+Adj
149	communauté	N	149	réseau socionumérique	N+Adj
150	page	N	150	système de recommandation sociale	N+Prep+N+Adj

Après avoir analysé le contenu de ces deux listes, nous remarquons que beaucoup de termes semblent appartenir directement au domaine des médias sociaux : *réseau socionumérique, hashtag, émoticône, Facebook*, etc. Cependant, la majorité des termes semble plutôt appartenir aux disciplines de la communication et de l'informatique : *algorithme, schéma publicitaire, vectoriel, relationniste*, etc. Ces premiers résultats nous ont quelque peu surpris, car nous nous attendions à ce que les résultats montrent plus de termes des médias sociaux.

#### **4.1.1 Les termes récurrents dans les deux corpus**

Le Tableau 8 *Les termes récurrents dans COMMU et INFOR* est une version abrégée du tableau présenté à la section 4.1. Il répertorie 33 termes qui sont présents dans les deux listes des 150 termes les plus spécifiques. Cela représente 22 % des termes.

Tableau 8 Les termes récurrents dans *COMMU* et *INFOR*

COMMU			INFOR		
Rang	Terme	PDD/SN	Rang	Terme	PDD/SN
1	média social	N+Adj	1	utilisateur	N
2	réseau social	N+Adj	2	réseau social	N+Adj
3	média	N	3	tweet	N
5	social	Adj	6	réseau	N
6	Web	N	7	média social	N+Adj
7	utilisateur	N	8	profil	N
8	réseau	N	9	hashtag	N
9	tweet	N	10	donnée	N
11	plateforme	N	11	utilisateur	Adj
14	twitter	V	12	social	Adj
17	contenu	N	17	information	N
19	Facebook	N	18	twitter	V
23	réseautage	N	24	plateforme	N
25	discursif	Adj	28	contenu	N
26	hashtag	N	31	Web	N
29	blog	N	32	interaction sociale	N+Adj
31	profil	N	34	réseautage	N
33	information	N	42	réseau social numérique	N+Adj+Adj
35	réseau socionumérique	N+Adj	49	blog	N
36	réseau social numérique	N+Adj+Adj	56	média	N
53	interface	N	57	fonctionnalité	N
56	URL	N	84	Facebook	N
63	conversationnel	Adj	87	URL	N
64	réseautage social	N+Adj	88	interface	N
72	utilisateur	Adj	91	lien	N
73	interaction sociale	N+Adj	92	réseautage social	N+Adj
79	fonctionnalité	N	103	site de réseautage	N+Prep+N
99	mise à jour	N+Prep N	106	discursif	Adj
105	lien	N	109	conversationnel	Adj
108	donnée	N	114	média traditionnel	N+Adj
127	Twitter	N	130	mise à jour	N+Prep+N
128	site de réseautage	N+Prep+N	134	Twitter	N
138	média traditionnel	N+Adj	149	réseau socionumérique	N+Adj

Nous remarquons que ces termes sont pour la plupart des termes des médias sociaux plus généraux et ne semblent pas nécessairement appartenir à la communication ou à l'informatique en particulier. Autrement dit, notre comparaison montre que certains termes sont inhérents à la discipline académique. Évidemment, les vedettes *média social* et *réseau social* font partie de ce groupe de termes. Une panoplie de termes quasi-synonymes est aussi présente : *réseau socionumérique*, *réseau social numérique* et *réseautage social*.

On peut remarquer que *réseau social* se positionne au deuxième rang dans les deux listes. Nous pensons que ce terme est sans doute l'un des plus importants du domaine. D'un autre côté, *média* est classé beaucoup plus haut dans le corpus de la communication que dans celui de l'informatique ; nous ne pensons pas que ce soit anodin. Nous posons l'hypothèse qu'en communication, on traite beaucoup plus des médias qu'en informatique et que les informaticiens se concentrent davantage sur des réseaux informatiques. Cette observation trouve une confirmation dans le fait que, dans la liste, *réseau social* a un rang plus bas que *média social* en informatique et vice versa pour la communication.

#### 4.1.2 Les termes présents dans un seul des deux corpus

Si 33 termes sont récurrents dans les deux corpus, ce sont 122 termes par corpus, donc 244 en tout, qui sont uniques. Nous pouvons nous demander pourquoi un si grand nombre de termes ne sont pas partagés dans des corpus traitant du même thème.

Tout d'abord, nous pouvons dire que certains termes apparaissent dans les deux corpus, mais à des rangs beaucoup plus éloignés, ce qui fait en sorte qu'ils n'ont pas été intégrés dans la liste des 150 termes les plus spécifiques. Par exemple *publication*, qui est présent dans la liste des 150 termes de COMMU, mais absent de celle de INFOR. Pourtant, *publication* est relevé par TermoStat dans INFOR, mais son score de spécificité de 16,21, nettement inférieur à son score de 31,01 dans COMMU, ne lui permet pas de figurer dans la liste des 150 termes de INFOR. C'est le cas également pour *retweet*, qui n'apparaît pas dans la liste de COMMU avec son score de spécificité de 18,98. Il fait pourtant partie de la liste de INFOR avec un score de 52,56.

En ce concerne les autres cas, nous pouvons émettre l'hypothèse que ces termes sont uniques à leur liste, car ils n'appartiennent véritablement qu'à la communication ou à l'informatique. Nous avons identifié quelques exemples que nous commentons ci-dessous.

Prenons d'abord le terme *ethos* (Tableau 7) qui n'est retrouvé que dans le corpus COMMU. Ce terme ne semble pas appartenir au domaine de l'informatique si nous nous fions à l'un de ses contextes dans COMMU : « Considérant que l'*ethos* constitue "une présentation

de soi au service du but recherché dans l'échange", Amossy (2010, p. 31) souligne que celui-ci s'élabore en interaction avec un ethos préalable » (Dupont et Perticoz, 2016, p. 2).

On peut dire la même chose du terme *centralité* (Tableau 7), qui, lui, n'appartient qu'à INFOR. Tout comme dans l'exemple précédent, en nous fiant à l'un des contextes de *centralité* repéré dans INFOR, nous pouvons présumer que ce terme ne relève pas de la communication : « La centralité représente la capacité à bloquer ou laisser filtrer l'accès à certaines parties du réseau » (Renaud, 2014, p. 94).

Les termes présents dans un seul corpus constituent la majorité des termes les plus spécifiques extraits par TernoStat. Nous ne nous attendions pas à une si grande disparité dans les résultats entre les deux corpus. Il semble bien que le point de vue emprunté pour parler des médias sociaux ait un impact majeur sur les résultats terminologiques de l'extraction. Du point de vue de la recherche en terminologie, nous pouvons maintenant supposer que si un thème (comme celui des médias sociaux) est abordé par plusieurs communautés d'experts (communication et informatique), la terminologie la plus spécifique de ce thème fera surface dans l'extraction (les 33 termes récurrents), mais la terminologie propre à ces communautés d'experts prendra une place toute aussi importante, sinon plus grande comme dans le cas de cette étude (les 122 termes uniques). Notre comparaison tend à indiquer que les thèmes à l'intersection de plusieurs disciplines sont bel et bien abordés avec une terminologie différente, selon le champ d'études. Elle montre également qu'ils possèdent un tronc commun qui est partagé par les différentes communautés d'experts.

#### **4.1.3 Les termes concurrents**

Il est intéressant de constater que certaines unités quasi synonymiques apparaissent dans le recensement de la terminologie. Ces quasi-synonymes peuvent indiquer la présence de CN, car cela voudrait dire que la dénomination de ces concepts ne fait pas consensus. Nous pouvons dire que ces termes peuvent être en *concurrency*. La concurrence peut être causée par le reflet de certains choix de l'auteur d'un texte (l'auteur ne veut pas répéter le même terme trop souvent par souci stylistique). Les quasi-synonymes présents dans les corpus différents, toutefois,

peuvent révéler que les experts de ces deux disciplines utilisent des termes concurrents lorsqu'ils écrivent sur les médias sociaux. Nous analysons ces termes dans les sous-sections suivantes.

#### 4.1.3.1 Les termes concurrents dans le même corpus

Comme nous venons de le mentionner, certains concepts sont exprimés sous différentes formes que nous considérons comme quasi synonymiques. Cette quasi-synonymie a été établie au moyen des contextes. Nous les présentons dans le Tableau 9.

Tableau 9 Les termes quasi synonymiques de COMMU

Les termes quasi synonymiques de COMMU			
Rang	Terme quasi synonymique A	Terme quasi synonymique B	Rang
6	<i>Web</i>	<i>Internet</i>	55
7	<i>utilisateur</i>	<i>usager</i>	13
22	<i>usage</i>	<i>utilisation</i>	39
23	<i>réseautage</i>	<i>networking</i>	131
43	<i>autoformation</i>	<i>autodidaxie</i>	114
50	<i>privacy</i>	<i>vie privée</i>	94
82	<i>site d'actualité</i>	<i>site d'information</i>	121

Les *quasi-synonymes A* sont ceux qui apparaissent plus haut la liste de termes : il s'agit donc des termes les plus spécifiques. Ce qui est étonnant dans ces résultats, c'est que le terme *usager* n'est pas plus spécifique que *utilisateur*, alors que *usage*, un mot de la même famille, est plus spécifique que le terme *utilisation*, appartenant à la même famille que *utilisateur*. Nous notons également que l'utilisation du terme d'origine anglaise *privacy* est plus importante que celle de son correspondant français *vie privée* ; à l'inverse, *networking* est moins spécifique que son quasi-synonyme *réseautage*.

Nous avons aussi réuni ces termes :

- *média social* ;
- *réseau social* ;
- *média socionumérique* ;
- *réseau socionumérique* ;
- *réseau social numérique*.

Nous avons pu constater dans les contextes que les termes *média social* et *réseau social* sont souvent interchangeables, sans qu'il n'y ait une distinction entre les deux. Pourtant, le terme *réseau social* est un hyponyme de *média social*. Il est donc juste de parler de *média social* dans tous les cas où l'on parle de *réseau social*, mais il est incorrect de faire l'inverse.

Nous constatons de plus que l'utilisation de l'adjectif *socionumérique* n'est pas facilement prévisible, comme si son emploi est parallèle à celui de l'adjectif *social* dans ce contexte. Qui plus est, l'adjectif *socionumérique* est également en concurrence avec le double adjectif *social numérique*. Les deux ont le même sens, mais nous pensons que le composé *socionumérique* pourrait être un néologisme. Notons que le terme *média social numérique* n'apparaît pas dans notre liste de 150 termes de COMMU, mais il a bel et bien été relevé par TermoStat plus loin dans la liste.

Nous n'avons pas relevé de termes concurrents au sein de INFOR, contrairement à COMMU, sauf la panoplie de termes (*réseau social numérique*, *média socionumérique*, etc.) également présente. Nous en déduisons que, dans le domaine de l'informatique, ces termes peuvent aussi être utilisés comme des quasi-synonymes.

Dans les deux corpus se trouve le terme *média traditionnel*. Les deux listes semblent indiquer que ce terme s'est implanté pour désigner quelque chose de différent de *média social*. Il semble être utilisé en contexte comme un antonyme de *média social*. Ce terme nous semble avoir un caractère néologique.



#### 4.1.3.2 Les termes concurrents dans des corpus différents

Les résultats montrent également des termes qui sont concurrents entre COMMU et INFOR. Les résultats sont peu nombreux. Le Tableau 10 répertorie ces termes, ainsi que leur rang.

Tableau 10 *Les termes quasi synonymiques de COMMU et INFOR*

Les termes quasi synonymiques de COMMU et INFOR			
Quasi-synonyme A : COMMU		Quasi-synonyme B : INFOR	
Rang	Terme	Terme	Rang
134	<i>site Web</i>	<i>page Web</i>	125
147	<i>binette</i>	<i>émoticône</i>	86

Commençons par l'analyse de *site Web/page Web*. Nous remarquons que ces deux termes sont utilisés de manière quasi synonymique dans les contextes. En nous fiant à nos résultats, nous supposons que les experts en communication et les experts en informatique ont privilégié un de ces quasi-synonymes lorsqu'ils rédigent des textes.

Pour ce qui est de *binette/émoticône*, leurs occurrences ne sont pas très élevées (neuf occurrences de *binette* dans COMMU contre 19 occurrences de *émoticône* dans INFOR). Toutes les occurrences de *binette* proviennent du même texte, alors que les occurrences de *émoticône* proviennent de neuf textes.

#### 4.1.4 La répartition des parties du discours et des syntagmes nominaux des termes de COMMU et INFOR

Comme nous pouvons le constater, la majorité des termes répertoriés dans les deux corpus sont des noms ou des termes complexes ayant pour tête un nom. Ces noms sont soit des noms d'entités (ex. : *tweet*) ou des noms d'action (ex. : *catégorisation*) (Tableau 7).

La deuxième catégorie grammaticale la plus présente dans les données est celle de l'adjectif. Les verbes sont presque absents des données, alors que les adverbes sont totalement absents (lors de la soumission des corpus à TermoStat, nous avons sélectionné des paramètres afin d'extraire des *noms*, *verbes*, *adverbes* et *adjectifs*, voir à la section 3.3.2).

Nous avons réparti les parties du discours pour chaque liste dans les graphiques de la Figure 7 et de la Figure 8.

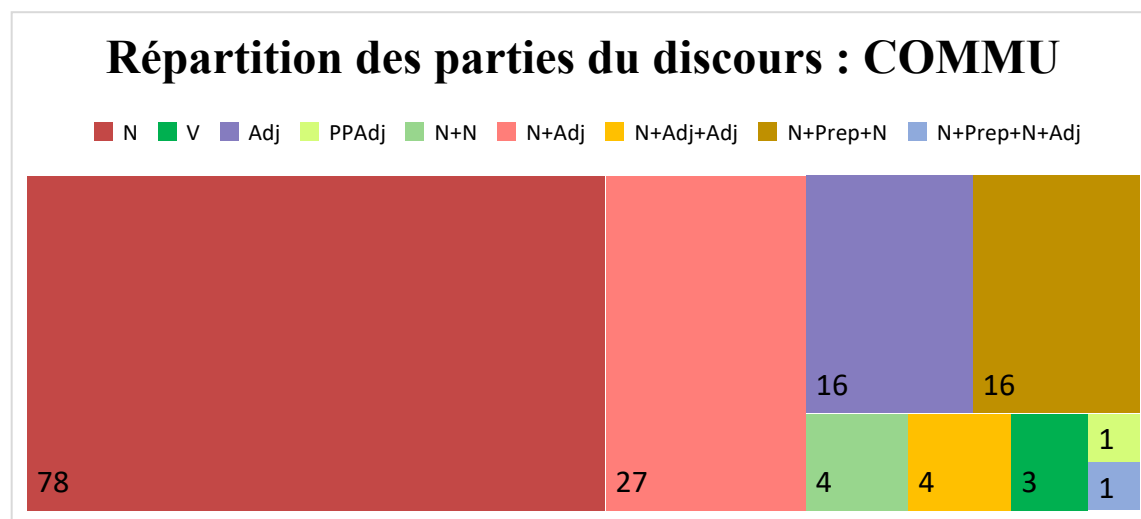


Figure 7 La répartition des parties du discours et des syntagmes nominaux dans COMMU

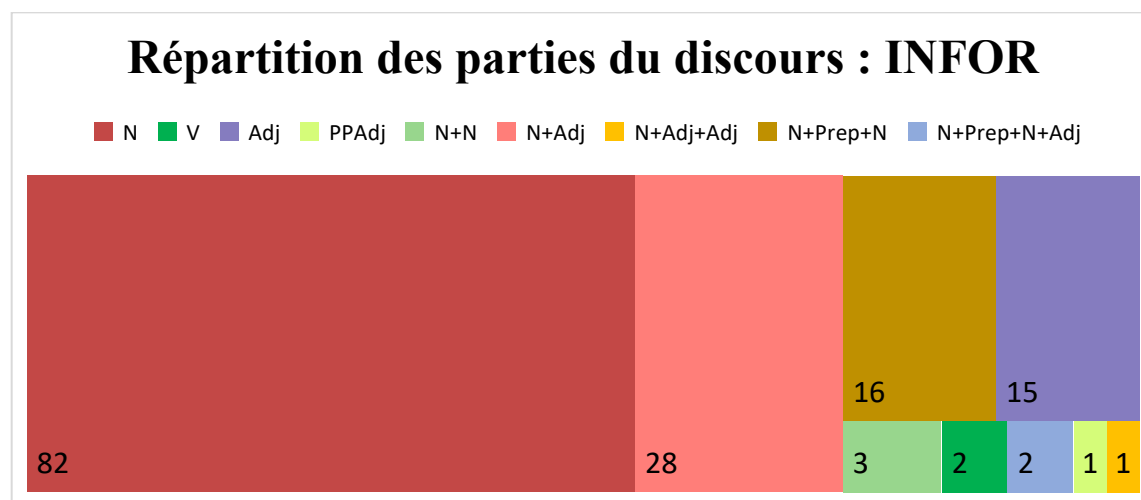


Figure 8 La répartition des parties du discours et des syntagmes nominaux dans INFOR

Comme nous pouvons le constater dans les graphiques de la Figure 7 et de la Figure 8, les résultats sont très semblables entre COMMU et INFOR. Dans les deux listes, nous trouvons les mêmes neuf parties du discours ou syntagmes nominaux :

Termes simples :

- Nom ;
- Verbe ;
- Adjectif ;
- Participe passé adjectival.

Termes complexes :

- Nom + Nom ;
- Nom + Adjectif ;
- Nom + Adjectif + Adjectif ;
- Nom + Préposition + Nom ;
- Nom + Préposition + Nom + Adjectif.

Les N représentent un peu plus de la moitié des termes dans les deux listes. Nous constatons que les PPAAdj et les N + Prep + N ont la même répartition dans les deux listes. Les Adj, les V, les N + N, les N + Adj et les N + Prep + N + Adj présentent tous une ou deux occurrences d'écart entre les deux listes. La donnée la plus marquante est celle des N + Adj + Adj : on en trouve quatre fois moins dans la liste INFOR que dans la liste COMMU. Toutefois les occurrences étaient déjà très faibles : une occurrence dans INFOR contre quatre dans COMMU.

Ces résultats semblent indiquer que la communication et l'informatique utilisent à peu près les mêmes parties du discours et syntagmes nominaux dans la terminologie des médias sociaux.

#### 4.1.5 La répartition des termes simples et complexes de COMMU et INFOR

Ensuite, nous avons observé la répartition des termes simples en la comparant à celle des termes complexes. Nous avons organisé les données dans les graphiques de la Figure 9 et de la Figure 10.

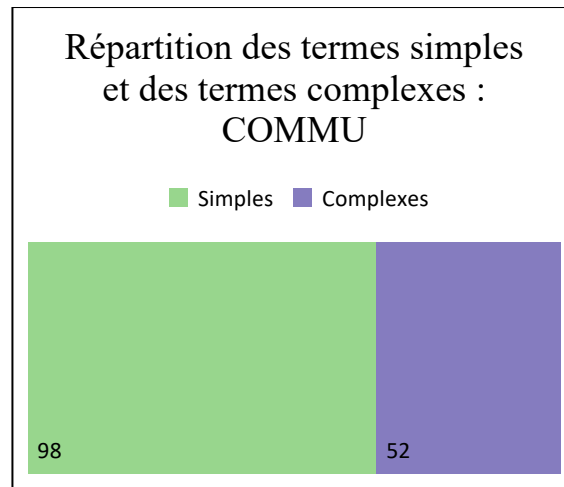


Figure 9 *La répartition des termes simples et complexes : COMMU*

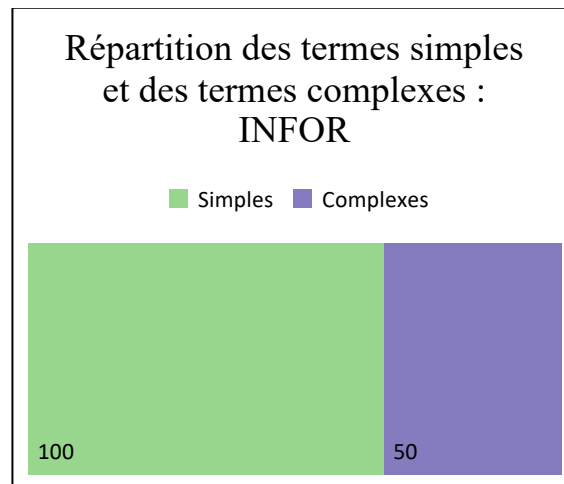


Figure 10 *La répartition des termes simples et complexes : INFOR*

Tout comme la répartition des parties du discours et des types de syntagmes nominaux, les données sont très similaires entre les deux listes. Nous constatons que près des deux tiers des termes sont des termes simples dans COMMU. Cette répartition est la même dans INFOR. Il est

toutefois important de mentionner que les termes complexes que nous avons relevés représentent tout de même une bonne partie de la terminologie des médias sociaux.

#### 4.1.6 L'influence des codes de recherches sur les résultats

Comme nous l'avons mentionné à la section 3.2.1, nous avons utilisé des codes de recherche pour cette étape de l'étude. Nous pensons que les termes faisant l'objet de codes de recherche sont plus susceptibles de se retrouver dans la liste de termes. Nous avons donc répertorié dans le Tableau 11 et dans le Tableau 12 les termes présents dans les deux listes de 150 termes qui ont fait l'objet d'un code de recherche lors du montage des corpus. Nous croyons également que cela peut expliquer pourquoi ils obtiennent un score de spécificité assez élevé pour faire partie des deux listes des 150 termes les plus spécifiques.

Tableau 11 *Les termes de COMMU utilisés dans les codes de recherche*

Termes présents dans la liste des 150 termes de COMMU et dans un code de recherche		
médias sociaux	communication	youtube
twitter	linkedin	

Tableau 12 *Les termes de INFOR utilisés dans les codes de recherche*

Termes présents dans la liste des 150 termes de INFOR et dans un code de recherche	
médias sociaux	booléen
algorithme	facebook

## 4.2 La validation des CN par les méthodes de filtrage

Les (300) termes des deux listes de COMMU et de INFOR sont maintenant considérés comme des CN. Les prochaines étapes servent à repérer les néologismes véritables et à exclure les termes qui ne sont pas des néologismes. Tous les CN sont soumis au corpus d'exclusion. Les sections suivantes présentent les résultats obtenus après avoir soumis les CN à trois méthodes de filtrage. Les CN qui ne seront pas exclus au terme de ces trois étapes seront considérés comme des néologismes. Les tests sont, dans cet ordre, la soumission aux dictionnaires spécialisés (4.2.1), le filtrage au moyen de l'outil de visualisation Google Books Ngram Viewer (Google, 2013) (4.2.2) et le repérage dans les dictionnaires de langue générale (4.2.3).

### 4.2.1 La soumission des CN aux dictionnaires spécialisés

Le Tableau 13 et le Tableau 14 présentent les résultats du filtrage des CN au moyen des dictionnaires spécialisés. Nous avons séparé les résultats pour COMMU et INFOR.

Nous donnons, de gauche à droite dans les tableaux, le rang, le candidat-néologisme et la partie du discours. Les six dernières colonnes sont réservées aux dictionnaires. Ce sont, dans cet ordre, le *Dictionnaire informatique & numérique* (informatique, DIN), le *Lexicom* (communication, L), le *Dictionnaire du journalisme et des médias* (communication, DJM), le *Lexique des médias sociaux* (médias sociaux, LMS), le *Vocabulaire des médias sociaux* (médias sociaux, VRS) et le *Dictionnaire de l'informatique et d'internet* (informatique, DII).

Le (x) indique la présence du terme dans le dictionnaire.

Les termes en blanc sont les termes absents des dictionnaires.

Les termes en noir (*facebook* (Adj) et *plateforme de réseau* (N+Prep+N)) sont des CN qui ont été rejetés à cette étape. Ils auraient dû être rejetés à l'étape de l'élagage (3.3.3.2). *Facebook* a été étiqueté comme un adjectif par le logiciel, alors qu'il agit comme un nom dans les contextes. Le terme *plateforme de réseau* a lui aussi été mal repéré. En parcourant les contextes, nous nous sommes rendu compte qu'il est toujours suivi de l'adjectif *social*. Le terme *plateforme de réseau social* fait déjà partie de la liste (Tableau 7).

Les CN de couleurs sont présents dans au moins un dictionnaire, donc ne seront pas traités comme néologismes. Plus la teinte d'un CN est foncée, plus nombreux sont les dictionnaires dans lesquels il a été repéré (il y a six teintes en tout).

Tableau 13 La soumission des CN de COMMU aux dictionnaires spécialisés

Soumission aux dictionnaires spécialisés : COMMU								
Rang	CN	PDD/SN	DIN	L	DJM	LMS	VRS	DII
1	média social	N+Adj		x		x	x	
2	réseau social	N+Adj	x	x	x	x	x	x
3	média	N		x	x			x
4	vidéo	N		x		x		
5	social	Adj	x					
6	Web	N	x		x			x
7	utilisateur	N				x		x
8	réseau	N	x	x	x	x		x
9	tweet	N	x		x			x
10	interaction	N			x	x		
11	plateforme	N	x	x	x	x		
12	communication	N		x	x			
13	usager	N						
14	twitter	V						
15	publicité rédactionnelle	N+Adj		x				
16	socionumérique	Adj						
17	contenu	N			x	x		
18	internaute	N	x		x	x		x
19	Facebook	N	x		x			x
20	ethos	N			x			
21	outil	N						
22	usage	N						
23	réseautage	N					x	
24	fictionnel	Adj						
25	discursif	Adj						
26	hashtag	N	x	x	x			x
27	numérique	Adj	x	x				x
28	rédactionnel	Adj		x	x			
29	blog	N	x	x	x	x	x	x
30	commentaire	N			x	x	x	
31	profil	N					x	x
32	média socionumérique	N+Adj						
33	information	N			x			x
34	usage professionnel	N+Adj						
35	réseau socionumérique	N+Adj						
36	réseau social numérique	N+Adj+Adj						
37	identité professionnelle	N+Adj						
38	Instagram	N						x
39	utilisation	N						x
40	trace numérique	N+Adj						
41	interacteur	N						
42	blogueur	N			x	x		

43	autoformation	N						
44	innovation	N						
45	informations personnelles	N+Adj						
46	discours dominant	N+Adj						
47	pratiques transplateformes	N+Adj						
48	interagir	V				x		
49	communicationnel	Adj						
50	privacy	N			x			
51	podcast	N	x	x	x			
52	ligne	N	x					
53	interface	N	x					x
54	visibilité	N		x				
55	Internet	N	x	x	x			x
56	URL	N	x		x			x
57	mode de réception	N+Prep+N						
58	trace	N			x			
59	relationniste	N						
60	marque employeur	N+N						
61	informationnel	Adj						
62	podcaster	N						
63	conversationnel	Adj						
64	réseautage social	N+Adj				x		
65	identité professionnelle numérique	N+Adj+Adj						
66	LinkedIn	N						x
67	relation	N				x		
68	youtuber	N						
69	cyberactiviste	N						
70	parrhésie	N						
71	page Facebook	N+N						
72	utilisateur	Adj				x		x
73	interaction sociale	N+Adj						
74	communication institutionnelle	N+Adj		x				
75	recrutement	N		x	x			
76	micromessage	N				x	x	
77	affect	N		x				
78	outil de communication	N+Prep+N						
79	fonctionnalité	N				x		
80	web social	N+Adj					x	
81	forumier	N						
82	site d'actualité	N+Prep+N						
83	SNS	N						
84	stratégie de communication	N+Prep+N						
85	site	N		x	x			x
86	message	N				x		
87	réseau virtuel	N+Adj						
88	cyberbullying	N						
89	RSNP	N						
90	plateforme de réseau	N+Prep+N						
91	espace numérique	N+Adj						
92	relationnel	Adj						x
93	communiquer	V						
94	vie privée	N+Adj				x		



95	identité numérique	N+Adj						
96	nom de domaine	N+Prep+N			x			x
97	infomédiation	N						
98	YouTube	N	x		x			x
99	mise à jour	N+Prep+N		x	x	x		
100	bidirectionnelle	Adj						x
101	dévoilement	N						
102	publicité	N		x	x			
103	inclusif	Adj						
104	capture d'écran	N+Prep+N			x			x
105	lien	N	x		x			
106	virtuel	Adj	x					
107	interpersonnel	Adj						
108	donnée	N			x			x
109	sujet éthique	N+Adj						
110	politique d'utilisation	N+Prep+N						
111	publication	N				x	x	
112	photo de profil	N+Prep+N				x		
113	communauté Web	N+N				x		
114	autodidaxie	N						
115	plateforme de réseau social	N+Prep+N+Adj						
116	gestionnaire de communauté	N+Prep+N					x	
117	parrhésiasite	N						
118	réseau social traditionnel	N+Adj+Adj						
119	curation	N		x				
120	appropriation	N						
121	site d'information	N+Prep+N						
122	recherche	N		x				
123	publicité trompeuse	N+Adj						
124	partage	N				x		
125	réseau social professionnel	N+Adj+Adj				x		
126	relation virtuelle	N+Adj			x			
127	Twitter	N	x					x
128	site de réseautage	N+Prep+N						
129	follower	N	x		x			
130	outil de publication	N+Prep+N						
131	networking	N						
132	proxys	N	x					x
133	sociabilité	N						
134	site web	N+N		x		x		
135	connecté	PPAdj	x					
136	échange	N						
137	sujet d'actualité	N+Prep+N						
138	média traditionnel	N+Adj						
139	référencement	N	x	x	x			
140	communication interpersonnelle	N+Adj						
141	transmédiatique	Adj						
142	organisation émettrice	N+Adj						
143	schéma publicitaire	N+Adj						
144	compte utilisateur	N+Adj	x		x		x	
145	vidéoblogueuse	N						
146	outil de discussion	N+Prep+N						

147	binette	N					x	
148	contact	N	x			x		
149	communauté	N				x		
150	page	N	x			x		

Tableau 14 *La soumission des CN de INFOR aux dictionnaires spécialisés*

Soumission aux dictionnaires spécialisés : INFOR								
Rang	Terme	PDD/SN	DIN	L	DJM	LMS	VRS	DII
1	utilisateur	N				x		x
2	réseau social	N+Adj	x	x	x	x	x	x
3	tweet	N	x		x			x
4	algorithme	N	x	x				x
5	graphe	N						
6	réseau	N		x	x	x		x
7	média social	N+Adj		x		x	x	
8	profil	N					x	x
9	hashtag	N	x	x	x			x
10	donnée	N	x		x			x
11	utilisateur	Adj						
12	social	Adj	x					
13	détection	N						
14	temporel	Adj						
15	similarité	N						
16	jeu de données	N+Prep+N						
17	information	N			x			x
18	twitter	V						
19	événement	N			x	x		x
20	profil utilisateur	N+Adj	x					
21	mème	N				x		
22	descripteur	N						
23	filtrage	N						x
24	plateforme	N	x	x	x	x		
25	application	N	x	x	x	x		x
26	polarité	N						
27	contextuel	Adj						
28	contenu	N				x		
29	filtrage social	N+Adj						
30	système de recommandation	N+Prep+N						
31	Web	N	x		x			x
32	interaction sociale	N+Adj						
33	mésusage	N						
34	réseautage	N					x	
35	pertinent	N						
36	catégorisation	N						
37	annotation	N						x
38	textuel	Adj						
39	itération	N	x					x
40	classification	N						
41	requête	N						x

42	réseau social numérique	N+Adj+Adj						
43	centre d'intérêt	N+Prep+N						
44	pertinence	N						x
45	confidentialité	N				x	x	
46	cluster	N						x
47	propagation	N						x
48	centralité	N						
49	blog	N	x	x		x	x	x
50	tranche temporelle	N+Adj						
51	détecter	V						
52	déecté	PPAdj						
53	paramètre	N				x	x	x
54	apprentissage automatique	N+Adj						
55	clustering	N						
56	média	N	x	x				x
57	fonctionnalité	N				x		
58	détection de communauté	N+Prep+N						
59	réseau égocentrique	N+Adj						
60	visualisation	N						
61	réinformation	N	x					
62	mot principal	N+Adj						
63	annotateur	N						
64	détection d'événements	N+Prep+N						
65	retweet	N				x		
66	collaboratif	Adj						
67	permission	N						x
68	vecteur	N						x
69	besoin d'information	N+Prep+N						
70	tâche	N						x
71	filtrage collaboratif	N+Adj						
72	thématique	N						
73	bigramme	N						
74	collection de test	N+Prep+N						
75	pondération	N			x			
76	égocentrique	Adj						
77	topique	N						
78	homophilie	N						
79	F-mesure	N						
80	analyse de sentiment	N+Prep+N						
81	magnitude d'impact	N+Prep+N						
82	co-auteur (coauteur)	N				x		
83	classifieur	N						
84	Facebook	N	x			x		x
85	n-grammes	N						
86	émoticône	N	x			x		x
87	URL	N	x			x		x
88	interface	N	x					x
89	profil social	N						
90	item	N	x	x				
91	lien	N	x			x		
92	réseautage social	N+Adj					x	
93	résumé automatique	N+Adj						

94	graphe social	N+Adj						
95	mot lié	N+Adj						
96	hyperlien	N	x				x	x
97	réplication	N						x
98	dérivation	N						
99	recommandation	N						
100	évènement redondant	N+Adj						
101	dimension utilisateur	N+N						
102	information sociale	N+Adj						
103	site de réseautage	N+Prep+N						
104	TAL	N						
105	topique discursif	N+Adj						
106	discursif	Adj						
107	redondant	Adj						
108	booléen	Adj	x					x
109	conversationnel	Adj						
110	macro-précision	N						
111	flux de message	N+Prep+N						
112	authentification	N						x
113	matrice	N	x	x				x
114	média traditionnel	N+Adj						
115	vectorel	Adj						x
116	node	N						
117	lien faible	N+Adj						
118	mesure de centralité	N+Prep+N						
119	hameçonnage	N	x		x			x
120	moniteur positif	N+Adj						
121	prédiction	N						
122	imitation	N						
123	prescription	N						
124	routage	N		x				x
125	page Web	N+N				x	x	
126	Facebook	Adj						
127	mesure de similarité	N+Prep+N						
128	modifieur	N						
129	chaîne mémétique	N+Adj						
130	mise à jour	N+Prep+N		x	x	x		
131	extraction	N						
132	minimisation	N						
133	application sociale	N+Adj						
134	Twitter	N	x		x			x
135	pertinence système	N+N						
136	plate-forme sociale	N+Adj						
137	recommandation sociale	N+Adj						
138	poids de pertinence	N+Prep+N						
139	sous-graphe	N						
140	identification	N						x
141	données personnelles	N+Adj						
142	saillant	Adj						
143	fréquence quotidienne	N+Adj						
144	framework	N						
145	vecteur de support	N+Prep+N						

146	utilisateur actif	N+Adj				x		
147	énonciatif	Adj						
148	site de réseautage social	N+Prep+N+Adj						
149	réseau socionumérique	N+Adj						
150	système de recommandation sociale	N+Prep+N+Adj						

Nous pouvons remarquer que les CN (exclus) qui font l'objet d'une entrée dans *tous les dictionnaires* sont peu nombreux. Dans COMMU, ce ne sont que *réseau social* et *blog* ; dans INFOR, ce n'est que *réseau social*. Il se peut que ces termes soient des termes fondamentaux dans la terminologie des médias sociaux.

Nous constatons également que les termes les plus présents dans les nomenclatures sont souvent ceux qui ont obtenu un rang moins élevé. Effectivement, que ce soit pour COMMU ou INFOR, la couleur se manifeste beaucoup plus haut dans le tableau.

Nous émettons l'hypothèse que les CN présents dans trois dictionnaires ou plus sont susceptibles d'être des termes importants dans le domaine des médias sociaux. Ce qui nous pousse à poser cette hypothèse, c'est que ces CN nous semblent, selon notre connaissance du domaine des médias sociaux, être des concepts clés. Il s'agit en majorité de noms : *blogue, Web, hashtag, numérique, Internet, URL, etc.*

#### ***4.2.1.1 La répartition des CN et des exclusions de COMMU et INFOR après la soumission aux dictionnaires spécialisés***

Les termes n'ayant pas été exclus à cette étape par le corpus d'exclusion sont toujours considérés comme des CN pour les prochaines étapes. Les termes qui ont été exclus sont maintenant renommés *exclusions*. Nous constatons que la répartition des CN et des exclusions dans COMMU et INFOR est différente. Pourtant, comme nous l'avons vu précédemment, les résultats entre COMMU et INFOR étaient jusqu'ici assez équilibrés, que ce soit dans la répartition des unités terminologiques simples/complexes ou la répartition des parties du discours. À ce stade, 69 termes seront traités comme CN dans COMMU contre 91 dans INFOR.

La Figure 11 et la Figure 12 sont des graphiques qui montrent la répartition des CN et des exclusions dans COMMU et INFOR. À cette étape, l'addition des données donne 149 termes pour chaque corpus, comme nous avons élagué un terme dans chaque corpus *a posteriori*.

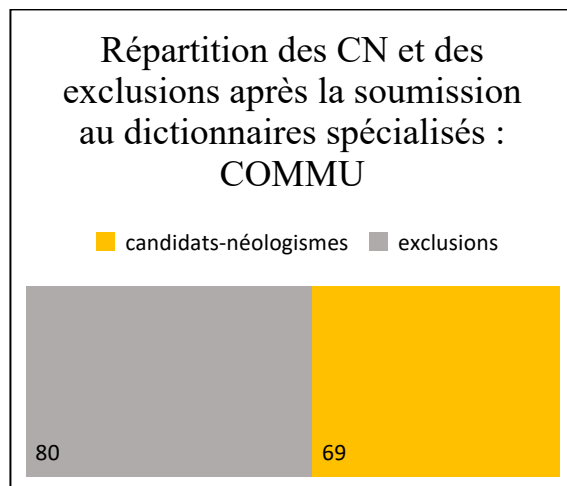


Figure 11 *La répartition des CN et des exclusions de COMMU après la soumission aux dictionnaires spécialisés*

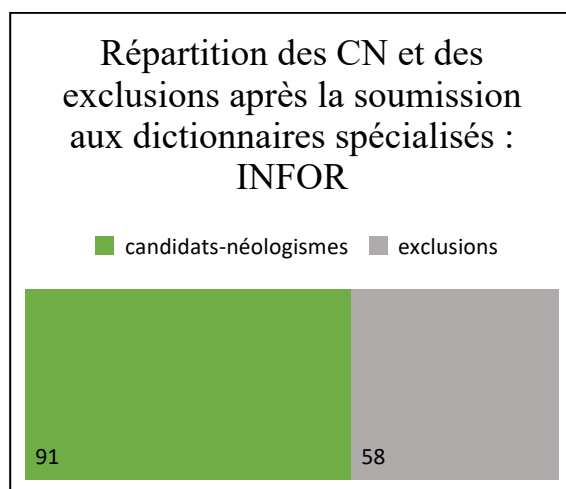


Figure 12 *La répartition des CN et des exclusions de INFOR après la soumission aux dictionnaires spécialisés*

Nous pouvons constater que les CN représentent 46 % des entrées de la liste des 150 termes dans COMMU, ce qui les place derrière les exclusions, qui sont plus nombreuses. Dans INFOR, les CN dépassent les exclusions avec 61 %. Cette répartition pourrait s'expliquer par le fait que les ressources du corpus d'exclusion répertorient plus de termes de la communication

que de l'informatique. Les premiers résultats risquent cependant de changer après la soumission des données aux méthodes subséquentes de validation des néologismes.

#### 4.2.1.2 La répartition des exclusions selon le type de dictionnaire

Nous nous sommes attardé sur résultats du premier test du corpus d'exclusion pour voir la répartition des exclusions selon le type de dictionnaires (dictionnaire de l'informatique, de la communication ou des médias sociaux).

La Figure 13 et la Figure 14 montrent en graphiques la répartition des CN trouvés dans les dictionnaires du corpus d'exclusion selon le type de dictionnaires. Il est à noter que nous comptons toutes les exclusions dans chaque dictionnaire : donc les CN qui sont présents dans plus d'un dictionnaire sont comptés pour *chaque* dictionnaire. Nous voulons voir combien de CN ont été exclus par dictionnaire. Les nombres dans les graphiques représentent les nombres de CN exclus. Les dictionnaires de la communication sont représentés par la couleur jaune, ceux de l'informatique par la couleur verte et ceux des médias sociaux par la couleur bleue.

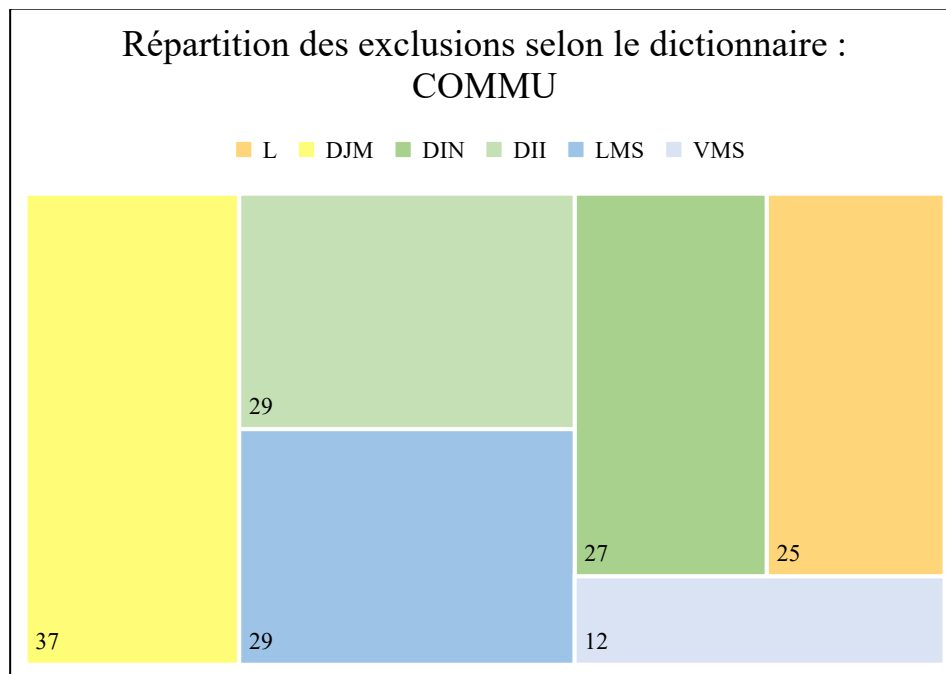


Figure 13 La répartition des exclusions de COMMU selon le dictionnaire

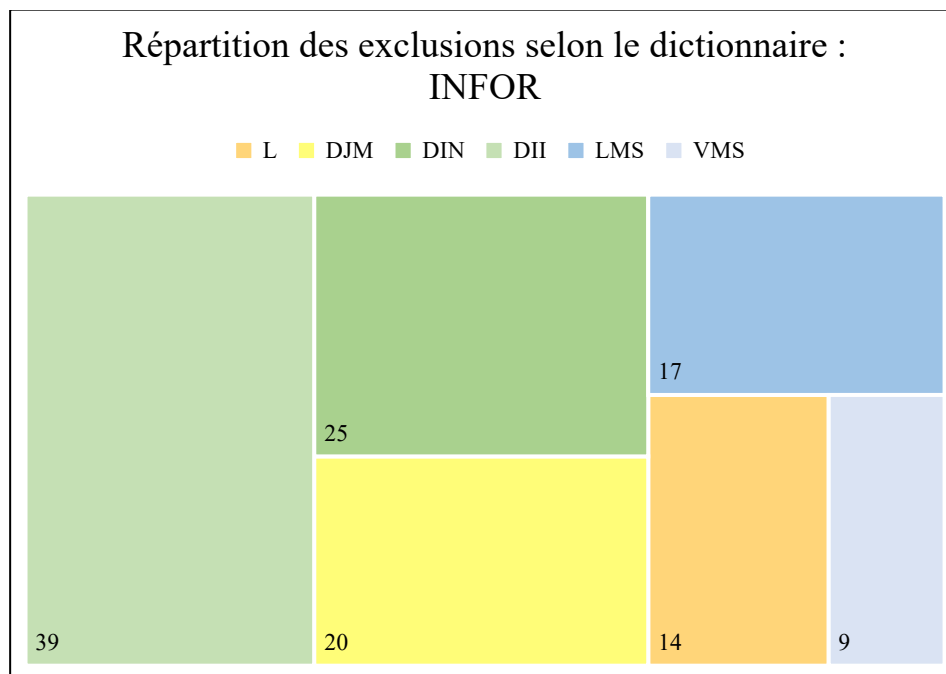


Figure 14 *La répartition des exclusions de INFOR selon le dictionnaire*

Ces graphiques montrent que les dictionnaires de l'informatique se sont montrés plus efficaces pour exclure les CN de COMMU que les dictionnaires de la communication pour exclure les CN de INFOR. En fait, le DIN et le DII (qui sont les dictionnaires de l'informatique) se sont retrouvés à être plus efficaces pour exclure des termes dans COMMU que le L (qui est un dictionnaire de la communication).

Même les dictionnaires des médias sociaux ont été plus utiles dans COMMU que dans INFOR. Cela pourrait expliquer pourquoi il reste plus de CN dans INFOR que dans COMMU à la fin de ce premier test.

#### **4.2.2 La recherche des CN dans l'outil de visualisation de n-grammes**

Le deuxième test de la validation des néologismes est celui de la recherche des CN qui n'ont pas été exclus par les dictionnaires spécialisés (4.2.1) dans l'outil en ligne de visualisation de n-grammes Google Books Ngram Viewer. Pour ce faire, nous avons cherché les 69 CN restants de COMMU et les 91 CN restants de INFOR. Il y a trois CN de COMMU (*twitter* (V),



*SNS* (N) et *RSNP* (N)) et trois CN de INFOR (*utilisateur* (Adj), *twitter* (V) et *TAL* (N)) que nous n'avons pas pu chercher dans la plateforme.

Nous n'avons pas pu chercher *twitter* et *utilisateur* (Adj), car la fonction sur Google Books Ngram Viewer pour rechercher par partie du discours est hors de notre portée. De ce fait, nous ne pouvons pas savoir si c'est le nom *Twitter* ou le verbe *twitter* qui apparaît dans les résultats. Le même problème se pose pour le nom *utilisateur* et l'adjectif *utilisateur*. Pour contourner cette difficulté pour *twitter* (V), nous avons essayé de trouver le verbe sous ses formes fléchies (*twittait, a twitté*), mais le logiciel ne nous a pas fourni de résultats. Par précaution, nous préférons soumettre ces CN à la troisième méthode du filtrage, à savoir les dictionnaires de langue générale.

Toujours par précaution, nous n'avons pas cherché les CN *SNS*, *RSNP* et *TAL*, car ces sigles et cet acronyme peuvent être utilisés dans plusieurs domaines et l'outil de visualisation de n-grammes ne nous permet pas de cibler la communication ou l'informatique. Malheureusement, nous ne pourrions pas non plus soumettre ces CN aux dictionnaires de langue générale, car ceux-ci ne répertorient pas vraiment de sigles ou les acronymes. De ce fait, ces trois CN sont considérés comme des néologismes potentiels si nous respectons l'application de notre méthode, car ils n'ont pas été exclus par le premier test et nous ne pouvons pas les soumettre au deuxième et au troisième. Cela révèle une limite de nos méthodes de filtrage. Pour être validés formellement comme néologismes, certains candidats devraient être soumis à un expert.

Le Tableau 15 et le Tableau 16 montrent les résultats de l'interrogation des 69 CN restants de COMMU et des 91 CN restants de INFOR dans l'outil de visualisation de n-grammes avec les rangs et les parties du discours.

Tableau 15 Liste des CN de COMMU après l'interrogation dans Google Books Ngram Viewer

Liste des CN de COMMU après l'interrogation dans Google Books Ngram Viewer					
Rang	CN	PDD/SN	Rang	CN	PDD/SN
13	usager	N	83	SNS	N
14	twitter	V	84	stratégie de communication	N+Prep+N
16	socionumérique	Adj	87	réseau virtuel	N+Adj
21	outil	N	88	cyberbullying	N
22	usage	N	89	RSNP	N
24	fictionnel	Adj	91	espace numérique	N+Adj
25	discursif	Adj	93	communiquer	V
32	média socionumérique	N+Adj	95	identité numérique	N+Adj
34	usage professionnel	N+Adj	97	infomédiation	N
35	réseau socionumérique	N+Adj	101	dévoilement	N
36	réseau social numérique	N+Adj+Adj	103	inclusif	Adj
37	identité professionnelle	N+Adj	107	interpersonnel	Adj
40	trace numérique	N+Adj	109	sujet éthique	N+Adj
41	interacteur	N	110	politique d'utilisation	N+Prep+N
43	autoformation	N	114	autodidaxie	N
44	innovation	N	115	plateforme de réseau social	N+Prep+N+Adj
45	informations personnelles	N+Adj	117	parrhésiaste	N
46	discours dominant	N+Adj	118	réseau social traditionnel	N+Adj+Adj
47	pratiques transplateformes	N+Adj	120	appropriation	N
49	communicationnel	Adj	121	site d'information	N+Prep+N
57	mode de réception	N+Prep+N	123	publicité trompeuse	N+Adj
59	relationniste	N	128	site de réseautage	N+Prep+N
60	marque employeur	N+N	130	outil de publication	N+Prep+N
61	informationnel	Adj	131	networking	N
62	podcaster	N	133	sociabilité	N
63	conversationnel	Adj	136	échange	N
65	identité professionnelle numérique	N+Adj+Adj	137	sujet d'actualité	N+Prep+N
68	youtuber	N	138	média traditionnel	N+Adj
69	cyberactiviste	N	140	communication interpersonnelle	N+Adj
70	parrhésie	N	141	transmédiatique	Adj
71	page Facebook	N+N	142	organisation émettrice	N+Adj
73	interaction sociale	N+Adj	143	schéma publicitaire	N+Adj
78	outil de communication	N+Prep+N	145	vidéoblogueuse	N
81	forumier	N	146	outil de discussion	N+Prep+N
82	site d'actualité	N+Prep+N			

Tableau 16 Liste des CN de INFOR après l'interrogation dans Google Books Ngram Viewer

Liste des CN de INFOR après l'interrogation dans Google Books Ngram Viewer					
Rang	CN	PDD/SN	Rang	CN	PDD/SN
5	graphe	N	89	profil social	N
11	utilisateur	Adj	93	résumé automatique	N+Adj
13	détection	N	94	graphe social	N+Adj
14	temporel	Adj	95	mot lié	N+Adj
15	similarité	N	98	dérivation	N
16	jeu de données	N+Prep+N	99	recommandation	N
18	twitter	V	100	évènement redondant	N+Adj
22	descripteur	N	101	dimension utilisateur	N+N
26	polarité	N	102	information sociale	N+Adj
27	contextuel	Adj	103	site de réseautage	N+Prep+N
29	filtrage social	N+Adj	104	TAL	N
30	système de recommandation	N+Prep+N	105	topique discursif	N+Adj
32	interaction sociale	N+Adj	106	discursif	Adj
33	mésusage	N	107	redondant	Adj
35	pertinent	N	109	conversationnel	Adj
36	catégorisation	N	110	macro-précision	N
38	textuel	Adj	111	flux de message	N+Prep+N
40	classification	N	114	média traditionnel	N+Adj
42	réseau social numérique	N+Adj+Adj	116	node	N
43	centre d'intérêt	N+Prep+N	117	lien faible	N+Adj
48	centralité	N	118	mesure de centralité	N+Prep+N
50	tranche temporelle	N+Adj	120	moniteur positif	N+Adj
51	détecter	V	121	prédiction	N
52	déTECTÉ	PPAdj	122	imitation	N
54	apprentissage automatique	N+Adj	123	prescription	N
55	clustering	N	127	mesure de similarité	N+Prep+N
58	détection de communauté	N+Prep+N	128	modifieur	N
59	réseau égocentrique	N+Adj	129	chaîne mémétique	N+Adj
60	visualisation	N	131	extraction	N
62	mot principal	N+Adj	132	minimisation	N
63	annotateur	N	133	application sociale	N+Adj
64	détection d'événements	N+Prep+N	135	pertinence système	N+N
66	collaboratif	Adj	136	plate-forme sociale	N+Adj
69	besoin d'information	N+Prep+N	137	recommandation sociale	N+Adj
71	filtrage collaboratif	N+Adj	138	poids de pertinence	N+Prep+N
72	thématique	N	139	sous-graphe	N
73	bigramme	N	141	données personnelles	N+Adj
74	collection de test	N+Prep+N	142	saillant	Adj
76	égocentrique	Adj	143	fréquence quotidienne	N+Adj
77	topique	N	144	framework	N
78	homophilie	N	145	vecteur de support	N+Prep+N
79	F-mesure	N	147	énonciatif	Adj
80	analyse de sentiment	N+Prep+N	148	site de réseautage social	N+Prep+N+Adj
81	magnitude d'impact	N+Prep+N	149	réseau socionumérique	N+Adj
83	classifieur	N	150	système de recommandation sociale	N+Prep+N+Adj
85	n-grammes	N			

Les CN en blanc sont les CN que nous n'avons pas pu chercher sur la plateforme pour les raisons que nous venons d'énumérer.

Les CN en rouge sont les CN qui ont été repérés par l'outil de visualisation de n-grammes avec une courbe stable entre 2000 et 2008 (la période de recherche choisie, voir à la section 3.5). Cependant, nous ne sommes pas prêts à tous les exclure. Comme plusieurs de ces CN sont des termes qui appartiennent aussi à la langue générale, nous les faisons passer au troisième test du corpus d'exclusion (4.2.3).

Les CN en vert sont les CN qui n'ont pas été repérés dans Google Books Ngram Viewer. Ils sont toujours considérés comme des CN.

Les CN en bleu sont les CN qui ont été repérés par Google Books Ngram Viewer avec une courbe dont l'évolution entre 2000 et 2008 est supérieure ou égale à 50 %. Selon les critères que nous avons établis à la section 3.5, nous les considérons toujours comme des CN.

Après ce test, 30 termes de COMMU et 34 termes de INFOR sont toujours considérés comme CN.

#### **4.2.3 La soumission des CN aux dictionnaires de langue générale**

Avant de dresser la liste définitive des néologismes, des CN doivent passer un troisième filtrage. Cela concerne des CN du Tableau 15 et du Tableau 16. Ces CN ont été sélectionnés indépendamment de leur présence ou de leur absence dans Google Books Ngram Viewer. Il s'agit de termes simples qui pourraient avoir une acception dans un dictionnaire de langue générale qui recoupe l'acception que nous observons dans nos corpus de communication ou d'informatique. Ces unités sont des termes simples (car les dictionnaires généraux répertorient très peu de termes complexes) qui pourraient ne pas être néologiques. Nous voulons valider leur statut néologique au moyen du troisième test.

Après le deuxième test, la méthode a fait évacuer certains termes qui nous semblaient avoir un caractère néologique, que ce soit pour leur forme atypique (ex. : *infomédiation*) ou qui pourraient avoir une nouvelle acception en langue de spécialité (ex. : *égocentrique*) (voir le

Tableau 15 et le Tableau 16). De plus, nous avons deux termes qui nous étaient impossible de vérifier dans l’outil de visualisation de n-grammes (*twitter* (V) et *utilisateur* (Adj)). Nous avons besoin d’un test de plus pour ces termes.

Dans certains cas, le troisième test a confirmé ce que les résultats de la soumission à Google Books Ngram Viewer montraient; dans d’autres cas, le troisième test les a contredits. Comme les outils de visualisation de n-grammes ne sont pas des dictionnaires, nous avons penché pour le résultat donné par ce filtrage plutôt que celui de Google Books Ngram Viewer dans le cas de ces termes. Cette décision est basée sur le fait que nous faisons plus confiance aux dictionnaires qu’aux n-grammes pour repérer les néologismes. Le deuxième test a principalement servi à évacuer les termes complexes non néologiques.

Ce dernier test est divisé en deux parties : d’abord, nous soumettons les CN sélectionnés au TLFi; puis, nous interrogeons Usito pour les CN qui n’ont pas été exclus par le TLFi.

#### ***4.2.3.1 La soumission des CN au TLFi***

Le Tableau 17 et le Tableau 18 montrent la liste des CN que nous avons soumis au TLFi pour COMMU (12 CN) et INFOR (13 CN). Les tableaux montrent les rangs, les CN, les parties du discours ainsi que l’année de la première attestation enregistrée.

Tableau 17 *Liste des CN de COMMU à revérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer*

Listes des CN de COMMU à revérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer			
Rang	CN	PDD	Année
13	usager	N	1956
14	twitter	V	x
21	outil	N	1857
22	usage	N	1792
41	interacteur	N	x
43	autoformation	N	x
49	communicationnel	Adj	x
97	infomédiation	N	x
101	dévoilement	N	1893
114	autodidaxie	N	x
133	sociabilité	N	1946
136	échange	N	1910

Tableau 18 *Liste des CN de INFOR à revérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer*

Listes des CN de INFOR à revérifier dans le TLFi après l'interrogation dans Google Books Ngram Viewer			
Rang	CN	PDD	Année
11	utilisateur	Adj	1932
18	twitter	V	x
22	descripteur	N	1964
35	pertinent	N	1792
38	textuel	Adj	1885
48	centralité	N	1792
60	visualisation	N	1981
63	annotateur	N	1948
76	égoцентриque	Adj	1955
78	homophilie	N	1971
99	recommandation	N	1839
122	imitation	N	1873
142	saillant	Adj	1823

Les (x) indiquent que les CN sont absents du TLFi. Ces CN sont en sarcelle. Ils sont considérés comme des néologismes.

Les CN en rose sont les CN qui sont exclus par le TLFi. Ils sont donc exclus de l'analyse des néologismes.

Il reste six CN pour COMMU à soumettre à la deuxième étape du troisième test. Il n'en reste qu'un seul pour INFOR. Lors du deuxième test, *communicationnel* (Adj), *autoformation* (N), *interacteur* (N), *infomédiation* (N) et *autodidaxie* (N) avaient été repérés par l'outil de visualisation de n-grammes. Comme nous priorisons les dictionnaires, à cette étape ils reprennent leur statut de CN.

Après avoir analysé tous les CN, le TLFi n'a pas permis de rejeter des CN de sens, donc nous aurions pu utiliser un autre dictionnaire.

#### 4.2.3.2 La soumission des CN à Usito

Un des défauts du TLFi est que sa nomenclature n'a pas été alimentée depuis 1996. Pour remédier à ce problème, nous soumettons les six CN qui n'ont pas été exclus par le TLFi au dictionnaire Usito. Les résultats sont compilés dans le Tableau 19. Le tableau indique les CN en question, leur rang et leur partie du discours.

Tableau 19 *Liste des CN de COMMU et INFOR à revérifier dans Usito*

Listes des CN de COMMU et INFOR à revérifier dans Usito		
Rang	CN	PDD
14	twitter	V
41	interacteur	N
43	autoformation	N
49	communicationnel	Adj
97	infomédiation	N
114	autodidaxie	N

Les CN en sarcelle sont ceux qui n'ont pas été exclus par Usito. Ils seront de ce fait considérés comme des néologismes.

Les CN en rose sont ceux qui ont été exclus par Usito. À partir de ce point, nous ne les considérons plus comme des néologismes potentiels.

### 4.3 Les néologismes des médias sociaux

Après toutes ces étapes, nous avons repéré 32 termes dans COMMU qui, selon la méthode appliquée, sont considérés comme des néologismes. Nous en avons également trouvé 32 dans INFOR. Ces néologismes sont présentés dans le Tableau 20 et le Tableau 21. Les tableaux sont divisés en trois colonnes : le rang, le néologisme et la partie du discours. Les néologismes en jaune sont ceux partagés par COMMU et INFOR.

Tableau 20 *Les néologismes des médias sociaux : COMMU*

Liste de néologismes de COMMU après l'application de toutes les étapes du corpus d'exclusion		
Rang	Néologisme	PDD/SN
14	twitter	V
16	socionumérique	Adj
32	média socionumérique	N+Adj
35	réseau socionumérique	N+Adj
36	réseau social numérique	N+Adj+Adj
40	trace numérique	N+Adj
41	interacteur	N
43	autoformation	N
47	pratiques transplateformes	N+Adj
60	marque employeur	N+N
62	podcaster	N
65	identité professionnelle numérique	N+Adj+Adj
68	youtuber	N
69	cyberactiviste	N
70	parrhésie	N
71	page Facebook	N+N
81	forumier	N
82	site d'actualité	N+Prep+N
83	SNS	N
88	cyberbullying	N
89	RSNP	N
97	infomédiation	N
115	plateforme de réseau social	N+Prep+N+Adj
117	parrhésiaste	N
118	réseau social traditionnel	N+Adj+Adj
128	site de réseautage	N+Prep+N
130	outil de publication	N+Prep+N
138	média traditionnel	N+Adj
142	organisation émettrice	N+Adj
143	schéma publicitaire	N+Adj
145	vidéoblogueuse	N
146	outil de discussion	N+Prep+N



Tableau 21 *Les néologismes des médias sociaux : INFOR*

Liste de néologismes de INFOR après l'application de toutes les étapes du corpus d'exclusion		
Rang	Néologisme	PDD/SN
18	twitter	V
42	réseau social numérique	N+Adj+Adj
58	détection de communauté	N+Prep+N
64	détection d'événements	N+Prep+N
66	collaboratif	Adj
71	filtrage collaboratif	N+Adj
74	collection de test	N+Prep+N
79	F-mesure	N
81	magnitude d'impact	N+Prep+N
83	classifieur	N
85	n-grammes	N
94	graphe social	N+Adj
95	mot lié	N+Adj
100	événement redondant	N+Adj
101	dimension utilisateur	N+N
103	site de réseautage	N+Prep+N
104	TAL	N
105	topique discursif	N+Adj
110	macro-précision	N
114	média traditionnel	N+Adj
118	mesure de centralité	N+Prep+N
120	moniteur positif	N+Adj
127	mesure de similarité	N+Prep+N
129	chaîne mémétique	N+Adj
135	pertinence système	N+N
138	poids de pertinence	N+Prep+N
139	sous-graphe	N
145	vecteur de support	N+Prep+N
147	énonciatif	Adj
148	site de réseautage social	N+Prep+N+Adj
149	réseau socionumérique	N+Adj
150	système de recommandation sociale	N+Prep+N+Adj

Il est surprenant que nous obtenions le même nombre de néologismes dans les deux listes de 150 termes (149 termes après avoir repéré les erreurs d'élagage *a posteriori*). Il est possible que la méthode utilisée soit à l'origine de cette similitude. Cela pourrait aussi être le fruit du hasard.

Quatre néologismes (*twitter* (V), *réseau social numérique* (N + Adj + Adj), *média traditionnel* (N + Adj) et *réseau socionumérique* (N + Adj)) sont présents dans les deux listes de néologismes. Cela représente 12,5 % des néologismes de COMMU et 12,5 % des néologismes de INFOR. Nous avons donc 28 néologismes exclusifs à COMMU, 28 néologismes exclusifs à INFOR et quatre néologismes présents dans les deux corpus. Cette étude a donc permis de repérer 60 néologismes.

#### 4.3.1 La répartition des néologismes et des exclusions de COMMU et INFOR

Après avoir soumis les deux listes de 150 termes au corpus d'exclusion dans le but d'y trouver des néologismes, nous avons classé ces termes en deux catégories : les néologismes et les exclusions. La Figure 15 et la Figure 16 montrent la répartition de ces deux catégories de termes par corpus.

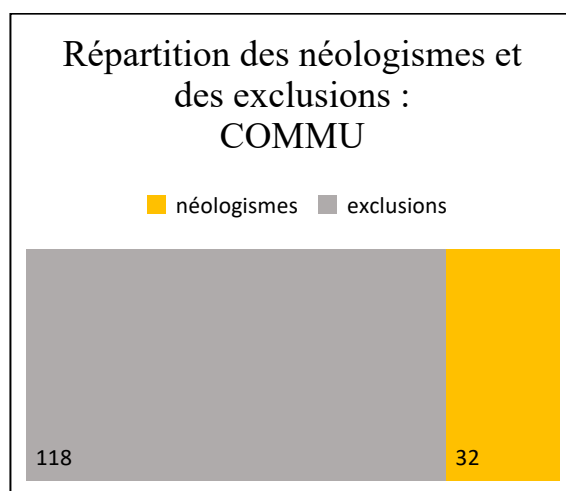


Figure 15 Répartition des néologismes et des exclusions : COMMU

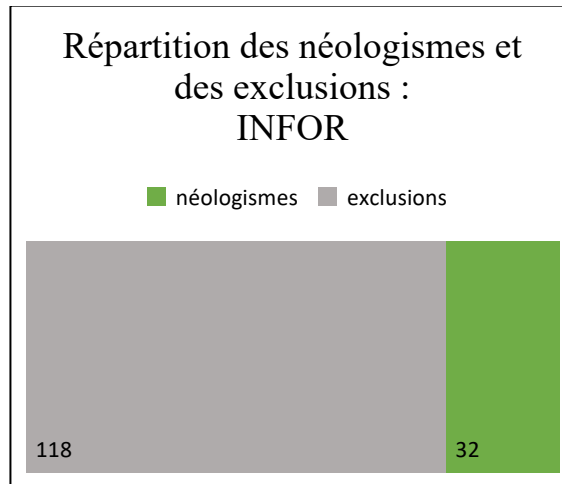


Figure 16 *Répartition des néologismes et des exclusions : INFOR*

Les résultats sont les mêmes pour COMMU et INFOR (dans les deux corpus nous avons repéré le même nombre de néologismes). Les néologismes représentent 21,3 % des termes, soit un peu plus du cinquième des termes. Il s'agit d'un nombre relativement élevé, ce qui nous permet de confirmer le fait que nous n'avions pas tort de croire que le domaine des médias sociaux était productif en matière de néologie, probablement à cause des importantes avancées technologiques dans le domaine dans les dernières années.

#### **4.3.2 La répartition des parties du discours et des syntagmes nominaux des néologismes de COMMU et INFOR**

Nous voulions savoir quelle est la répartition des parties du discours pour les néologismes. Nous avons fait cette comparaison pour les deux listes de 150 termes (4.1.4). Les graphiques de la Figure 17 et de la Figure 18 montrent les résultats seulement pour les néologismes de COMMU et INFOR.

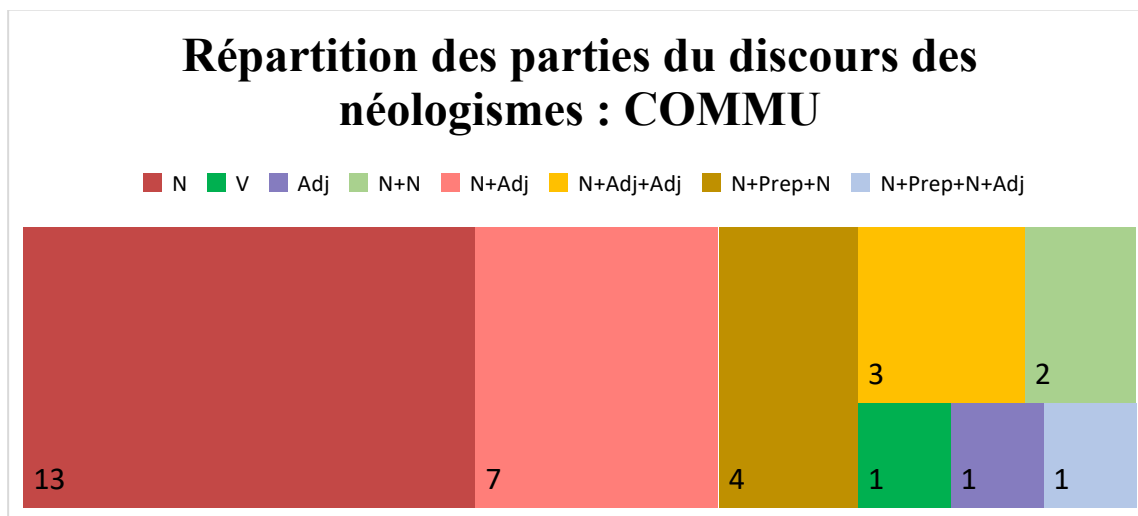


Figure 17 Répartition des parties du discours et des syntagmes nominaux des néologismes :  
COMMU

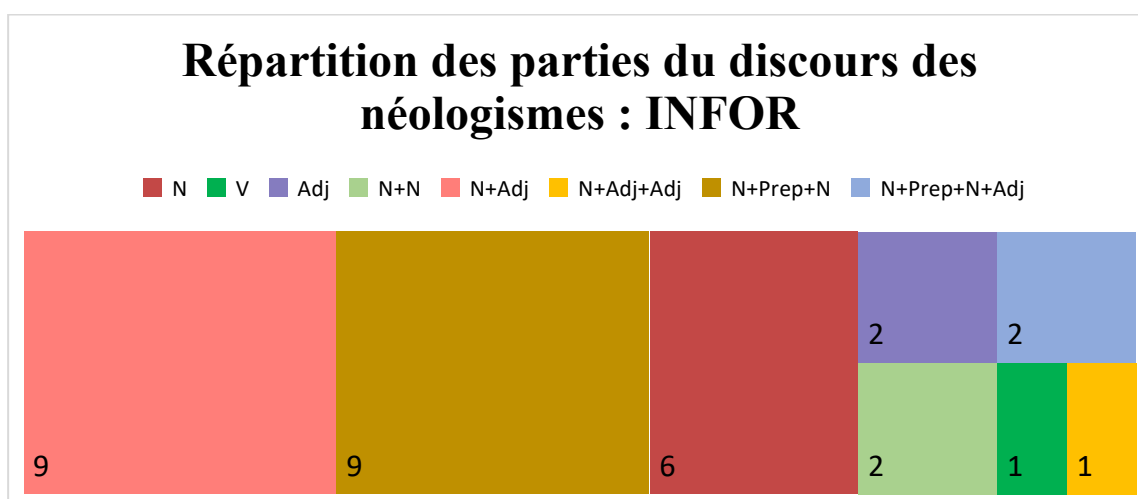


Figure 18 Répartition des parties du discours et des syntagmes nominaux des néologismes :  
INFOR

D'entrée de jeu, nous constatons une différence entre les néologismes de COMMU et INFOR. Les noms occupent la première place dans les néologismes de COMMU, alors qu'ils sont en troisième place dans les néologismes de INFOR derrière les termes complexes composés de N + Adj et N + Prep + Adj.

Les cinq autres types de parties du discours ou syntagmes nominaux (Adj, N + Prep + N + Adj, N + Adj + Adj, N + N et V) sont en moins grand nombre et les résultats sont très semblables dans les deux disciplines.

Si nous comparons cette répartition à celle observée pour les listes de 150 termes complètes, les résultats divergent également. D'abord, les N sont au premier rang pour INFOR dans la liste de 150 termes alors que, dans les néologismes, ils sont au troisième rang. Ensuite, les Adj sont beaucoup moins présents dans le cas des néologismes que dans le cas des listes complètes. Finalement, la répartition des N + Prep + N est beaucoup plus grande dans les graphiques des néologismes que dans les graphiques des listes complètes.

Nous pouvons dire que la répartition des parties du discours ou syntagmes nominaux des néologismes de COMMU est plutôt comparable à la liste de 150 termes de COMMU (mis à part les Adj). Néanmoins, ce n'est pas le cas pour INFOR. On peut voir des changements dans les Adj, mais aussi une plus grande répartition des N + Adj et N + Prep + Adj. Nous pouvons supposer que les néologismes en informatique sont plus souvent des termes complexes et ont tendance à être produits selon ces modèles combinatoires.

### **4.3.3 La répartition des termes simples et complexes néologiques de COMMU et INFOR**

Nous analysons maintenant la répartition des termes simples et des termes complexes pour les néologismes issus de COMMU et de INFOR. Nous avons fait cette analyse pour les listes de 150 termes de COMMU et INFOR à la section 4.1.5. Nous voulons voir si les néologismes se répartissent différemment. Les graphiques de la Figure 19 et de la Figure 20 montrent ces répartitions.

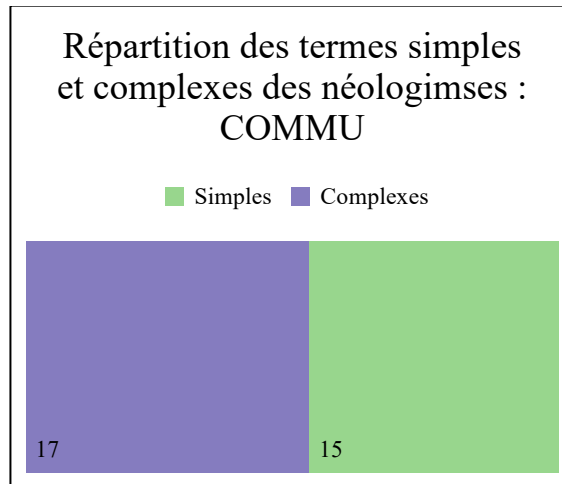


Figure 19 Répartition des termes simples et complexes des néologismes : *COMMU*

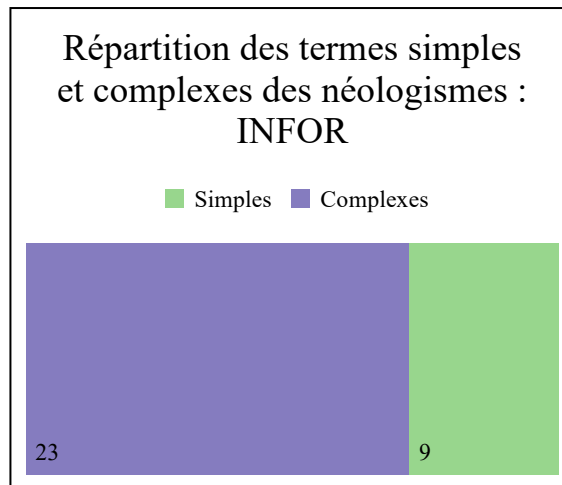


Figure 20 Répartition des termes simples et complexes des néologismes : *INFOR*

Les résultats sont très différents de la répartition des termes simples et complexes des listes de 150 termes de COMMU et INFOR.

Dans le cas des 150 termes de COMMU, les termes complexes représentaient 34,7 % des résultats. Les données montrent qu'ils représentent 53,1 % des néologismes de COMMU.

Pour les 150 termes de INFOR, les termes complexes comptaient pour 33,3 % des résultats. Dans le cas des néologismes, ils représentent 71,9 % au total.

Ces données indiquent que la création de termes complexes est un mode de création néologique important, particulièrement en informatique.

#### **4.4 Les procédés néologiques utilisés pour la création des néologismes de COMMU et INFOR**

La dernière étape de notre recherche consiste à classer les néologismes répertoriés. Nous étiquèterons les 32 néologismes de COMMU et les 32 néologismes de INFOR.

##### **4.4.1 Les néologismes étiquetés**

Pour présenter nos résultats (sections 4.4.1.1 et 4.4.1.2), nous avons utilisé un tableau qui comporte 5 éléments d'information par entrée néologique : le rang, le néologisme, la partie du discours, le type de néologisme selon Dubuc (2002), ainsi qu'un contexte issu de l'un de nos corpus. Pour rendre la lecture plus facile, nous avons également ajouté un code de couleur aux types des néologismes. Les tons de vert sont des dérivations, les tons de jaune sont des compositions, les tons de violet sont des emprunts. Nous avons prévu les tons de rouge pour les extensions sémantiques et les tons de bleus pour les changements grammaticaux, cependant ces deux catégories ne se sont pas manifestées dans nos listes de néologismes. Nous en reparlerons à la section 4.4.3.

##### ***4.4.1.1 Les néologismes de COMMU étiquetés***

Les 32 néologismes de COMMU sont étiquetés et présentés dans le Tableau 22.

Tableau 22 Les néologismes étiquetés : COMMU

La typologie des néologismes : COMMU			
Rang	Néologisme	PDD/SN	Type de néologisme (Dubuc 2002)
14	twitter	V	emprunt naturalisé
Contexte	« Azy j'kiff Hitler, yakwa?? », <b>twittait</b> de son côté Stephen des Aulnois, qui vient de mettre en pause son blogue sur la pornographie. (ligedulol.txt)		
16	socionumérique	Adj	composition par soudure d'éléments français
Contexte	Sans oublier que selon le type d'information et le média <b>socionumérique</b> considérés, l'analyse ne prendra pas la même forme. (usagesprofessionnels.txt)		
32	média socionumérique	N+Adj	composition par subordination
Contexte	L'essor des <b>médias socionumériques</b> depuis quelques années a donné lieu à de nouveaux usages professionnels des TIC liés à la circulation de l'information (veille, sélection, diffusion, etc.). (usagesprofessionnels.txt)		
35	réseau socionumérique	N+Adj	composition par subordination
Contexte	Selon les auteurs les « <b>réseaux socionumériques</b> » doivent fonder leur attractivité essentiellement sur « l'opportunité de retrouver ses « amis » et d'interagir avec eux par le biais de profils, de listes de contacts et d'applications à travers une grande variété d'activités » (2011, p 13) et non sur une activité particulière. (erecruitment.txt)		
36	réseau social numérique	N+Adj+Adj	composition par subordination
Contexte	Par ailleurs, Twitter, bien qu'originellement défini comme un réseau <b>social</b> numérique de contenu, est de plus en plus considéré comme un réseau <b>social</b> numérique de contact, les fonctionnalités de réseaux prenant une place grandissante dans les usages (contact interpersonnel, conversations). (wikipedia.txt)		
40	trace numérique	N+Adj	composition par subordination
Contexte	Notre étude renforce ce constat en établissant que les tweets présentent effectivement les traits caractéristiques des <b>traces</b> numériques. (tweetstraces.txt)		
41	interacteur	N	dérivation propre ou normale
Contexte	Un <b>interacteur</b> peut, par exemple, être en train d'écrire un tweet personnel à un ami et recevoir au même moment un message d'un personnage fictionnel qui s'insérera dans son fil d'actualités Twitter et qui l'incitera peut-être à interagir avec le personnage une fois le message personnel envoyé. (personnagesfictifs.txt)		
43	autoformation	N	composition par préfixe non autonome
Contexte	Le sujet apprenant en situation d' <b>autoformation</b> associe donc clairement « celui qui sait » à son parcours formatif. (autoformation.txt)		
47	pratiques transplateformes	N+Adj	composition par subordination
Contexte	L'un des acteurs interrogés récemment, qui figure comme le plus actif au niveau des <b>pratiques transplateformes</b> parmi nos informateurs, affirmait d'ailleurs se préoccuper d'éviter de créer de la « dissonance » par rapport à son image dans les médias sociaux, tout en étant très appliqué à ne pas trop « se répéter » d'une plateforme d'Instagram à Facebook. (transplateformes.txt)		
60	marque employeur	N+N	composition par coordination
Contexte	Ces discussions existent déjà «hors ligne», mais les médias sociaux permettent de toucher une audience bien plus large, augmentant ainsi les conséquences pour la <b>marque</b> employeur. (wikipedia.txt)		
62	podcaster	N	emprunt intégral
Contexte	Cependant, les <b>podcasters</b> ne représentent qu'une frange des internautes, probablement parmi les plus compétents en termes techniques et communicationnels par rapport aux usagers des médias sociaux en général. (transplateformes.txt)		



65	identité professionnelle numérique	N+Adj+Adj	composition par subordination
Contexte	Ainsi, si tous les médias socionumériques n'ont pas pour vocation de favoriser des usages professionnels, chacun d'entre eux peut être un espace de développement de tels usages, lesquels participent à la construction d'une <b>identité professionnelle numérique</b> . (usagesprofessionnels.txt)		
68	youtuber	N	emprunt intégral
Contexte	Les collégiens qui se montrent plus concernés par le débat éthique posé par la pratique de la publicité rédactionnelle sur youtube adoptent la plupart du temps une position téléologique de défense des droits du <b>youtuber</b> à réaliser ce type d'alliance. (publicitetrompeuse.txt)		
69	cyberactiviste	N	composition par préfixe non autonome
Contexte	D'après nos échanges avec des <b>cyberactivistes</b> tunisiens et nos observations en ligne, il apparaît cependant que les internautes les plus mobilisés, prêts à chercher des solutions de protection et de contournement de la censure plus sophistiquées et efficaces que les proxys basiques, n'ont pour la plupart pas eu besoin de cette aide des Anonymous. (revolutiontunisienne.txt)		
70	parrhésie	N	emprunt naturalisé
Contexte	La vidéo de chubbygirl27 a touché Katlunn (voilà ce qui l'a incitée à produire une vidéo) et, de prime abord, cette dernière semble pratiquer la <b>parrhésie</b> . (youtubeobesite.txt)		
71	page Facebook	N+N	composition par subordination
Contexte	En plus de leur blog, Lada et Georges possèdent une <b>page Facebook</b> publique dédiée à leur podcast. (tranplateformes.txt)		
81	forumier	N	dérivation propre ou normale
Contexte	Beaucoup de <b>forumiers</b> reprennent des textes et des vidéos consultées sur Facebook, et citent notamment les chaînes satellitaires Al Jazeera et Al Arabiya. (revolutiontunisienne.txt)		
82	site d'actualité	N+Prep+N	composition par subordination
Contexte	Cependant, la comparaison devient plus intéressante si on renverse la perspective en examinant quels sont les sujets qui ont été le plus surreprésentés sur Twitter par rapport au nombre d'articles produits par les <b>sites d'actualité</b> . (infomédiation.txt)		
83	SNS	N	emprunt intégral
Contexte	L'infomédiation sociale implique deux composantes : premièrement des plates-formes numériques de sociabilité, qui doivent remplir les trois conditions d'un <b>SNS</b> ou Social Networking Site (Boyd & Ellison, 2007) 1 ; (infomédiation.txt)		
88	cyberbullying	N	emprunt intégral
Contexte	Cette forme de harcèlement constitue le <b>cyberbullying</b> . (cyberbullying.txt)		
89	RSNP	N	composition par soudure d'éléments français
Contexte	En d'autres termes, l'usage des <b>RSNP</b> ne répond pas à une volonté de rencontrer de nouveaux « amis » ou d'entretenir son réseau affectif mais s'inscrit dans une démarche globale d'utilisation des moyens les plus efficaces afin d'atteindre un objectif précis (trouver un emploi, de nouveaux collaborateurs, etc.). (linkedincommunication.txt)		
97	infomédiation	N	composition par soudure d'éléments français
Contexte	Malgré les précautions devant accompagner l'interprétation de ces études, il n'en demeure pas moins qu'elles dessinent les contours de modes d'accès à l'actualité émergents, fondés sur l' <b>infomédiation</b> sociale. (infomédiation.txt)		
115	plateforme de réseau social	N+Prep+N+Adj	composition par subordination
Contexte	Pour l'expliquer, il convient d'abord de revenir sur certaines caractéristiques de cette <b>plateforme de réseau social</b> , que nous avons déjà resituée dans l'évolution, depuis une bonne décennie, des usages citoyens et militants d'Internet par des Tunisiens ordinaires. (revolutiontunisienne.txt)		
117	parrhésiaste	N	dérivation propre ou normale
Contexte	Nous l'avons retenue aux fins d'analyse parce que, à l'instar de Katlunn, la vidéoblogueuse qui l'a mise en ligne « cristinamariex3 » a été touchée par une vidéo vue sur YouTube et elle semblait reprendre le rôle de <b>parrhésiaste</b> . (youtubeobesite.txt)		

<b>118</b>	réseau social traditionnel	N+Adj+Adj	composition par subordination
<b>Contexte</b>	En effet, le <b>réseau social traditionnel</b> , dans une conception africaine est une organisation sociale qui rassemble des individus ayant des objectifs communs, des valeurs communes et généralement basée sur le respect de ces valeurs avec 1 une possibilité de contact physique. (cotedivoire.txt)		
<b>128</b>	site de réseautage	N+Prep+N	composition par subordination
<b>Contexte</b>	L'intégration de ces boutons à même le blog invite les auditeurs à publiciser le podcast sur Facebook en y ajoutant un marqueur symbolique de leur appréciation, soit le « Like » du <b>site de réseautage</b> , soit un commentaire personnel qui sera simultanément diffusé à tous leurs « amis » Facebook. (transplateformes.txt)		
<b>130</b>	outil de publication	N+Prep+N	composition par subordination
<b>Contexte</b>	Grâce à son <b>outil de publication</b> , LinkedIn est également un très bon canal d'information et un relais de campagnes publicitaires. (oeilmoderne.txt)		
<b>138</b>	média traditionnel	N+Adj	composition par subordination
<b>Contexte</b>	La grande différence d'avec les <b>médias traditionnels</b> , c'est que les médias sociaux sont un espace d'expression et de diffusion de l'information où les contenus peuvent être personnalisés et interactifs. (oeilmoderne.txt)		
<b>142</b>	organisation émettrice	N+Adj	composition par subordination
<b>Contexte</b>	Notre questionnement a ainsi visé à établir si le tweet pouvait être considéré comme une trace numérique, avec toutes les implications induites pour les <b>organisations émettrices</b> . (tweetstraces.txt)		
<b>143</b>	schéma publicitaire	N+Adj	composition par subordination
<b>Contexte</b>	La publicité rédactionnelle permet aux annonceurs de contourner le <b>schéma</b> publicitaire et ses effets en activant des <b>schémas</b> liés à la production éditoriale, induisant par là le consommateur en erreur (Hoofnagle et Meleshinsky, 2015). (publicitetrompeuse.txt)		
<b>145</b>	vidéoblogueuse	N	emprunt naturalisé
<b>Contexte</b>	Notre analyse porte sur les mots ou les énoncés des <b>vidéoblogueuses</b> sur YouTube. (youtubeobesite.txt)		
<b>146</b>	outil de discussion	N+Prep+N	composition par subordination
<b>Contexte</b>	Le leader mondial en matière d' <b>outil de discussion</b> est le logiciel Skype [réf. nécessaire]. (wikipedia.txt)		

#### 4.4.1.1.1 La répartition des néologismes de COMMU

La Figure 21 est un graphique qui présente la répartition des types de néologismes de COMMU. Là encore, les dérivations sont en tons de vert, les compositions en ton de jaune et les emprunts en tons de violet.

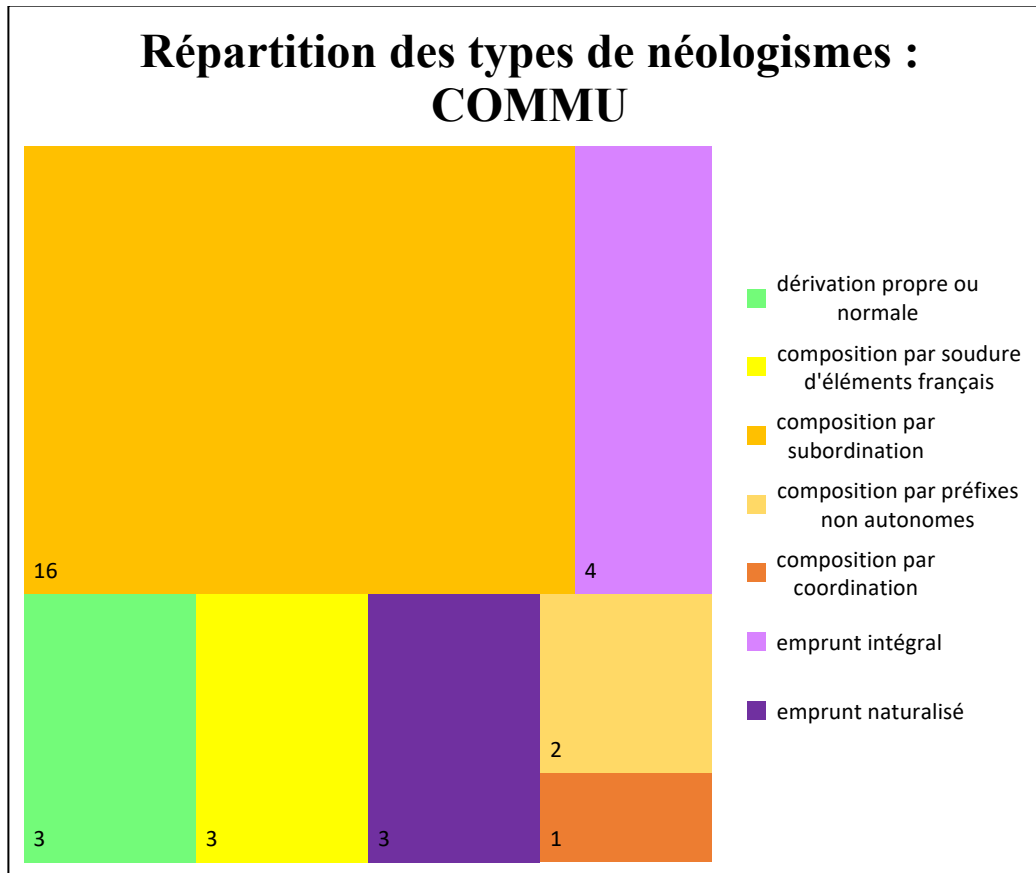


Figure 21 *La répartition des types de néologismes : COMMU*

Comme nous pouvons le constater, la composition (jaune) est le procédé néologique le plus présent dans notre échantillon de néologismes (22 néologismes sur 32). La composition par subordination est de loin le procédé le plus courant avec près de la moitié des résultats qui tombe dans cette catégorie.

Le deuxième procédé le plus représenté est l'emprunt (violet). L'écart entre l'emprunt intégral et l'emprunt naturalisé est négligeable.

Finalement, les dérivations (vert) représentent moins d'un dixième des résultats, ce qui les place en troisième place, derrière la composition et l'emprunt.

#### 4.4.1.2 Les néologismes répertoriés dans INFOR

Cette section présente maintenant les 32 néologismes étiquetés de INFOR. Ils sont répertoriés dans le Tableau 23. La présentation des résultats est la même qu'à la section 4.4.1.1.

Tableau 23 Les néologismes étiquetés : INFOR

INFOR			
Rang	Terme	PDD	Type de néologisme (Dubuc 2002)
18	twitter	V	emprunt naturalisé
Contexte	<b>Twitter</b> est très important pour les gens de tous les horizons de la vie et de toutes les nationalités avec toutes les langues de la planète (Mendoza et al., 2010). (introductiontwitter.txt)		
42	réseau social numérique	N+Adj+Adj	composition par subordination
Contexte	Cependant, comme introduit précédemment, la prise en compte du temps est essentielle pour un <b>réseau social numérique</b> . (profilutilisateur.txt)		
58	détection de communauté	N+Prep+N	composition par subordination
Contexte	L'objectif de la <b>détection de communautés</b> est de créer une partition des sommets de telle sorte que les sommets au sein d'une même communauté soient fortement reliés entre eux et faiblement connectés au reste du réseau. (intelligenceartificielle.txt)		
64	détection d'événements	N+Prep+N	composition par subordination
Contexte	Matrice de comparaison des méthodes existantes pour la <b>détection d'événements</b> . (diffusioninformation.txt)		
66	collaboratif	Adj	dérivation propre ou normale
Contexte	Ils distinguent ainsi six types de médias sociaux : les projets <b>collaboratifs</b> , les blogs et micro-blogs, les communautés de contenu, les sites de réseaux sociaux, les mondes de jeux virtuels et les mondes sociaux virtuels. (cartographie.txt)		
71	filtrage collaboratif	N+Adj	composition par subordination
Contexte	Cependant, dans ce mémoire, nous considérons le sens premier de cette expression qui implique forcément l'usage d'un réseau social à la différence de l'utilisation des utilisateurs similaires dans le <b>filtrage collaboratif</b> . (reseauxegocentriques.txt)		
74	collection de test	N+Prep+N	composition par subordination
Contexte	La <b>collection de test</b> est composée d'un ensemble statique d'articles Wikipédia, de Tweets et de phrases contextuelles de référence sélectionnées par les organisateurs. (contextualisationtweets.txt)		
79	F-mesure	N	composition par soudure d'éléments français
Contexte	La <b>mesure</b> appelée <b>F-mesure</b> permet de combiner la précision et le rappel en une seule et unique <b>mesure</b> . (rechercheinformation.txt)		
81	magnitude d'impact	N+Prep+N	emprunt intérieur
Contexte	Enfin, lorsque k événements distincts (i.e. en n'incluant pas les éventuels événements redondants) ont été traités, le troisième composant fusionne les éventuels événements redondants et renvoie la liste L contenant les k événements ayant eu les plus grandes <b>magnitudes d'impact</b> sur les utilisateurs. (diffusioninformation.txt)		
83	classifieur	N	dérivation propre ou normale
Contexte	Pour chaque <b>classifieur</b> , nous avons étudié l'apport de chaque groupe de traits (cf. Section 3.1) au processus d'apprentissage. (ironietweets.txt)		
85	n-grammes	N	composition par soudure d'éléments français
Contexte	Les 1000 <b>n-grammes</b> les plus discriminants, selon leur gain d'information (Mitchell, 1997), sont alors retenus pour constituer le vocabulaire de description, et chaque message est donc représenté par un vecteur de dimension 1000. (diffusioninformation.txt)		

94	graphe social	N+Adj	composition par subordination
Contexte	Dès lors, ils s'intéressent plus aux évaluations des mécanismes de filtrage social qu'à l'optimisation de l'usage du réseau social ( <b>graphe</b> social) de l'utilisateur. (reseauxegocentrique.txt)		
95	mot lié	N+Adj	composition par subordination
Contexte	Chaque <b>mot lié</b> t 0 q est représenté par un n'ud et possède un arc dirigé vers le mot principal, dont le poids est égal à wq. (diffusioninformation.txt)		
100	évènement redondant	N+Adj	composition par subordination
Contexte	Le mot principal devient l'agrégation des mots principaux des <b>évènements redondants</b> . (diffusioninformation.txt)		
101	dimension utilisateur	N+N	composition par subordination
Contexte	Nous partons des travaux existants pour construire les dimensions sociales et la <b>dimension utilisateur</b> du profil de l'utilisateur. (profilutilisateur.txt)		
103	site de réseautage	N+Prep+N	composition par subordination
Contexte	Une rumeur via les réseaux sociaux est une nouvelle qui est diffusée via un <b>site de réseautage</b> . (detectionrumeurs.txt)		
104	TAL	N	composition par soudure d'éléments français
Contexte	Ce numéro spécial montre de quelle manière les méthodes de <b>TAL</b> contribuent à l'analyse des réseaux sociaux. (defistal.txt)		
105	topique discursif	N+Adj	composition par subordination
Contexte	Cependant, lorsque le processus d'imitation s'intensifie et que se produit une transformation significative du <b>topique discursif</b> lui-même, alors c'est à une recréation que l'on commence à avoir affaire. (propagationmemes.txt)		
110	macro-précision	N	composition par préfixe non autonome
Contexte	Sur cette tâche, les résultats en <b>macro-précision</b> (tableau 5) varient de 0,029 à 0,613, avec une moyenne de 0,514, une médiane de 0,217 et un écart-type de 0,029. (emotionstweets.txt)		
114	média traditionnel	N+Adj	composition par subordination
Contexte	Les messages sont labélisés en fonction de leur provenance, c'est-à-dire selon que le groupe est jugé comme <b>média traditionnel</b> ou de réinformation. (reinformation.txt)		
118	mesure de centralité	N+Prep+N	composition par subordination
Contexte	Ces <b>mesures de centralité</b> sont appliquées aux groupes d'individus dans le réseau social. (reseauxegocentriques.txt)		
120	moniteur positif	N+Adj	composition par subordination
Contexte	Pour identifier la source d'une rumeur, nous utilisons l'intuition selon laquelle la source doit être proche des <b>moniteurs positifs</b> mais loin des moniteurs négatifs. (detectionrumeurs.txt)		
127	mesure de similarité	N+Prep+N	composition par subordination
Contexte	Nous obtenons ainsi des tweets avec plus d'informations ce qui permet aux <b>mesures de similarité</b> de mieux détecter les termes en commun entre une paire de tweets. (detectiontwitter.txt)		
129	chaîne mémétique	N+Adj	composition par subordination
Contexte	Le repérage de ces points de contact n'est possible, encore une fois, que si l'utilisateur a une pleine connaissance de la <b>chaîne mémétique</b> même qui leur sert de référence (à la manière d'une sorte de « répertoire » de memes). (propagationmemes.txt)		
135	pertinence système	N+N	composition par subordination
Contexte	Le score de <b>pertinence système</b> de chaque document est proportionnel à la distance entre les représentations vectorielles du document et de la requête. (rechercheinformation.txt)		
138	poids de pertinence	N+Prep+N	composition par subordination
Contexte	A partir des informations que partage chaque individu du réseau égo-centrique de l'utilisateur, nous proposons de pondérer leur poids de <b>pertinence</b> en intégrant des informations temporelles. (profilutilisateur.txt)		
139	sous-graphe	N	composition par préfixe autonome

<b>Contexte</b>	Bien que ces figures représentent une représentation du sous- <b>graphe</b> global que l'on aurait obtenu à partir de l'ensemble des conversations, ces observations sont confirmées par la valeur de mesure de cohésion qui estime le ratio entre liens «intra» et liens «inter» (une valeur supérieure à 1 souligne la prépondérance des liens «intra»). (statistiquesinformation.txt)		
<b>145</b>	vecteur de support	N+Prep+N	composition par subordination
<b>Contexte</b>	Les résultats ont révélé que l'algorithme machine à <b>vecteurs de support</b> est l'algorithme le plus précis.(bigsocialdata.txt)		
<b>147</b>	énonciatif	Adj	dérivation propre ou normale
<b>Contexte</b>	Ce contrat <b>énonciatif</b> peut s'appliquer à tout type de média [Ibid. p. 206] et Veron analysera de la sorte aussi bien la radio que la presse, la télévision ou encore les différents formats publicitaires. (systemesderecommandation.txt)		
<b>148</b>	site de réseautage social	N+Prep+N+Adj	composition par subordination
<b>Contexte</b>	Toute personne doit être en mesure de créer et d'organiser un ou plusieurs profils différents en utilisant un site de <b>réseautage</b> social de confiance de choix, y compris l'hébergement de leur propre site sur lequel ils se sont lancés soit sur un serveur ou localement dans leur navigateur. (donneesmanquantes.txt)		
<b>149</b>	réseau socionumérique	N+Adj	composition par subordination
<b>Contexte</b>	Enfin, nous nous appuyerons sur trois ans d'observations menées sur les principaux <b>réseaux socionumériques</b> de 2008 à fin 2010 ainsi que sur une monographie de Facebook détaillant son modèle économique, ses fonctionnalités et le discours accompagnant ses services. (systemesderecommandation.txt)		
<b>150</b>	système de recommandation sociale	N+Prep+N+Adj	composition par subordination
<b>Contexte</b>	Afin de déterminer quel réseau social serait le plus approprié dans un système de <b>recommandation</b> sociale, (Guy et al., 09) réalisent une expérimentation dans laquelle pour un utilisateur u, et un item i, le poids prédit en fonction du réseau social de u sur l'item i est déterminé par la formule (5). (reseauxegocentriques.txt)		

#### 4.4.1.2.1 La répartition des néologismes de INFOR

Comme pour COMMU (4.4.1.1.1), nous avons analysé la répartition des types de néologismes de INFOR. Ils sont présentés sous forme de graphique à la Figure 22.

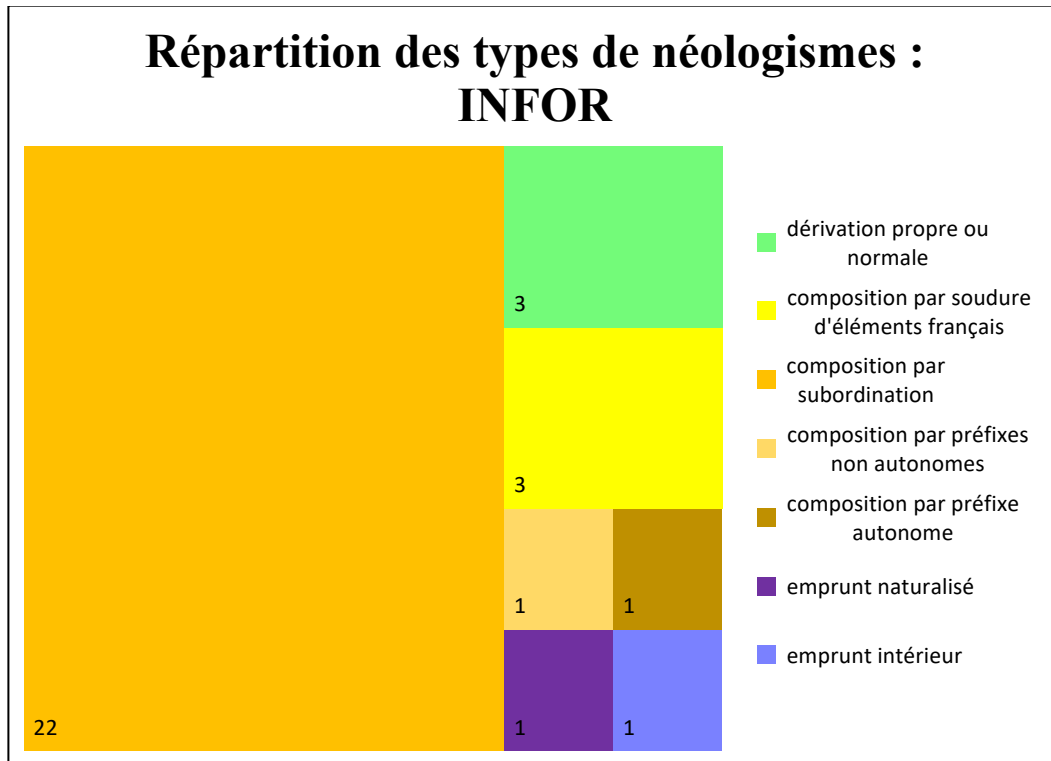


Figure 22 *La répartition des types de néologismes : INFOR*

Nous constatons que la composition (jaune) est toute aussi importante dans INFOR que dans COMMU (84,4 %). Les compositions par subordination sont en tête dans ce corpus.

La dérivation propre ou normale, seule représentante de sa catégorie (vert), occupe la deuxième place à égalité avec la composition par soudure d'éléments français.

Les emprunts (violet) occupent la dernière place dans INFOR.

#### 4.4.2 La comparaison de la répartition des néologismes entre COMMU et INFOR

La composition par subordination est de loin le procédé néologique le plus souvent observé dans les deux corpus. Les compositions en général sont les procédés les plus importants dans COMMU et INFOR.

L'emprunt est nettement moins important dans INFOR que dans COMMU. Effectivement, la catégorie se place au deuxième rang dans COMMU alors qu'elle est au dernier rang dans INFOR.

Deux procédés sont exclusifs à un des corpus : la composition par coordination n'est trouvée que dans COMMU, alors que la composition par préfixe autonome n'est présente que dans INFOR.

Finalement, la dérivation n'a qu'un seul représentant dans les deux corpus, soit la dérivation propre ou normale. Celle-ci se manifeste trois fois dans les néologismes de COMMU, tout comme INFOR.

Le Tableau 24 montre les résultats de la répartition des néologismes en termes de pourcentages. La colonne de gauche présente tous les procédés néologiques répertoriés dans les deux corpus, les colonnes du centre indiquent les pourcentages obtenus pour chaque type de néologismes dans COMMU et dans INFOR, puis, les colonnes de droite indiquent la répartition par catégories de néologismes dans COMMU et INFOR.

Tableau 24 *La répartition des procédés néologiques dans COMMU et INFOR*

PROCÉDÉS DE NÉOLOGISATION	COMMU	INFOR	COMMU	INFOR
dérivation propre ou normale	9,4 %	9,4 %	9,4 %	9,4 %
composition par soudure d'éléments français	9,4 %	9,4 %	68,7 %	84,4 %
composition par subordination	50 %	68,8 %		
composition par préfixe non autonome	6,2 %	3,1 %		
composition par préfixe autonome	0 %	3,1 %		
composition par coordination	3,1 %	0 %		
emprunt intégral	12,5 %	0 %	21,9 %	6,2 %
emprunt naturalisé	9,4 %	3,1 %		
emprunt intérieur	0 %	3,1 %		
Total	100 %	100 %	100 %	100 %

Concentrons-nous sur les colonnes du centre. Nous pouvons constater que la composition propre ou normale représente 9,4 % des résultats dans les deux corpus. C'est le seul procédé pour lequel nous avons obtenu des résultats identiques.

Nous pouvons voir que la composition par subordination est plus importante dans INFOR que COMMU. D'un autre côté, l'emprunt intégral et l'emprunt naturalisé sont plus importants dans INFOR.

Les emprunts ne sont pas vraiment importants dans INFOR, mais nous constatons qu'un des procédés d'emprunt lui est exclusif : l'emprunt intérieur.



Si nous nous concentrons sur les colonnes de droite, nous pouvons voir que les données obtenues pour la dérivation sont toujours identiques. On remarque, en revanche, une distinction quant à la répartition des compositions et des emprunts dans les deux corpus. Le corpus de la communication présente plus de diversité que le corpus de l'informatique, car les néologismes sont mieux répartis entre la composition et l'emprunt. Ces données semblent indiquer que la communication emprunte plus à l'anglais que l'informatique, si nous nous fions aux répartitions pour l'emprunt intégral. De plus, les résultats en informatique montrent qu'il n'y a aucun emprunt intégral, et que l'emprunt à l'anglais a été naturalisé pour s'adapter au français.

#### **4.4.3 Les procédés de néologisation de Dubuc (2002) absents de COMMU et INFOR**

La typologie des néologismes de Dubuc (2002) comptait 28 procédés différents. Pourtant, les résultats de COMMU et INFOR combinés ne révèlent que 9 procédés, soit environ le tiers. Le Tableau 25 présente les types de procédés néologiques qui n'ont pas été utilisés dans l'étiquetage des néologismes de cette étude.

Tableau 25 *Types de procédés néologiques absents des résultats par catégorie*

Types de procédés néologiques de Dubuc (2002) absents des résultats par catégorie		
extensions sémantiques		
glissement de la relation logique du concret à l'abstrait	glissement de la relation logique de la matière à l'objet	
glissement de la relation logique du signe au signifié	glissement de la relation logique de la partie au tout	
glissement de la relation logique de la cause à l'effet	glissement de la relation logique du contenant au contenu	
analogie de forme	assimilation	analogie de fonction
changements grammaticaux		
changement de genre	changement de nombre	
changement de statut	changements de catégorie	
emprunts		
calque	faux emprunt	xénisme
dérivations		
La dérivation régressive ou impropre		
compositions		
composition par onomatopée	composition par racine empruntée aux langues anciennes (grecque ou latine)	

Deux catégories entières qui n'ont pas été observées : les extensions sémantiques et les changements grammaticaux. Il est possible que la méthode utilisée n'ait pas été favorable au repérage de ces catégories. En effet, l'extracteur de termes se concentre sur les unités qui présentent un caractère spécifique par rapport à un corpus de référence. Les extensions sémantiques et les changements grammaticaux sont peut-être plus difficiles pour le logiciel à

repérer. Il est également possible que ces types de procédés soient moins productifs que la dérivation, la composition ou l'emprunt.

Les autres types des autres catégories n'ont pas de représentants dans nos corpus. Nous émettons comme première hypothèse qu'ils sont peut-être moins productifs que ceux qui sont représentés dans nos corpus. Il est aussi possible que les textes que nous avons choisis ne rendent pas compte de ces types de néologismes. Ce qui est le plus probable, toutefois, c'est que TermoStat ne les a pas repérés à cause des fréquences et de la polysémie.

## Conclusion

Cette étude avait pour objectif de faire le repérage de néologismes dans le domaine des médias sociaux à partir de deux corpus de textes en français qui empruntent des points de vue différents sur ce domaine.

Deux corpus de la communication et de l'informatique d'environ 200 000 mots chacun nous ont permis d'établir une terminologie des médias sociaux pour ces deux disciplines. L'extraction des termes s'est faite au moyen de TermoStat (Drouin, 2003), et grâce à son interface qui présente les candidats-termes en rang de spécificité, nous avons pu filtrer les termes pour ne garder que les 150 plus spécifiques de chaque corpus. Après avoir analysé ces deux listes de termes, nous avons pu faire ressortir les ressemblances et les différences entre les terminologies des médias sociaux de la communication et de l'informatique.

Grâce à une méthode s'appuyant principalement sur un corpus d'exclusion, nous avons pu valider les néologismes de ces listes, pour ensuite leur accoler une étiquette décrivant leur type de néologisme. Finalement, les résultats ont pu mettre de l'avant la contribution des domaines de la communication et de l'informatique à la néologie.

Nos résultats montrent que les terminologies des médias sociaux de la communication et de l'informatique se ressemblent sur plusieurs plans. En effet, la répartition des parties du discours et des termes simples et complexes était pratiquement la même. À partir des deux listes de 150 termes, nous avons repéré le même nombre de néologismes (32).

Notre travail a également confirmé l'intérêt d'une méthode de repérage des néologismes faisant appel à l'utilisation d'un extracteur de termes comme TermoStat et qu'une méthode de validation des néologismes comprenant plusieurs tests se révèle efficace pour repérer de la néologie. En effet, l'utilisation d'un simple corpus d'exclusion comme méthode de validation des néologismes n'était pas suffisante pour notre étude. Nous avons montré que la combinaison d'un corpus d'exclusion à un outil de visualisation de n-grammes est probante. Nous avons également atteint notre objectif qui visait à repérer différents types de néologismes, ce qui nous a permis d'établir que, pour les domaines de la communication et de l'informatique, ce sont les compositions par subordination qui ont été les plus utilisées entre 2010 et 2019 pour des textes

traitant des médias sociaux. La comparaison entre deux langues spécialisées nous permet de dire que deux domaines assez éloignés peuvent utiliser une terminologie commune si elles ont des thèmes en commun. Effectivement, de nos deux listes de 150 termes, 33 étaient communs à la communication et à l'informatique, alors que 122 étaient exclusifs à l'une et à l'autre. L'utilisation d'un thème ciblé et nouveau comme le nôtre se montre utile pour la recherche en néologie, car nous avons réussi à repérer 60 néologismes à partir de seulement deux listes de 150 termes issues de deux corpus de 200 000 mots.

L'étude comportait plusieurs défis et certaines limites. Tout d'abord, le montage des corpus s'est révélé difficile, particulièrement pour l'informatique. En effet, les textes en français étaient un peu plus ardues à trouver. De ce fait, les genres représentés dans chacun de nos corpus ne sont pas exactement les mêmes. Ils ne comportent aussi qu'environ 200 000 mots chacun ; de plus grands corpus auraient fourni plus de résultats. Le corpus de référence utilisé par l'extracteur de termes auquel nous avons fait appel n'était pas optimal. Il est compilé à partir de textes de 2002 et la plupart des termes des médias sociaux n'existaient pas à cette époque. De plus, cela ne coïncidait pas avec notre période cible qui était de 2010 à 2019. La période ne coïncidait pas non plus pour Google Books Ngram Viewer (Google, 2013) qui ne cherche pas dans les e-books publiés après 2008. Nous avons limité notre étude aux 150 termes les plus spécifiques de chacun de nos corpus en tenant pour acquis qu'ils étaient les plus représentatifs. Cependant, nous pourrions enrichir notre étude en analysant des termes n'ayant pas obtenu un score de spécificité aussi élevé. Il est possible que nos méthodes de filtrage ne soient pas parvenues à éliminer tous les termes non néologismes. Une dernière validation pourrait être faite par un expert.

Cette étude peut être reproduite à partir d'une autre langue, par exemple l'anglais. Elle pourrait également être reprise en choisissant d'autres disciplines qui pourraient venir enrichir la terminologie néologique que nous avons déjà repérée. Nous pourrions de même étendre le corpus, ce qui pourrait faire ressortir plus de néologismes. Elle pourrait de plus être appliquée à un autre thème que les médias sociaux à partir de deux autres disciplines.

## Références

- Barbu, A.-M. (2011). Monitoring neologisms in newspapers within the NEOROM project. *Memoirs of the Scientific Sections of the Romanian Academy*, 34.
- Biskri, I., Meunier, J.-G. et Joyal, S. (2004). *L'extraction des termes complexes: une approche modulaire semiautomatique*. Communication présentée aux 7<sup>es</sup> Journées internationales d'analyse statistique des données textuelles, Louvain La Neuve, Belgique.
- Boulanger, J.-C. (1979). Problématique d'une méthodologie de l'identification des néologismes en terminologie in *Neologie et Lexicologie*. Dans R. Adda (dir.), *Néologie et lexicologie : Hommage à Louis Guilbert*. Paris, France: Larousse, p. 36-46.
- Boulanger, J.-C. (2010). Sur l'existence des concepts de "néologie" et de "néologisme." Propos sur un paradoxe lexical et historique. *Actes del I congrès internacional de neologia de les llengües romàniques*, 31-72.
- Bourassa, N. et Drouin, P. (2005). Dépouillement terminologique assisté par ordinateur de sites Web spécialisés. *Meta: Journal des traducteurs*, 50(4). doi: 10.7202/019911ar
- Cabré, M. et de Yzaguirre, L. (1995). Stratégie pour la détection semi-automatique des néologismes de presse. *TTR: traduction, terminologie, rédaction*, 8(2), 89-100.
- Cartier, E. (2016). Neoveille, système de repérage et de suivi des néologismes en sept langues. *Neologica : Revue internationale de néologie*, 10, 101-131.
- Centre national de ressources textuelles et lexicales. (2012). Morphalou. Repéré à <https://www.cnrtl.fr/lexiques/morphalou/>
- Charaudeau, P., Maingueneau, D. et Adam, J.-M. (2002). *Dictionnaire d'analyse du discours*. Paris, France: du Seuil.
- Département d'informatique et de recherche opérationnelle. (2019). Expertises de recherche. Repéré à <https://diro.umontreal.ca/recherche/interets/>
- Département de communication. (2019). Expertises de recherche. Repéré à <https://com.umontreal.ca/recherche/interets/>
- dexonline. (2019). dexonline: dicționare ale limbii române. Repéré à <https://dexonline.ro>
- Drouin, P. (2003). TermoStat (Version 3.0) [Logiciel]. Repéré à <http://termostat.ling.umontreal.ca/index.php>
- Drouin, P. (2010). TermoStat 3.0 : Présentation du logiciel. Repéré à [http://termostat.ling.umontreal.ca/doc\\_termostat/doc\\_termostat.html](http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html)
- Dubuc, R. (2002). *Manuel pratique de terminologie*. (4e éd.). Brossard: Linguattech.
- Dupont, O. et Perticoz, L. (2016). Les ethos des jeunes professionnels de la communication sur LinkedIn. *Itinéraires. Littérature, textes, cultures*, 2015(3), 1-18.
- Dury, P. (2004). Building a bilingual diachronic corpus of ecology: The long road to completion. *Icame*, 28, 5-16.
- Estopà, R. (2009). Les composés populaires de la langue catalane : étude diachronique de la néologie. *Meta*, 54(3), 551-565. doi: <https://doi.org/10.7202/038314ar>
- Falk, I., Bernhard, D. et Gérard, C. (2014). *From Non Word to New Word: Automatically Identifying Neologisms in French Newspapers*. Paris: European Language Resources Assoc-Elra.
- Gérard, C. (2018). Le contexte, méconnu célèbre des études de néologie. Dans D. Bernhard, M. Hutchins-Boisseau, C. Gérard, T. Grass & A. Todirașcu-Courtier (dir.), *La néologie en contexte : cultures, situations, textes* (p. 9-21). Limoges, France: Lambert-Lucas.

- Goguey, E. (2019). *Dictionnaire de l'informatique et d'internet*. Repéré à <http://www.dicofr.com>
- Google. (2013). Google Books Ngrams Viewer. Repéré à <https://books.google.com/ngrams>
- Guilbert, L. (1973). Théorie du néologisme. *Cahiers de l'AIEF*, 25(1), 9-29.
- Humbley, J. (2006). La néologie : interface entre ancien et nouveau. *Langues et cultures : une histoire d'interface*, 91-103. doi: halshs-00168574
- Humbley, J. (2018). *La néologie en terminologie*. Limoges, France: Lambert-Lucas.
- Jalam, R. et Chauchat, J.-H. (2002). *Pourquoi les n-grammes permettent de classer des textes? Recherche de mots-clefs pertinents à l'aide des n-grammes caractéristiques*. Communication présentée aux 6<sup>es</sup> Journées internationales d'analyse statistique des données textuelles, Saint-Malo, France.
- Korenchuk, Y. (2018). Identification automatique de familles morphologiques et de néologismes. Dans D. Bernhard, M. Hutchins-Boisseau, C. Gérard, T. Grass & A. Todiraşcu-Courtier (dir.), *La néologie en contexte : cultures, situations, textes* (p. 265-279). Limoges: Lambert-Lucas.
- L'Homme, M.-C. (2004). *Sélection de termes dans un dictionnaire d'informatique: comparaison de corpus et critères lexicosémantiques*. Communication présentée à Euralex Lorient, France. Repéré à <https://pdfs.semanticscholar.org/9cad/9f686a89dcb8588790c2c19dcc64f09b1005.pdf>
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots. Les langages du politique*, 1(1), 127-165.
- Lafrance, A. (2016). Étude de l'évolution dans la terminologie de l'informatique en anglais avant et après 2006. (Mémoire de maîtrise, Université de Montréal). Repéré à <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/18768>
- Le Bohec, J. (2010). *Dictionnaire du journalisme et des médias*. Rennes: Rennes, France : Presses universitaires de Rennes.
- Lilen, H. (2014). *Dictionnaire informatique & numérique*. (2<sup>e</sup> éd.). Paris: Paris, France : First Interactive.
- lipa. (2018). Logoscope. Repéré à <http://logoscope.unistra.fr>
- Marshman, E. (2003). *Construction et gestion des corpus : Résumé et essai d'uniformisation du processus pour la terminologie*. Montréal, Québec: Observatoire de linguistique Sens-Texte, Université de Montréal.
- Milon, A. et Saint-Michel, S.-H. (2015). *Lexicom : les 3500 mots du marketing publicitaire, de la communication et des techniques de production*. (3<sup>e</sup> éd.). Levallois-Perret : Bréal.
- néologie. (1972). Dans *Trésor de la langue française informatisé*. Repéré à <http://stella.atilf.fr/Dendien/scripts/tlfiv5/advanced.exe?8;s=601643940>
- Néoveille. (2019). Néoveille, plateforme de repérage, analyse et suivi des néologismes en sept langues. Repéré à <https://tal.lipn.univ-paris13.fr/neoveille/html/login.php?action=login>
- Office québécois de la langue française. (1985). Néologie. Dans *Grand dictionnaire terminologique*. Repéré à [http://www.granddictionnaire.com/ficheOqlf.aspx?Id\\_Fiche=2199471](http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=2199471)
- Office québécois de la langue française. (2008). Néologisme. Dans *Grand dictionnaire terminologique*. Repéré à [http://www.granddictionnaire.com/ficheOqlf.aspx?Id\\_Fiche=8348998](http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=8348998)

- Office québécois de la langue française. (2018). Vocabulaire des réseaux sociaux. Repéré à <https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/vocabulaire-reseaux-sociaux.aspx>
- Paquin, A. (2007). Étude de la néologie dans la terminologie du terrorisme avant et après septembre 2001: une approche lexicométrique. (Mémoire de maîtrise, Université de Montréal). Repéré à <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/16581>.
- Pruvost, J. et Sablayrolles, J.-F. (2016). *Les néologismes*. (Troisième édition mise à jour). Paris: Presses universitaires de France.
- Renaud, C. (2014). *Conception d'un outil d'analyse et de visualisation des mèmes Internet: le cas du réseau social chinois Sina Weibo*. (ParisTech Telecom, Paris, France).
- Sablayrolles, J.-F. (2000). *La néologie en français contemporain : examen du concept et analyse de productions néologiques récentes*. Paris: H. Champion.
- Sablayrolles, J.-F. (2002). Fondements théoriques des difficultés pratiques du traitement des néologismes. *Revue française de linguistique appliquée*, VII(1), 97-111.
- Sablayrolles, J.-F. (2006). La néologie aujourd'hui. Dans G. Claude (dir.), *A la recherche du mot : De la langue au discours* (p. 141-157): Lambert-Lucas.
- Schmid, H. (1995). TreeTagger. Repéré à <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Termium plus. (2018). *Lexique des médias sociaux*. Repéré à <https://www.btb.termiumplus.gc.ca/publications/media-soc-fra.html>
- Usito, D. (2017). Les avantages d'Usito. Repéré à <https://www.usito.com/avantages.html>



## Annexe 1 : Le corpus de la communication

Corpus		Nom	
communication		abcpourosbl	
Nombre de mots	Date de la recherche	Date de publication	
7934	27/02/2019	non daté	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	initié	non-initié
Titre du document		Titre du texte	
L'ABC des médias sociaux pour les organismes sans but lucratif		L'ABC des médias sociaux pour les organismes sans but lucratif	
Nom de l'auteur		Retraits	
Kim Fuller		page couverture, table des matières	
URL			
<a href="https://www.fondationbombardier.ca/images/pds_ressources/LABC_des_medias_sociaux.pdf">https://www.fondationbombardier.ca/images/pds_ressources/LABC_des_medias_sociaux.pdf</a>			
Code de recherche		Thème	
médias sociaux communication		Organisation sans but lucratif	
Commentaire			
Document d'information pour les organisations sans but lucratif pour leur utilisation des médias sociaux.			
Référence			
Fuller, K. <i>L'ABC des médias sociaux pour les organismes sans but lucratif</i> . (n. d.) Repéré à <a href="https://www.fondationbombardier.ca/images/pds_ressources/LABC_des_medias_sociaux.pdf">https://www.fondationbombardier.ca/images/pds_ressources/LABC_des_medias_sociaux.pdf</a>			

Corpus		Nom	
communication		administrationquebecoise	
Nombre de mots	Date de la recherche	Date de publication	
8575	07/03/2019	2017	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
Rapport officiel	spécialisé	spécialiste	non-initié
Titre du document		Titre du texte	
Addenda sur Twitter : écrire pour l'État à l'ère du Web 2.0		Addenda sur Twitter : écrire pour l'État à l'ère du Web 2.0	
Nom de l'auteur		Retraits	
Julie Bérubé		page couverture, remerciements, table des matières, tableaux, liste des tableaux, liste des abréviations et des sigles, caractéristiques, bibliographie, annexe	
URL			
<a href="http://numerique.banq.qc.ca/patrimoine/details/52327/2930716">http://numerique.banq.qc.ca/patrimoine/details/52327/2930716</a>			
Code de recherche		Thème	
médias sociaux		communication politique	
Commentaire			
Rapport du Conseil supérieur de la langue française sur la langue utilisée par l'État sur Twitter.			
Référence			
Bérubé, J. (2017). <i>Addenda sur Twitter : écrire pour l'État à l'ère du Web 2.0</i> . Repéré à <a href="http://numerique.banq.qc.ca/patrimoine/details/52327/2930716">http://numerique.banq.qc.ca/patrimoine/details/52327/2930716</a>			

Corpus		Nom	
communication		affectsfacebook	
Nombre de mots	Date de la recherche	Date de publication	
4775	27/02/2019	2017	
Genre du texte	Spécialisation	Type d'auteur	Destinataire

revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Les affects en partage sur Facebook entre supposé secret et déclaration		Les affects en partage sur Facebook entre supposé secret et déclaration	
Nom de l'auteur		Retraits	
Sansra Lemeilleur		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/2896">https://journals.openedition.org/rfsic/2896</a>			
Code de recherche		Thème	
médias sociaux		médias sociaux, technologies interactives	
Commentaire			
La venue des billets d'humeur numérique en 2016 change la dynamique sur Facebook. Code de recherche : médias sociaux			
Référence			
<b>Sandra</b> Lemeilleur, « Les affects en partage sur Facebook entre supposé secret et déclaration d'amour », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 11   2017, mis en ligne le 01 août 2017, consulté le 27 février 2019. URL : <a href="http://journals.openedition.org/rfsic/2896">http://journals.openedition.org/rfsic/2896</a> ; DOI : 10.4000/rfsic.2896			

Corpus		Nom	
communication		algerie	
Nombre de mots		Date de la recherche	
6961		27/02/2019	
Date de publication		2018	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Les réseaux sociaux et la participation politique en Algérie		Les réseaux sociaux et la participation politique en Algérie	
Nom de l'auteur		Retraits	
Tahar Ouchiha		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
URL			
<a href="https://journals.openedition.org/communication/8846">https://journals.openedition.org/communication/8846</a>			
Code de recherche		Thème	
médias sociaux		technologies interactives	
Commentaire			
La place des médias sociaux dans la politique algérienne.			
Référence			
<b>Tahar</b> Ouchiha, « Les réseaux sociaux et la participation politique en Algérie », <i>Communication</i> [En ligne], vol. 35/2   2018, mis en ligne le 07 décembre 2018, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/communication/8846">http://journals.openedition.org/communication/8846</a> ; DOI : 10.4000/communication.8846			

Corpus		Nom	
communication		antivaccins	
Nombre de mots		Date de la recherche	
434		27/02/2019	
Date de publication		25/02/2019	
Genre du texte		Spécialisation	
journal		non spécialisé	
Type d'auteur		Destinataire	
journaliste		non-initié	
Titre du document		Titre du texte	

Pinterest et YouTube disent agir contre les messages anti-vaccins		Pinterest et YouTube disent agir contre les messages anti-vaccins	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photo	
URL			
<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/25/01-5216058-pinterest-et-youtube-disent-agir-contre-les-messages-anti-vaccins.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/25/01-5216058-pinterest-et-youtube-disent-agir-contre-les-messages-anti-vaccins.php</a>			
Code de recherche		Thème	
médias sociaux		Controverses publiques, communication et santé	
Commentaire			
Article sur le retrait de Pinterest et Facebook de messages anti-vaccins.			
Référence			
Agence France-Presse. (2019). Pinterest et YouTube disent agir contre les messages anti-vaccins . <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/25/01-5216058-pinterest-et-youtube-disent-agir-contre-les-messages-anti-vaccins.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/25/01-5216058-pinterest-et-youtube-disent-agir-contre-les-messages-anti-vaccins.php</a>			

Corpus		Nom	
communication		apprentissage langues	
Nombre de mots		Date de la recherche	
5518		27/02/19	
Date de publication		2012	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
De l'attrait des médias sociaux pour l'apprentissage des langues : regard sur l'état de l'art		De l'attrait des médias sociaux pour l'apprentissage des langues : regard sur l'état de l'art	
Nom de l'auteur		Retraits	
Katerina Zourou		remerciements, résumés, nom, mots-clés, plan	
URL			
<a href="https://journals.openedition.org/alsic/2485">https://journals.openedition.org/alsic/2485</a>			
Code de recherche		Thème	
médias sociaux		Environnement d'apprentissage	
Commentaire			
Les médias sociaux sont utilisés pour l'apprentissage des langues. Traduction de l'anglais.			
Référence			
<b>Katerina</b> Zourou, « De l'attrait des médias sociaux pour l'apprentissage des langues – Regard sur l'état de l'art », <i>Alsic</i> [En ligne], Vol. 15, n°1   2012, mis en ligne le 30 mars 2012, Consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/alsic/2485">http://journals.openedition.org/alsic/2485</a>			

Corpus		Nom	
communication		autoformation	
Nombre de mots		Date de la recherche	
5189		28/02/2019	
Date de publication		2013	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Autoformation et réseaux virtuels : ou à la recherche des traces autoformatives dans les méandres réticulaires de YouTube		Autoformation et réseaux virtuels : ou à la recherche des traces autoformatives dans les méandres réticulaires de YouTube	
Nom de l'auteur		Retraits	

Pascal Cyrot et Christophe Jeunesse	titre, date, nom, résumé, table des matières, bibliographie, notes, citations, index, références		
URL			
<a href="https://journals.openedition.org/dms/137">https://journals.openedition.org/dms/137</a>			
Code de recherche		Thème	
youtube		Environnement d'apprentissage	
Commentaire			
Le média social YouTube permet maintenant l'autoformation par ses tonnes de vidéos explicatives.			
Référence			
<b>Pascal Cyrot et Christophe Jeunesse</b> , « Autoformation et réseaux virtuels », <i>Distances et médiations des savoirs</i> [En ligne], 1   2012, mis en ligne le 12 décembre 2012, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/dms/137">http://journals.openedition.org/dms/137</a> ; DOI : 10.4000/dms.137			

Corpus		Nom	
communication		avantages/inconvénients	
Nombre de mots		Date de la recherche	Date de publication
2201		27/02/2019	2019
Genre du texte	Spécialisation	Type d'auteur	Destinataire
blogue	non spécialisé	blogueur	non-initié
Titre du document		Titre du texte	
20 avantages et inconvénients des médias sociaux		20 avantages et inconvénients des médias sociaux	
Nom de l'auteur		Retraits	
Elise T. Turcotte et Sarah Laou		images, numéros de page, références	
URL			
<a href="https://www.selection.ca/sante/vivre-sainement/les-avantages-et-les-inconvénients-des-reseaux-sociaux/">https://www.selection.ca/sante/vivre-sainement/les-avantages-et-les-inconvénients-des-reseaux-sociaux/</a>			
Code de recherche		Thème	
médias sociaux		Réseaux sociaux	
Commentaire			
Capsules ludiques sur les avantages et les désavantages de l'utilisation des médias sociaux.			
Référence			
T. Turcotte, E. et S. Laou. (2019). 20 avantages et inconvénients des médias sociaux. <i>Sélection</i> . Repéré à <a href="https://www.selection.ca/sante/vivre-sainement/les-avantages-et-les-inconvénients-des-reseaux-sociaux/">https://www.selection.ca/sante/vivre-sainement/les-avantages-et-les-inconvénients-des-reseaux-sociaux/</a>			

Corpus		Nom	
communication		batirstrategie	
Nombre de mots		Date de la recherche	Date de publication
1107		27/02/2019	2017
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	spécialiste	non-initié
Titre du document		Titre du texte	
Comment bâtir votre stratégie sur les médias sociaux ?		Comment bâtir votre stratégie sur les médias sociaux ?	
Nom de l'auteur		Retraits	
anthedesign		images	
URL			
<a href="https://www.anthedesign.fr/sociaux/reseaux-sociaux-guide-pratique-community-manager/">https://www.anthedesign.fr/sociaux/reseaux-sociaux-guide-pratique-community-manager/</a>			
Code de recherche		Thème	
médias sociaux communication		Communication tactiques d'entreprises	
Commentaire			
Stratégies de mises en place de médias sociaux pour les spécialistes du design.			
Référence			

Anthedesign. (2017). *Comment bâtir votre stratégie sur les médias sociaux ?* Repéré à <https://www.anthedesign.fr/sociaux/reseaux-sociaux-guide-pratique-community-manager/>

Corpus		Nom	
communication		biographisation	
Nombre de mots	Date de la recherche	Date de publication	
3619	28/02/2019	2016	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Facebook, Snapchat : Instances de biographisation partagées		Facebook, Snapchat : Instances de biographisation partagées	
Nom de l'auteur		Retraits	
Anne Dizerbo		titre, date, nom, résumé, table des matières, bibliographie, notes, citations, index, références	
URL			
<a href="https://www.cairn.info/revue-le-sujet-dans-la-cite-2016-1-page-129.htm">https://www.cairn.info/revue-le-sujet-dans-la-cite-2016-1-page-129.htm</a>			
Code de recherche		Thème	
snapchat		Interactions sociales	
Commentaire			
Les réseaux sociaux Facebook et Snapchat servent de biographies en lignes pour les jeunes de la nouvelle génération.			
Référence			
Dizerbo, A. (2016). Facebook, Snapchat : Instances de biographisation partagées. <i>Le sujet dans la cité</i> .5(1). 129-142. Repéré à <a href="https://www.cairn.info/revue-le-sujet-dans-la-cite-2016-1-page-129.htm">https://www.cairn.info/revue-le-sujet-dans-la-cite-2016-1-page-129.htm</a>			

Corpus		Nom	
communication		communication digitale	
Nombre de mots	Date de la recherche	Date de publication	
1771	27/02/2019	2019	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	spécialiste	non-initié
Titre du document		Titre du texte	
Les médias sociaux : c'est quoi et pourquoi les utiliser ?		Les médias sociaux : c'est quoi et pourquoi les utiliser ?	
Nom de l'auteur		Retraits	
Ludovic Salenne		images, commentaires, date, nom	
URL			
<a href="https://www.ludosln.net/medias-sociaux-definition-et-utilites/">https://www.ludosln.net/medias-sociaux-definition-et-utilites/</a>			
Code de recherche		Thème	
médias sociaux communication		Études des médias	
Commentaire			
Présentation des médias sociaux, de leur fonctionnement et de leur utilisation.			
Référence			
Salenne, L. (2019). <i>Les médias sociaux : c'est quoi et pourquoi les utiliser ?</i> Repéré à <a href="https://www.ludosln.net/medias-sociaux-definition-et-utilites/">https://www.ludosln.net/medias-sociaux-definition-et-utilites/</a>			

Corpus		Nom	
communication		communiquer	
Nombre de mots	Date de la recherche	Date de publication	

3374	27/02/2019	2012	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Changement de logique et des Arts de faire dans les pratiques communicationnelles avec les médias sociaux		Changement de logique et des Arts de faire dans les pratiques communicationnelles avec les médias sociaux	
Nom de l'auteur		Retraits	
Francine Charest et Anne-Marie Gauthier		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
URL			
<a href="https://journals.openedition.org/communicationorganisation/3696">https://journals.openedition.org/communicationorganisation/3696</a>			
Code de recherche		Thème	
médias sociaux		Époque contemporaine	
Commentaire			
À l'ère des médias sociaux, les techniques du monde de la communication ont changé.			
<p><b>Francine Charest et Anne-Marie Gauthier</b>, « Changement de logique et des Arts de faire dans les pratiques communicationnelles avec les médias sociaux », <i>Communication et organisation</i> [En ligne], 41   2012, mis en ligne le 01 juin 2015, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/communicationorganisation/3696">http://journals.openedition.org/communicationorganisation/3696</a> ; DOI : <a href="https://doi.org/10.4000/communicationorganisation.3696">10.4000/communicationorganisation.3696</a></p>			

Corpus		Nom	
communication		controlesocial	
Nombre de mots	Date de la recherche	Date de publication	
2632	27/02/2019	2011	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
L'usage des réseaux socionumériques : une intériorisation douce et progressive du contrôle social		L'usage des réseaux socionumériques : une intériorisation douce et progressive du contrôle social	
Nom de l'auteur		Retraits	
Serge Proulx et Mary Jane Kwok Choon		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
URL			
<a href="https://www.cairn.info/revue-hermes-la-revue-2011-1-page-105.htm?contenu=resume">https://www.cairn.info/revue-hermes-la-revue-2011-1-page-105.htm?contenu=resume</a>			
Code de recherche		Thème	
médias sociaux		Études sur la surveillance	
Commentaire			
Les réseaux socionumériques constituent une figure emblématique de la « société contrôle ».			
<p>Proulx, S. &amp; Kwok Choon, M. (2011). L'usage des réseaux socionumériques : une intériorisation douce et progressive du contrôle social. <i>Hermès, La Revue</i>, 59(1), 105-111. <a href="https://www.cairn.info/revue-hermes-la-revue-2011-1-page-105.htm">https://www.cairn.info/revue-hermes-la-revue-2011-1-page-105.htm</a>.</p>			

Corpus		Nom	
communication		cotedivoire	
Nombre de mots	Date de la recherche	Date de publication	
4809	27/02/2019	2018	

Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Du réseau social traditionnel au réseau social numérique : pistes de réflexion pour une éducation aux médias sociaux numériques en Côte d'Ivoire		Du réseau social traditionnel au réseau social numérique : pistes de réflexion pour une éducation aux médias sociaux numériques en Côte d'Ivoire	
Nom de l'auteur		Retraits	
Gnéré Blama Dagnogo		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/3495">https://journals.openedition.org/rfsic/3495</a>			
Code de recherche		Thème	
médias sociaux		Interactions sociales	
Commentaire			
L'implantation de l'utilisation des médias sociaux par la population de la Côte d'Ivoire.			
Référence			
<p><b>Gnéré Blama Dagnogo</b>, « Du réseau social traditionnel au réseau social numérique : pistes de réflexion pour une éducation aux médias sociaux numériques en Côte d'Ivoire », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 12   2018, mis en ligne le 05 février 2018, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/3495">http://journals.openedition.org/rfsic/3495</a> ; DOI : 10.4000/rfsic.3495</p>			

Corpus	Nom		
communication	cyberbullying		
Nombre de mots	Date de la recherche	Date de publication	
3505	28/02/2019	2017	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
mémoire	spécialisé	expert	initié
Titre du document		Titre du texte	
L'utilisation des réseaux sociaux (Snapchat, Whatsapp et Instagram) et le cyberbullying		L'utilisation des réseaux sociaux (Snapchat, Whatsapp et Instagram) et le cyberbullying	
Nom de l'auteur		Retraits	
Leticia Ribeiro De Carvalho, Saima Nasratulla et Eric Tardif		Nous n'avons gardé que le chapitre 1 (introduction) et le chapitre 2 (méthodologie)	
URL			
<a href="http://doc.rero.ch/record/306056">http://doc.rero.ch/record/306056</a>			
Code de recherche		Thème	
snapchat		Controverses sociotechniques	
Commentaire			
Mémoire sur le cyberbullying à travers les réseaux sociaux.			
Référence			
<p>Ribeiro De Carvalho, L. Nasratulla, S. et Tardif E. (Dir.). (2017). <i>L'utilisation des réseaux sociaux (Snapchat, Whatsapp et Instagram) et le cyberbullying</i> (Mémoire professionnel, Haute école pédagogique du canton de Vaud, Suisse). Repéré à <a href="http://doc.rero.ch/record/306056">http://doc.rero.ch/record/306056</a>.</p>			

Corpus	Nom		
communication	erecruitment		
Nombre de mots	Date de la recherche	Date de publication	
9080	27/02/2019	2011	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	

L'apparition des médias sociaux dans l'e-GRH : gestion de la marque employeur et e-recrutement		L'apparition des médias sociaux dans l'e-GRH : gestion de la marque employeur et e-recrutement	
Nom de l'auteur		Retraits	
Aurélie Girard, Bernard Fallery et Florence Rondhain		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, titre	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-00843689/">https://hal.archives-ouvertes.fr/hal-00843689/</a>			
Code de recherche		Thème	
médias sociaux		Pratiques situées et institutionnalisées en milieu de travail	
Commentaire			
Cas concrets de l'utilisation des médias sociaux dans l'e-recrutement			
Référence			
Aurélie Girard, Bernard Fallery, Florence Rodhain. L'apparition des médias sociaux dans l'e-GRH : gestion de la marque employeur et e-recrutement . <i>16 ème Congrès de l'AIM</i> , 2011, St Denis de La Réunion, France. pp.1-18, 2011. Repéré à <a href="https://hal.archives-ouvertes.fr/hal-00843689/">https://hal.archives-ouvertes.fr/hal-00843689/</a> .			

Corpus		Nom	
communication		fearofmissingout	
Nombre de mots		Date de la recherche	
1193		27/02/2019	
Date de publication		2019	
Genre du texte		Spécialisation	
blogue		non spécialisé	
Type d'auteur		Destinataire	
blogueur		non-initié	
Titre du document		Titre du texte	
Réseaux sociaux : 8 façons de dominer votre syndrome de FOMO		Réseaux sociaux : 8 façons de dominer votre syndrome de FOMO	
Nom de l'auteur		Retraits	
Rachel Sokol		images, numéros de page, références	
URL			
<a href="https://www.selection.ca/sante/vivre-sainement/8-facons-de-dominer-votre-syndrome-fomo/">https://www.selection.ca/sante/vivre-sainement/8-facons-de-dominer-votre-syndrome-fomo/</a>			
Code de recherche		Thème	
médias sociaux communication		Cyberculture, Interaction humain-ordinateur	
Commentaire			
Article ludique sur comment gérer son syndrome de <i>fear of missing out</i> .			
Référence			
Sokol, R. (2019). Réseaux sociaux : 8 façons de dominer votre syndrome de FOMO. <i>Sélection</i> . Repéré à <a href="https://www.selection.ca/sante/vivre-sainement/8-facons-de-dominer-votre-syndrome-fomo/">https://www.selection.ca/sante/vivre-sainement/8-facons-de-dominer-votre-syndrome-fomo/</a> .			

Corpus		Nom	
communication		feuxdeforet	
Nombre de mots		Date de la recherche	
8774		27/02/2019	
Date de publication		2010	
Genre du texte		Spécialisation	
thèse		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Communication relation et médias sociaux : étude du cas de la sensibilisation aux feux de forêt		Introduction	
Nom de l'auteur		Retraits	
Bruno Asdourian		Nous n'avons gardé que l'introduction .	
URL			
<a href="https://www.theses.fr/2010AIX22300">https://www.theses.fr/2010AIX22300</a>			
Code de recherche		Thème	



médias sociaux	Communication en situation de crise
Commentaire	
Thèse doctorale sur l'utilisation des médias sociaux dans la sensibilisation aux feux de forêt.	
Référence	
Asdourian, B. (2010). <i>Communication relation et médias sociaux : étude du cas de la sensibilisation aux feux de forêt</i> (Thèse de doctorat, Aix-Marseille Université, France). Repéré à <a href="https://www.theses.fr/2010AIX22300">https://www.theses.fr/2010AIX22300</a> .	

Corpus		Nom	
communication		gangsternumerique	
Nombre de mots		Date de la recherche	
286		18/02/2019	
Date de la recherche		Date de publication	
27/02/2019		18/02/2019	
Genre du texte		Spécialisation	
journal		non spécialisé	
Type d'auteur		Destinataire	
journaliste		non-initié	
Titre du document		Titre du texte	
Facebook qualifié de «gangster numérique» par des parlementaires britanniques		Facebook qualifié de «gangster numérique» par des parlementaires britanniques	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photo	
URL			
<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215228-facebook-qualifie-de-gangster-numerique-par-des-parlementaires-britanniques.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215228-facebook-qualifie-de-gangster-numerique-par-des-parlementaires-britanniques.php</a>			
Code de recherche		Thème	
médias sociaux		Controverses sociotechniques, controverses publiques	
Commentaire			
Facebook sert à la désinformation.			
Référence			
Agence France-Presse. (2019). Facebook qualifié de «gangster numérique» par des parlementaires britanniques. <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215228-facebook-qualifie-de-gangster-numerique-par-des-parlementaires-britanniques.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215228-facebook-qualifie-de-gangster-numerique-par-des-parlementaires-britanniques.php</a> .			

Corpus		Nom	
communication		homosexuelmusulman	
Nombre de mots		Date de la recherche	
469		13/02/2019	
Date de la recherche		Date de publication	
27/02/2019		13/02/2019	
Genre du texte		Spécialisation	
journal		non spécialisé	
Type d'auteur		Destinataire	
journaliste		non-initié	
Titre du document		Titre du texte	
Disparition d'un compte Instagram narrant la vie d'un homosexuel musulman en Indonésie		Disparition d'un compte Instagram narrant la vie d'un homosexuel musulman en Indonésie	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photo	
URL			
<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/13/01-5214536-disparition-dun-compte-instagram-narrant-la-vie-dun-homosexuel-musulman-en-indonesie.php?utm_categorieinterne=trafficedrivers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS4">https://www.lapresse.ca/techno/reseaux-sociaux/201902/13/01-5214536-disparition-dun-compte-instagram-narrant-la-vie-dun-homosexuel-musulman-en-indonesie.php?utm_categorieinterne=trafficedrivers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS4</a>			
Code de recherche		Thème	
médias sociaux		Controverses sociotechniques, controverses publiques	

Commentaire	
Un le compte Instagram d'un musulman indonésien a été supprimé de la plateforme	
Référence	
Agence France-Presse. (2019). Disparition d'un compte Instagram narrant la vie d'un homosexuel musulman en Indonésie. <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/13/01-5214536-disparition-dun-compte-instagram-narrant-la-vie-dun-homosexuel-musulman-en-indonesie.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS4">https://www.lapresse.ca/techno/reseaux-sociaux/201902/13/01-5214536-disparition-dun-compte-instagram-narrant-la-vie-dun-homosexuel-musulman-en-indonesie.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS4</a> .	

Corpus		Nom					
communication		hopitaux					
Nombre de mots		Date de la recherche		Date de publication			
4382		27/02/2019		2017			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		spécialisé		expert		initié	
Titre du document				Titre du texte			
L'impact des réseaux sociaux sur la communication hospitalière aux États-Unis				L'impact des réseaux sociaux sur la communication hospitalière aux États-Unis			
Nom de l'auteur				Retraits			
Pablo Medina				références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index			
URL							
<a href="https://journals.openedition.org/rfsic/3080">https://journals.openedition.org/rfsic/3080</a>							
Code de recherche			Thème				
médias sociaux			Communication et santé, communication organisationnelle				
Commentaire							
La communication par les médias sociaux dans les hôpitaux américains a changé les pratiques professionnelles.							
Référence							
<b>Pablo Medina</b> , « L'impact des réseaux sociaux sur la communication hospitalière aux États-Unis », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 11   2017, mis en ligne le 01 août 2017, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/3080">http://journals.openedition.org/rfsic/3080</a> ; DOI : 10.4000/rfsic.3080							

Corpus		Nom					
communication		infomédiation					
Nombre de mots		Date de la recherche		Date de publication			
10350		16/03/2019		2012			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		Spécialisé		expert		initié	
Titre du document				Titre du texte			
Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter				Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter			
Nom de l'auteur				Retraits			
Bernhard Rieder et Nikos Smyrnaio				page couverture, en-têtes, pieds de pages, notes en marge, références			
URL							
<a href="https://www.cairn.info/revue-reseaux-2012-6-page-105.htm">https://www.cairn.info/revue-reseaux-2012-6-page-105.htm</a>							
Code de recherche			Thème				
twitter			Époque contemporaine, journalisme				
Commentaire							

L'information maintenant présente sur Internet grâce à Twitter.		
Référence		
Rieder, B. & Smyrniotis, N. (2012). Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter. <i>Réseaux</i> , 176(6), 105-139. doi:10.3917/res.176.0105.		

Corpus		Nom	
communication		liguedulol	
Nombre de mots		Date de la recherche	
768		27/02/2019	
Date de publication			
		12/02/2019	
Genre du texte		Spécialisation	
journal		non spécialisé	
Type d'auteur		Destinataire	
journaliste		non-initié	
Titre du document		Titre du texte	
« Ligue du LOL » : une violence bien ancrée sur les médias sociaux		« Ligue du LOL » : une violence bien ancrée sur les médias sociaux	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photos	
URL			
https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214409-ligue-du-lol-une-violence-bien-ancree-sur-les-reseaux-sociaux.php?utm_categorieinterne=traffidriviers&utm_cotentuinterne=cyberpresse_cinemaStandard_3852482_section_POS6			
Code de recherche		Thème	
médias sociaux		Controverses sociotechniques	
Commentaire			
En France, après les années 2000, le harcèlement sur le Web s'est bien installé.			
Référence			
Agence France-Presse. (2019). « Ligue du LOL » : une violence bien ancrée sur les médias sociaux. <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214409-ligue-du-lol-une-violence-bien-ancree-sur-les-reseaux-sociaux.php?utm_categorieinterne=traffidriviers&amp;utm_cotentuinterne=cyberpresse_cinemaStandard_3852482_section_POS6">https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214409-ligue-du-lol-une-violence-bien-ancree-sur-les-reseaux-sociaux.php?utm_categorieinterne=traffidriviers&amp;utm_cotentuinterne=cyberpresse_cinemaStandard_3852482_section_POS6</a> .			

Corpus		Nom	
communication		linkedincommunication	
Nombre de mots		Date de la recherche	
7404		14/03/2019	
Date de publication			
		2016	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Les ethos des jeunes professionnels de la communication sur LinkedIn		Les ethos des jeunes professionnels de la communication sur LinkedIn	
Nom de l'auteur		Retraits	
Olivier Dupont et Lucien Perticoz		résumés, entrées d'index, plan, bibliographie	
URL			
https://journals.openedition.org/itineraires/3023			
Code de recherche		Thème	
linkedin		Identité occupationnelle, pratiques situées et institutionnalisées en milieu de travail	
Commentaire			
L'utilisation de LinkedIn par les jeunes professionnels de la communication.			

Référence		
<b>Olivier Dupont et Lucien Perticoz</b> , « Les ethos des jeunes professionnels de la communication sur LinkedIn », <i>Itinéraires</i> [En ligne], 2015-3   2016, mis en ligne le 01 juillet 2016, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/itineraires/3023">http://journals.openedition.org/itineraires/3023</a> ; DOI : 10.4000/itineraires.3023		

Corpus		Nom	
communication		marketingterritorial	
Nombre de mots		Date de la recherche	Date de publication
2064		28/02/2019	2010
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Les médias sociaux au service du marketing territorial : une approche exploratoire		Les médias sociaux, des nouveaux outils au service des territoires ?	
Nom de l'auteur		Retraits	
Michel Barabel, Samuel Mayol et Olivier Meier		nous n'avons gardé que le chapitre 2 de l'article	
URL			
<a href="https://www.cairn.info/revue-management-et-avenir-2010-2-page-233.htm">https://www.cairn.info/revue-management-et-avenir-2010-2-page-233.htm</a>			
Code de recherche		Thème	
médias sociaux		Communication politique	
Commentaire			
Les territoires utilisent les médias sociaux pour faire leur marketing.			
Référence			
Barabel, M., Mayol, S. & Meier, O. (2010). Les médias sociaux au service du marketing territorial : une approche exploratoire. <i>Management &amp; Avenir</i> , 32(2), 233-253. doi:10.3917/mav.032.0233.			

Corpus		Nom	
communication		militairesrusses	
Nombre de mots		Date de la recherche	Date de publication
351		27/02/2019	12/02/2019
Genre du texte	Spécialisation	Type d'auteur	Destinataire
journal	non spécialisé	journaliste	non-initié
Titre du document		Titre du texte	
Russie: les militaires appelés à la discrétion numérique		Russie: les militaires appelés à la discrétion numérique	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photos	
URL			
<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214422-russie-les-militaires-appelles-a-la-discretion-numerique.php?utm_categorieinterne=traffiddrivers&amp;utm_contenuinterne=cyberpresse_cinemaS_tandard_3852482_section_POS5">https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214422-russie-les-militaires-appelles-a-la-discretion-numerique.php?utm_categorieinterne=traffiddrivers&amp;utm_contenuinterne=cyberpresse_cinemaS_tandard_3852482_section_POS5</a>			
Code de recherche		Thème	
médias sociaux		Études sur la surveillance	
Commentaire			
L'armée russe demande à ses militaires de rester discret sur les médias sociaux.			
Référence			

Agence France-Presse. (2019). Russie: les militaires appelés à la discrétion numérique. *La Presse*. Repéré à [https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214422-russie-les-militaires-appeles-a-la-discretion-numerique.php?utm\\_categorieinterne=traficdrivers&utm\\_contenuinterne=cyberpresse\\_cinemaStandard\\_3852482\\_section\\_POS5](https://www.lapresse.ca/techno/reseaux-sociaux/201902/12/01-5214422-russie-les-militaires-appeles-a-la-discretion-numerique.php?utm_categorieinterne=traficdrivers&utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS5).

Corpus		Nom	
communication		multicanal	
Nombre de mots	Date de la recherche	Date de publication	
1109	27/02/2019	2011	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	spécialiste	non-initié
Titre du document		Titre du texte	
Les médias sociaux dans la stratégie marketing communication et multi canal		Les médias sociaux dans la stratégie marketing communication et multi canal	
Nom de l'auteur		Retraits	
Jacques Barbier		image	
URL			
<a href="http://www.actuentreprise.com/nos-articles/les-medias-sociaux-dans-la-strategie-marketing-communication-et-multi-canal/">http://www.actuentreprise.com/nos-articles/les-medias-sociaux-dans-la-strategie-marketing-communication-et-multi-canal/</a>			
Code de recherche		Thème	
médias sociaux communication		Communications tactiques d'entreprises	
Commentaire			
Stratégies pour les entreprises.			
Référence			
Barbier, J. (2011). Les médias sociaux dans la stratégie marketing communication et multi canal. <i>Actu'entreprise</i> . Repéré à <a href="http://www.actuentreprise.com/nos-articles/les-medias-sociaux-dans-la-strategie-marketing-communication-et-multi-canal/">http://www.actuentreprise.com/nos-articles/les-medias-sociaux-dans-la-strategie-marketing-communication-et-multi-canal/</a> .			

Corpus		Nom	
communication		netart	
Nombre de mots	Date de la recherche	Date de publication	
4156	27/02/2019	2012	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Net art : créer à l'ère des médias numériques		Net art : créer à l'ère des médias numériques	
Nom de l'auteur		Retraits	
Jean-Paul Fourmentraux		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/179">https://journals.openedition.org/rfsic/179</a>			
Code de recherche		Thème	
médias sociaux		Arts médiatiques	
Commentaire			
La place de l'art numérique dans les médias sociaux.			
Référence			
<b>Jean-Paul</b> Fourmentraux, « Net art – Créer à l'ère des médias numériques », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 1   2012, mis en ligne le 01 septembre 2012, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/179">http://journals.openedition.org/rfsic/179</a> ; DOI : 10.4000/rfsic.179			

Corpus		Nom	
communication		oeilmoderne	
Nombre de mots	Date de la recherche	Date de publication	
1984	27/02/2019	2019	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	spécialiste	non-initié
Titre du document		Titre du texte	
Le défi de la transition numérique		communiquer sur les médias sociaux	
Nom de l'auteur		Retraits	
Un œil moderne		images	
URL			
<a href="https://www.un-oeil-moderne.com/Publications/communiquer-sur-les-reseaux-sociaux">https://www.un-oeil-moderne.com/Publications/communiquer-sur-les-reseaux-sociaux</a>			
Code de recherche		Thème	
médias sociaux communication		Cyberculture	
Commentaire			
Revue des différents types de médias sociaux de la manière de communiquer sur ceux-ci.			
Référence			
Un œil moderne. (2019). Le défi de la transition numérique : communiquer sur les médias sociaux. <i>Un œil moderne</i> . Repéré à <a href="https://www.un-oeil-moderne.com/Publications/communiquer-sur-les-reseaux-sociaux">https://www.un-oeil-moderne.com/Publications/communiquer-sur-les-reseaux-sociaux</a> .			

Corpus		Nom	
Communication		personnages fictifs	
Nombre de mots	Date de la recherche	Date de publication	
3947	28/02/2019	2018	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Immersion, médias sociaux et récit transmédiatique		Immersion, médias sociaux et récit transmédiatique	
Nom de l'auteur		Retraits	
Philippe Gauthier		résumés, entrées d'index, plan, bibliographie, notes	
URL			
<a href="https://journals.openedition.org/communication/8081">https://journals.openedition.org/communication/8081</a>			
Code de recherche		Thème	
médias sociaux		Culture populaire	
Commentaire			
Quel est l'impact de la création de pages de médias sociaux de personnages fictifs.			
Référence			
<b>Philippe</b> Gauthier, « Immersion, médias sociaux et récit transmédiatique », <i>Communication</i> [En ligne], vol. 35/2   2018, mis en ligne le 30 novembre 2018, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/communication/8081">http://journals.openedition.org/communication/8081</a> ; DOI : 10.4000/communication.8081			

Corpus		Nom	
communication		policing	
Nombre de mots	Date de la recherche	Date de publication	
10000	07/03/2019	2016	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
mémoire	spécialisé	expert	initié
Titre du document		Titre du texte	

Innovation et « policing » : L'adoption des médias sociaux par les membres des organisations policières du Québec	variation et « policing » : L'adoption des médias sociaux par les membres des organisations policières du Québec
Nom de l'auteur	Retraits
Julie Delle Donne	Nous n'avons conservé que les 10 000 premiers mots du Chapitre 1.
URL	
<a href="https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18926/Delle_Donne_Julie_2017_memoire.pdf?sequence=2&amp;isAllowed=y">https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18926/Delle_Donne_Julie_2017_memoire.pdf?sequence=2&amp;isAllowed=y</a>	
Code de recherche	Thème
médias sociaux	Études sur la surveillance
Commentaire	
Les policiers utilisent maintenant les médias sociaux dans le cadre de leur travail.	
Référence	
Delle Donne, J. (2016). <i>Innovation et « policing » : L'adoption des médias sociaux par les membres des organisations policières du Québec</i> (Mémoire, Université de Montréal, Canada). Repéré à <a href="https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18926/Delle_Donne_Julie_2017_memoire.pdf?sequence=2&amp;isAllowed=y">https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18926/Delle_Donne_Julie_2017_memoire.pdf?sequence=2&amp;isAllowed=y</a>	

Corpus		Nom	
communication		professionnelscommunication	
Nombre de mots	Date de la recherche	Date de publication	
4081	27/02/2019	2013	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Appropriation et stratégies d'intégration des médias sociaux par les professionnels de la communication		Appropriation et stratégies d'intégration des médias sociaux par les professionnels de la communication	
Nom de l'auteur		Retraits	
Francine Charest, Anne-Marie Gauthier et François Grenon		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
URL			
<a href="https://journals.openedition.org/communicationorganisation/4259">https://journals.openedition.org/communicationorganisation/4259</a>			
Code de recherche		Thème	
médias sociaux		Communication organisationnelle	
Commentaire			
Une démarche d'intégration des médias sociaux détaillée pour les professionnels de la communication.			
Référence			
<b>Francine Charest, Anne-Marie Gauthier et François Grenon</b> , « Appropriation et stratégies d'intégration des médias sociaux par les professionnels de la communication », <i>Communication et organisation</i> [En ligne], 43   2013, mis en ligne le 01 juin 2015, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/communicationorganisation/4259">http://journals.openedition.org/communicationorganisation/4259</a> ; DOI : 10.4000/communicationorganisation.4259			

Corpus		Nom	
communication		publicitetrompeuse	
Nombre de mots	Date de la recherche	Date de publication	
7046	28/02/2019	2016	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	

Éthique de la publicité rédactionnelle sur YouTube : faut-il réagir en déontologiste face à la publicité trompeuse ?		Éthique de la publicité rédactionnelle sur YouTube : faut-il réagir en déontologiste face à la publicité trompeuse ?	
Nom de l'auteur		Retraits	
Anthony Galluzo	titre, date, nom, résumé, table des matières, bibliographie, notes, citations, index, références, en-têtes, pieds de page		
URL			
<a href="https://www.cairn.info/revue-management-et-avenir-2016-4-page-101.htm">https://www.cairn.info/revue-management-et-avenir-2016-4-page-101.htm</a>			
Code de recherche		Thème	
youtube	Communications tactiques d'entreprises, controverses publiques		
Commentaire			
La publicité pour les jeux vidéos peut s'avérer trompeuse sur YouTube.			
Référence			
Galluzzo, A. (2016). Éthique de la publicité rédactionnelle sur <i>YouTube</i> – Faut-il réagir en déontologiste face à la publicité trompeuse ?. <i>Management &amp; Avenir</i> , 86(4), 101-118. doi:10.3917/mav.086.0101.			

Corpus		Nom	
communication		recrutement	
Nombre de mots		Date de la recherche	
5543	27/02/2019		Date de publication
			2012
Genre du texte		Spécialisation	
revue scientifique	spécialisé		Type d'auteur
			expert
Titre du document		Destinataire	
Les médias sociaux dans les stratégies de recrutement : quelle compatibilité avec la vie privée ?		initié	
Titre du document		Titre du texte	
Les médias sociaux dans les stratégies de recrutement : quelle compatibilité avec la vie privée ?		Les médias sociaux dans les stratégies de recrutement : quelle compatibilité avec la vie privée ?	
Nom de l'auteur		Retraits	
Laïla Benraïss-Noailles	références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés		
URL			
<a href="https://www.cairn.info/revue-francaise-de-gestion-2012-5-page-125.htm">https://www.cairn.info/revue-francaise-de-gestion-2012-5-page-125.htm</a>			
Code de recherche		Thème	
médias sociaux	Pratiques situées et institutionnalisées en milieu de travail		
Commentaire			
Les médias sociaux sont maintenant utilisés lors du recrutement par les ressources humaines.			
Référence			
Benraïss-Noailles, L. & Viot, C. (2012). Les médias sociaux dans les stratégies de recrutement: Quelle compatibilité avec la vie privée ?. <i>Revue française de gestion</i> , 224(5), 125-138. <a href="https://www.cairn.info/revue-francaise-de-gestion-2012-5-page-125.htm">https://www.cairn.info/revue-francaise-de-gestion-2012-5-page-125.htm</a> .			

Corpus		Nom	
communication		revolutiontunisienne	
Nombre de mots		Date de la recherche	
13166	27/02/2019		Date de publication
			2011
Genre du texte		Spécialisation	
revue scientifique	spécialisé		Type d'auteur
			expert
Titre du document		Destinataire	
Révolution Tunisienne et Internet : le rôle des médias sociaux		initié	
Titre du document		Titre du texte	
Révolution Tunisienne et Internet : le rôle des médias sociaux		Révolution Tunisienne et Internet : le rôle des médias sociaux	



Nom de l'auteur		Retraits	
Romain Lecomte		références, nom, notes, bibliographie, en-têtes, pieds de pages	
URL			
<a href="https://journals.openedition.org/anneemaghreb/1288">https://journals.openedition.org/anneemaghreb/1288</a>			
Code de recherche		Thème	
médias sociaux		Communication en situation de crise, communication politique	
Commentaire			
Les médias sociaux et leur rôle dans la révolution tunisienne.			
Référence			
<p><b>Romain Lecomte</b>, « Révolution tunisienne et Internet : le rôle des médias sociaux », <i>L'Année du Maghreb</i> [En ligne], VII   2011, mis en ligne le 01 janvier 2013, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/anneemaghreb/1288">http://journals.openedition.org/anneemaghreb/1288</a> ; DOI : 10.4000/anneemaghreb.1288</p>			

Corpus		Nom	
communication		rtrusse	
Nombre de mots		Date de la recherche	Date de publication
534		27/02/2019	18/02/2019
Genre du texte	Spécialisation	Type d'auteur	Destinataire
journal	non spécialisé	journaliste	non-initié
Titre du document		Titre du texte	
La télévision russe RT proteste après le blocage de ses pages par Facebook		La télévision russe RT proteste après le blocage de ses pages par Facebook	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédits photo	
URL			
<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215114-la-television-russe-rt-proteste-apres-le-blocage-de-ses-pages-par-facebook.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215114-la-television-russe-rt-proteste-apres-le-blocage-de-ses-pages-par-facebook.php</a>			
Code de recherche		Thème	
médias sociaux		Controverses sociotechniques	
Commentaire			
RT, une chaîne de télévision russe, s'est fait bloquer sa page Facebook.			
Référence			
<p>Agence France-Presse. (2019). La télévision russe RT proteste après le blocage de ses pages par Facebook. <i>La Presse</i>. Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215114-la-television-russe-rt-proteste-apres-le-blocage-de-ses-pages-par-facebook.php">https://www.lapresse.ca/techno/reseaux-sociaux/201902/18/01-5215114-la-television-russe-rt-proteste-apres-le-blocage-de-ses-pages-par-facebook.php</a>.</p>			

Corpus		Nom	
communication		saintamable	
Nombre de mots		Date de la recherche	Date de publication
3969		27/02/2019	2015
Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	non spécialisé	initié	non-initié
Titre du document		Titre du texte	
Politique d'utilisation des médias sociaux		Politique d'utilisation des médias sociaux	
Nom de l'auteur		Retraits	
Municipalité de Saint-Amable		contacts, menu	
URL			
<a href="https://www.st-amable.qc.ca/politique-utilisation-medias-sociaux">https://www.st-amable.qc.ca/politique-utilisation-medias-sociaux</a>			
Code de recherche		Thème	
médias sociaux communication		Communication organisationnelle	

Commentaire	
Les politiques d'utilisation des médias sociaux établies par la municipalité de Saint-Amable.	
Référence	
Municipalité de Saint-Amable. (2015). <i>Politique d'utilisation des médias sociaux</i> . Repéré à <a href="https://www.st-amable.qc.ca/politique-utilisation-medias-sociaux">https://www.st-amable.qc.ca/politique-utilisation-medias-sociaux</a> .	

Corpus		Nom	
communication		sante	
Nombre de mots		Date de la recherche	
475		27/02/2019	
Date de publication			
2014			
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
S'informer via des médias sociaux de santé : quelle place pour les experts ?		S'informer via des médias sociaux de santé : quelle place pour les experts ?	
Nom de l'auteur		Retraits	
Céline Paganelli et Viviane Clavier		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
URL			
<a href="https://www.cairn.info/revue-le-temps-des-medias-2014-2-page-141.htm?contenu=resume">https://www.cairn.info/revue-le-temps-des-medias-2014-2-page-141.htm?contenu=resume</a>			
Code de recherche		Thème	
médias sociaux		Communication et santé	
Commentaire			
Les usagers d'Internet peuvent maintenant s'informer sur leur santé par des médias sociaux.			
Référence			
Paganelli, C., Clavier, V. & Duetto, A. (2014). S'informer via des médias sociaux de santé : quelle place pour les experts ?. <i>Le Temps des médias</i> , 23(2), 141-143. doi:10.3917/tm.023.0141.			

Corpus		Nom	
communication		sixetapes	
Nombre de mots		Date de la recherche	
898		27/02/2019	
Date de publication			
2013			
Genre du texte		Spécialisation	
vulgarisation scientifique		semi-spécialisé	
Type d'auteur		Destinataire	
spécialiste		non-initié	
Titre du document		Titre du texte	
Médias sociaux : 6 étapes à suivre pour élaborer votre stratégie		Médias sociaux : 6 étapes à suivre pour élaborer votre stratégie	
Nom de l'auteur		Retraits	
Espresso Communication		image	
URL			
<a href="http://www.espressocommunication.com/communication/8353/medias-sociaux-6-etapes-a-suivre-pour-elaborer-votre-strategie">http://www.espressocommunication.com/communication/8353/medias-sociaux-6-etapes-a-suivre-pour-elaborer-votre-strategie</a>			
Code de recherche		Thème	
médias sociaux communication		Communications tactiques d'entreprises	
Commentaire			
Six étapes faciles dans la gestion des médias sociaux pour les entreprises.			
Référence			
Espresso Communication. (2013). Médias sociaux : 6 étapes à suivre pour élaborer votre stratégie. <i>Espresso</i> . Repéré à <a href="http://www.espressocommunication.com/communication/8353/medias-sociaux-6-etapes-a-suivre-pour-elaborer-votre-strategie">http://www.espressocommunication.com/communication/8353/medias-sociaux-6-etapes-a-suivre-pour-elaborer-votre-strategie</a> .			

Corpus		Nom	
communication		talibans	
Nombre de mots		Date de la recherche	
782		27/02/2019	
Genre du texte		Date de publication	
journal		15/02/2019	
Spécialisation		Type d'auteur	
non spécialisé		journaliste	
Destinataire		Retraits	
non-initié		date, nom, crédits photos	
Titre du document		URL	
Les talibans adeptes des réseaux sociaux dans leur guerre de propagande		<a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/15/01-5214836-les-talibans-adeptes-des-reseaux-sociaux-dans-leur-guerre-de-propagande.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS2">https://www.lapresse.ca/techno/reseaux-sociaux/201902/15/01-5214836-les-talibans-adeptes-des-reseaux-sociaux-dans-leur-guerre-de-propagande.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS2</a>	
Code de recherche		Thème	
médias sociaux		Controverses politiques, controverses publiques	
Commentaire			
Les talibans utilisent les médias sociaux pour faire de la propagande de fake news.			
Référence			
Agence France-Presse. (2019). Les talibans adeptes des réseaux sociaux dans leur guerre de propagande. <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/reseaux-sociaux/201902/15/01-5214836-les-talibans-adeptes-des-reseaux-sociaux-dans-leur-guerre-de-propagande.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS2">https://www.lapresse.ca/techno/reseaux-sociaux/201902/15/01-5214836-les-talibans-adeptes-des-reseaux-sociaux-dans-leur-guerre-de-propagande.php?utm_categorieinterne=traffidriviers&amp;utm_contenuinterne=cyberpresse_cinemaStandard_3852482_section_POS2</a> .			

Corpus		Nom	
communication		transplateforme	
Nombre de mots		Date de la recherche	
3781		27/02/2019	
Genre du texte		Date de publication	
revue scientifique		2013	
Spécialisation		Type d'auteur	
spécialisé		expert	
Destinataire		Retraits	
initié		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés	
Titre du document		URL	
Pratiques transplateformes et convergence dans l'usage des médias sociaux		<a href="https://journals.openedition.org/communicationorganisation/4116">https://journals.openedition.org/communicationorganisation/4116</a>	
Code de recherche		Thème	
médias sociaux		Études des médias	
Commentaire			
Regard sur le podcasting et les nouvelles pratiques transplateformes.			
Référence			
<b>Mélanie</b> Millette, « Pratiques transplateformes et convergence dans les usages des médias sociaux », <i>Communication et organisation</i> [En ligne], 43   2013, mis en ligne le 01 juin 2015, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/communicationorganisation/4116">http://journals.openedition.org/communicationorganisation/4116</a> ; DOI : 10.4000/communicationorganisation.4116			

Corpus		Nom	
communication		tweetstraces	
Nombre de mots	Date de la recherche	Date de publication	
5738	27/02/2019	2018	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Les tweets comme traces numériques : anticiper les enjeux réputationnels des publications sur Twitter		Les tweets comme traces numériques : anticiper les enjeux réputationnels des publications sur Twitter	
Nom de l'auteur		Retraits	
Tama Rchika		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/3602">https://journals.openedition.org/rfsic/3602</a>			
Code de recherche		Thème	
médias sociaux		Conflit organisationnel, communication organisationnelle	
Commentaire			
Les entreprises doivent prendre des précautions lors de la publication de tweets, car ces publications resteront toujours accessibles.			
Référence			
<b>Tama Rchika</b> , « Les tweets comme traces numériques », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 13   2018, mis en ligne le 01 juin 2018, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/3602">http://journals.openedition.org/rfsic/3602</a> ; DOI : 10.4000/rfsic.3602			

Corpus		Nom	
communication		twittermariage	
Nombre de mots	Date de la recherche	Date de publication	
4494	27/02/2019	2014	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
#mariagepourtous : Twitter et la politique affective des hashtags		#mariagepourtous : Twitter et la politique affective des hashtags	
Nom de l'auteur		Retraits	
Maxime Cervulle et Fred Pailler		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/717">https://journals.openedition.org/rfsic/717</a>			
Code de recherche		Thème	
médias sociaux		Controverses politiques, controverses publiques	
Commentaire			
Le rôle de Twitter lors du débat sur le mariage de même sexe.			
Référence			
<b>Maxime Cervulle et Fred Pailler</b> , « #mariagepourtous : Twitter et la politique affective des hashtags », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 4   2014, mis en ligne le 01 janvier 2014, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/717">http://journals.openedition.org/rfsic/717</a> ; DOI : 10.4000/rfsic.717			

Corpus		Nom	
communication		usagesprofessionnels	
Nombre de mots	Date de la recherche	Date de publication	
4793	27/02/2019	2015	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Usages professionnels et figures d'usagers des médias socionumériques		Usages professionnels et figures d'usagers des médias socionumériques	
Nom de l'auteur		Retraits	
Jean-Claude Domenget		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/1325">https://journals.openedition.org/rfsic/1325</a>			
Code de recherche		Thème	
médias sociaux		Identité occupationnelle	
Commentaire			
Les médias sociaux sont devenus un outils pour les professionnels et les entreprises.			
Référence			
<p><b>Jean-Claude Domenget</b>, « Usages professionnels et figures d'usager des médias socionumériques », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 6   2015, mis en ligne le 23 janvier 2015, consulté le 19 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/1325">http://journals.openedition.org/rfsic/1325</a> ; DOI : 10.4000/rfsic.1325</p>			

Corpus		Nom	
communication		vieprivée	
Nombre de mots	Date de la recherche	Date de publication	
5058	27/02/2019	2013	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Contre l'hypothèse de la «fin de la vie privée : la négociation de la privacy dans les médias sociaux		Contre l'hypothèse de la «fin de la vie privée : la négociation de la privacy dans les médias sociaux	
Nom de l'auteur		Retraits	
Antonio A. Casilli		références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés, index	
URL			
<a href="https://journals.openedition.org/rfsic/630">https://journals.openedition.org/rfsic/630</a>			
Code de recherche		Thème	
médias sociaux		Études sur la surveillance	
Commentaire			
Thèse contre l'idée que les relations internationales sont maintenant dépourvues de discrétion.			
Référence			
<p><b>Antonio A. Casilli</b>, « Contre l'hypothèse de la « fin de la vie privée » », <i>Revue française des sciences de l'information et de la communication</i> [En ligne], 3   2013, mis en ligne le 31 juillet 2013, consulté le 20 mars 2019. URL : <a href="http://journals.openedition.org/rfsic/630">http://journals.openedition.org/rfsic/630</a> ; DOI : 10.4000/rfsic.630</p>			

Corpus		Nom	
communication		wikipedia	
Nombre de mots	Date de la recherche	Date de publication	
2536	27/02/2019	2019	

Genre du texte	Spécialisation	Type d'auteur	Destinataire
vulgarisation scientifique	semi-spécialisé	inconnu	non-initié
Titre du document		Titre du texte	
Médias sociaux		Médias sociaux	
Nom de l'auteur		Retraits	
inconnu		références, articles connexes, bibliographie	
URL			
<a href="https://fr.wikipedia.org/wiki/Médias_sociaux">https://fr.wikipedia.org/wiki/Médias_sociaux</a>			
Code de recherche		Thème	
médias sociaux communication		Médias, réseaux sociaux	
Commentaire			
Page wikipédia sur les médias sociaux.			
Référence			
Médias sociaux. (s. d.). Dans <i>Wikipédia, l'encyclopédie libre</i> . Repéré le 27/02/2019 à <a href="https://fr.wikipedia.org/wiki/Médias_sociaux">https://fr.wikipedia.org/wiki/Médias_sociaux</a>			

Corpus	Nom		
communication	youtubedebats		
Nombre de mots	Date de la recherche	Date de publication	
4698	28/02/2019	2012	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
La force des mots : les mécanismes sémantiques de production et l'interprétation des actes de parole « menaçants »		La force des mots : les mécanismes sémantiques de production et l'interprétation des actes de parole « menaçants »	
Nom de l'auteur		Retraits	
Sara Amadori		titre, date, nom, résumé, table des matières, bibliographie, notes, citations	
URL			
<a href="http://www.revue-signes.info/document.php?id=2853">http://www.revue-signes.info/document.php?id=2853</a>			
Code de recherche		Thème	
youtube		Controverses publiques	
Commentaire			
Les débats sur YouTube peuvent vire à la violence verbale.			
Référence			
Amadori, Sara. Le débat d'idées en ligne : formes de la violence polémique sur YouTube. <i>Signes, Discours et Sociétés</i> [en ligne], 9. La force des mots : les mécanismes sémantiques de production et l'interprétation des actes de parole « menaçants », 30 juillet 2012. Disponible sur Internet : <a href="http://www.revue-signes.info/document.php?id=2853">http://www.revue-signes.info/document.php?id=2853</a>			

Corpus	Nom		
communication	youtubeobesite		
Nombre de mots	Date de la recherche	Date de publication	
7430	28/02/2019	2010	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
La contestation des discours sur l'obésité : YouTube et la femme « grosse » <i>parrhésiaste</i>		La contestation des discours sur l'obésité : YouTube et la femme « grosse » <i>parrhésiaste</i>	
Nom de l'auteur		Retraits	

Stephanie Mackay et Christine Dallaire	références, nom, notes, bibliographie, en-têtes, pieds de pages, résumés
URL	
<a href="https://www.erudit.org/en/journals/rf/2010-v23-n2-rf4006/045664ar/abstract/">https://www.erudit.org/en/journals/rf/2010-v23-n2-rf4006/045664ar/abstract/</a>	
Code de recherche	Thème
youtube	Mouvements sociaux
Commentaire	
La prise de paroles de femme obèse sur leur condition sur YouTube.	
Référence	
Mackay, S. & Dallaire, C. (2010). La contestation des discours sur l'obésité : YouTube et la femme « grosse » parrhésiaste. <i>Recherches féministes</i> , 23(2), 7–24. <a href="https://doi.org/10.7202/045664ar">https://doi.org/10.7202/045664ar</a>	

## Annexe 2 : Le corpus de l'informatique

Corpus		Nom	
informatique		apevents	
Nombre de mots	Date de la recherche	Date de publication	
445	27/02/2019	2019	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
article de périodique	non spécialisé	journaliste	non initié
Titre du document		Titre du texte	
Des applications enverraient des données personnelles sensibles à Facebook		Des applications enverraient des données personnelles sensibles à Facebook	
Nom de l'auteur		Retraits	
Agence France-Presse		date, nom, crédit photo	
URL			
<a href="https://www.lapresse.ca/techno/applications/201902/22/01-5215806-des-applications-enverraient-des-donnees-personnelles-sensibles-a-facebook.php?utm_categorieinterne=trafficdrivers&amp;utm_contenu_interne=cyberpresse_B13b_applications_3851290_section_POS1">https://www.lapresse.ca/techno/applications/201902/22/01-5215806-des-applications-enverraient-des-donnees-personnelles-sensibles-a-facebook.php?utm_categorieinterne=trafficdrivers&amp;utm_contenu_interne=cyberpresse_B13b_applications_3851290_section_POS1</a>			
Code de recherche		Thème	
médias sociaux		Sécurité informatique	
Commentaire			
L'application App Events recueille des informations personnelles sur les usagers, notamment sur la santé.			
Référence			
Agence France-Presse. (2019). Des applications enverraient des données personnelles sensibles à Facebook. <i>La Presse</i> . Repéré à <a href="https://www.lapresse.ca/techno/applications/201902/22/01-5215806-des-applications-enverraient-des-donnees-personnelles-sensibles-a-facebook.php?utm_categorieinterne=trafficdrivers&amp;utm_contenu_interne=cyberpresse_B13b_applications_3851290_section_POS1">https://www.lapresse.ca/techno/applications/201902/22/01-5215806-des-applications-enverraient-des-donnees-personnelles-sensibles-a-facebook.php?utm_categorieinterne=trafficdrivers&amp;utm_contenu_interne=cyberpresse_B13b_applications_3851290_section_POS1</a>			

Corpus		Nom	
informatique		bigsocialdata	
Nombre de mots	Date de la recherche	Date de publication	
6969	17/04/2019	2019	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Transformer les big social data en prévisions – méthodes et technologies : application à l'analyse de sentiments		Chapitre 1	
Nom de l'auteur		Retraits	
Imane El Alaoui		Seul le chapitre 1 a été conservé.	
URL			
<a href="https://tel.archives-ouvertes.fr/tel-02060594">https://tel.archives-ouvertes.fr/tel-02060594</a>			
Code de recherche		Thème	
booléen médias sociaux		Circuits booléens	



Commentaire	
L'utilisation des données des réseaux sociaux aux fins de la recherche d'information.	
Référence	
El Alaoui, I. (2019). <i>Transformer les big social data en prévisions – méthodes et technologies : application à l'analyse de sentiments</i> . (Maîtrise, Université d'Angers, France). Repéré à <a href="https://tel.archives-ouvertes.fr/tel-02060594">https://tel.archives-ouvertes.fr/tel-02060594</a>	

Corpus		Nom					
informatique		capturededonnes					
Nombre de mots		Date de la recherche		Date de publication			
6230		15/04/2019		2016			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		spécialisé		expert		initié	
Titre du document				Titre du texte			
Bandit contextuel pour la capture de données temps réel sur les médias sociaux				Bandit contextuel pour la capture de données temps réel sur les médias sociaux			
Nom de l'auteur				Retraits			
Thibault Gisselbrecht, Sylvain Lamprier et Patrick Gallinari				Page titre, remerciements, bibliographie, résumé en anglais			
URL							
<a href="https://hal.archives-ouvertes.fr/hal-01355408/">https://hal.archives-ouvertes.fr/hal-01355408/</a>							
Code de recherche			Thème				
informatique médias sociaux			Science des données				
Commentaire							
La collecte de données en temps réel sur les médias sociaux se complexifie.							
Référence							
Thibault Gisselbrecht, Sylvain Lamprier, Patrick Gallinari. Bandit Contextuel pour la Capture de Données Temps Réel sur les Médias Sociaux. Semaine du Document Numérique et de la Recherche d'Information (SDNRI 2016), Mar 2016, Toulouse, France. pp.57-72.							

Corpus		Nom					
informatique		cartographie					
Nombre de mots		Date de la recherche		Date de publication			
3532		27/02/2019		2013			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		spécialisé		expert		initié	
Titre du document				Titre du texte			
Médias sociaux : clarification et cartographie pour une approche sociotechnique				Médias sociaux : clarification et cartographie pour une approche sociotechnique			
Nom de l'auteur				Retraits			
Thomas Stenger et Alexandre Coutant				auteurs, résumé en anglais, mots-clés en anglais, contacts			
URL							
<a href="https://www.jstor.org/stable/24582919?seq=1#metadata_info_tab_contents">https://www.jstor.org/stable/24582919?seq=1#metadata_info_tab_contents</a>							
Code de recherche			Thème				
médias sociaux			Réseaux sociaux				
Commentaire							
Définition des types de médias sociaux.							
Référence							
Stenger, T., & Coutant, A. (2013). Médias sociaux : Clarification et cartographie Pour une approche sociotechnique. <i>Décisions Marketing</i> , (70), 107-117. Retrieved from <a href="http://www.jstor.org/stable/24582919">http://www.jstor.org/stable/24582919</a>							

Corpus	Nom
--------	-----

informatique		contextualisationtweets	
Nombre de mots		Date de la recherche	Date de publication
5800		09/04/2019	2013
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Contextualisation automatique de Tweets à partir de Wikipédia		Contextualisation automatique de Tweets à partir de Wikipédia	
Nom de l'auteur		Retraits	
Romain Deveaud et Florian Boudin		page titre, bibliographie en-tête, résumé en anglais, mots-clés	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-00816351/">https://hal.archives-ouvertes.fr/hal-00816351/</a>			
Code de recherche		Thème	
traitement automatique des langues médias sociaux		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Les tweets n'ont que 140 caractères et sont difficiles à contextualiser. Cette recherche utilise Wikipédia pour le faire.			
Référence			
Romain Deveaud, Florian Boudin. Contextualisation automatique de Tweets à partir de Wikipédia. Conférence en Recherche d'Information et Applications (CORIA), Apr 2013, Neuchatel, Suisse. 2013.			

Corpus		Nom	
informatique		dangersreseauxsociaux	
Nombre de mots		Date de la recherche	Date de publication
12066		16/04/2019	2012
Genre du texte	Spécialisation	Type d'auteur	Destinataire
thèse	spécialisé	expert	initié
Titre du document		Titre du texte	
Plateforme pour se protéger tant de soi-même que de ses "amis" sur Facebook		Introduction ; Chapitre 1 ; Chapitre 2 ; Chapitre 3.	
Nom de l'auteur		Retraits	
Charles Hérou		Seuls l'introduction et les chapitres 1, 2 et 3 ont été conservés.	
URL			
<a href="https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9168">https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9168</a>			
Code de recherche		Thème	
infographie facebook		Sécurité informatique	
Commentaire			
Regard sur les problèmes de sécurité dans les réseaux sociaux.			
Référence			
Hérou, C. (2012). <i>Plateforme pour se protéger tant de soi-même que de ses "amis" sur Facebook</i> . (Doctorat, Université de Montréal, Canada). Repéré à <a href="https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9168">https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9168</a>			

Corpus		Nom	
informatique		defistal	
Nombre de mots		Date de la recherche	Date de publication
3084		22/03/2019	2013
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Les défis de l'analyse des médias sociaux pour le traitement automatique des langues		Les défis de l'analyse des médias sociaux pour le traitement automatique des langues	

Nom de l'auteur		Retraits	
Atefeh Farzindar et Mathieu Roche		résumé en anglais, mots-clés en anglais, remerciements, bibliographie, pieds de page, en-têtes	
URL			
<a href="http://agritrop.cirad.fr/575911/">http://agritrop.cirad.fr/575911/</a>			
Code de recherche		Thème	
atefeh farzindar		Traitement automatique des langages naturels (TALN)	
Commentaire			
C'est une analyse du discours dans les médias sociaux pour le traitement automatique des langues. L'étude se concentre sur l'infobésité dans les médias sociaux.			
Référence			
Farzindar, A. et Roche, M. (2013). Les défis de l'analyse des médias sociaux pour le traitement automatique des langues. <i>TAL</i> 54(3). 7-16. Repéré à <a href="http://agritrop.cirad.fr/575911/">http://agritrop.cirad.fr/575911/</a>			

Corpus		Nom	
informatique		detectionpointsdevue	
Nombre de mots		Date de la recherche	
13825		15/04/2019	
Date de publication		2018	
Genre du texte		Spécialisation	
thèse		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Détection de points de vue sur les médias sociaux numériques		Chapitre 2 ; Chapitre 5.	
Nom de l'auteur		Retraits	
Ophélie Fraisier		Seuls les chapitres 2 et 5 ont été conservés.	
URL			
<a href="http://thesesups.ups-tlse.fr/4064/">http://thesesups.ups-tlse.fr/4064/</a>			
Code de recherche		Thème	
algorithme médias sociaux		Algorithmique, réseaux sociaux	
Commentaire			
Analyse algorithmique des points de vue sur les médias sociaux.			
Référence			
Fraisier, O. (2018). <i>Détection de points de vue sur les médias sociaux numériques. (Doctorat, Université Toulouse, France.)</i> Repéré à <a href="http://thesesups.ups-tlse.fr/4064/">http://thesesups.ups-tlse.fr/4064/</a>			

Corpus		Nom	
informatique		detectionrumeurs	
Nombre de mots		Date de la recherche	
10075		16/04/2019	
Date de publication		2017	
Genre du texte		Spécialisation	
mémoire		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Détection des rumeurs dans les médias sociaux		Chapitre 1 ; Chapitre 2 ; Chapitre 3.	
Nom de l'auteur		Retraits	
Khadijia Bensalem et Samiha Soualmi		Seuls les chapitres 1, 2 et 3 ont été conservés.	
URL			
<a href="http://www.univ-bejaia.dz/dspace/handle/123456789/5736">http://www.univ-bejaia.dz/dspace/handle/123456789/5736</a>			
Code de recherche		Thème	
algorithme médias sociaux		Algorithmique, réseaux sociaux	

Commentaire		
Mémoire qui fait un état de l'art sur les algorithmes utilisés pour localiser les rumeurs dans les médias sociaux ainsi que présentation d'une nouvelle méthode.		
Référence		
Bensalem, K. (2017). <i>Détection des rumeurs dans les médias sociaux</i> . (Maîtrise, Université Abderrahmane Mira –Bejaia, Algérie). Repéré à <a href="http://www.univ-bejaia.dz/dspace/handle/123456789/5736">http://www.univ-bejaia.dz/dspace/handle/123456789/5736</a>		

Corpus		Nom	
informatique		detectiontwitter	
Nombre de mots	Date de la recherche	Date de publication	
8832	09/04/2019	2013	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Réseaux Sociaux		Détection d'événements à partir de Twitter	
Nom de l'auteur		Retraits	
Houssem Eddine Dridi et Guy Lapalme		Résumé en anglais, mots-clés en anglais, page couverture, titres, auteurs, bibliographie, en-têtes, pieds de page	
URL			
<a href="http://agritrop.cirad.fr/576352/1/Roche-Traitement_automatique_des_langues-Coord.revue_2013.pdf">http://agritrop.cirad.fr/576352/1/Roche-Traitement_automatique_des_langues-Coord.revue_2013.pdf</a>			
Code de recherche		Thème	
atefe farzindar		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Détection d'événements populaires à partir de patterns sur Twitter. Di			
Référence			
Dridi, H. E., Lapalme, G. (2013). Détection d'événements à partir de Twitter. <i>Traitement automatique des langues</i> , 54(3), 11-33.			

Corpus		Nom	
informatique		diffusioninformation	
Nombre de mots	Date de la recherche	Date de publication	
12909	15/04/2019	2014	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
thèse	spécialisé	expert	initié
Titre du document		Titre du texte	
Diffusion de l'information dans les médias sociaux : modélisation et analyse		Chapitre 3	
Nom de l'auteur		Retraits	
Adrien Guille		Seul le chapitre 3 de la thèse a été conservé	
URL			
<a href="https://tel.archives-ouvertes.fr/tel-01100255/">https://tel.archives-ouvertes.fr/tel-01100255/</a>			
Code de recherche		Thème	
algorithme médias sociaux		Médias sociaux	
Commentaire			
Nouvelle approche sur la modélisation de la diffusion de l'information dans les médias sociaux.			
Référence			
Adrien Guille. Diffusion de l'information dans les médias sociaux : modélisation et analyse. Informatique [cs]. Université Lumière Lyon 2, 2014. Français.			

Corpus		Nom	
informatique		donneesmanquantes	
Nombre de mots	Date de la recherche	Date de publication	

7839	16/04/2019	2014	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
mémoire	spécialisé	expert	initié
Titre du document		Titre du texte	
Estimation des données manquantes et catégorisation des identités sociaux		Introduction ; Chapitre 1.	
Nom de l'auteur		Retraits	
Naim El Hosseyn		Seuls l'introduction et le chapitre 1 ont été conservés.	
URL			
<a href="http://dspace.univ-tlemcen.dz/bitstream/112/6265/1/Estimation-des-donnees-manquantes-et-categorisation-des-identites-sociaux.pdf">http://dspace.univ-tlemcen.dz/bitstream/112/6265/1/Estimation-des-donnees-manquantes-et-categorisation-des-identites-sociaux.pdf</a>			
Code de recherche		Thème	
programmation médias sociaux		Programmation	
Commentaire			
Recherche des données manquantes dans les médias sociaux.			
Référence			
El Hosseyn, N. (2014). <i>Estimation des données manquantes et catégorisation des identités sociaux</i> . (Mémoire, Université Abou Bakr Belkaid–Tlemcen, Algérie.) Repéré à <a href="http://dspace.univ-tlemcen.dz/bitstream/112/6265/1/Estimation-des-donnees-manquantes-et-categorisation-des-identites-sociaux.pdf">http://dspace.univ-tlemcen.dz/bitstream/112/6265/1/Estimation-des-donnees-manquantes-et-categorisation-des-identites-sociaux.pdf</a>			

Corpus		Nom	
informatique		emotionstweets	
Nombre de mots	Date de la recherche	Date de publication	
4565	09/04/2019	2017	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT)		Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT)	
Nom de l'auteur		Retraits	
Thierry Hamon, Amel Fraise, Patrick Paroubek, Pierre Zweigenbaum et Cyril Grouin		Page titre, bibliographie, résumés, en-têtes, pieds de page	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-01617180/">https://hal.archives-ouvertes.fr/hal-01617180/</a>			
Code de recherche		Thème	
traitement automatique des langues médias sociaux		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Analyse des émotions présentes sur Twitter à l'aide du traitement automatique des langues.			
Référence			
Thierry Hamon, Amel Fraise, Patrick Paroubek, Pierre Zweigenbaum, Cyril Grouin. Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT). Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015), Jun 2015, Caen, France.			

Corpus		Nom	
informatique		integritecontextuelle	
Nombre de mots	Date de la recherche	Date de publication	
9333	16/04/2019	2018	
Genre du texte	Spécialisation	Type d'auteur	Destinataire

mémoire	spécialisé	expert	initié
Titre du document		Titre du texte	
Défauts d'intégrité contextuelle liés à la collecte de données personnelles par des applications de médias sociaux sur Android		Introduction ; Chapitre 1 ; Chapitre 2 ; Chapitre 3.	
Nom de l'auteur		Retraits	
Bouganim Taïna		Seuls l'introduction et les chapitres 1, 2 et 3 ont été conservés.	
URL			
<a href="http://espace.etsmtl.ca/2161/">http://espace.etsmtl.ca/2161/</a>			
Code de recherche		Thème	
programmation médias sociaux		Programmation	
Commentaire			
Analyse de sécurité reliée aux applications de médias sociaux sur la plateforme mobile Android.			
Référence			
Boganim, T. (2018). <i>Défauts d'intégrité contextuelle liés à la collecte de données personnelles par des applications de médias sociaux sur Android</i> . (Maîtrise, Université du Québec, Canada). Repéré à <a href="http://espace.etsmtl.ca/2161/">http://espace.etsmtl.ca/2161/</a>			

Corpus		Nom	
informatique		intelligence artificielle	
Nombre de mots		Date de la recherche	
10677		15/04/2019	
Date de publication		2019	
Genre du texte		Spécialisation	
acte de conférence		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Intelligence artificielle et réseaux sociaux		Intelligence artificielle et réseaux sociaux	
Nom de l'auteur		Retraits	
Arnaud Martin		Page titre, références, sommaire	
URL			
<a href="https://afia.asso.fr/wp-content/uploads/2019/02/103_jan19.pdf#page=15">https://afia.asso.fr/wp-content/uploads/2019/02/103_jan19.pdf#page=15</a>			
Code de recherche		Thème	
algorithme médias sociaux		Science des données	
Commentaire			
Analyse des données brutes dans les médias sociaux			
Référence			
Martin, A. (2019, janvier). <i>Bulletin de l'AFIA n°103 : Intelligence artificielle et réseaux sociaux</i> . Rennes, France. Résumé repéré à <a href="https://afia.asso.fr/wp-content/uploads/2019/02/103_jan19.pdf#page=15">https://afia.asso.fr/wp-content/uploads/2019/02/103_jan19.pdf#page=15</a>			

Corpus		Nom	
informatique		introduction twitter	
Nombre de mots		Date de la recherche	
953		17/04/2019	
Date de publication		2017	
Genre du texte		Spécialisation	
mémoire		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Classification supervisée de textes courts et bruités : application au domaine des médias sociaux		Chapitre 3	
Nom de l'auteur		Retraits	
Billal Belainine		Seul le chapitre 3 a été conservé	
URL			
<a href="https://archipel.uqam.ca/9944/">https://archipel.uqam.ca/9944/</a>			
Code de recherche		Thème	
modèles statistiques médias sociaux		Modèles statistiques	

Commentaire	
Utilisation des statistiques pour classer des tweets.	
Référence	
Billal, B. (2017). <i>Classification supervisée de textes courts et bruités : application au domaine des médias sociaux</i> . (Maîtrise, Université du Québec, Canada). Repéré à <a href="https://archipel.uqam.ca/9944/">https://archipel.uqam.ca/9944/</a>	

Corpus		Nom					
informatique		ironietweets					
Nombre de mots		Date de la recherche		Date de publication			
3157		09/04/2019		2015			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		spécialisé		expert		initié	
Titre du document				Titre du texte			
Détection automatique de l'ironie dans les tweets français				Détection automatique de l'ironie dans les tweets français			
Nom de l'auteur				Retraits			
Jihen Karoui, Farah Benamara Zitoune, Véronique Moriceau, Nathalie Aussenac-Gilles et Lamia Hadrich Belguith				Page titre, résumés, mots-clés, bibliographie			
URL							
<a href="https://hal.archives-ouvertes.fr/hal-01334721/">https://hal.archives-ouvertes.fr/hal-01334721/</a>							
Code de recherche				Thème			
traitement automatique des langues médias sociaux				Traitement automatique des langues naturelles (TALN)			
Commentaire							
Détection de l'ironie sur Twitter grâce à un classifieur binaire.							
Référence							
Jihen Karoui, Farah Benamara Zitoune, Véronique Moriceau, Nathalie Aussenac-Gilles, Lamia Hadrich Belguith. Détection automatique de l'ironie dans les tweets en français. <i>22eme Conference sur le Traitement Automatique des Langues Naturelles (TALN 2015)</i> , Jun 2015, Caen, France. Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles, pp. 1-6, 2015. <a href="https://hal-01334721/">hal-01334721</a>							

Corpus		Nom					
informatique		meltwater					
Nombre de mots		Date de la recherche		Date de publication			
1687		17/04/2019		2015			
Genre du texte		Spécialisation		Type d'auteur		Destinataire	
revue scientifique		spécialisé		expert		initié	
Titre du document				Titre du texte			
Meltwater, un grand de la surveillance médias sociaux et presse web				Meltwater, un grand de la surveillance médias sociaux et presse web			
Nom de l'auteur				Retraits			
Anne-Marie Libmann				En-tête			
URL							
<a href="https://www.fla-consultants.com/images/Meltwater_surveillance_medias_sociaux_et_web.pdf">https://www.fla-consultants.com/images/Meltwater_surveillance_medias_sociaux_et_web.pdf</a>							
Code de recherche				Thème			
booléen médias sociaux				circuits booléens			
Commentaire							
Revue d'un logiciel de veille informatique (Meltwater).							
Référence							

Libmann, A.-M. (2015). Meltwater, un grand de la surveillance médias sociaux et presse web. *Netsources*, 118, 6-9. Repéré à [https://www.fla-consultants.com/images/Meltwater\\_surveillance\\_medias\\_sociaux\\_et\\_web.pdf](https://www.fla-consultants.com/images/Meltwater_surveillance_medias_sociaux_et_web.pdf)

Corpus		Nom	
informatique		memesreseauxsociaux	
Nombre de mots	Date de la recherche	Date de publication	
12856	15/02/2019	2014	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Conception d'un outil d'analyse et de visualisation des mêmes internet : le cas du réseau social chinois Sina Weibo		Chapitre 3	
Nom de l'auteur		Retraits	
Clément Renaud		Seul le chapitre 3 de la thèse a été conservé	
URL			
<a href="https://www.theses.fr/2014ENST0070">https://www.theses.fr/2014ENST0070</a>			
Code de recherche		Thème	
algorithme médias sociaux		Algorithmique, réseaux sociaux	
Commentaire			
Une analyse de la diffusion des mêmes.			
Référence			
Clément Renaud. Conception d'un outil d'analyse et de visualisation des mêmes internet : le cas du réseau social chinois Sina Weibo. Sociologie. Télécom ParisTech, 2014. Français.			

Corpus		Nom	
informatique		mesusagemedicaments	
Nombre de mots	Date de la recherche	Date de publication	
2789	17/04/2019	2018	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Détection de mésusages de médicaments dans les réseaux sociaux		Détection de mésusages de médicaments dans les réseaux sociaux	
Nom de l'auteur		Retraits	
Elise Bigeard, Natalia Grabar et Frantz Thiessard		Page titre, remerciements, références	
URL			
<a href="https://halshs.archives-ouvertes.fr/halshs-01968335/">https://halshs.archives-ouvertes.fr/halshs-01968335/</a>			
Code de recherche		Thème	
bayésien médias sociaux		Inférence bayésienne	
Commentaire			
Recherche des commentaires sur les forums qui amènent les patients à mal utiliser leurs prescriptions.			
Référence			
Elise Bigeard, Natalia Grabar, Frantz Thiessard. Détection de mésusages de médicaments dans les réseaux sociaux. <i>TALN</i> , May 2018, Rennes, France. ( <a href="https://halshs-01968335/">halshs-01968335</a> )			

Corpus		Nom	
informatique		opinionsblogosphere	
Nombre de mots	Date de la recherche	Date de publication	
5226	09/04/2019	2010	
Genre du texte	Spécialisation	Type d'auteur	Destinataire



acte de conférence	spécialisé	expert	initié
Titre du document		Titre du texte	
Classification automatique d'opinions dans la blogosphère		Classification automatique d'opinions dans la blogosphère	
Nom de l'auteur		Retraits	
Jacques Savoy et Olena Zubaryeva		Titres, résumé en anglais, bibliographie	
URL			
<a href="http://doc.rero.ch/record/31109">http://doc.rero.ch/record/31109</a>			
Code de recherche		Thème	
traitement automatique des langues médias sociaux		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Le système traite le vocabulaire spécifique afin de déterminer si les phrases publiées dans la blogosphère sont des opinions ou pas.			
Référence			
Savoy, J., Zubaryeva, O. (2010). <i>Classification automatique d'opinions dans la blogosphère</i> . Actes 10 <sup>e</sup> journées Analyse statistiques des Données Textuelles. Neuchâtel, Suisse.			

Corpus		Nom	
informatique		plateformesondy	
Nombre de mots		Date de la recherche	Date de publication
1126		15/04/2019	2013
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Sondy : une plateforme open-source d'analyse et fouille pour les réseaux sociaux en ligne		Sondy : une plateforme open-source d'analyse et fouille pour les réseaux sociaux en ligne	
Nom de l'auteur		Retraits	
Adrien Guille, C. Favre et Djamel Abdelkader Zighed		page titre, remerciements, bibliographie	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-00770557/">https://hal.archives-ouvertes.fr/hal-00770557/</a>			
Code de recherche		Thème	
algorithme médias sociaux		Algorithmique, réseaux sociaux	
Commentaire			
Présentation de la plateforme Sondy.			
Référence			
Adrien Guille, C. Favre, Djamel Abdelkader Zighed. <i>SONDY : une plateforme open-source d'analyse et de fouille pour les réseaux sociaux en ligne</i> . 13e Conférence en Extraction et Gestion des Connaissances, Jan 2013, France. 2013, pp.45-48, 2013.			

Corpus		Nom	
informatique		profilutilisateur	
Nombre de mots		Date de la recherche	Date de publication
6132		15/04/2019	2015
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Enrichissement du profil utilisateur à partir de son réseau social dans un contexte dynamique : application d'une méthode de pondération temporelle		Enrichissement du profil utilisateur à partir de son réseau social dans un contexte dynamique : application d'une méthode de pondération temporelle	
Nom de l'auteur		Retraits	

Marie-Françoise Canut, Sirinya On-at, André Péninou et Florence Sèdes	résumé en anglais, remerciements, bibliographie
URL	
<a href="https://oatao.univ-toulouse.fr/15310/">https://oatao.univ-toulouse.fr/15310/</a>	
Code de recherche	Thème
Algorithme médias sociaux	Science des données
Commentaire	
Les profils des utilisateurs peuvent être enrichis automatiquement par les intérêts.	
Référence	
Canut, Marie-Françoise and On-At, Sirinya and Péninou, André and Sèdes, Florence Enrichissement du profil utilisateur à partir de son réseau social dans un contexte dynamique : application d'une méthode de pondération temporelle. (2015) In: 33eme congrès INFormatique des Organisations et Systemes d'Information et de Decision (INFORSID 2015), 26 May 2015 - 29 May 2015 (Biarritz, France).	

Corpus		Nom	
informatique		propagationmemes	
Nombre de mots	Date de la recherche	Date de publication	
8977	16/04/2019	2018	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Pour une sémiotique de la propagation : invention et imitation sur les réseaux sociaux		Pour une sémiotique de la propagation : invention et imitation sur les réseaux sociaux	
Nom de l'auteur		Retraits	
Yvana Fechine	En-tête, références		
URL			
<a href="https://www.unilim.fr/actes-semiotiques/5953">https://www.unilim.fr/actes-semiotiques/5953</a>			
Code de recherche		Thème	
programmation médias sociaux		Programmation	
Commentaire			
Partage des memes sur les médias sociaux.			
Référence			
Yvana FECHINE, « Pour une sémiotique de la propagation : invention et imitation sur les réseaux sociaux », <i>Actes Sémiotiques</i> [En ligne], 121, 2018, consulté le 16/04/2019, URL : <a href="https://www.unilim.fr/actes-semiotiques/5953">https://www.unilim.fr/actes-semiotiques/5953</a>			

Corpus		Nom	
informatique		rechercheinformation	
Nombre de mots	Date de la recherche	Date de publication	
7288	17/04/2019	2017	
Genre du texte	Spécialisation	Type d'auteur	Destinataire
thèse	spécialisé	expert	initié
Titre du document		Titre du texte	
Impact des réseaux sociaux sur le processus de recherche d'information		Chapitre 2	
Nom de l'auteur		Retraits	
Chahrazed Bouhini	Seul le chapitre 2 a été conservé.		
URL			
<a href="https://tel.archives-ouvertes.fr/tel-01528583/">https://tel.archives-ouvertes.fr/tel-01528583/</a>			
Code de recherche		Thème	
booléen médias sociaux		Circuits booléens	
Commentaire			

Les changements apportés par les réseaux sociaux dans la recherche d'information.	
Référence	
Bouhini, C. (2017). <i>Impact des réseaux sociaux sur le processus de recherche d'information</i> . (Doctorat, Université de Lyon, France). Repéré à <a href="https://tel.archives-ouvertes.fr/tel-01528583/">https://tel.archives-ouvertes.fr/tel-01528583/</a>	

Corpus		Nom	
informatique		reinformation	
Nombre de mots		Date de la recherche	
2900		09/04/2019	
Date de publication		2016	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Médias traditionnels, médias sociaux : caractériser la réinformation		Médias traditionnels, médias sociaux : caractériser la réinformation	
Nom de l'auteur		Retraits	
Cédric Maigrot, Ewa Kijak et Vincent Claveau		Résumés, mots-clés, page couverture, bibliographie	
URL			
<a href="https://hal.inria.fr/hal-01349871/">https://hal.inria.fr/hal-01349871/</a>			
Code de recherche		Thème	
traitement automatique des langues médias sociaux		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Analyse des messages sur Facebook pour montrer le changement de l'information.			
Référence			
Cédric Maigrot, Ewa Kijak, Vincent Claveau. Médias traditionnels, médias sociaux : caractériser la réinformation. TALN 2016 - 23ème Conférence sur le Traitement Automatique des Langues Naturelles, Jul 2016, Paris, France. 2016.			

Corpus		Nom	
informatique		reseauxegocentriques	
Nombre de mots		Date de la recherche	
13141		17/04/2019	
Date de publication		2013	
Genre du texte		Spécialisation	
revue scientifique		spécialisé	
Type d'auteur		Destinataire	
expert		initié	
Titre du document		Titre du texte	
Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés réseaux k-égocentriques		Chapitre 2	
Nom de l'auteur		Retraits	
Dieudonné Tchunte		Seul le chapitre 2 a été conservé.	
URL			
<a href="http://thesesups.ups-tlse.fr/1972/1/2013TOU30002.pdf">http://thesesups.ups-tlse.fr/1972/1/2013TOU30002.pdf</a>			
Code de recherche		Thème	
bayésien médias sociaux		Inférence bayésienne	
Commentaire			
Analyse des profils utilisateurs sur les réseaux sociaux.			
Référence			
Tchunte, D. (2013). <i>Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés réseaux k-égocentriques</i> . (Doctorat, Université de Toulouse, France). Repéré à <a href="http://thesesups.ups-tlse.fr/1972/1/2013TOU30002.pdf">http://thesesups.ups-tlse.fr/1972/1/2013TOU30002.pdf</a>			

Corpus	Nom
--------	-----

informatique		securisationreseauxsociaux	
Nombre de mots		Date de la recherche	Date de publication
7274		16/04/2019	2014
Genre du texte	Spécialisation	Type d'auteur	Destinataire
thèse	spécialisé	expert	initié
Titre du document		Titre du texte	
La sécurisation des réseaux sociaux mobiles		Chapitre 4	
Nom de l'auteur		Retraits	
Mohamed Amine Ferrag		Seul le chapitre 4 a été conservé.	
URL			
<a href="https://www.researchgate.net/profile/Mohamed_Amine_Ferrag/publication/279289396_La_sécurisation_des_reseaux_sociaux_mobiles/links/559180ee08ae15962d8e2598.pdf">https://www.researchgate.net/profile/Mohamed_Amine_Ferrag/publication/279289396_La_sécurisation_des_reseaux_sociaux_mobiles/links/559180ee08ae15962d8e2598.pdf</a>			
Code de recherche		Thème	
sécurité informatique réseaux sociaux		Sécurité informatique	
Commentaire			
Réponse aidant à la sécurisation de la confidentialité sur les réseaux sociaux mobiles.			
Référence			
Ferrag, M. A. (2014). <i>La sécurisation des réseaux sociaux mobiles</i> . (Doctorat, Université Badji Mokhtar-Annaba, Algérie).			

Corpus		Nom	
informatique		sentimentstwitter	
Nombre de mots		Date de la recherche	Date de publication
959		15/04/2019	2016
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Intégration du contexte spatio-temporel et social pour l'analyse de sentiments sur Twitter		Intégration du contexte spatio-temporel et social pour l'analyse de sentiments sur Twitter	
Nom de l'auteur		Retraits	
Ophélie Fraisier		références, remerciements, page titre, mots-clés	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-01475015/document">https://hal.archives-ouvertes.fr/hal-01475015/document</a>			
Code de recherche		Thème	
traitement automatique des langues médias sociaux		Traitement automatique des langues naturelles (TALN)	
Commentaire			
Ce texte traite de l'analyse des sentiments sur Twitter.			
Référence			
Fraisier, Ophélie Intégration du contexte spatio-temporel et social pour l'analyse de sentiments sur Twitter. (2016) In: 8eme Forum Jeunes Chercheurs du congrès INformatique des Organisations et Systemes d'Information et de Decision (INFORSID 2016), 31 May 2016 (Grenoble, France).			

Corpus		Nom	
informatique		statistiquesinformation	
Nombre de mots		Date de la recherche	Date de publication
3935		15/04/2019	2015
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Accès à l'information dans les réseaux sociaux : quelles formes de collaboration ?		Accès à l'information dans les réseaux sociaux : quelles formes de collaboration ?	
Nom de l'auteur		Retraits	

Lynda Tamine et al.	page titre, mots-clés, résumés, bibliographie, remerciements		
URL			
<a href="http://oatao.univ-toulouse.fr/15390/">http://oatao.univ-toulouse.fr/15390/</a>			
Code de recherche		Thème	
algorithme médias sociaux		Algorithmique, réseaux sociaux	
Commentaire			
Analyse statistique de la diffusion de l'information sur Twitter.			
Référence			
Tamine, Lynda and Soulier, Laure and Ben Jabeur, Lamjed and Amblard, Frédéric and Hanachi, Chihab and Hubert, Gilles and Roth, Camille Accès à l'information dans les réseaux sociaux : quelles formes de collaboration ? (2015) In: 6eme Conférence sur les Modèles et l'Analyse des Réseaux : Approches Mathématiques et Informatique (MARAMI 2015), 14 October 2015 - 16 October 2015 (Nîmes, France).			

Corpus		Nom	
informatique		systemesderecommandation	
Nombre de mots		Date de la recherche	Date de publication
7421		15/04/2019	2016
Genre du texte	Spécialisation	Type d'auteur	Destinataire
revue scientifique	spécialisé	expert	initié
Titre du document		Titre du texte	
Partir des usages pour analyser les systèmes de recommandation : le cas des médias sociaux		Partir des usages pour analyser les systèmes de recommandation : le cas des médias sociaux	
Nom de l'auteur		Retraits	
Alexandre Coutant et Jean-Claude Domenget		Page titre, remerciements, bibliographie	
URL			
<a href="https://hal.archives-ouvertes.fr/hal-01352921/">https://hal.archives-ouvertes.fr/hal-01352921/</a>			
Code de recherche		Thème	
algorithme médias sociaux		Science des données	
Commentaire			
Présentation de l'analyse des systèmes de recommandation			
Référence			
Alexandre Coutant, Jean-Claude Domenget. Partir des usages pour analyser les systèmes de recommandation : le cas des médias sociaux. Gérald Kembellec, Ghislaine Chartron, Imad Saleh. Les moteurs et systèmes de recommandation, ISTE Editions, 2014, 978-1-78405-041-2. .			

### Annexe 3 : Les thèmes de recherche du Département de communication de l'Université de Montréal

<b>Expertises de recherche en communication du guide de disciplines de l'Université de Montréal</b>		
Agentivité	Amérique du Nord	Amériques
Approches constitutives de la communication	Approches rhétoriques de la communication politique	Archéologie des médias
Arts médiatiques	Bioarts	Canada
Cinéma documentaire	Cinéma expérimental	Collaboration interprofessionnelle
Communautés virtuelles	Communication écrite et technologie	Communication en situation de crise
Communication en philosophie politique	Communication et santé	Communication internationale
Communication organisationnelle	Communication politique	Communications tactiques d'entreprise
Conflit organisationnel	Controverses politiques	Controverses publiques
Controverses sociotechniques	Culture numérique	Culture populaire
Culture visuelle	Culture du vieillissement	Cyberculture
Design participatif / Conception participative	Édition	Environnement d'apprentissage
Époque contemporaine	Esthétique de la communication	États-Unis
Éthique discursive	Ethnographie	Études américaines
Études culturelles	Études culturelles de la mémoire	Études de la mobilité
Études des médias	Études des phénomènes organisés	Études sur la surveillance
Gouvernementalité algorithmique	Guerre dans les médias	Identité occupationnelle
Imagerie politico-culturelle des contenus télévisuels	Industries culturelles	Industries musicales
Intelligence artificielle	Interaction humain-ordinateur	Interactions politiques et organisationnelles
Interactions sociales	Jeux sérieux	Journalisme
Logique naturelle	Matérialisation des connaissances	Matérialités et infrastructures
Médias	Médias alternatifs	Médias et culture populaire
Médias et musique populaire	Médiations de vieillissement	Méthode d'étude sur le terrain
Méthodes numériques	Mobilités	Mondialisation des médias
Mouvements sociaux	Organisation sans but lucratif	Politiques culturelles
Politiques de radiodiffusion	Pragmatique du langage	Pratiques de travail en contexte de changement
Pratiques situées et institutionnalisées en milieu de travail	Recherche-création	Récits pratiques
Réseau de communication	Réseaux sociaux	Sémiotique
Sociologie de la culture	Sociologie des sciences et techniques	Sociologie du journalisme
Technologies de l'information et de la communication	Technologies interactives	Technologies numériques et intervention politique
Technologies numériques et intervention sociale	Théorie et analyse de l'image en mouvement	Théorie féministe

Théorie queer	Théories de l'espace public	Théories de la communication
Théories de la sexualité	Théories décoloniales	Théories du discours social
Théories féministes	Théories médiatiques	Travail collaboratif

## Annexe 4 : Les thèmes de recherche du Département d'informatique de l'Université de Montréal

<b>Expertises de recherche en informatique du guide des disciplines de l'Université de Montréal</b>		
Acquisition des connaissances (Systèmes experts)	Affectation du trafic	Algorithmes de graphes
Algorithme de routage	Algorithmes génétiques	Algorithmique
Alignement des séquences	Analyse combinatoire	Analyse des logiciels
Animation par ordinateur	Apprentissage analogique	Apprentissage automatique
Apprentissage de représentations	Apprentissage profond	Apprentissage statistique
Architecture des ordinateurs	Assistant de preuve	Automates finis
Automatisation du génie logiciel	Base de données temporelle	Bio-informatique
Calcul quantique	Capture de mouvements immersive	Centres d'appels
Circuits booléens	Classification non supervisée	Coloration des graphes
Combinatoire des mots	Commerce électronique	Compilation
Conception assistée par ordinateur	Conception de logiciels	Conception des réseaux
Conception paramétrique	Cryptographie	Cryptographie quantique
Déconvolution	Détection des contours	Données volumineuses
Économétrie	Enseignement à distance	Estimateur (statistique)
Estimation	Estimation des paramètres modèles	Évolution (biologie)
Extraction de caractéristiques visuelles	Fab labs	Famille multigénique
Finances	Fondements de la théorie quantique	Forage de données textuelles
Fusion	Générateur de nombres aléatoires	Génération automatique de textes
Génie biomédical	Génie logiciel	Génie logiciel empirique
Génomique	Génomique comparative	Génomique évolutive
Géométrie combinatoire	Gérontechnologie	Gestion de la mémoire
Gestion des réseaux	Gestion du revenu	Graphe de Caley
Graphes infinis	Homomorphismes de graphes	Imagerie médicale
Imagerie multispectrale	Immersion (réalité virtuelle)	Inéquation variationnelle
Inférence bayésienne	Infographie	Informatique bio-inspirée
Informatique quantique	Informatique théorique	Ingénierie de langages logiciels
Ingénierie dirigée par les modèles	Intelligence artificielle	Intelligence émotionnelle
Interface neuronale directe	Internet	Intrication quantique
Jeux poursuite évasion	Langage formel	Langage de programmation de haut niveau
Langages de programmation dynamiques	Langages de programmation fonctionnels	Lignes des produits logiciels
Logique combinatoire	Mathématiques discrètes	Mécanique quantique
Médias sociaux	Métaheuristique	Méthode de Monte-Carlo
Méthodes de décomposition	Méthodes formelles métriques (logiciel)	Modèles d'équilibre
Modèles probabilistes	Modèles statistiques	Modélisation
Modélisation de l'éclairage	Modélisation procédurale	Moteurs de recherche
Multi-projection simultanée	Optimisation combinatoire	Optimisation des opérations ferroviaires



Optimisation linéaire en nombres entiers	Optimisation mathématique	Optimisation réseau
Optimisation Stochastique	Optimisation des systèmes de transports	Parallélisme (informatique)
Phénomènes naturels	Prédiction de demande	Problèmes combinatoires
Problèmes de planification de grande taille	Processeurs graphiques (GPU)	Processus d'apprentissage
Programmation à deux niveaux	Programmation fonctionnelle	Programmation non linéaire
Programmation stochastique	Protection de la vie privée	Protocole de distribution quantique des clés
Protocoles de communication	Pseudotélépathie	Réarrangement de génomes
Recherche d'informations	Recherche de motifs biologiques	Recherche heuristique en génie logiciel
Recherche locale (optimisation)	Recherche opérationnelle	Reconfiguration dynamique des FPGAs
Reconnaissance d'activités en domotique	Reconnaissance de formes	Reconstruction 3D à partir d'images
Reconstruction de formes	Réduction de la variance	Rendu
Réseaux ad hoc	Réseaux de neurones	Réseaux de téléphonie mobile
Réseaux de transports	Réseaux futés	Réseaux sans fil
Réseaux sociaux	Résumé automatique de texte	Rétro-ingénierie
Robotique	Robustesse d'algorithmes	Science des données
Sécurité de l'information	Sécurité des systèmes d'information	Sécurité des technologies de l'information et de la communication
Sécurité informatique	Segmentation d'image	Séquençage de génomes
Simulation	Simulation à événements discrets	Simulation numérique
Simulation stochastique	Statistiques non paramétriques	Stéréoscopie
Stratégies d'apprentissage	Stratégies pédagogiques	Structures de données
Surfaces de subdivision	Symétrie des graphes	Synthèse d'images
Systèmes de types	Systèmes d'information de gestion	Tarifification optimale
Techniques hiérarchiques de relaxation	Technologies de l'information et de la communication	Technologie de l'information et de la communication pour l'éducation
Téléportation quantique	Théorie combinatoire des ensembles	Théorie combinatoire des groupes
Théorie combinatoire des nombres	Théorie de l'information quantique	Théorie de la complexité (informatique théorique)
Théorie des choix discrets	Théorie des graphes	Théorie des jeux
Topologie combinatoire	Traduction assistée par ordinateur	Traitement automatique des langages naturels (TALN)
Traitement d'images	Traitement du signal	Traitement parallèle (parallélisme)
Transports	Tutoriels	Types dépendants
Véhicules autonomes	Vidéosurveillance	Vision par ordinateur
Visualisation du logiciel		

## Annexe 5 : Les CT de COMMU et INFOR rejetés de l'analyse

Élagage COMMU	Élagage INFOR
corpus	al
osbl	n
al	uds
rs	ud
analyse	q
type d'écart	t
écart	k
individu	i
pratique	tfd
ms	v
policing	u
internauts tunisiens	http
personnage fictionnel	https
monde fictionnel	é
orthographe lexical	p
caractéristique	cf
apprentissage	dft
addenda	ex
tunisien	ion
lexical	ri
organisations policières	corpus
ethos adopté	analyse
ukulélé	méthode
relations publiques	évènement détecté
participation politique	utilisées
formes identitaires	phrase
commentaire laissé	individu
analysées	modèle
ponctuation	recherche d'information
déontologiste	dimension sociale
apprenant	approche
étude de cas	fréquence
cadre théorique	big
processusde recrutement	sémantique
pratiques policières	document pertinent
organisation	contexte
professionnel	caractéristique
jeunes professionnels	big social
utilisées	phrase candidate
ethos préalable	contenu textuel
feux de forêt	requête q
type de texte	dater
adoption	mot
tsunami	normalisées
analyse linguistique	probabiliste
recruteur	section
mouvement de contestation	ref

identité	document d
type	ints
reconfiguration	contextualisation
campagne de boycottage	lemmatisé
municipalité	lexique
contexte	calculer
locuteur	apprentissage
langue franc	attribut
usages contestataires	k
senoufos	n
éthique égoïste	é
ibid	type de réseau
médias sociaux utilisés	idft
régime éthique	maxw
analyse de contenu	assorec
stratégie	avgw
usage d'internet	citeall
cause d'intérêt général	supervisé
stakeholder	système
utilisé	vocable
typographie	chapitre
niveau d'analyse	ii
soit	po
personnel	manuellement
professionnel	méthode proposée
contre	terme t
orthographe	évaluation
révolution tunisienne	call
construction discursive	sous-ensemble
usage contestataire d'internet	utiliser
logique d'usage	document
signification sociale	identifier
déontologiste face	définis
pratique communicationnel	automatique
proulx	basées
orthographe grammatical	uvre
pratiques discursives	annotée
entretien collectif	terme
syntaxe	utilisateurs-trices
sensibilisation	cit
processus	z
forme	score
étude	learning
catégorie	sémantiquement
paradox	fréquence d'occurrence
énonciation	isotopie
marketing	section précédente
discours	texte lemmatisé
différent	nombre de tweets
sous-catégorie	lexicale
cit	caractérisation
	figure

sémiotique
x
contenir
classifier
za
négation
outil
wikipédia
critère temporel
call rec
communauté détectée
lemme
campagne d'évaluation
analyse de réseaux
k
citerec
tf
syntaxique
précision
extraire
description
opinion
différent
données issues
modélisation
iii
implémentation
proposés
type
collectée
méthode de détection
délect
ion de po
modèle de langue
mot t
assoall
ique
vi
nombre de n
region
traitement automatique