

Univerzita Karlova

Filosofická fakulta

Ústav románských studií



BAKALÁŘSKÁ PRÁCE

**L'anglais comme le déterminant de la terminologie française
contemporaine en informatique**

**English as a determinant of the contemporary French IT
terminology**

Ing. Jakub Šlechta
Praha 2024

Vedoucí práce:
Mgr. et Mgr. Radka Mudrochová, Ph.D.

Poděkování

Tímto bych rád poděkoval Mgr. et Mgr. Radce Mudrochové, Ph.D. za odbornou pomoc, hodnotné rady a čas, který mi věnovala při zpracování této bakalářské práce.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne

.....
Jakub Šlechta

Abstrakt

Bakalářská práce pojednává o jazykové situaci v rámci terminologie oboru informačních technologií v současné francouzštině. V rámci teoretické části je poskytnut exkurz do historického kontaktu francouzštiny a angličtiny, francouzského purismu a geneze a partikularit oboru informačních technologií, zejména ve vztahu k jejich terminologii. Praktická část posléze navazuje analýzou výskytu vybraných IT termínů skrze dotazování paralelního korpusu ParaCrawl.

Klíčová slova:

francouzština, anglicismus, terminologie, informatique, purismus

Abstract

The bachelor thesis deals with the linguistic situation within the terminology of the field of information technology in contemporary French. The theoretical part provides an excursus into the historical contact between French and English, French purism and the genesis and particularities of the field of information technology, especially in relation to its terminology. The practical part is followed by an analysis of the occurrence of selected IT terms through the interrogation of the ParaCrawl parallel corpora.

Key Words:

French, anglicism, terminology, information technology, purism

Table des matières

Introduction.....	8
1 Recherche théorique.....	8
1.1 La perspective historique sur l'influence de l'anglais sur la langue française	8
1.1.1 Le contact anglo-français en dehors de l'Europe.....	9
1.2 Jargon versus terminologie	11
1.2.1 Les outils de la création de terminologie	12
1.2.2 Les spécificités de la terminologie technique	13
1.3 Le purisme linguistique français	13
1.4 Le domaine des technologies de l'information et les particularités de sa terminologie.....	15
1.4.1 Le rôle de l'anglais dans la genèse de la terminologie de l'IT	17
1.4.2 Le rôle du français dans la genèse de la terminologie de l'IT	18
2 Partie méthodologique	21
2.1 La méthodologie de l'analyse	21
2.2 Les moyens et procédure d'analyse du corpus analysé	21
2.3 La méthode de sélection de termes	26
3 Analyse	28
3.1 Analyse des termes choisis	28
3.1.1 Téléverser.....	28
3.1.2 Pare-feu.....	29
3.1.3 Logiciel et Matériel.....	30
3.1.4 Bogue	32
3.1.5 Dialogueur.....	33
3.1.6 Module d'extension	33
3.1.7 Mémoire vive	35
3.1.8 Serveur mandataire	36
3.1.9 Témoin de connexion.....	37
3.1.10 Code à octets	38
3.1.11 Pseudo-code Java	39
3.1.12 IUG	40
3.1.13 EDI.....	41
3.1.14 Informatique en nuage	41
3.1.15 Application dorsale, application frontale	42
3.1.16 Boucle de programmation.....	43
3.1.17 Développeur généraliste.....	44

3.1.18	Cadre d'application.....	45
3.1.19	Entrepôt de données.....	46
3.1.20	Outil d'alimentation.....	47
3.1.21	Logiciel local.....	47
3.1.22	Métadonnées.....	48
3.1.23	Informatique décisionnelle.....	49
4	Résultats et synthèse de l'analyse.....	51
	Conclusion.....	54
	Résumé (en tchéque).....	55
	Bibliographie.....	56
	Liste des graphiques.....	59
	Liste des tableaux.....	59
	Liste des images.....	60

Introduction

L'influence de l'anglais contemporain sur la terminologie des autres langues indo-européennes est incontestable. Dans le même temps, les technologies de l'information sont un domaine en constante évolution dont le développement est loin d'éteindre la demande de dénomination de nouveaux concepts.

Par ailleurs, l'espace linguistique français est très puriste et la demande de terminologie propre est relativement significative. Tout aussi remarquable est le fait que l'histoire de l'interaction entre le français et l'anglais est très riche en raison des aspirations historiques concurrentes des formations étatiques de ce qui sont aujourd'hui la France et la Grande-Bretagne, du passé colonial et du partage de l'espace géographique.

Tous ces faits constituent un contexte très stimulant pour l'élaboration d'un mémoire de licence sur le sujet: L'anglais comme le déterminant de la terminologie française contemporaine en informatique.

Le travail sera divisé en deux parties: une partie théorique et une partie pratique.

La partie pratique du mémoire sera principalement consacrée à une excursion historique de base sur le contact anglo-français, le purisme français et la genèse du domaine des technologies de l'information. En même temps, des concepts linguistiques clés pertinents pour la mémoire seront thématés et déconstruits dans cette partie.

Dans la partie pratique, une analyse de l'utilisation de termes sélectionnés dans le domaine des technologies de l'information sera effectuée. Le corpus parallèle ParaCrawl sera utilisé pour cette analyse.

1 Recherche théorique

1.1 La perspective historique sur l'influence de l'anglais sur la langue française

L'influence mutuelle du français et de l'anglais est perceptible essentiellement dès le début du Moyen Âge. À ce moment-là on peut commencer à observer un contact plus intense entre les entités étatiques sur le territoire francophone et les îles Britanniques. Comme un point significatif concernant les relations culturelles, il est possible de marquer la conquête normande de l'Angleterre qui s'est déroulée sur une période de 1066 à 1071. Par cette conquête menée par Guillaume I^{er} appelé « le Conquérant », les élites anglo-saxonnes ont été remplacées par celles de Normandie. La nouvelle noblesse n'a emmené seulement un nouvel ordre social et politique, mais aussi toute la culture de la France continentale contemporaine, y compris la langue. (Cartwright, 2019) Depuis, le français a commencé à être le moyen de communication dominante de la haute classe.

Ce fait a significativement influencé le développement de l'anglais pour les siècles suivants. Pendant la période de XII^e au XV^e siècle la culture linguistique anglo-française a été établie. Tandis que, l'usage du français prévalait dans la communication des cercles sociaux supérieurs, l'anglais est resté la langue des gens ordinaires. Cet état de fait a causé que les deux cultures se mélangeaient pendant le moyen-âge entier. Le fait que l'anglais était parlé principalement par les classes inférieures ne signifie pas qu'aucune production écrite n'existait à l'époque. Comme plusieurs auteurs mentionnent, Il est possible de voir

que l'existence de l'anglais écrit a continué même après la conquête. Ça évoque que l'anglais a été parallèlement utilisé parmi ceux qui étaient responsables pour sa production écrite et ainsi les gens éduqués, très probablement d'origine sociale plus élevée. D'après cela, on peut bien voir que la situation en Angleterre représentait un environnement dans lequel les deux langues existaient parallèlement dans tous les aspects linguistiques, même la réalité d'usage et le contexte sociale ont été tout différents. (Wilson, 1943, p.40-44)

On peut bien voir que la dynamique de l'influence mutuelle entre les deux langues était en faveur du français à l'époque. La première époque historique du contact anglo-français culturel direct était déterminée plutôt par l'élément francophone. Ce fait a laissé une marque assez importante sur l'anglais contemporain. Certains spécialistes affirment que plus d'un tiers du vocabulaire anglais est d'origine française. Il est évident que cela ne provient seulement du fait que la noblesse francophone s'est établie dans les îles britanniques après la bataille d'Hastings. (Cartwright, 2019) Le français puis a joué le rôle dominant dans le monde politique pendant l'époque qui a duré depuis la fin du Moyen Âge jusqu'à la fin de la Première Guerre mondiale. Les facteurs comme la puissance militaire et politique du Royaume de France ont déterminés l'importance du français dans nombreux de domaines. En plus de la politique, surtout dans les arts et les sciences, par exemple. On peut dire qu'en particulier dans les domaines où des contacts internationaux ou interlinguistiques ont eu lieu. Le français de cette période pourrait donc être considéré comme une *lingua franca* contemporaine. Après la Première Guerre mondiale l'anglais a commencé à être de plus en plus significatif dans la diplomatie et la communication internationale en générale, une tendance qui se poursuit encore aujourd'hui.

D'une part, Il est évident que l'époque pendant laquelle le français influençait l'anglais a été beaucoup plus longue que l'inverse. D'autre part le dernier siècle, au cours duquel la langue de Shakespeare a pris le pas sur celle de Voltaire en termes d'importance mondiale, a été plus novateur que jamais. Le nouveau domaine des technologies de l'information a connu un essor sans précédent. Le progrès, non seulement technologique mais aussi politique par exemple à travers du développement des organisations internationales, et aussi l'émergence des médias de masse ont signifié l'impact rapide et significatif que l'anglais a pu avoir comme la langue dominante. Cette dynamique accélérée du développement au XX^e siècle a donc donné lieu à la situation contemporaine. (Phillipson, 2018, p.1-5)

1.1.1 Le contact anglo-français en dehors de l'Europe

Dans le sillage des découvertes outre-mer et de la pénétration des Européens dans les territoires nouvellement découverts, les ambitions coloniales de la France et de l'Angleterre, et par extension de l'Empire britannique, se sont accrues. Ces ambitions, dans le cas de ces puissances européennes, se traduisent par de véritables conquêtes et l'exercice d'une influence notamment sur le continent Afrique et Asiatique et aux Amériques. A partir de l'essor colonial au tournant des XVI^e et XVII^e siècles, la compétition de pouvoir mutuelle entre Français et Britanniques a commencé à se transposer dans ces territoires non européens. Ce fait a causé que les deux entités culturelles et linguistiques s'influencent parallèlement dans plusieurs régions du monde. Par conséquent, l'influence culturelle anglo-française mutuelle a pu exister dans plusieurs contextes socio-économiques et locaux très différents, ce qui a donné lieu à un échange culturel et linguistique incomparablement plus riche que dans le cas du contact

unidimensionnel traditionnel entre deux nations et leurs cultures. C'est le cas, par exemple, du contact entre les cultures polonaise et tchèque.

En ce qui concerne, les territoires francophones voisines des colonies britanniques, ces situations ont été nombreuses au cours de l'histoire. Les cas les plus significatives d'un tel contact avec un chevauchement notable dans le présent sont sans aucun doute la concurrence entre les Britanniques et les Français sur le continent africain et en Amérique du Nord. Comme on a déjà indiqué, ce ne sont pas les seuls exemples, néanmoins dans ces deux cas les européennes ont dominés les territoires sur le plan linguistique. Les deux cas sont différents ainsi les moyens utilisés pour y parvenir. Alors que en Amérique du nord il s'agissait d'établir une domination culturelle complète en soumettant par la force la population indigène, en Afrique, la domination linguistique des colonisateurs découlait principalement de la restructuration territoriale et politique du territoire, où la langue des colonisateurs est devenue le seul moyen linguistique commun pour les entités tribales incluses dans la formation administrative/étatique créée par les Européens. Ces deux cas ont des implications qui se recoupent aujourd'hui, car l'anglais et le français ont conservé leur statut de langues officielles ou utilisées indépendamment de la décolonisation des territoires susmentionnés. (Higonnet, 1968, p.57-60); (Griffiths, 1986, p.204-207)

1.1.1.1 Le cas du Canada

L'un des territoires où les Anglais et les Français ont mené leurs campagnes coloniales antagonistes a été la région située au nord-est de ce qui est aujourd'hui les États-Unis et le nord-ouest du Canada. Ces campagnes, associées aux luttes des grandes puissances sur le continent européen et dans d'autres territoires coloniaux, se sont transformées en conflits armés répétés à plus ou moins grande échelle. Ces dynamiques ont défini la lutte d'influence entre les hégémons européens en Amérique du Nord pendant une grande partie des XVII^e et XVIII^e siècles. Pourtant, le résultat de La Guerre de sept ans, souvent appelée « le 1^{er} conflit mondial » a changé cette dichotomie. Le royaume de France a perdu ses territoires au Canada. Il s'agissait en fait d'un signe avant-coureur de l'expulsion imminente de la France du continent nord-américain. Car, le grand territoire de Louisiane française est devenue insoutenable en étant coupé de la côte atlantique et a été pragmatiquement vendu aux États-Unis d'Amérique par la France de Napoléon en 1803. (Scott, 2011, p.422-425)

Bien que les Français aient été chassés du continent, le Canada francophone soumis à la Grande-Bretagne a réussi à maintenir son identité linguistique et culturelle française jusqu'à aujourd'hui. L'existence du Canada francophone a essentiellement préservé une zone parallèle de contact linguistique entre le français et l'anglais, au sein de laquelle les deux langues ont été et continuent d'être influencées mutuellement et de manière très particulière. (Blattberg, 2013)

1.1.1.2 Le cas de l'Afrique

L'Afrique n'a pas été colonisée à grande échelle par les puissances européennes avant le dernier quart du XIX^e siècle. Jusque-là, les Européens n'étaient généralement intéressés que par le contrôle des ports stratégiques qui leur permettaient de contrôler le commerce des marchandises et des esclaves. Jusqu'en 1870, les Européens ne contrôlaient qu'environ 10 % du territoire africain. Parallèlement à l'industrialisation croissante du continent

européen dans la seconde moitié du XIX^e siècle, l'idée d'exploiter les ressources minérales du continent africain a commencé à germer.

L'approche coloniale de l'Afrique a commencé à s'imposer dans les années 1880 et a été approuvée diplomatiquement lors de la conférence de Berlin. Elle aborde diplomatiquement la question des sphères d'influence et des conditions de légitimité de la colonisation afin d'éliminer les conflits militaires entre les grandes puissances découlant d'aspirations coloniales concurrentes. Le résultat a été le partage du continent africain par les États européens. Selon ce principe, 90 % de l'Afrique était occupée au début de la Première Guerre mondiale.

Le partage du territoire africain ne reflétait pas beaucoup la situation ethnique, linguistique et étatique des régions colonisées. Les frontières administratives écrites par les diplomates européennes ont causé que ces nouveaux territoires coloniaux ont très souvent inclus les dizaines de tribus ou des autres formations quasi-étatiques avec leurs propres langues et leur propre droit coutumier. En construisant des formations administratives de cette manière, les colonisateurs ont souvent créé les conditions d'une intolérance ethnique au sein de ces formations après leur décolonisation à la suite de la Seconde Guerre mondiale.

Quand la plupart de l'Afrique a été décolonisée au début des années 1960, les nouveaux états africains ont été proclamés précisément à l'intérieur des frontières administratives créées par les Européens. On peut dire que le concept d'État et ses institutions ont été adoptés à l'identique par les nouveaux États africains. Cela concerne aussi la langue officielle, car la langue des colonisateurs a souvent été le seul moyen de communication pour tous les habitants autochtones.

Les colonisations française et belge ont abouti à la création d'une ceinture consolidée de territoires francophones qui représentent près de la moitié de la superficie de l'Afrique et le français est parlé par environ 140 000 000 personnes ce qui en fait le continent avec le plus locuteurs du monde.

L'Afrique anglophone est plus petite géographiquement, mais elle est presque deux fois plus peuplée en termes de nombre de locuteurs. Les entités anglophones se trouvent principalement dans les parties occidentale, orientale et méridionale du continent. Les contacts entre le français et l'anglais sont observables principalement dans le centre-ouest de l'Afrique, où se déroulent les activités commerciales et les migrations transfrontalières. Le contact de loin le plus intense semble être celui entre l'anglais et le français au Cameroun, ces deux langues étant les langues officielles et utilisées dans le pays. (Griffiths, 1986,p.204-207); (Observatoire de la langue française, 2018 ,s.1-2)

1.2 Jargon versus terminologie

Tout d'abord, il est souhaitable de définir ce que l'on entend par jargon et terminologie dans le discours du domaine.

Le terme «*jargon*» semble être plutôt polysémique. Le Petit Robert le définit de deux manières. La définition première est «*langage déformé, fait d'éléments disparates ; par extension langage incompréhensible.* » et la deuxième est «*façon de s'exprimer propre à un groupe, une profession, une activité, difficilement compréhensible pour le profane.*» Au sein de ce mémoire on se concentre surtout sur ce qui est représenté par la deuxième phrase.

Conformément à la deuxième définition présentée, le concept peut être décrit plus en détail selon Nadziejka. D'après lui presque chaque métier ou profession a son propre langage. « *Les personnes exerçant un métier ou travaillant sur des chaînes de production l'appellent «langage d'atelier» ; celles qui travaillent dans des universités ou des organismes de recherche l'appellent «la terminologie acceptée dans le domaine» ; celles qui travaillent dans des bureaux d'édition l'appellent généralement «jargon».* (Nadziejka,1992 ,p. 663) ». Nadziejka mentionne que l'utilisation des idiomes et d'autres expressions typiques peut faciliter la communication dans un groupe lié à un domaine compliqué. Bien évidemment, cela rend l'accès des informations de certain domaine plus exclusive.

Ici on peut clairement voir deux dimension dans lesquelles les deux définitions présentées se chevauchent. Il s'agit de l'exclusivité de compréhension. Soit dans un environnement professionnel soit dans certaines cercles sociales. Même dans le cas de la première définition, qui correspond mieux à à l'acception courante du terme argot dans la terminologie linguistique, cet élément fonctionnel est le plus essentiel pour le jargon.

En ce qui concerne le terme « *terminologie* » Le Petit Robert le décrit comme suit. « *ensemble des désignations et des notions appartenant à un domaine spécial (science, technique, etc.)* ». Contrairement au jargon, la terminologie vise à la clarté et à la précision, souvent standardisée par des organismes officiels. C'est-à-dire, par rapport à jargon terminologie est formelle et rigoureusement définie, destinée à un usage plus large et souvent académique.

Une propriété très importante de la terminologie est qu'elle part du concept et procède ensuite à la désignation. La terminologie doit être absolument sûre qu'elle nomme un concept spécifique et non un concept similaire. C'est pourquoi les dictionnaires terminologiques privilégient les définitions descriptives exhaustives des concepts, qui indiquent souvent aussi les relations entre les concepts apparentés. (Cabré, 1999, p.34)

1.2.1 Les outils de la création de terminologie

La création de terminologie est un processus dont la méthode dépend naturellement du fait qu'il s'agit de la genèse de concepts existant dans une autre langue source, ou qu'il s'agit essentiellement de la création universelle de significations ou de concepts entièrement nouveaux.

En ce qui concerne le deuxième cas, il s'agit des situations où il faut établir des significations ou des concepts entièrement nouveaux sans possibilité d'utiliser les moyens d'expression présents dans d'autres langues naturelles. C'est-à-dire, il s'agit des cas tout sans précédent où il est nécessaire de nommer les réalités nouvellement construites. Il faut avouer que dans le premier cas susmentionné , le modus operandi du néologisme sémantique peut être le même, car la création lexicale dans une langue naturelle humaine est toujours unique, quel que soit le nombre d'autres langues humaines ayant une équivalence sémantique directe. Dans de tels cas, cependant, il est possible et souvent approprié d'utiliser des concepts existants et des équivalences sémantiques en langues étrangères.

En ce qui concerne la création de vocabulaire dans les domaines techniques, des outils de formation de mots utilisant des associations avec des concepts déjà existants dans le vocabulaire général sont souvent utilisés. « *La création d'un langage sectoriel implique, donc, un large éventail de moyens, allant de la simple extension sémantique aux processus*

complexes impliquant des métaphores, des métonymies, des associations représentationnelles » (Costăchescu, 2012, p.142-143).

Comme on a déjà indiqué, le contact interlinguistique au niveau des langages sectoriels peut aussi représenter non seulement un moyen de la néologie très intuitif mais aussi une approche très avantageuse. Telle création des lexèmes à travers de l'emploi des emprunts lexicaux est souvent le processus le plus naturel et continu de création de vocabulaire.

1.2.2 Les spécificités de la terminologie technique

L'essence du langage et du vocabulaire technique repose d'après Louis Gilbert sur une opposition vers le lexique général. Cela va de pair avec la pratique historique du refus de l'Académie française d'intégrer les termes artistiques et scientifiques de Thomas Corneille qui constituaient l'essentiel du contenu de son Dictionnaire des arts et des sciences finit en 1709. « *Cette décision reflétait l'idéologie dominante dans la société monarchique; il y avait d'une part le langage de la cour et des écrivains bien en cour, la plus saine partie des auteurs de ce temps selon Vaugelas, d'autre part le langage des métiers et des sciences qui ne relevait pas de la culture de l'honnête homme* ». La création des néologismes, qui décriraient la réalité accessible aux élites intellectuelles mais pas forcément à l'aristocratie, donc a été exclue du courant principal de normalisation et de codification du français de l'époque.

Telle approche vers le lexique de la langue française avait déjà été contesté et critiqué par les grands encyclopédistes des Lumières. Par exemple, Denis Diderot percevait le lexique d'une langue comme une ensemble cohérente des sens décrivant toutes les connaissances de ce peuple. C'est-à-dire que l'exclusion d'une couche de la langue n'est pas souhaitable dans ce contexte. Pourtant, la pratique de création lexicale, qui s'est instauré pour les siècles suivantes peut être qualifiée d'ambivalente, oscillant entre l'inclusion partielle et l'exclusion d'un vocabulaire spécifique. Cette exclusion à degrés divers, persiste jusqu'à l'époque moderne pour la simple raison que l'avance technologique décompose le détail de la connaissance en parties de plus en plus petites. Avec la vitesse de développement de certains domaines technologiques, un tel détail semble presque impossible à couvrir en temps réel dans le cadre des vocabulaires généraux de la langue. « *Le développement accéléré de la science et de la technologie au cours de la période récente s'est accompagné de l'apparition d'un grand nombre de nouveaux concepts, voire de nouveaux champs conceptuels, qui nécessitent de nouvelles dénominations.* » (Cabré, 1999, p.4) Cet aspect, en particulier depuis le début du XX^e siècle, a provoqué une plus grande prolifération des dictionnaires propres à chaque activité scientifique. Parallèlement, au cours de la seconde moitié du XX^e siècle, la terminologie technique (souvent informatique) est devenue une sorte d'outil universel de modernisation du langage commun. (Cabré, 1999, p.5-6)

1.3 Le purisme linguistique français

Le concept de purisme existe au sein de la langue française depuis le XVII^e siècle. Il s'agit d'une approche qui cherche à préserver la pureté de la langue en éliminant les emprunts étrangers, les tournures de langage jugées impures et en promouvant un usage conforme à des normes considérées comme idéales.

L'idée fondamentale du purisme revêt une importance cruciale. En substance, ce concept comporte deux aspects distincts : le terme « *pureté* » fait référence à un état idéal envisagé ou à un point de départ, tandis que les expressions « *purisme* » et « *purification* » décrivent l'action entreprise pour parvenir ou préserver cet état idéal. La pureté de la langue française est atteinte lorsque son utilisation actuelle concorde avec l'idéal envisagé. (Funk, 2018, p.111); (Houdebine, 2016, p.39-40)

Au XVII^e siècle, l'Académie française, créée en 1635, était chargée de créer des directives linguistiques et de les nettoyer des influences étrangères. Cette méthode visait à maintenir l'authenticité du français tout en établissant une norme pour le contact avec d'autres langues. (Houdebine, 2016, p.37-38)

Au XX^e siècle, les débats autour du purisme linguistique se poursuivaient en raison de la mondialisation. Certains linguistes et écrivains ont suggéré que le français nécessite une plus grande ouverture aux influences étrangères afin de maintenir sa pertinence dans un monde en évolution rapide. Cependant, certains persistent à défendre l'héritage linguistique français, considérant le purisme comme essentiel au maintien de l'identité de la langue. (Burke, 1998)

Différents niveaux de purisme français sont présents dans la langue. La promotion des mots français se fait par des moyens lexicaux plutôt que par des emprunts étrangers. Les alternatives françaises aux mots anglais sont fréquemment utilisées dans le discours puriste. En matière de grammaire, le purisme peut s'exprimer à travers la défense de constructions syntaxiques plus conformes aux conventions établies. « *Comme le dit la linguiste Adriana Costăchescu, « Cette situation s'explique, au moins partiellement, par les particularités de la politique linguistique des autorités françaises. Déjà en 1966 Georges Pompidou, alors Premier ministre, a créé sous son autorité le Haut comité pour la défense et l'expansion de la langue française. Cette organisation a subi plusieurs transformations, s'appelant actuellement la Délégation générale à la langue française et aux langues de France (DGLFLF), dénomination qui date depuis 2001. La DGLFLF est rattachée au ministère chargé de la Culture et de la Communication et elle a pour mission d'élaborer une politique de la langue française dans le but de garantir (i) la primauté du français sur le territoire national, (ii) la cohésion sociale, (iii) la promotion de la diversité culturelle et (iv) l'enrichissement et la modernisation de la langue française. Son activité est appuyée par d'autres organismes partenaires, comme le Conseil supérieur de la langue française, l'Académie française et la Commission générale de terminologie et néologie, cette dernière placée auprès du Premier ministre ».* (Costăchescu, 2012 ,p.142-143)

Certains voient au purisme une forme de discrimination linguistique, dans laquelle la langue se développe elle-même plutôt que d'évoluer par le biais de la diversité culturelle. Ceux qui rejettent le purisme soulignent qu'être prêt à accepter les influences étrangères peut améliorer la langue et lui permettre de s'adapter aux circonstances contemporaines. Le purisme français est un concept complexe et évolutif et il a été une composante essentielle de la préservation et de l'évolution de la langue française.

Dans la tendance au purisme linguistique français, on peut voir les sentiments vers le passé glorieux de la culture francophone du côté des Français contemporains. Dans ce contexte, la réalité est que toutes les tendances à la diversification linguistique sont perçues comme des distorsions de la langue. Le purisme semble avoir le soutien de la majorité de la société française. Néanmoins ce sont des Français qui sont confrontés à des néologismes de plus

en plus nombreux. Le fait que, dans beaucoup de nouveaux domaines, tel que celui des technologies de l'information la terminologie anglaise s'internationalise et par conséquent est globalement acceptée, cause une demande pour la terminologie française. Comme on a déjà dit l'une des origines les plus significatives de la pensée puriste se trouve plutôt dans la réminiscence de l'époque de la domination francophone. C'est-à-dire à l'époque où la grande majorité de la terminologie a été d'origine française. Par contre la situation contemporaine cause qu'il faut créer beaucoup de nouveau vocabulaire équivalent aux termes qui existent déjà au sein du discours internationale et avec lesquels les Français d'aujourd'hui sont déjà en contact parallèle. (Funk, 2018, p.113-115) Cela bien évidemment peut provoquer une certaine réticence ou la simple incapacité d'utiliser des lexiques redondantes. Un tel phénomène est très significativement présent dans la pratique des autres langues que on pourrait marquer comme puristes. Par exemple dans le cas du croate, où le vocabulaire officiel codifié ne remplace pas une grande partie du lexique original serbo-croate dans la communication quotidienne, même si c'est l'un de ses principaux objectifs. (Bradarić, 2023)

En ce qui concerne le domaine de l'IT, la vitesse de la genèse de nouveaux termes est énorme. C'est parce que on ne parle pas uniquement du vocabulaire directement lié aux technologies de l'information en tant que telles, mais de celui des sous-domaines qui se sont développés dans l'univers des technologies de l'information et qui devient de plus en plus utilisé. Dans nombreux cas, on peut même percevoir ces domaines comme nouvelles technologies. D'ailleurs, en 2023, il y a très peu nouvelles technologies tout indépendantes de ce qu'on considère l'IT. Ces domaines sont donc précisément ceux qui créent constamment une demande pour de nouvelles créations puristes au sein de la langue française. (Merriam-Webster, 2023)

1.4 Le domaine des technologies de l'information et les particularités de sa terminologie

Le véritable début de la genèse du domaine des technologies de l'information peut être daté du tournant des années 1930 et 1940. Il s'agit de l'époque où les premières machines à calculer ont été inventées. Ils ont souvent été développés à des fins militaires et de décryptage. L'un d'entre eux a été par exemple l'ordinateur binaire d'Alan Turing ou aussi «Colossus Mark I», l'ordinateur électronique qui a été créé pour déchiffrer les messages produits par le dispositif de codage Enigma. (Statistique Canada, 2021) La découverte de la semi-conductivité du silicium à partir d'une tension électrique et l'invention de la micro-transistor en 1947 constitue une étape importante. Au cours des deux décennies suivantes, le développement des ordinateurs centraux s'est principalement concentré sur les ordinateurs transistorisés. Le premier ordinateur de ce type est développé à l'Institut de technologie du Massachusetts en 1957.

Pendant la deuxième moitié du XX^e siècle la cybernétique s'est imposée comme une discipline technologique stratégique. Les ordinateurs et les systèmes de contrôle automatisés ont été de plus en plus utilisés dans de nombreuses industries stratégiques.

À la fin du XX^e siècle la cybernétique a commencé à jouer le rôle significatif en ce qui concerne la connectivité croissante dans un monde de plus en plus globalisé. Cette tendance s'est maintenue en principe jusqu'à aujourd'hui. Pendant les années 0 et les années 10 les technologies de l'information et de la communication n'ont seulement commencées à changer la façon de travailler et de la pratique dans beaucoup de domaines mais elles ont

aussi pénétrées la vie de la grande majorité des habitants de la planète. Les nouvelles plateformes électroniques ont non seulement modifié l'efficacité du travail et la gestion de l'information, mais elles ont également redéfini la manière dont les relations sociales se construisent dans la société. Les plateformes telles que les réseaux sociaux, etc. ont sans aucun doute provoqué un tremblement de terre considérable à cet égard. (Haigh, 2013, p.448-465)

D'après ce qu'on a mentionné il semble être évident que dans le cas du domaine des technologies de l'information il s'agit à vrai dire d'un des domaines les plus significatives et influents de toute l'histoire de l'humanité. Cette discipline est parmi «les autres significatives» toujours un peu plus spéciale et extraordinaire. Son extraordinaire réside avant tout dans l'influence qu'elle a exercée sur toutes les autres disciplines et domaines. Cela a sans doute eu un impact sur la vitesse de développement de ces autres disciplines et en général sur le développement civilisationnel de l'époque dans laquelle on vit.

Aujourd'hui, les technologies et les processus informatiques représentent un outil stratégique indispensable de la gestion de l'efficacité. Cela sans aucun doute constitue l'une des raisons pour lesquels le progrès dans ce domaine détermine le mouvement des autres secteurs. Le domaine des technologies de l'information ne cesse pas à évoluer.

Pour mémoire, grâce à la célèbre loi de Moore, qui stipule que tous les 18 mois, il est possible de doubler le nombre de transistors sur un circuit intégré, au prix actuel, on peut supposer que la performance de calcul d'un ordinateur commercial a le potentiel d'au moins doubler à cet horizon. Si l'on considère la vitesse de développement ne serait-ce qu'à travers ce paradigme informatique, on obtient un message assez clair sur le potentiel et la nature révolutionnaire de ces changements continus. (Haigh, 2013, p.444-465)

Toutefois, ces dernières années, les programmes informatiques et l'intelligence artificielle ont commencé à pénétrer cette discipline. L'intelligence artificielle sous la forme d'algorithmes d'apprentissage automatique, en particulier, semble être un élément qui a le potentiel de changer radicalement non seulement notre façon de travailler, mais probablement la société dans son ensemble. (Hodková, 2024)

D'après certains experts l'invention de l'intelligence artificielle représente le même jalon que l'invention de l'électricité. *« Elle va transformer l'ensemble des secteurs économiques et créer énormément de valeur. Des technologies telles que l'apprentissage supervisé démultiplient la puissance de l'automatisation. Or l'automatisation aura des conséquences sur toutes les branches de l'économie, de la santé à l'industrie manufacturière, en passant par logistique et le commerce de détail. »* (Jewell, 2019) Les algorithmes d'apprentissage automatique jouent un rôle important sur le marché depuis plus de deux décennies. Ces algorithmes dits « supervisé » se sont imposés comme des outils puissants pour la construction de modèles statistiques prédictifs dans de nombreux domaines travaillant avec des données massives, mais ce n'est que lorsque les grands modèles de langage ont été mis à la disposition du public que la technologie a vraiment suscité l'intérêt du grand public et a pris de l'ampleur. (Jewell, 2019)

Les technologies de l'information sont susceptibles non seulement de maintenir leur impact actuel sur la société dans son ensemble, mais aussi de l'approfondir, la dynamique du changement progressif continuera à s'intensifier. Sous ce contexte, le langage naturel

humain subira cette influence continue, ce qui constitue un excellent domaine pour l'analyse et la recherche sociolinguistiques. Des analyses et des recherches similaires à celles qui font l'objet de ce mémoire.

1.4.1 Le rôle de l'anglais dans la genèse de la terminologie de l'IT

L'influence de l'anglais sur la terminologie de l'IT a été éminente. Cette affirmation repose sans aucun doute sur le fait que les premiers développements ont eu lieu principalement dans le monde anglophone. Il s'agit en particulier de la période allant de la Seconde Guerre mondiale jusqu'à la fin des années 1950. C'est-à-dire de la construction des premiers ordinateurs à transistors. C'est-à-dire depuis le développement des machines de décryptage dans le camp allié pendant la guerre à la construction des premiers ordinateurs à transistors. Bien évidemment, l'aspect cryptographique de la guerre avec l'Allemagne nazie a causé que ce type de connaissances et de savoir-faire ont été cumulés principalement aux États-Unis et au Royaume-Uni, autrement dit sur le territoire anglophone. (Haigh, 2013, p.456-457) Le progrès du domaine après la guerre a été le plus soutenu et significatif dans ce pays. Cela est dû, entre autres, au fait que les pays victorieux de la guerre recrutaient les scientifiques allemands, souvent les anciens nazis. (Watkins, 2014, s.1-2) En outre, Non seulement les Soviétiques n'ont pas donné la priorité à la recherche dans ce types des domaines comme l'électronique, cybernétique ou génétique, mais ils les ont directement marginalisés. Peu après la fin de la Seconde Guerre mondiale, ces disciplines ont été qualifiées par les autorités soviétiques, parmi d'autres, comme des pseudo-sciences bourgeoises . La position de la cybernétique en URSS n'a commencé à changer progressivement qu'avec la condamnation du stalinisme après le XXe congrès du Parti communiste de l'Union soviétique en 1956. (Kvapilík, 2010)

Toutefois, compte tenu des faits susmentionnés, il est évident que c'est avant tout le monde occidental et anglophone qui a occupé la position la plus privilégiée dans le développement de la cybernétique et des technologies de la communication au cours des deux premières décennies de l'après-guerre. Les industries des pays alliés, notamment celle aux États-Unis, n'ont pas été détruits pendant la Second Guerre mondiale, au contraire, la croissance de leurs économies était stimulée par la production de guerre. (Tassava,2008)

Les premiers ordinateurs véritablement commerciaux ont été par des entreprises américaines ou britanniques. Cette tendance s'est poursuivie jusqu'au milieu des années 1980. Entre-temps, des entreprises américaines telles qu'Apple et Microsoft ont introduit des systèmes d'exploitation et d'autres logiciels pour les ordinateurs personnels. C'est dans les pays anglophones, en particulier aux États-Unis, que divers éléments des technologies de l'information ont été introduits au cours des premières décennies de boom de l'informatique. Ainsi, l'impact du monde occidental, surtout des pays anglophones, était don très éminent à cette époque. Bien entendu, cette influence comporte également une dimension culturelle. (Haigh, 2013, s.443-451)

Tout le nouveau vocabulaire du domaine, basé sur la terminologie technique, a été établi dans un environnement anglophone. L'hégémonie croissante de l'anglosphère dans le monde des affaires dont il a été question dans les chapitres précédents de cette mémoire s'est traduite par une adoption majoritaire de cette nouvelle terminologie informatique dans la plupart du monde. Il est possible de marquer le domaine comme celle qui a été littéralement inventé et développé dans la langue anglaise. Les raisons principales en sont alors doubles. Il s'agit premièrement de l'hégémonie technologique déjà mentionnée et deuxièmement de l'hégémonie culturelle émergente du monde anglophone occidental.

Il faut se rendre compte, que grâce à cela le domaine n'est devenu pas seulement un objet de cette « anglophonisation » mais il a servi comme son instrument direct. En effet, il ne fait aucun doute que les technologies de l'information et de la communication ont rendu l'ensemble du processus de mondialisation infiniment plus efficace. (Lawlor, 2008, p.45-46)

1.4.2 Le rôle du français dans la genèse de la terminologie de l'IT

Comme la partie du monde occidental la France et la plupart de l'espace francophone ont bénéficié du progrès dans le domaine de l'IT. Toutefois, la France et d'autres pays occidentaux francophones, n'ont pas été aussi pionniers dans l'IT comme les pays du camp allié pendant la Seconde Guerre mondiale. Cela est très compréhensible, car l'économie française et d'autres pays européens francophones qui lui sont liés ont été beaucoup plus touchés par tous les impacts de la guerre. Les économies de ces pays ont été soit gravement affectées, soit restructurées en des économies de guerre. *«En Grande-Bretagne et en Amérique, de grands projets de défense comme les machines à déchiffrer et l'ENIAC ont réuni des experts de tous les domaines et de toutes les disciplines, qui ont ensuite partagé des objectifs et des parcours communs. Pendant ce temps, dans la France occupée par les nazis, la principale obsession de la plupart des gens était de survivre ; ceux qui ont continué à se battre l'ont fait avec les moyens les plus grossiers, non avec les technologies les plus avancées.»* (Musée Informatique, 2012)

Des années 1950 aux années 1960, l'informatique s'est imposée sur le territoire français. Même si les inventions et les tendances aient eu tendance à provenir du monde anglophone, la France et d'autres marchés francophones sont devenus des marchés très importants pour les TIC. L'écart technologique entre surtout les États-Unis et d'autres parties du monde développé en informatique au début des années 1960 a provoqué la création du lobby informatique, qui a été actif au sein des institutions comme le Commissariat général du Plan ou la Délégation Générale à la Recherche Scientifique. Au milieu des années 1960, l'informatique est devenue un domaine stratégique pour l'état français. En ce qui concerne le véritable développement français, il y avait certains déterminants qui ont compliqués l'émergence du domaine depuis le début. Par exemple, la plupart des universités n'avaient pas et, pour diverses raisons, ne voulait pas avoir des départements d'ingénierie électrique. En même temps, le domaine a été mal perçu par la plupart des chercheurs aux universités et matières techniques et cette approche de la communauté professionnelle a resté presque tout inchangé jusqu'au milieu des années 1970.

Le rôle de victime pendant la guerre susmentionnée et le fait que toute l'économie française a été considérablement touchée à l'époque n'explique pas entièrement toutes les années de retard. Il est possible d'y voir les raisons on peut dire plutôt culturelles qui déterminaient le progrès social bien avant la Seconde Guerre mondiale. Il faut se rendre compte du fait que, avec « l'anglophonisation » croissante après la Première Guerre mondiale, la recherche française, toutes disciplines confondues, a commencé à montrer certaines tendances isolationnistes vers des influences étrangères. En outre, il y avait depuis toujours une affinité pour l'absence de travail en équipe dans la tradition universitaire française. (Musée Informatique, 2012)

En raison du retard, causé par les aspects susmentionnés, le rôle culturel de la France et du français dans la genèse du domaine n'était aussi éminent que celui de l'anglais. Malgré cela, la nature puriste du français provoqué la création immédiate des néologismes

équivalents aux nouveaux termes anglais décrivant les outils, processus et rôles présents en IT. Un exemple est la codification est un terme clé utilisé non seulement par la communauté professionnelle, mais aussi par le grand public pour désigner un appareil pour utilisation informatique. Le mot « *ordinateur* » équivalent à l'origine anglais « *computer* » a été adopté en 1955 et ducoup presque immédiatement après la sortie du premier ordinateur produit en masse. Le mot a été choisi d'après sa racine latine « *ordinat* » signifiante l'ordre. Le nouveau terme crée qui déjà existait au sein du vocabulaire théologique désignant selon le Littré « *Dieu qui met de l'ordre dans le monde* » a été préféré à la traduction littérale de l'origine anglais. Pour le créateur de ce néologisme, le philologue, latiniste et professeur à la Faculté des lettres de Paris Jacques Perret, le mot « *calculateur* » ne transmettait pas suffisamment au public l'essence de cet appareil révolutionnaire. Il a choisi le terme « *ordinateur* » parmi ses autres prototypes originaux comme « *systèmeur* » ou encore « *combineur* », mais aussi « *digesteur* » et d'autres termes qui, d'un point de vue actuel, semblent plutôt bizarres. Le nom qui a été déposé par la distribution française de la marque IBM a passé très vite dans le langage parlé. (Bern, 2020) Cet exemple de formation des mots au sein du domaine des technologies de l'IT illustre très bien l'essence de l'approche particulière française de la genèse du vocabulaire d'une discipline nouvellement établie. Bien évidemment, les néologismes équivalents au terme « *computer* » ont été adopté dans plusieurs langues, cependant le processus complexe derrière le choix du mot « *ordinateur* » est un exemple de certain besoin de bien et précisément nommer et comprendre les éléments au tour de nous. Ce qu'on marque comme une conséquence de du purisme linguistique peut également être concomitant du besoin social d'intégrer dans le langage uniquement les concepts dont le signifiant capturera suffisamment la substance du signifié.

Comme l'affirme la linguiste Adriana Costăchescu, « *Quand il s'agit d'adapter dans une certaine langue la terminologie scientifique ou technique, d'un domaine dont la terminologie est consacrée dans une autre langue, théoriquement deux attitudes antithétiques sont possibles : il y a des langues où les mots anglais ont été repris tels quels, par exemple l'italien, le roumain, le russe. Dans ces cas l'emprunt, plus ou moins adapté, est le principal moyen employé. L'autre procédure est celle de traduire et d'adapter les mots étrangers. Le français favorise ce second moyen, puisque le vocabulaire sectoriel de l'informatique et de l'Internet contient peu d'emprunts.* » (Costăchescu, 2012, p.142)

La genèse du vocabulaire de l'informatique français a été rendue possible grâce au complexe de moyens d'enrichissement lexical, tandis que l'établissement celui de l'anglais s'est appuyé surtout sur les moyens rhétoriques comme les métaphores et les métonymies avec l'appui desquels l'expressivité et la vivacité du langage du domaine a été maintenu. Quand il y a un besoin dans la société d'établir une nouvelle notion, en plus de l'emprunt ou de l'adaptation des originaux la langue peut recourir au développement au sein du système lexical propre, l'invention d'un mot ou l'extension de celui qui déjà existe. À cause des raisons politiques et culturelles susmentionnées, dans le cas du vocabulaire français informatique la voie de développement lexicale a été choisie. (Costăchescu, 2012, p.142)

Ainsi, si l'on parle de l'influence du français sur le vocabulaire actuel de l'informatique, il semble inconcevable que celle-ci soit nulle. La vérité est que l'approche puriste et donc le développement du champ lexical a eu pour conséquence l'enrichissement global du vocabulaire français du domaine considéré. Cette information n'est pas surprenante, néanmoins il faut la percevoir dans le contexte du fait que les inventions des autorités

linguistiques français ont un impact potentiel sur tout l'espace francophone. Le développement du champ lexical français applicable aux sciences des technologies de l'information ne représente pas seulement une création des équivalents déjà existants qui viennent de l'environnement linguistique différent mais une fondation du contact parmi la culture francophone et un nouveau domaine technique qui a le potentiel de tout changer. De ce point de vue, le rôle du français et, en principe, des autres grandes langues peut être qualifié de non marginal. Toutefois, un tel point de vue est, comme déjà indiqué précédemment, fortement déterminé par la dimension hégémonique du français basée sur le fait que la mode de perception linguistique de la réalité de millions de locuteurs vivants en dehors de l'Hexagone est souvent construit à travers des décisions des autorités à Paris.

En même temps, il ne faut pas oublier l'aspect des influences linguistiques anglo-français historiques qui a été mentionnée au sein d'un des chapitres précédents. Il existe un nombre conséquent de mots anglais empruntés à l'ancien français. En outre, certains d'entre eux se sont devenus la partie intégrale du vocabulaire de l'IT original. Comme exemple il est possible de mentionner un mot français ancien « *défaut* » avec le sens de « *carence, pénurie ; péché* ». « *En français moderne le mot dérive de l'ancien participe passé de défailir au sens de « manquer » (fin XI^e siècle, cf. PR). A son tour défailir provient du latin populaire fallire « tromper, échapper à ». L'emploi du mot se retrouve non seulement dans la langue courante, mais aussi dans plusieurs domaines techniques: droit (jugement par contumace/ par défaut, emploi qui se retrouve en anglais aussi, « judgement by default »), chasse (pour désigner la perte de la piste de l'animal pourchassé), en psychologie (défaut d'expérience, de mémoire), en mathématiques et en physique nucléaire (où le mot signifie « différence »).» (Costăchescu, 2012, p.142) L'utilisation du mot en anglais contemporain, en revanche, est largement dominée par la signification « *absence, manque* ». L'expression au sein de laquelle le mot est présent dans le vocabulaire de l'IT d'aujourd'hui « *by default* ». Ce terme désigne le plus souvent l'état dans lequel le logiciel utilise la valeur qui est attribué automatiquement sans aucune indication de l'utilisateur. Sur la base de ce qui précède, il est possible de dire que grâce à ce contact culturel anglo-français très intensif qui a duré plus que millénaire, un mot qui a été adopté par la langue anglaise a fait son retour dans le français moderne après des centaines d'années dans le vocabulaire informatique. Tout cela bien illustre non seulement l'impact du français sur le vocabulaire du domaine des technologies de l'information mais aussi la dynamique générale de l'interaction entre l'anglais et le français.*

2 Partie méthodologique

2.1 La méthodologie de l'analyse

L'analyse au sein de ce mémoire se servira d'une sous-analyse lexicographique à travers de laquelle les termes analysés seront choisis. Au total nous choisirons 25 termes français de l'IT, et la sélection des mots spécifiques se fera selon trois catégories de base. Il s'agira de mots utilisés quotidiennement par le grand public dans le contexte des technologies de l'information, puis de mots liés au sous-domaine BI/DWH, et la troisième catégorie sera composée de termes relatifs au développement d'applications. Les deux derniers sous-domaines seront sélectionnés en raison de la réalisation des projets de l'auteur dans ces domaines. Ce fait promet une connaissance approfondie de l'utilisation spécifique du langage et, par extension, la sélection d'un échantillon de qualité de termes. La première sélection des termes permettra ensuite d'établir un contraste entre les termes informatiques généraux qui sont déjà largement entrés dans le vocabulaire du grand public et le jargon largement technique existant dans les domaines de la BI/DWH et du développement d'applications.

Les termes dans le cadre de trois catégories ci-dessus seront sélectionnés sur la base d'une recherche de sources expertes et de l'expérience pratique de l'auteur évoluant dans un environnement authentique de développement de logiciels.

La sélection de base des termes se fera en fonction de l'utilisation de la langue, en particulier dans un environnement international. L'utilisation des équivalents pertinents dans l'environnement francophone contemporain sera ensuite analysée. L'analyse se concentrera sur l'utilisation parallèle du vocabulaire informatique anglais international en français.

L'analyse comprendra deux éléments clés :

Dans un premier temps, elle sera mise en œuvre par la recherche dans le Grand dictionnaire terminologique et dans les dictionnaires monolingues. Plus précisément, il s'agira principalement du dictionnaire français Larousse et du dictionnaire anglais Merriam-Webster. En l'absence de termes analysés, d'autres dictionnaires généraux monolingues tels que Le petit Robert ou le dictionnaire de Cambridge seront consultés. Si nécessaire, des sources telles que des encyclopédies techniques comme Techopedia, etc. seront consultées pour obtenir une définition de base des termes.

Dans un second temps, l'analyse sera effectuée à travers de l'utilisation du corpus parallèle ParaCrawl ce qui est un ensemble de données comprenant le téléchargement de documents, le prétraitement et la normalisation, et l'alignement des documents et des segments. Ce corpus parallèle a été créé à l'aide de l'outil de recherche de données parallèles Bitextor permettant de récolter automatiquement des bitextes, c'est-à-dire les textes parallèles en deux ou plusieurs langues, à partir de sites web multilingues. (ParaCrawl). L'analyse se concentrera individuellement sur chacun des termes analysés et tentera de fournir un aperçu qualifié de la pratique de l'usage du vocabulaire de l'IT.

2.2 Les moyens et procédure d'analyse du corpus analysé

Comme nous l'avons mentionné, les termes seront analysés à l'aide du corpus parallèle ParaCrawl. L'interrogation du corpus sera effectuée par l'intermédiaire de l'interface web

OPUS - Corpus query (CWB). Dans cette interface web, les corpus xml originaux, qui sont codés selon la norme CWB (Corpus Workbench), peuvent être interrogés par le biais de scripts CQP (Corpus query processor). Il s'agit d'une plateforme d'interrogation qui permet d'interroger les données de manière linéaire. (OPUS) L'interface web susmentionnée ne permet pas l'utilisation de requêtes analytiques et utilise également un ordre aléatoire tout en maintenant une typologie représentative des sous-corpus. Les scripts d'interrogation seront construits de manière à ce que toutes les formes possibles du terme puissent être recherchées, voir l'exemple de requête CQP pour le terme « cloud computing » : [word="cloud"] [word="comput(es|ed|ing)?"]. De même, la construction de la requête permettra de sélectionner le contexte correct en cas de polysémie.

OPUS - Corpus query (CWB)

L'image 1 CWB

La sortie interrogée sera limitée à cent occurrences en raison de la compacité de l'ensemble de données pour l'analyse. La sortie de l'interface web sous la forme d'un tableau sera ensuite extraite du code source html de la page.

```

<option selected value="1">1</option>
<option value="2">2</option>
<option value="3">3</option>
<option value="4">4</option>
<option value="5">5</option>
</select><select name="bc" >
<option value="file">file</option>
<option value="p">p</option>
<option selected value="s">s</option>
</select><select name="acs" >
<option selected value="1">1</option>
<option value="2">2</option>
<option value="3">3</option>
<option value="4">4</option>
<option value="5">5</option>
</select><select name="ac" >
<option value="file">file</option>
<option value="p">p</option>
<option selected value="s">s</option>
</select></td> <td>positional annotation<br><input type="checkbox" name="attr" value="word" checked >word<input type="chec
</form></div></td></tr></table><p><hr><div class="result">Query string: '[word=&quot;cloud&quot;] [word=&quot;comput(es|e
</body>

```

L'image 2 Code HTML

Ensuite, ce tableau html sera converti en fichier csv en langage de programmation python à l'aide de la bibliothèque BeautifulSoup4.

```
analyza.py x analyza2.py
1 from bs4 import BeautifulSoup
2 import pandas as pd
3
4 filename = 'C:\\Users\\sleeh\\Downloads\\clou_comp2.html'
5
6 with open(filename, 'r', encoding='utf-8') as f:
7     html_file = f.read()
8
9 usage
10 def html_to_csv(html_file, output_file):
11     # Analyze HTML content
12     soup = BeautifulSoup(html_file, features='html.parser')
13     # Find all tables in the HTML
14     tables = soup.find_all('table')
15
16     if len(tables) == 0:
17         print("No tables found in the HTML.")
18         return
19
20     table = tables[0]
21
22     rows = []
23
24     for row in table.find_all('tr'):
25         # Initialize list for cells
26         cells = []
27         # Iterate through all cells in the row
28         for cell in row.find_all(['td', 'th']):
29             # Append cell text to the list
30             cells.append(cell.get_text(strip=True))
31         rows.append(cells)
32
33     df = pd.DataFrame(rows)
34
35     df.to_csv(output_file, sep=';', index=False, header=False)
36
37     print(f"CSV file '{output_file}' has been created successfully.")
38
39 print(html_file)
40 output_file = 'output.csv'
41 html_to_csv(html_file, output_file)
```

L'image 3 Code python

Dans l'étape suivante, l'analyse proprement dite sera effectuée sur le nouvel ensemble de données csv à l'aide des bibliothèques pandas et seaborn.

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df=pd.read_csv(filepath_or_buffer: 'C:\\Users\\sleeh\\PycharmProjects\\pythonProjectdiplomka\\output.csv', delimiter=',')
pd.set_option('display.max_colwidth', None)
word='cloud'
result = df[df['fr'].str.contains(word, na=False, case=False)]

word2='nuage'
result2 = df[df['fr'].str.contains(word2, na=False, case=False)]

inner_join = pd.merge(result, result2, on='result', how='inner')
print(inner_join)
inner_join_count=inner_join.shape
common_results = inner_join['result'].tolist()

result = result[~result['result'].isin(common_results)]
result2 = result2[~result2['result'].isin(common_results)]
result_count=result.shape
result2_count=result2.shape
union_all = pd.concat([result['result'], result2['result']])

value_counts = union_all.value_counts()
print(value_counts)
result3 = df[~df['result'].isin(value_counts.index)]
result3 = result3[~result3['result'].isin(common_results)]

print(result3)
result3_count=result3.shape
print(result3_count)
print(result_count)
print(result2_count)
print(inner_join_count)
overall=result_count[0]+result2_count[0]+result3_count[0]+inner_join_count[0]
print(overall)

```

L'image 4 Code python 2

Le code ci-dessus divise l'ensemble de données en sous-ensembles en fonction des variantes des termes de recherche et calcule leur fréquence. La logique de recherche des chaînes ou de leur concaténation sera conçue dans chaque cas pour couvrir toutes les formes d'un terme donné, voir le résultat suivant.


```

[3 rows x 5 columns]
result
1055863126 1
1055879075 1
1079003888 1
1079027646 1
1136038390 1
..
1491621114 1
1493799440 1
1504456768 1
1518389495 1
1539522124 1
Name: count, Length: 82, dtype: int64
   result ...
18 1109912829 ... Le géant français de l' infonuagique OVH investira 32,9M$ pour mettre sur
26 1147730448 ...
32 1218490599 ...
37 1218553682 ...
45 1337536854 ... Associé
50 1343241535 ...
54 1429432029 ... Xpertdoc conclut un partenariat avec OneShield ; embauche pour
55 1429432967 ...
56 1429489387 ...
81 1518308712 ... Aussi , nous avons consacré 1 % de nos avoirs , 1 % du temps de nos employés et 1 % de nos
84 1518389678 ...
89 1554499052 ...
90 1555457184 ... Ce programme avancé que nous avons conçu pour
91 1555521218 ... Pa
97 1617764346 ...

[15 rows x 3 columns]
(15, 3)
(49, 3)
(33, 3)
(3, 5)
100

```

L'image 5 Sortie du programme

Les résultats de l'analyse seront placés dans un tableau distinct pour chaque terme dans le format suivant.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Terme no.1	'Terme no.1'	51	Terme anglais	[word="cloud"] [word="comput(es)(ed ing)?"]
Terme no.2	'Terme no.1'	35	Terme anglais	-,,-
Les autres	x	14	Terme anglais	-,,-

Le tableau 1 L'exemple de tableau résultant

Un graphique sera généré sur les fréquences résultantes des catégories mentionnées pour chaque terme. Les bibliothèques python Seaborn et Matplotlib seront utilisées pour générer le graphique.

```

0 data1={
1     'Terme': [word_word2,'les deux','les autres'],
2     'Nombre': [result_count[0],result2_count[0],inner_join_count[0],result3_count[0]]
3 }
4
5 sns.barplot(x='Terme', y='Nombre', data=data1, palette='Set3')
6 plt.xlabel('Terme')
7 plt.ylabel('Nombre')
8 plt.title('Représentation en données')
9 plt.show()

```

L'image 6 Code python 3

Outre les chaînes de recherche, le graphe contiendra également des informations dans la catégorie Les deux sur les cas dans lesquels les deux termes de recherche se chevauchent. En cas de pertinence (si l'une des chaînes n'est pas un nom, etc.), ce chevauchement sera également enregistré dans le tableau résultant.

2.3 La méthode de sélection de termes

Comme il était déjà susmentionné la sélection est composée de trois sous-sélection. Chaque de ces groupes thématiques se concentre sur le sous-domaine différent. Alors que le premier groupe couvre principalement le vocabulaire avec lequel le grand public a un certain contact et dont on peut s'attendre à ce qu'il figure dans les dictionnaires généraux. Les autres groupes traitent de termes qui peuvent être considérés comme faisant partie du langage professionnel dans le contexte donné.

Les termes ont été choisis principalement en raison de l'existence d'un terme international/anglais équivalent, largement établi, avec lequel le terme français analysé sera mis en opposition. Cette opposition permettra d'obtenir un matériel d'analyse illustrant la situation réelle de l'utilisation contemporaine des purismes et des anglicismes dans le domaine de l'IT.

Par rapport à l'intention initiale, l'échantillon a été réduit de cinquante à vingt-cinq dans une distribution de sous-sélection de 10, 10, 5 (1^{er} groupe 10, 2^e groupe contenant des termes du développement d'applications 10 et 3^e groupe contenant des termes de BI/DWH 5). Cette réduction a été effectuée principalement en raison des limites du corpus analysé. Ces limitations découlent de la typologie des pages web qui sont utilisées pour créer le corpus ParaCrawl. En effet, il s'agit de données dans lesquelles la probabilité d'occurrence de termes décrivant la nature ou l'état de solutions résultant du développement d'applications ou d'entrepôts de données est plus élevée. Toutefois, la probabilité d'apparition de termes décrivant des activités et des processus dans le cadre de ces développements est beaucoup plus faible. Pour capturer ces termes, le corpus devrait couvrir des plateformes de développement telles que Stack Overflow, Git Hub, etc.

La sélection des termes pour l'analyse finale est comme suit:

Le groupe de sélection	Terme
1	Téléverser
1	Pare-feu
1	Logiciel
1	Matériel
1	Bogue
1	Dialogueur
1	Module d'extension
1	Mémoire vive
1	Serveur mandataire
1	Témoin de connexion
2	Code à octets
2	Pseudo-code Java
2	IUG
2	EDI
2	Informatique en nuage
2	Application dorsale
2	Application frontale
2	Boucle de programmation
2	Développeur généraliste
2	Cadre d'application
3	Entrepôt de données
3	Outil d'alimentation
3	Logiciel local

Le tableau 2 La liste des termes sélectionnés

3 Analyse

3.1 Analyse des termes choisis

3.1.1 Téléverser

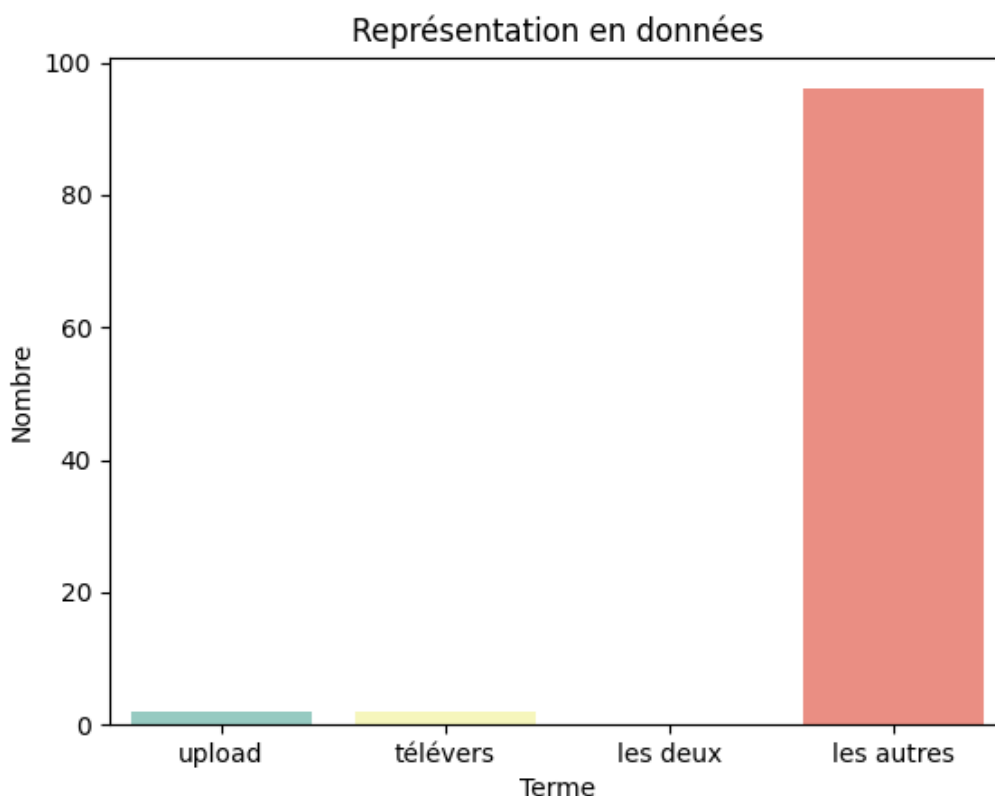
En sus de l'apparition du mot « *télécharger* », le verbe « *téléverser* » représente une variante moins fréquente équivalente au terme anglais original « *upload* », qui est défini par Merriam-Webster comme « *acte ou cas de transfert de quelque chose (comme des données ou des fichiers) d'un ordinateur ou d'un autre appareil numérique vers la mémoire d'un autre appareil (comme un ordinateur plus grand ou distant).* » (Merriam Webster) D'après le dictionnaire Larousse, le verbe téléverser est décrit comme le transfert des données d'un ordinateur local vers un autre qui est distant utilisant un réseau de télécommunications. (Larousse)

Comme indiqué ci-dessus le terme est dans le même sens utilisé dans le monde francophone en parallèle avec le mot « *télécharger* ». L'aspect problématique du terme est représenté par le fait qu'il s'agit d'un exemple de la contronymie car il a aussi le sens équivalent au terme anglais « *download* », par conséquent du terme décrivant presque la même action du point de vue de l'observateur mais totalement opposée du point de vue son acteur.

Le Grand dictionnaire terminologique recommande l'usage du mot « *téléverser* » plutôt que « *télécharger* » pour la raison évidente qu'il permet d'éviter la confusion. Le dictionnaire aussi indique la pratique de l'utilisation de l'anglicisme « *uploader* » et ne la recommande pas. « *L'emprunt hybride uploader est déconseillé puisqu'il est formé du terme anglais upload et du suffixe verbal français -er.* » (Grand dictionnaire terminologique)

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
téléverser	'télévers'	2	upload	[word="upload(s)(ed ing)?"]
upload/er	'upload'	2	upload	-,,-
Les autres	x	96	upload	-,,-

Le tableau 3 L'analyse du terme « téléverser »



Le graphique 1 L'analyse du terme « téléverser »

Parmi les 96 occurrences de chaînes autres que les chaînes de recherche, le sous-ensemble de loin le plus nombreux (61) est celui qui contient le terme standard télécharger ou charger. Aucun autre terme significatif n'apparaît dans les instances de données restantes (31). Il convient toutefois de noter la présence de termes tels que « *mettre en ligne* » et « *vitesse montante* ».

3.1.2 Pare-feu

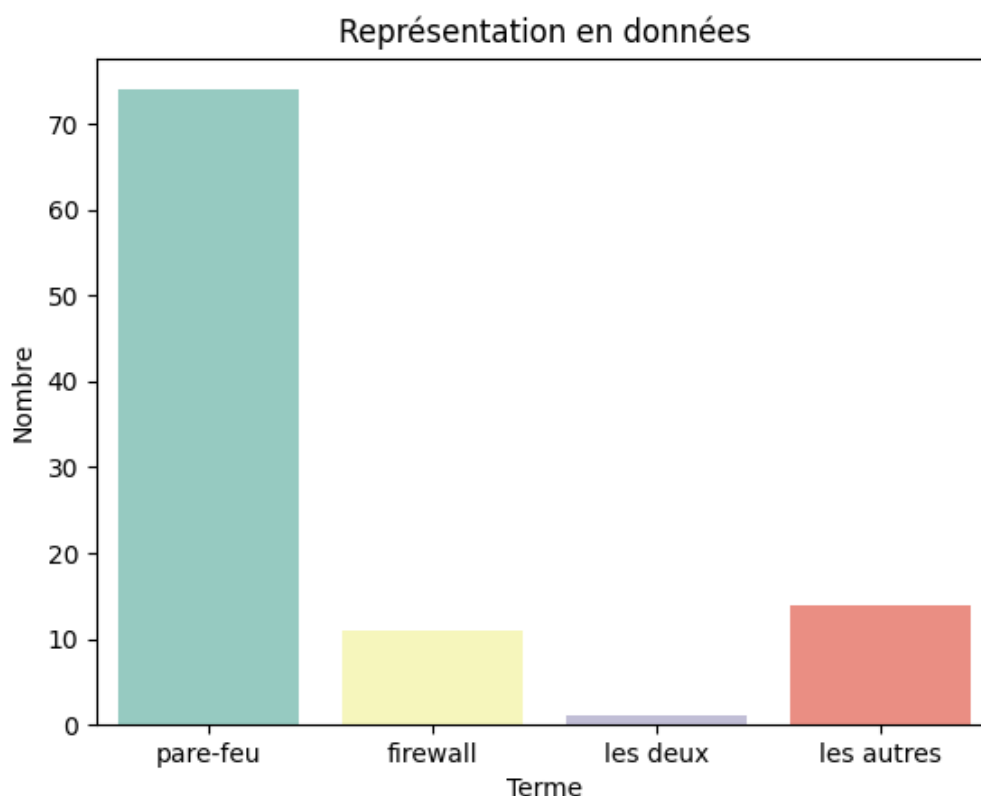
Pare-feu est un terme signifiant un outil logiciel qui protège le réseau privé contre les menace électroniques présentes sur le réseau internet. « *Équipement situé entre le réseau Internet et le réseau privé d'une entreprise pour accroître la sécurité de ce dernier en filtrant le trafic en provenance ou à destination d'Internet.* » (Larousse) Il s'agit d'une traduction qui reste relativement fidèle au terme original anglais « *firewall* ». Tel scénario semble être logique, car la signification informatique n'est pas la seule qui est partagée par les mots susmentionnés. Principalement, ces mots partagent le sens liés au domaine des pompiers et de lutte contre l'incendie. D'après Merriam-Webster « *un mur construit pour empêcher la propagation du feu.* » (Merriam-Webster) La signification pertinente pour l'informatique est simplement le résultat de l'extension sémantique du sens initial qui s'est produite dans les deux langues.

Le Grand dictionnaire terminologique définit la signification du domaine informatique de « *pare-feu* » comme suit: « *Système de sécurité conçu pour filtrer les flux de données entre un réseau et un autre réseau de confiance moindre, le plus souvent Internet, selon une politique d'accès préétablie.* » En outre, il mentionne en plus du terme analysé des autre équivalents à utiliser dans ce sens. Il s'agit des termes coupe-feu, barrière de sécurité et garde-barrière. Néanmoins, l'usage de deux dernières expressions est évalué

comme assez rare. Le dictionnaire aussi indique la pratique de l'utilisation fréquente de l'anglicisme « *firewall* » et il la fortement déconseillé.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Pare_feu	'pare-feu'	74	firewall software	[word="firewall(s)(ed ing)?"] [word="software"]
Firewall	'firewall'	14	firewall software	-, -
Les autres	x	11	firewall software	-, -

Le tableau 4 L'analyse du terme « pare-feu »



Le graphique 2 L'analyse du terme « pare-feu »

Parmi les 11 résultats qui ne renvoient pas aux chaînes de recherche, le terme concurrentiel « *coupe-feu* » est de loin le plus répandu. En outre, il existe des unités de cas avec la forme incorrecte « *pare feu* » sans trait d'union au milieu. Dans un cas, on a également trouvé le terme « *garde-barrière* » annoncée plus haut. Dans un cas, le terme rare « *murs à l'épreuve du feu* » a également été enregistré.

3.1.3 Logiciel et Matériel

En anglais les significations des termes analysés sont fondées sur l'opposition binaire, c'est-à-dire tous les éléments de l'IT sont divisés en deux dimension, l'une matérielle qu'on peut toucher ou mesurer et l'autre immatérielle et virtuelle. En ce qui concerne la deuxième dimension, » l'approche de la délimitation du terme français « *matériel* » qui est d'après Larousse décrit comme « *Ensemble des éléments physiques employés pour le traitement de l'information (dispositifs d'entrées-sorties, organes de liaison,*

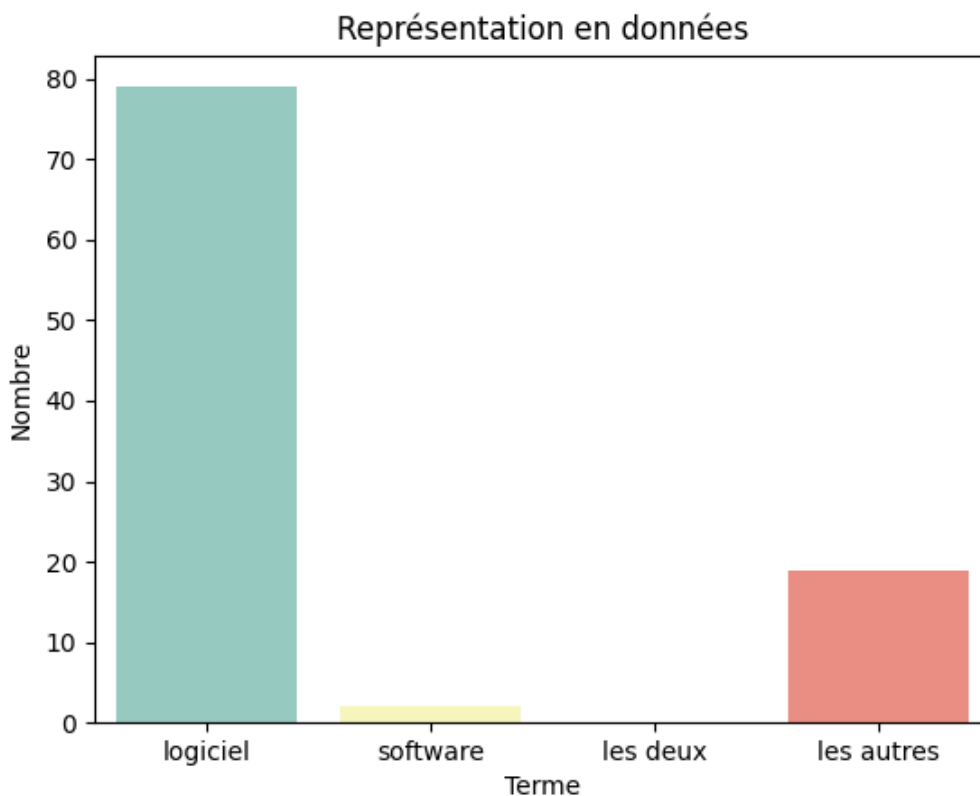
mémoires, circuits de traitement) » reste la même que dans le cas du terme original « *the physical components (such as electronic and electrical devices) of a vehicle (such as a spacecraft) or an apparatus (such as a computer)* » (Merriam-Webster).

Même si, l'équivalent du terme anglais « *software* », le mot « *logiciel* » décrit un ensemble identique (tout ce qui n'est pas perceptible comme matériel) il a été créé sur la base des mots « *logique* » et « *matériel* ». Autrement dit, le terme implique déjà dans son nom le fait que logiciel est un élément qui est basé sur les opérations logiques construites au sein d'un langage de programmation. Alors, tandis que le premier terme « *matériel* » a été créé par la traduction plus ou moins exacte dans le cas du « *logiciel* » il s'agissait de la néologie par invention (Costăchescu,148).

Le Grand dictionnaire terminologique définit le terme « *logiciel* » comme « *Ensemble des programmes constituant une unité destinée à effectuer un traitement particulier sur un ordinateur* » et le terme comme « *Ensemble des éléments physiques d'une installation informatique* ». Dans les deux cas le dictionnaire ne suggère pas explicitement l'utilisation fréquente des emprunts « *software* » et « *hardware* » en français.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Logiciel	'logiciel'	79	software	[word="software"]
Software	'software'	2	software	-,,-
Les autres	x	19	software	-,,-

Le tableau 5 L'analyse du terme « logiciel »

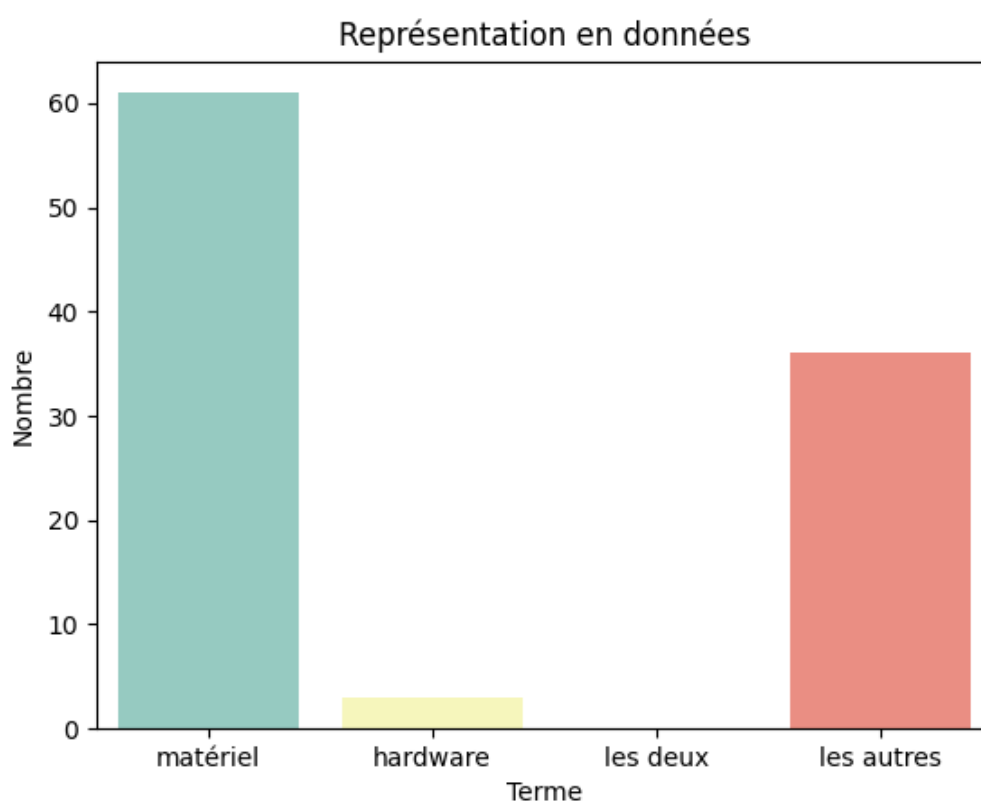


Le graphique 3 L'analyse du terme « logiciel »

Dans le cas de catégorie les autres il s'agit de traductions utilisant des termes légèrement différents qui fonctionnent dans leur contexte particulier. Par exemple, l'utilisation du mot « *programme* », « *système* », etc. Dans d'autres cas, il s'agit manifestement de données erronées.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Matériel	'matériel'	61	hardware	[word="hardware"] :FR [word!="quincaillerie"]
Hardware	'hardware'	3	hardware	-,,-
Les autres	x	36	hardware	-,,-

Le tableau 6 L'analyse du terme « matériel »



Le graphique 4 L'analyse du terme « matériel »

Presque tous les résultats au sein de la catégorie « Les autres » sont les exemples des dispositifs externes connectés à un ordinateur, tels que les imprimantes, les claviers, les souris, les écrans, etc. ou les équipements de réseau (routeurs, switches), les équipements audio-visuels, etc. Ces deux sous-catégories de matériel sont couvertes en français par les termes périphériques et équipement, qui constituent la majorité des résultats de la catégorie susmentionnée de l'ensemble de données analysé.

3.1.4 Bogue

Le mot « *bogue* » représente un emprunt lexical avec l'adaptation phonologique du terme anglais « *bug* ». Le terme anglais dont le sens originel « *insecte* » a été transféré à la signification du domaine informatique « *un défaut inattendu, une faute, un vice* »

ou une imperfection » sur la base d'une expérience réelle lorsque les relais des premiers ordinateurs ont été obstrués par de vrais insectes. (Merriam-Webster).

Larousse et Le Grand dictionnaire terminologique définissent tous les deux le terme comme suit. « *Défaut de conception d'un logiciel ou d'un matériel se manifestant par des anomalies de fonctionnement* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Bouge	'bouge'	54	bug	[word="bug"] :FR [word!="insecte(s x)?"] [word!="punaise(s x)?"]
Bug	'bug'	29	bug	-,,-
Les autres	x	17	bug	-,,-

Le tableau 7 L'analyse du terme « bogue »

Les résultats non classés sont dans la plupart des cas le résultat de l'utilisation de termes plus généraux pour traduire le sens du terme « *bouge* ». Par exemple, les mots « *erreur* », « *problème* », etc.

3.1.5 Dialogueur

Pas seulement le terme analysé mais la signification générale est relativement nouveau et elle a trouvé son chemin à la langue codifiée il y a moins que vingt ans avec les progrès des algorithmes autonomes et des modèles linguistiques.

Le terme « *dialogueur* » représente un équivalent de l'anglicisme « *chatbot* », dont l'utilisation a été recommandée par la Commission d'enrichissement de la langue française. Le dictionnaire Larousse n'a intégré que l'emprunt anglais, tandis que le néologisme français est présent dans le dictionnaire Le Petit Robert. Le grand dictionnaire terminologique a adopté la définition suivante fourni par la CELF « *Logiciel spécialisé dans le dialogue en langage naturel avec un humain, qui est capable notamment de répondre à des questions ou de déclencher l'exécution de tâches.* » Il aussi implique, en accord avec le Petit Robert et la recommandation de la CELF, l'usage d'un autre néologisme concurrentiel « *agent de dialogue* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Dialogueur	'dialogueur'	0	chatbot	[word="chatbot"]
Chatbot	'chatbot'	4	chatbot	-,,-
Les autres	x	0	chatbot	-,,-

Le tableau 8 L'analyse du terme « dialogueur »

Dans le corpus Para Crawl, seuls 4 enregistrements ont été trouvés pour la chaîne de recherche « *chatbot* ». Dans tous ces cas, les anglicismes ont été repris dans le corpus parallèle français.

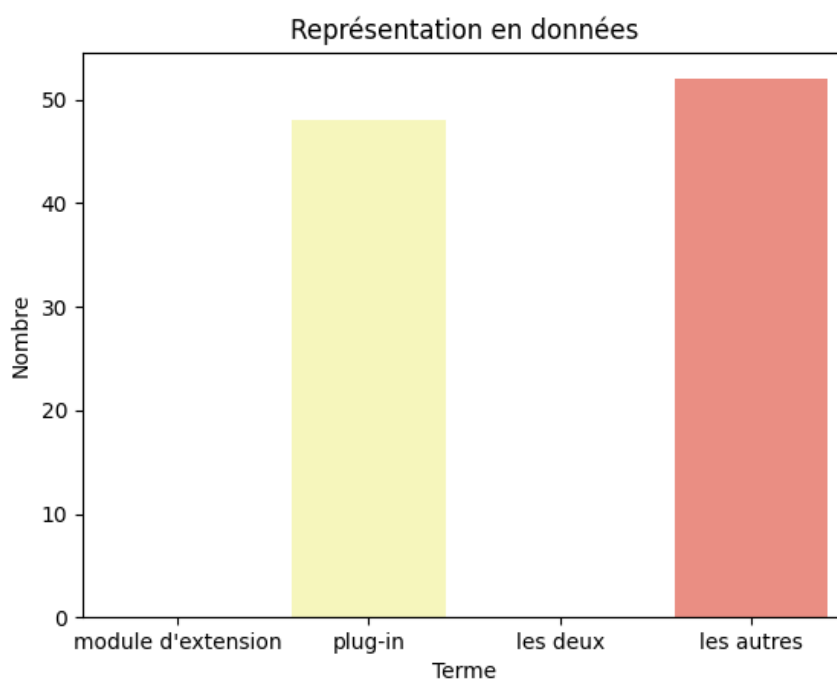
3.1.6 Module d'extension

Il s'agit du terme équivalent à l'anglicisme « *plug-in* » (« *un petit logiciel qui complète un programme plus important* » d'après Merriam-Webster) dont l'utilisation en français

est déconseillée, mais sa pratique effective au Québec et en Europe francophone est néanmoins reconnue par le Grand dictionnaire terminologique : « *L'emprunt à l'anglais plug-in n'est pas acceptable. Dans les ouvrages de référence québécois et européens où il est relevé, il est parfois accompagné de précisions normatives quant à son usage. De surcroît, il ne s'intègre pas naturellement dans la langue. En effet, on remarque notamment que dans les médias écrits, où il est largement répandu au Québec et en Europe francophone, il est écrit en italique, encadré de guillemets ou suivi d'un équivalent français, ce qui traduit les réserves relatives à son emploi.* ». L'usage du terme analysé a été recommandé par la CELF dès la fin des années 1990. Pourtant le néologisme n'est pas présent dans les dictionnaires généraux français tel que le Larousse et Le Petit Robert, alors que son homologue, l'anglicisme susmentionné, l'est. Le Larousse mentionne notamment le néologisme québécois « *plugiciel* » qui fait naturellement partie du Grand dictionnaire terminologique.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Module d'extension	'module d'extension'	0	plug-in	[word="plug-in"] [word!="hybrid(s x)?"]:FR [word!="(H h)?ybride(s x)?"]
Plug-in	'plug-in'	48	plug-in	-, -
Les autres	x	52	plug-in	-, -

Le tableau 9 L'analyse du terme « module d'extension »



Le graphique 5 L'analyse du terme « module d'extension »

Il y avait aucune occurrence du terme français recherché tandis que l'anglicisme « *plug-in* » a été présent presque dans une moitié des résultats. Les données non classées sont de loin les plus nombreuses. Parmi ces résultats l'autre terme mentionné « *plugiciel* » a été enregistré 12 fois et la forme du terme « *plug-in* » sans trait d'union au milieu dix fois.

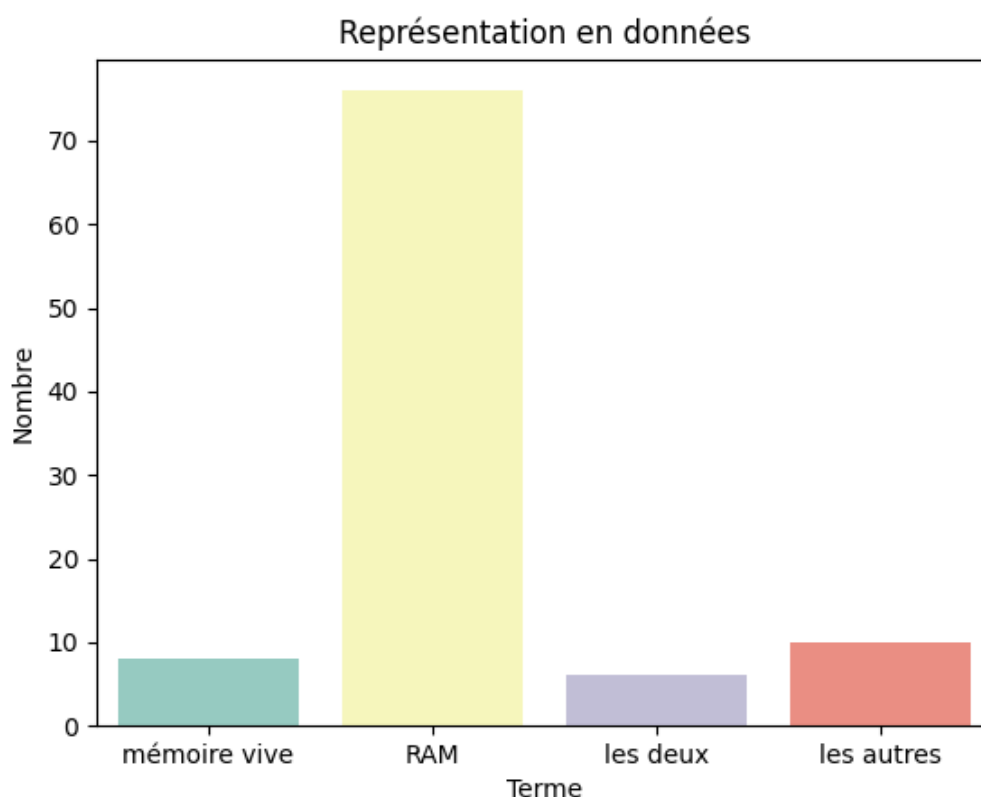
3.1.7 Mémoire vive

Il s'agit du terme officiellement recommandé à utiliser au lieu de l'abréviation anglaise « RAM » signifiant l'expression « *random-access memory* », c'est-à-dire littéralement mémoire à accès aléatoire ou d'après Merriam-Webster « *une mémoire d'ordinateur sur laquelle les données peuvent être lues et écrites et sur laquelle l'emplacement des données n'affecte pas la vitesse de leur récupération* ». Dans le dictionnaire Larousse le terme est décrit comme « *mémoire qui peut être écrite ou lue* » avec référence à l'abréviation anglaise : « *désignée souvent sous le sigle anglais RAM* » admettant son usage répandu en français moderne.

Le Grand dictionnaire terminologique définit le terme comme suit « *Mémoire primaire d'un ordinateur, rapidement accessible, dans laquelle les données peuvent être lues, écrites ou effacées.* » et ne déconseille pas l'utilisation de l'abréviation « RAM ». En outre, le dictionnaire mentionne également la possibilité d'utiliser un acronyme équivalent français « MEV ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Mémoire vive	'mémoire vive'	6	RAM	[word="RAM"]
RAM	'RAM'	84	RAM	-,,-
Les autres	x	18	RAM	-,,-

Le tableau 10 L'analyse du terme «mémoire vive»



Le graphique 6 L'analyse du terme «mémoire vive»

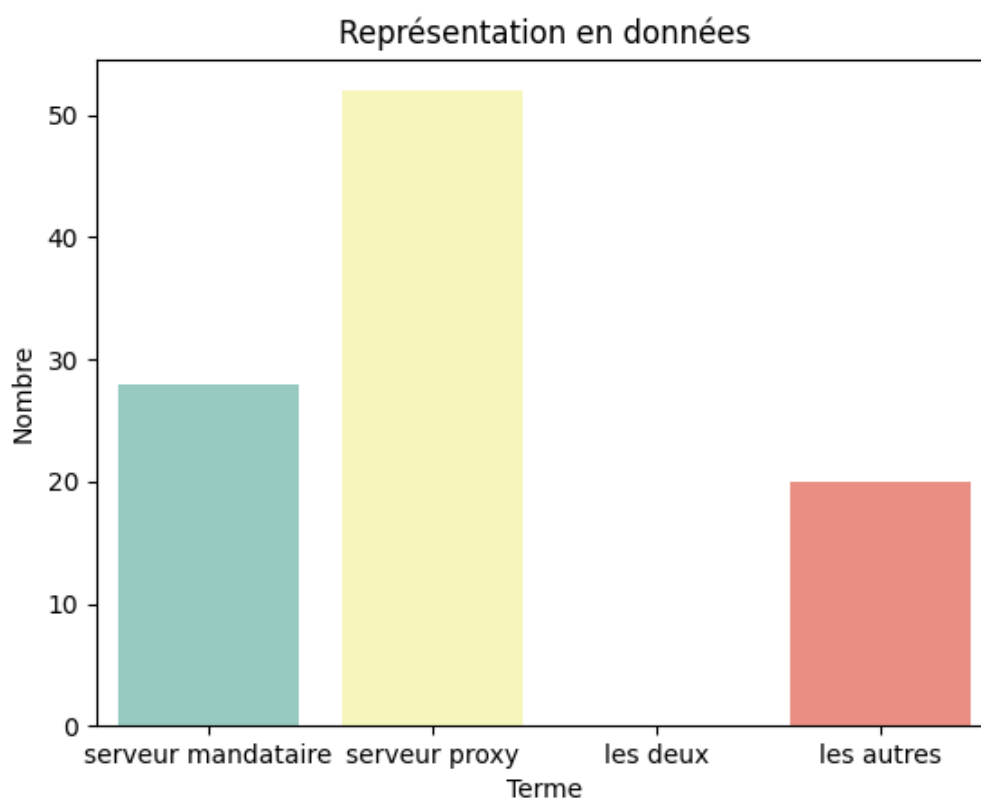
Parmi « *les autres* » sont deux exemples de plus de mémoire vive dont le codage semble corrompu et certains caractères sont manquants.

3.1.8 Serveur mandataire

Le terme « *serveur mandataire* » représente un équivalent de l'anglicisme « *proxy server* », dont l'utilisation est recommandée par la Commission d'enrichissement de la langue française. Le Grand dictionnaire terminologique définit le terme comme suit : « *serveur qui s'interpose entre un utilisateur et un serveur Web pour effectuer des opérations en leur nom* ». Contrairement au terme anglais, « *serveur mandataire* » n'a pas encore trouvé sa place dans les dictionnaires généraux français comme Larousse et Le Petit Robert, bien que son usage soit de plus en plus courant dans les publications spécialisées en informatique et en réseau.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Serveur mandataire	'serveur mandataire'	28	proxy server	[word="proxy"] [word="server"]
Serveur proxy	'serveur proxy'	52	proxy server	-,,-
Les autres	x	20	proxy server	-,,-

Le tableau 11 L'analyse du terme «serveur mandataire»



Le graphique 7 L'analyse du terme «serveur mandataire»

Parmi les autres, plus que cinq occurrences d'une autre variante du terme français « *serveur mandaté* » ont été enregistrés. En outre, dans presque dix cas la variante abrégée de l'anglicisme co-analysé « *proxy* » est apparu.

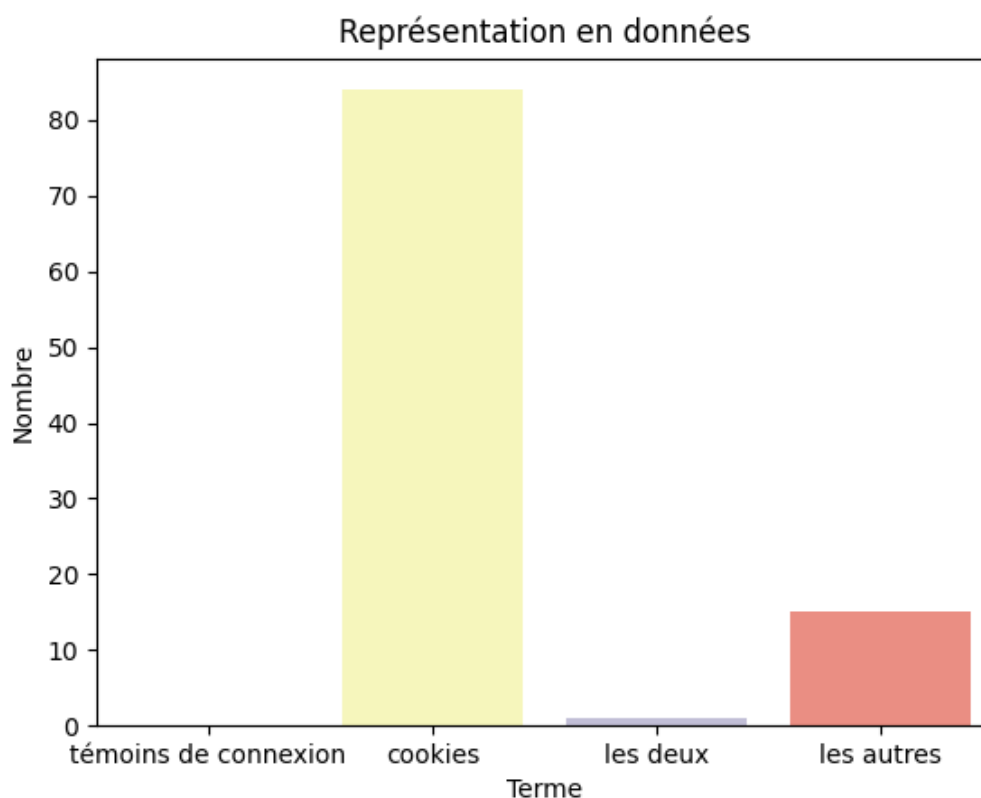
3.1.9 Témoin de connexion

Il s'agit du terme équivalent à l'anglicisme « *cookie* » (« *un petit fichier stocké sur l'ordinateur de l'utilisateur par un site Web pour conserver des informations sur les préférences de l'utilisateur* » d'après Merriam-Webster). L'utilisation de ce terme est encouragée par la Commission d'enrichissement de la langue française pour remplacer l'anglicisme, qui est largement utilisé dans le domaine de l'informatique. Le Grand dictionnaire terminologique définit le terme comme suit : « *Petit fichier texte qu'un serveur Web envoie à l'ordinateur d'un utilisateur et qui est renvoyé par le navigateur de l'utilisateur à chaque fois que celui-ci accède au même serveur Web.* »

Contrairement à l'anglicisme, « *témoin de connexion* » n'est pas encore intégré dans les dictionnaires généraux français comme Larousse et Le Petit Robert, mais son usage est reconnu et recommandé par les autorités linguistiques françaises (France Terme).

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Témoins de connexion	'témoins de connexion'	1	cookies	[word="cookies"] :FR [word!="biscuits"]
Cookies	'cookies'	85	cookies	-,,-
Les autres	x	15	cookies	-,,-

Le tableau 12 L'analyse du terme «témoins de connexion»



Le graphique 8 L'analyse du terme «témoins de connexion»

Presque une moitié des résultats appartenant au groupe « les autres » représentent la variante abrégée du terme français analysé « *témoin* ».

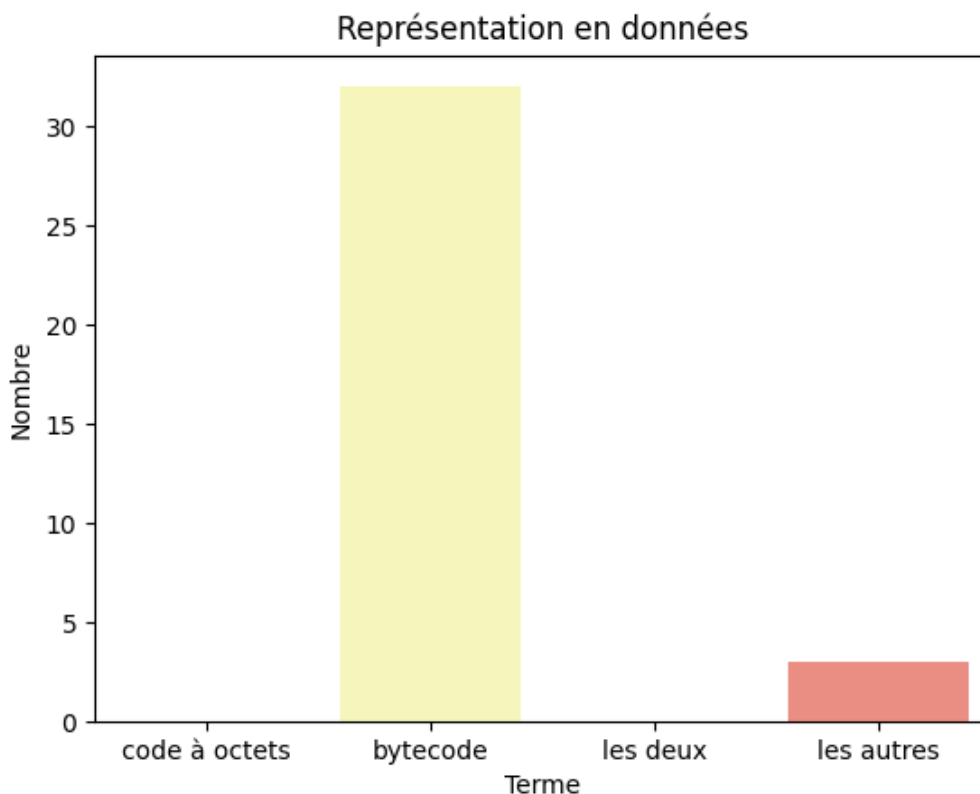
3.1.10 Code à octets

Le terme « *code à octets* » représente une traduction directe du terme anglais « *bytecode* », utilisé fréquemment dans le domaine de la programmation informatique. Merriam-Webster définit « *bytecode* » comme une forme intermédiaire de code qui est plus abstraite que le code machine, souvent utilisé pour améliorer la portabilité des programmes en permettant leur exécution sur différentes plateformes via une machine virtuelle.

Selon le dictionnaire Larousse, « *code à octets* » désigne « une représentation intermédiaire entre le code source et le code machine, exécutée par une machine virtuelle ». Cette définition met en lumière la nature de « *code à octets* » comme une étape de transformation du code source écrit par le programmeur en un format plus facile à interpréter par les systèmes informatiques, tout en restant indépendant des spécificités du matériel. Dans le monde francophone, le terme « *code à octets* » est utilisé parallèlement au terme « *bytecode* » en raison de l'influence omniprésente de l'anglais dans le domaine informatique. Le Grand dictionnaire terminologique recommande l'usage du terme « *code à octets* » pour éviter l'anglicisme et rester conforme aux recommandations de terminologie française. Il décrit le terme comme suit : « *Une suite d'instructions encodées sous forme binaire, souvent utilisée pour permettre l'exécution de programmes sur des machines virtuelles indépendamment du système matériel sous-jacent.* »

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Code à octets	'code à octets'	0	bytecode	[word="bytecode"]
Bytecode	'bytecode'	32	bytecode	-,,-
Les autres	x	3	bytecode	-,,-

Le tableau 13 L'analyse du terme «code à octets»



Le graphique 9 L'analyse du terme «code à octets»

Parmi les trois instances de la catégorie « les autres », il y a deux termes concurrentiel. Il s'agit d'une variante du terme français analysé avec un trait d'union au lieu de la conjonction « à » et de l'expression « pseudo-code ».

3.1.11 Pseudo-code Java

Le terme « pseudo-code Java » fait référence à une représentation intermédiaire du code Java, souvent sous forme de « *bytecode* » ou « *code à octets* ». En d'autres termes, il s'agit d'un exemple concret du terme analysé précédemment.

Selon le dictionnaire Larousse, le « *pseudo-code Java* », ou « *bytecode Java* », est « *une représentation intermédiaire du code Java, générée par le compilateur Java et exécutée par la machine virtuelle Java (JVM)* ». Cette définition met en lumière la nature du pseudo-code Java comme une étape essentielle dans le processus de compilation et d'exécution des programmes Java. Le Grand dictionnaire terminologique recommande l'usage du terme « pseudo-code Java » pour désigner cette représentation intermédiaire, tout en précisant que c'est une forme de code optimisée pour être exécutée par la JVM, permettant ainsi une grande portabilité et efficacité d'exécution. Le dictionnaire déconseille l'utilisation de l'anglicisme « *bytecode* ».

Aucune entrée correspondante n'a été trouvée pour ce terme dans le corpus ParaCrawl. La commande suivante a été utilisée pour rechercher le terme : [word="JAVA"] [word="bytecode"].

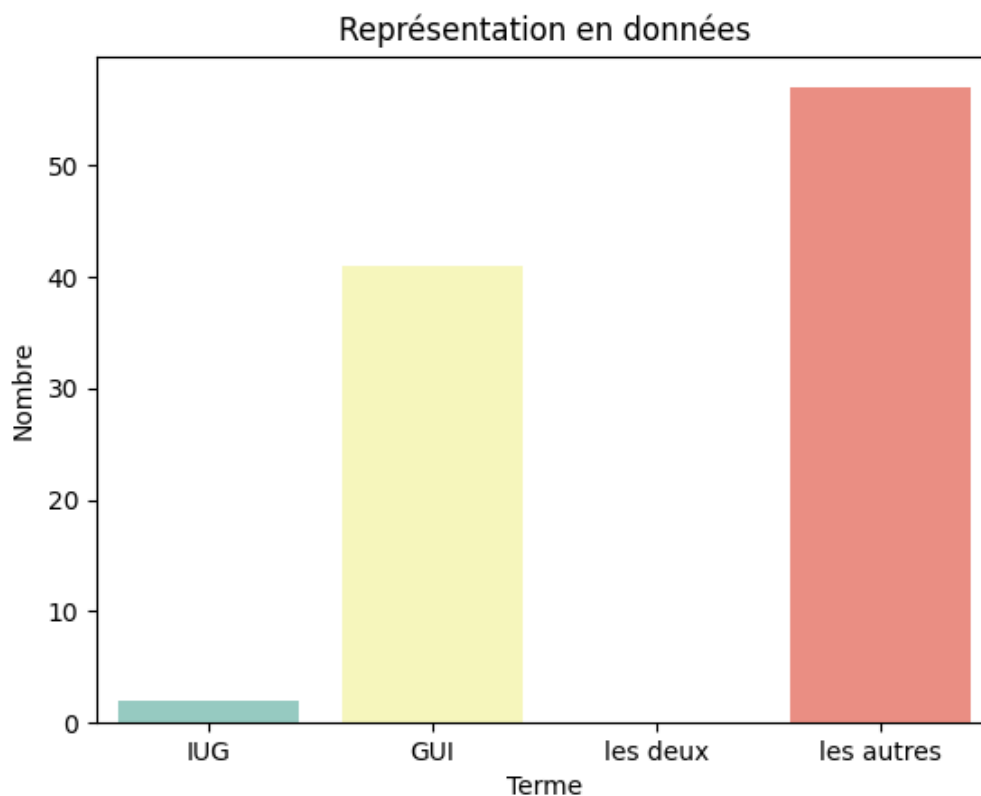
3.1.12 IUG

Le terme « *IUG* » (Interface Utilisateur Graphique), correspondant à l'anglais « *GUI* » (Graphical User Interface), désigne une interface permettant aux utilisateurs d'interagir avec des ordinateurs et des logiciels à travers des éléments graphiques tels que des icônes, des boutons et des fenêtres, plutôt que par des commandes textuelles. Selon Merriam-Webster, une GUI est une « *interface permettant aux utilisateurs d'interagir avec un appareil électronique à travers des éléments graphiques plutôt qu'une interface en ligne de commande textuelle* ». Le dictionnaire Larousse définit une IUG comme « *un ensemble de dispositifs graphiques permettant à l'utilisateur de dialoguer avec un système informatique de manière intuitive et visuelle* ».

Le Grand dictionnaire terminologique confirme cette définition en décrivant l'IUG comme « *un dispositif de dialogue homme-machine utilisant des éléments graphiques (fenêtres, icônes, menus, etc.) pour rendre l'interaction avec le système informatique plus conviviale et intuitive* ». Le GDT recommande l'usage du terme « Interface Utilisateur Graphique » plutôt que l'anglicisme « GUI »

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
IUG	'IUG'	2	GUI	[word=" GUI "]
GUI	'GUI'	41	GUI	-,,-
Les autres	x	57	GUI	-,,-

Le tableau 14 L'analyse du terme «IUG»



Le graphique 10 L'analyse du terme «IUG»

Dans deux cas la variante longue du mot recherché « *Interface Utilisateur Graphique* » est apparu. L'utilisation partielle du terme « *GUI* » a été néanmoins présent plus souvent, car les expressions « *Interface Utilisateur* » sont apparues dans l'ensemble des données vingt-huit fois au total.

3.1.13 EDI

L'abréviation « *EDI* » (Environnement de Développement Intégré) est l'équivalent français de l'anglicisme « *IDE* » (Integrated Development Environment). Il désigne un logiciel regroupant plusieurs outils pour faciliter le développement de logiciels, tels que l'édition de code, le débogage, la compilation et parfois le déploiement. Selon Merriam-Webster, un IDE est « *un ensemble d'outils logiciels intégrés permettant aux programmeurs de développer des logiciels plus efficacement* ». Le dictionnaire Larousse définit « *EDI* » comme « *un logiciel regroupant un ensemble d'outils facilitant le développement, la compilation et le débogage de programmes informatiques* ».

Le Grand dictionnaire terminologique (GDT) décrit un EDI comme « *un environnement logiciel intégré regroupant plusieurs outils nécessaires au développement, à la compilation, au débogage et à la gestion de projets informatiques* ». Le GDT recommande l'usage du terme « *EDI* » pour promouvoir l'utilisation de la terminologie française et éviter l'anglicisme.

Aucune donnée n'a pu être extraite du corpus ParaCrawl dans le contexte ciblé. La commande suivante a été utilisée pour rechercher le terme : [word="IDE"]

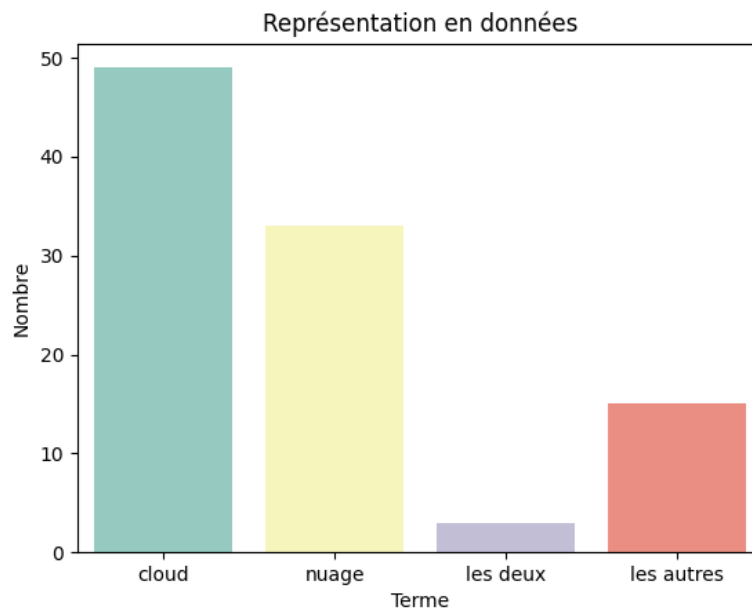
3.1.14 Informatique en nuage

Le terme « *informatique en nuage* » représente une adaptation française de l'anglicisme « *cloud computing* », une technologie largement utilisée dans le domaine de l'informatique. Selon Larousse et GDT définit presque à l'identique comme « *informatique en nuage* », ducoup comme « *un modèle informatique qui permet un accès à la demande et via Internet à des ressources informatiques partagées configurables, comme des réseaux, des serveurs, des stockages, des applications et des services, qui peuvent être rapidement provisionnés et libérés avec un effort minimal de gestion ou d'interaction avec le fournisseur de services.* »

En outre, le Grand Dictionnaire Terminologique mentionne d'autres variantes utilisées principalement au Canada francophone, tels que « *nuagique* », ou encore « *informatique infonuagique* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Informatique en nuage; Nuage	'nuage'	36	cloud computing	[word="cloud"] [word="comput(es)(ed ing)?"]
Cloud computing; Cloud	'cloud'	50	cloud computing	-, -
Les autres	x	15	cloud computing	-, -

Le tableau 15 L'analyse du terme «informatique en nuage»



Le graphique 11 L'analyse du terme «informatique en nuage»

Parmi les données non pertinentes, le terme canadien infonuagique était de loin le plus populaire dans la catégorie les autres. Dans un cas, le terme de rareté l' informatique dématérialisée était présent.

3.1.15 Application dorsale, application frontale

Le terme «*dorsale*» ou «*application dorsale* » est l'équivalent français de l'anglicisme «*backend* », couramment utilisé en informatique pour désigner les parties d'un système logiciel qui gèrent la logique de l'application, la base de données et l'intégration des services, sans être directement accessibles aux utilisateurs finaux.(Merriam-webster) Le dernier terme «*frontal*» ou «*application frontale* », en revanche, couvre le sens de la partie pertinente de l'application, qui est vue par l'utilisateur moyen et avec laquelle il interagit qui est décrit par Merriam-Webster comme «*une interface logicielle (telle qu'une interface utilisateur graphique) conçue pour permettre une interaction conviviale avec un ordinateur* ». Bien que certains dictionnaires français comme Larousse et Le petit Robert n'aient pas encore intégré ce terme, «*application dorsale* » et «*application frontale* » figurent dans le Grand Dictionnaire Terminologique.

Le Grand Dictionnaire Terminologique définit les termes comme suit 1) «*application dorsale* » : «*programme qui, dans une architecture client-serveur, traite les commandes en provenance de l'application frontale.*» 2) «*application frontale* » : «*Programme qui, dans une architecture client-serveur, reçoit les commandes de l'utilisateur et les envoie à l'ordinateur serveur qui les traite.*»

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Application dorsale	'dorsal'	0	backend application	[word="backend"] [word="application"]
Backend application	'backend'	0	backend application	-,,-
Les autres	x	6	backend application	-,,-

Le tableau 16 L'analyse du terme «application dorsale»

En ce qui concerne le terme « *application dorsale* » seulement six résultats ont été trouvés dans le corpus analysé pour ce terme dans le contexte donné. Les termes utilisés dans les phrases de partie française de ParaCrawl sont « *application sous-jacente* », « *application principale* » et « *application d'arrière-plan* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Application frontale	'frontal'	2	Frontend	[[word="frontend"] :FR [word="programme"]
Frontend application	'frontend'	0	Frontend	-,,-
Les autres	x	5	Frontend	-,,-

Le tableau 17 L'analyse du terme «application frontale»

À l'instar du cas précédent, seulement six résultats apparaissent dans le corpus analysé pour le terme « *application frontale* » dans le contexte donné.

En outre, les résultats « *interface* » et « *interface graphique* » étaient présentes dans l'ensemble de données obtenu.

3.1.16 Boucle de programmation

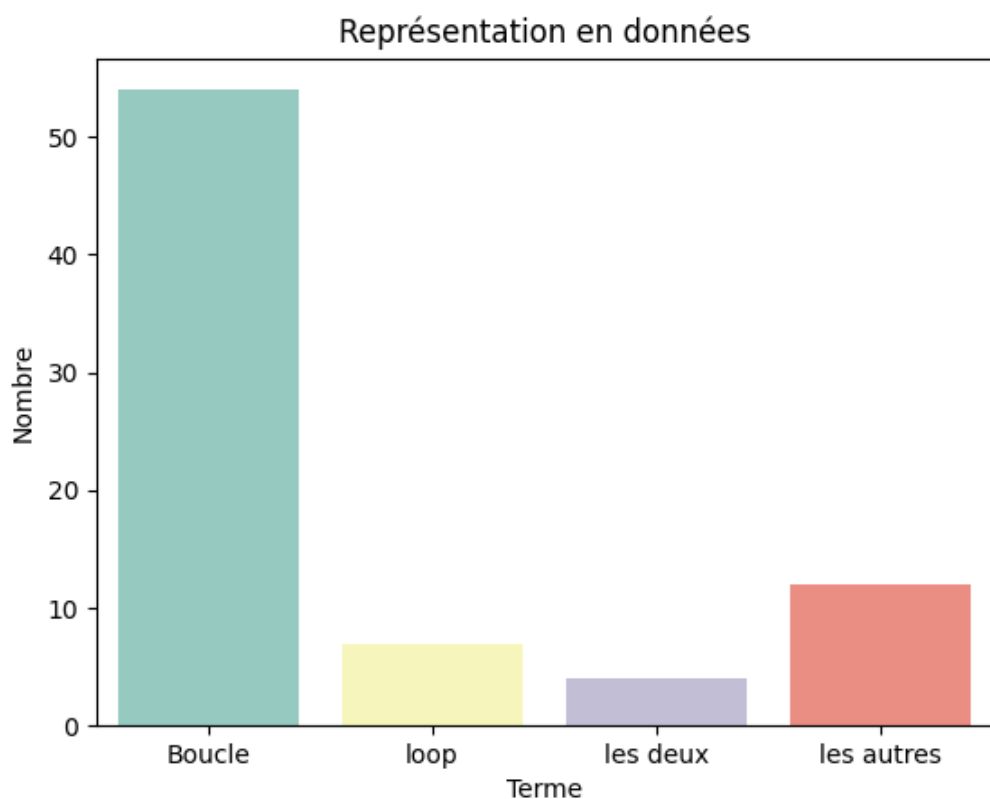
Il s'agit du terme équivalent à l'anglicisme « *loop* » (« *une structure de programmation qui répète une séquence d'instructions jusqu'à ce qu'une condition spécifique soit remplie* » d'après Merriam-Webster). Le terme représente par conséquent une des opérations logiques indispensables pour le codage des algorithmes modernes et efficaces.

Larousse le décrit comme « *une suite de relations de cause à effet telle que le dernier effet réagisse sur la première cause* » Le Grand dictionnaire terminologique définit le terme comme suit : « *Élément de programme qui peut être exécuté de façon répétée, tant qu'une certaine condition est vérifiée, ou jusqu'à ce qu'on rencontre une condition qui y mette fin* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Boucle de programmation	'boucle'	58	backend application	[word="backend"] [word="application"]
Loop	'loop'	11	backend application	-,-
Les autres	x	7	backend application	-,-

Le tableau 18 L'analyse du terme «boucle de programmation»

77 résultats ont été trouvés dans le corpus ParaCrawl pour le terme « boucle (de programmation) ».



Le graphique 12 L'analyse du terme «boucle de programmation»

3.1.17 Développeur généraliste

Le terme « *développeur généraliste* » est l'équivalent français de l'anglicisme « *full-stack developer* », couramment utilisé dans le domaine de l'informatique pour désigner un développeur capable de travailler à la fois sur les parties frontales (front-end) et dorsales (back-end) d'une application. Le terme semble ne pas être intégré dans les dictionnaires généraux français comme Larousse et Robert. Ni son homologue « *full-stack developer* » comme un rôle plus spécifique de métier développeur/programmeur est présent dans le Merriam-Webster. Même le dictionnaire terminologique France Terme fourni par la Ministère de la Culture de la République Française ne contient le terme.

Seule Le Grand dictionnaire terminologique inclut le terme français sus-mentionné et le décrit comme suit. « *Développeur de logiciels qui possède les connaissances nécessaires pour concevoir et réaliser toutes les composantes d'un système informatique.* »

Dans le corpus ParaCrawl, aucune donnée correspondant au terme analysé n'a été trouvée. Le corpus a été examiné avec la commande suivante : [word="full-stack"] [word="developer "]

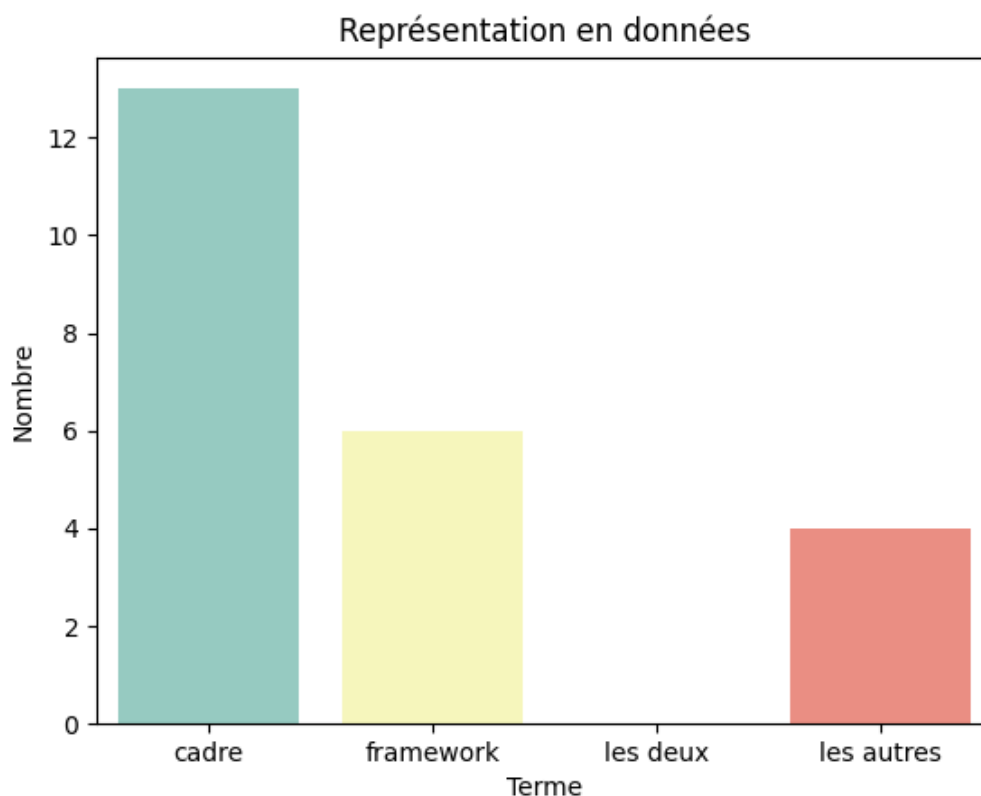
3.1.18 Cadre d'application

Le terme « *cadre d'applications* » est une traduction de l'anglicisme « *application framework* » : « *une bibliothèque logicielle qui fournit une structure fondamentale pour soutenir le développement d'applications dans un environnement spécifique* » d'après Technopedia. Son utilisation a été recommandée par l'Office québécois de la langue française. Il s'agit du coup du terme canadien qui n'a pas toujours trouvé sa place dans les dictionnaires français pertinents, qu'ils soient généraux ou techniques. Selon le Grand dictionnaire terminologique, un « *Infrastructure logicielle permettant la conception de programmes.*»

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Cadre d'application	'cadre'	13	application framewok	[word="application"] [word ="framework"]
Application framework	'framework'	6	application framewok	-,,-
Les autres	x	4	application framewok	-,,-

Le tableau 19 L'analyse du terme «cadre d'application»

Seulement vingt-trois résultats apparaissent dans le corpus ParaCrawl pour le terme.



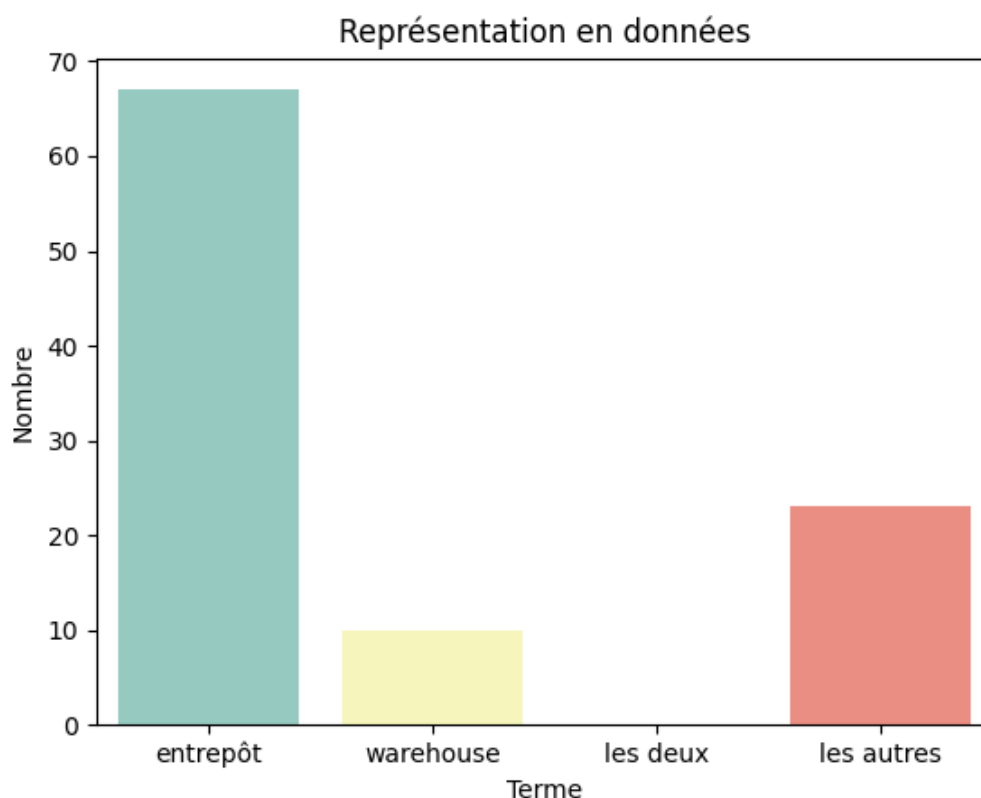
Le graphique 13 L'analyse du terme «cadre d'application»

3.1.19 Entrepôt de données

Le terme « *Entrepôt de données* » représente l'équivalent du terme original anglais « *Data Warehouse* ». Entrepôt de données est un système de stockage de grands volumes de données, qui est souvent mis en œuvre sur la base de la technologie des bases de données relationnelles. Bien que le terme anglais soit très largement utilisé, non seulement au niveau des détails techniques, mais aussi dans un certain nombre de domaines dont il constitue un outil efficace, il ne figure pas dans le dictionnaire général de la langue anglaise, tel que le Merriam-Webster. Il est néanmoins présent au sein de son homologue britannique « Cambridge Dictionary » sous la définition suivante « *une grande quantité d'informations stockées sur un ordinateur ou sur un certain nombre d'ordinateurs au même endroit* ».

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Entrepôt de données	'entrepôt'	67	data warehouse	[word="data"] [word="warehouse"]
Data Warehouse	'warehouse'	10	data warehouse	-, -
Les autres	x	23	data warehouse	-, -

Le tableau 20 L'analyse du terme «entrepôt de données»



Le graphique 14 L'analyse du terme «entrepôt de données»

Parmi 23 résultats appartenant à la catégorie de recherche « les autres », la plupart est couvert par l'utilisation du terme « *base de données* ». En outre, Les terme « *depôt de données* », « *entreposage de données* » et « *stockage de données* » sont apparus.

3.1.20 Outil d'alimentation

Le terme « *Outil d'alimentation* » correspond au sens de l'abréviation anglaise « *ETL* » contenant les mots « *extract* », « *transform* » et « *load* ». L'acronym ne fait toujours pas partie des dictionnaires générales anglais et il est seulement présent et décrit dans les publications et sur les sites web techniques tels que Techopedia: « *L'extraction-transformation-chargement (ETL) est le processus d'extraction, de transformation et de chargement lors de l'utilisation d'une base de données, mais plus particulièrement lors de l'utilisation du stockage des données.* ». La même situation est du côté français. L'équivalent susmentionné français est offert par le Grand dictionnaire terminologique sous la définition. « *Outil informatique qui est destiné à extraire des données de diverses sources (bases de données de production, fichiers, Internet, etc.), de les transformer et de les charger dans un entrepôt de données.* » Cette définition du terme fourni par l'Office québécois de la langue française est en même temps accompagné par la note indiquant qu'il est possible de trouver de multiples variantes dans le discours contemporain comme « *outil d'extraction* », « *outil d'extraction et de transformation de données* » ou parfois même « *outil ETL* » et que cette ambiguïté terminologique est principalement causé par le fait que « *Tous les auteurs ne découpent pas cette phase du processus de construction et d'entretien d'un entrepôt de données de la même manière, car les outils qui existent, sur le marché, ne possèdent pas tous les mêmes fonctionnalités* »

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Outil d'alimentation	'outil d'alimentation'	0	ETL	[word = "ETL"] :FR [word = "données"]
ETL	'ETL'	15	ETL	-, -
Les autres	x	3	ETL	-, -

Le tableau 21 L'analyse du terme «outil d'alimentation»

Seulement 18 résultats pour le terme analysé ont été trouvé au sein du corpus ParaCrawl. Parmi les résultats non classés l'acronym français « *ETC (extraction, transformation et chargement)* » a été enregistré.

3.1.21 Logiciel local

Le terme « *logiciel local* » correspond au logiciel installé et exécuté directement sur un ordinateur ou un serveur local, par opposition aux solutions de logiciel en nuage. Ce terme n'est pas encore largement reconnu dans les dictionnaires généraux, que ce soit en anglais ou en français, mais il est couramment utilisé dans les publications et sur les sites web techniques tels que Techopedia. En anglais, le terme correspondant est souvent « *on premises software* » ou « *on prem software* », défini par Techopedia comme: « *Les logiciels sur site sont un type de modèle de fourniture de logiciels qui sont installés et exploités à partir du serveur et de l'infrastructure informatique internes du client.* »

Alors que, le Grand dictionnaire terminologique définit le terme comme : « *Logiciel installé directement sur un ordinateur ou accessible par le réseau local d'une entreprise.* » ,il en même temps fortement déconseille l'utilisation du terme « *logiciel sur site* » et le considère inintégré au système linguistique du français.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Logiciel local	'local'	1	on premises	[word = "on"][word = "premises"] :FR [word = "cloud"] [word = "nuage"]
On premises software	'on prem'	0	on premises	-,-
Les autres	x	10	on premises	-,-

Le tableau 22 L'analyse du terme «logiciel local»

Dans ce cas seulement onye résultats paertinents ont été generés dans le contexte donné. La grande majorité des résultats d'interrogation sont apparu avec le terme français susmentionné et déconseillé «*sur site*».

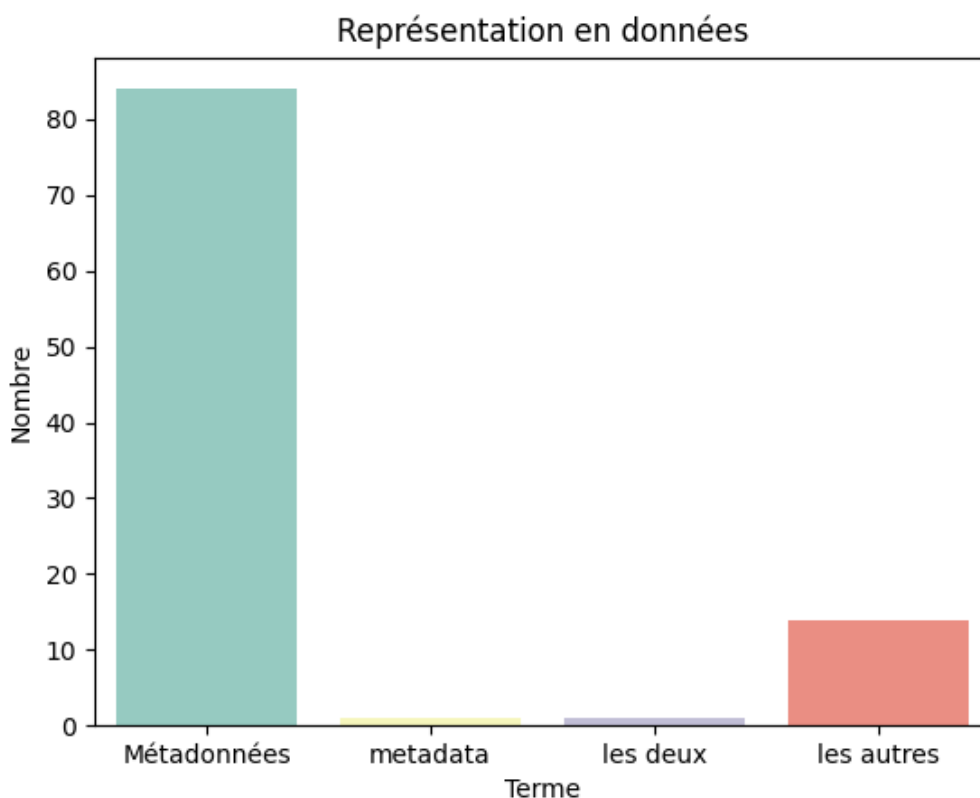
3.1.22 Métadonnées

Il s'agit du terme équivalent à l'anglicisme « *metadata* » (« *des données qui fournissent des informations sur d'autres données* » d'après Merriam-Webster). Le terme métadonnées représente une des opérations essentielles pour la gestion et l'organisation des informations dans les systèmes informatiques, bases de données et les entrepôts de données modernes et efficaces. Techopedia

Tandis que le terme n'apparaît pas dans le dictionnaire Larousse, le Petit Robert donne cette définition générale : « *Caractéristique formelle normalisée et structurée utilisée pour la description et le traitement des contenus des ressources numériques.* » Le Grand dictionnaire terminologique définit le terme comme suit : « *ensemble structuré de données accompagnant un ouvrage et servant notamment à en décrire le contenu et le format, à assurer son indexation dans les moteurs de recherche et les bases de données, et à faciliter la gestion des droits d'auteur qui y sont liés.* »

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Métadonnées	'métadonnées'	85	metadata	[word = " metadata "]
Metadata	'metadata'	1	metadata	-,-
Les autres	x	14	metadata	-,-

Le tableau 23 L'analyse du terme «métadonnées»



Le graphique 15 L'analyse du terme «métadonnées»

Parmi les autres termes trouvés , il y a neuf occurrences d'autres formes du mot analysé comme par exemple « *métadonnée* », « *metadonnées* » ou « *méta-données* ».

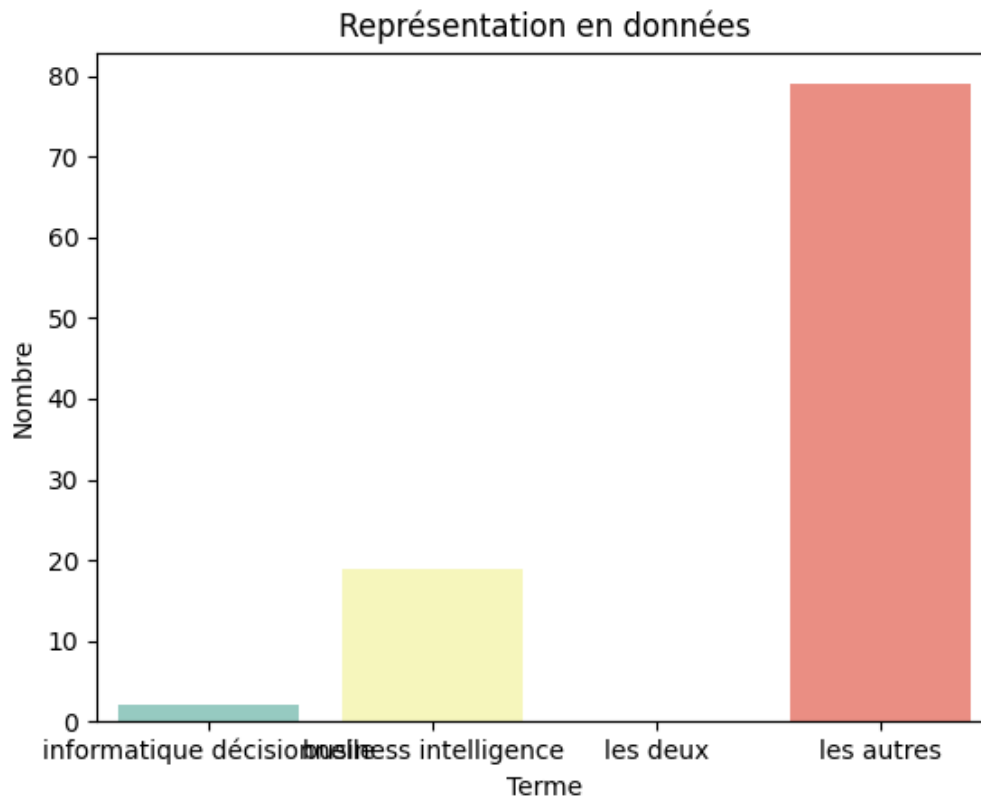
3.1.23 Informatique décisionnelle

Le terme « *informatique décisionnelle* » est l'équivalent du terme anglais original « *business intelligence* » qui est très répandu dans le discours pas forcément technique mais notamment dans le langage d'affaires de nombreux domaines d'aujourd'hui. Bien que la fréquence d'utilisation du terme dans l'environnement international augmente parallèlement à l'importance croissante de l'analyse des données, le terme anglais ne fait toujours pas partie des dictionnaires généraux prestigieux tels que le Merriam-Webster ou le dictionnaire « Cambridge Dictionary ». Techopedia définit le terme « *business intelligence* » comme suit : « *La Business Intelligence (BI) est l'utilisation des technologies informatiques pour l'identification, la découverte et l'analyse des données de l'entreprise - comme le chiffre d'affaires, les produits, les coûts et les revenus.* »

Le Grand dictionnaire terminologique décrit le terme français comme « *Branche de l'informatique qui fait appel à l'extraction, à la compilation et à l'analyse des données afin d'assister le personnel de direction dans la prise de décision stratégique.* » Outre le terme analysé, le dictionnaire admet également l'utilisation de l'expression « *informatique de décision* » et déconseille l'utilisation de la variante « *intelligence d'affaires* » calqué sur son homologue anglais susmentionnés, car le mot « *intelligence* » ne partage pas le sens de « *recherche et traitement d'informations* » avec son équivalent anglais.

Terme	Chaînes interrogées	Nombre	Terme anglais original interrogé	CQP
Informatique décisionnelle	'informatique décisionnelle'	2	business intelligence	[word = "business"] [word = "intelligence"]
Business intelligence	'business intelligence'	19	business intelligence	-,-
Les autres	x	79	business intelligence	-,-

Le tableau 24 L'analyse du terme «informatique décisionnelle»



Le graphique 16 L'analyse du terme «informatique décisionnelle

Dans la catégorie des autres résultats trouvés, les termes pertinents ayant une occurrence significative sont les suivants: « *BI* » (16), « *intelligence d'affaires* » (6), « *renseignements commerciaux* » (6) et « *renseignements commerciaux* » (4) . La partie restante de cet ensemble consistait principalement en des descriptions générales des attributs du concept analysé.

4 Résultats et synthèse de l'analyse

Le tableau suivant a été conçu pour les besoins de l'analyse finale. Dans ce tableau, les termes sont classés selon les groupes de sélection déjà présentés. En outre, au sein de ces groupes, ils sont triés en fonction de leur pertinence (une sous-analyse de terme est considérée comme pertinente si au moins cinquante résultats pour le terme de recherche ont été trouvés dans le corpus ParaCrawl). Les termes qui ne répondent pas à ce critère sont indiqués en rouge. Dans ce qui suit, les sous-analyses des termes sont classées par ordre décroissant de l'occurrence des termes préférés en français. Les lignes présentant une prévalence de ces termes préférés sont indiquées en vert. La prévalence des anglicismes, quant à elle, est indiquée en bleu. Les cas de recherche dans lesquels un groupe statistiquement significatif d'un autre terme principalement non analysé a été identifié dans les données sont indiqués en jaune. Un terme de quinze occurrences a été considéré comme statistiquement significatif dans le contexte de ce travail. Dans l'attribut « *Signification.d'occ.des autres termes* », si un tel terme apparaît, il est également indiqué s'il s'agit d'un anglicisme (Oui|EN) ou non (Oui|FRE).

Le groupe de sélection	Terme	No. d'occurrences	No. d'occ. d terme	No. d'occ. de l'anglicisme	Signification. d'occ. des autres termes	Pertinence
1	Logiciel	100	79	2	Non	Oui
1	Pare-feu	100	74	14	Non	Oui
1	Matériel	100	61	3	Non	Oui
1	Bogue	100	54	29	Non	Oui
1	Serveur mandataire	100	28	52	Non	Oui
1	Mémoire vive	100	6	84	Non	Oui
1	Téléverser	100	2	2	Oui FR	Oui
1	Module d'extension	100	0	48	Oui EN	Oui
1	Témoin de connexion	100	0	85	Non	Oui
1	Dialogueur	4	0	4	Non	Non
2	Boucle de programmation	77	58	11	Non	Oui
2	Informatique en nuage	100	36	50	Non	Oui
2	IUG	100	2	41	Oui EN	Oui
2	Cadre d'application	23	13	6	Non	Non
2	Code à octets	35	0	32	Non	Non
2	Pseudo-code Java	0	0	0	Non	Non
2	EDI	0	0	0	Non	Non
2	Application dorsale	6	0	0	Non	Non
2	Application frontale	6	0	0	Non	Non
2	Développeur généraliste	0	0	0	Non	Non
3	Métadonnées	100	85	1	Non	Oui
3	Entrepôt de données	100	67	10	Non	Oui
3	Informatique décisionnelle	100	2	19	Oui EN	Oui
3	Logiciel local	11	1	0	Non	Non
3	Outil d'alimentation	18	0	15	Non	Non

Le tableau 25 Les résultats de l'analyse

Dans le premier groupe, un seul des termes a été identifié comme non pertinent sur la base des critères susmentionnés. Dans le groupe 2, il y avait déjà 7 termes et dans le groupe 3, il y avait 2 termes. Dans le premier groupe, la prévalence de termes français a été observée dans 5 sorties, tandis que la prévalence d'anglicismes a été observée dans 4 cas. Dans le deuxième groupe, seuls un cas de prévalence de termes français et deux cas de prévalence d'anglicismes ont été identifiés dans les résultats pertinents. Dans le groupe 3, il y avait 2 cas avec une prévalence de termes français et un cas avec une prévalence d'anglicismes.

Les résultats montrent donc une utilisation réelle d'expressions françaises préférées et d'expressions anglaises originales. Les termes français tendent à prédominer dans des expressions et des concepts de nature plus générale et qui sont, en principe, plus établis dans la communication ordinaire non professionnelle. Les résultats de l'analyse montrent en partie que plus un terme est technique, plus il a tendance à s'unifier dans le langage courant avec des anglicismes utilisés dans le monde entier. En même temps, l'échantillon montre que les termes d'un seul mot français sont les plus courants dans l'usage réel.

Les résultats de l'analyse semblent illustrer le fait que le purisme linguistique et l'équivalence dans la nomenclature jouent sans aucun doute un rôle très important dans la codification de nouveaux termes. Cependant, il est évident que l'évolution rapide de la technologie et de son langage dépasse de loin l'intégration de nouveaux termes dans les vocabulaires généraux ou même techniques, sans parler de l'établissement d'un terme équivalent français dans la langue quotidienne. Un certain nombre de termes initialement très techniques commenceront progressivement à pénétrer le vocabulaire du profane. Au cours de cette pénétration, une préférence pour les termes français peut se produire et se produit probablement. A ce stade, le terme est déjà diffusé massivement par les médias qui s'efforcent de suivre un style imposé. L'objet de la diffusion de masse est donc déjà le plus souvent un néologisme français. Si le terme n'existe que dans un contexte purement technique, son usage tend à se limiter aux communautés professionnelles, souvent internationales, pour lesquelles le plus simple est d'utiliser l'anglicisme original.

Conclusion

Ce mémoire de licence traitait l'impact de l'anglais sur le vocabulaire français du domaine des technologies de l'information. La tâche principale de cette mémoire était d'analyser l'usage réel d'une sélection de termes informatiques français par rapport à leurs équivalents anglais. Cette analyse, illustrant l'influence susmentionnée de l'anglais sur le vocabulaire français du domaine en question, a été abordée dans la partie pratique grâce à l'utilisation du corpus parallèle ParaCrawl. Cette partie pratique a été soutenue d'une part par une excursion dans l'histoire de l'interaction entre le français et l'anglais et d'autre part par une recherche sur les questions de la formation des néologismes et de la terminologie technique.

Les objectifs définis dans l'annotation originale du mémoire étaient les suivants : 1) Fournir des définitions linguistiques générales des termes pertinents 2) Introduction à la genèse du domaine des technologies de l'information et à sa terminologie 3) Décrire l'influence générale de la langue anglaise sur le français contemporain 4) Analyse de cinquante termes informatiques françaises sélectionnées à l'aide de corpus métalinguistiques et de la littérature pertinente.

Il est possible de déclarer que les trois premiers objectifs ont été pleinement atteints. En ce qui concerne le dernier objectif, qui a fait l'objet de la partie pratique, on peut dire qu'il a été partiellement atteint puisque la sélection des termes a été limitée à vingt-cinq pour les raisons pertinentes décrites dans le chapitre méthodologique.

Le principal résultat de l'analyse est le constat d'une certaine ambivalence dans l'utilisation des termes français préférés. Les résultats indiquent que la tendance à utiliser des néologismes français pourrait aller de pair avec l'établissement d'un terme particulier dans un contexte linguistique plus large et sa diffusion massive. En revanche, les termes plus techniques qui sont utilisés principalement au sein des communautés professionnelles montrent plutôt une tendance à l'utilisation d'anglicismes originaux.

Résumé (en tchéque)

Názvem práce je „Angličtina jako determinant současné francouzské terminologie v IT“. Realizace práce tedy byla naplánována s ohledem na předpoklad poměrně zajímavého střetu při genezi terminologie oboru informačních technologií. Tento předpoklad vychází na jedné straně z faktu, že obor IT je nejen velmi progresivním (potřeba nových termínů, pojmů a konceptů), ale i velmi internacionalizovaným (potřeba si rozumět) a na straně druhé z reality v podobě trvalé francouzské tendence k purifikaci jazyka a etablování ekvivalentní svébytné terminologie. Tyto zmíněné faktory zakládají určitý předpoklad poměrně zajímavých výstupů případné analýzy.

Hlavní náplní teoretické části práce se stalo v zásadě stanovení informačního základu použitelného pro interpretaci samotné analýzy, jenž je předmětem praktické části. Konkrétně šlo o exkurzi historie kontaktu francouzštiny a angličtiny, francouzského purismu a geneze a partikularit oboru informačních technologií, zejména ve vztahu k jejich terminologii.

Praktická část následně zaštitila realizaci analýzy vybraných IT termínů skrze dotazování zvoleného paralelního korpusu ParaCrawl a následnou analýzu exploratorní vedenou skrze dostupné a relevantní knihovny programu python.

Výsledky analýzy byly v následné finální části práce integrovány do výsledné matice, v rámci níž byla vyhodnocena relevance a hodnota dat. V rámci evaluace dat byl brán hlavní zřetel na to, zda byl výsledný vzorek pro konkrétní termín dominantě francouzský, či byl z většiny zastoupen konkurenčním anglicismem.

Hlavním výsledkem analýzy je zjištění určité ambivalence v používání preferovaných francouzských výrazů. Výsledky naznačují, že tendence k užívání francouzských neologismů může jít ruku v ruce s etablováním určitého termínu v širším jazykovém kontextu a jeho masovým rozšířením. Na druhou stranu u odbornějších termínů, které se používají především v rámci odborných komunit, se projevuje tendence k užívání původních anglicismů.

Bibliographie

BERN, Stéphane. D'où vient le mot "ordinateur" ? Europe 1. en ligne. 2020. Disponible sur: <https://www.europe1.fr/culture/dou-vient-lorigine-du-mot-ordinateur-3991972>. [consulté le 22 juillet 2024].

BLATTBERG, Charles. *Canadian Identity and Language*. en ligne. The Canadian Encyclopedia. 2013. Disponible sur: <https://www.thecanadianencyclopedia.ca/en/article/canadian-identity-and-language>. [consulté le 21 juillet 2024].

BRADARIĆ, Branimir. Borba za čisti hrvatski jezik: Zna li šta su prestrujnik, šećernice, bakroza? Aljazeera. en ligne. 2023. Disponible sur: <https://balkans.aljazeera.net/teme/2023/3/21/pokusaj-prociscavanja-hrvatskog-jezika-od-tudjica>. [consulté le 22 juillet 2024].

BURKE, Peter. Langage de la pureté et pureté du langage. en ligne. *Terrain*. 1998, no. 31, p. 103-112. ISSN 0760-5668. Disponible sur: <https://doi.org/10.4000/terrain.3142>. [consulté le 21 juillet 2024].

CABRÉ, M.T. Terminology. Theory, Methods and Applications. Amsterdam: Benjamins, 1999.

CAMBRIDGE DICTIONARY. en ligne. Disponible sur: <https://dictionary.cambridge.org/>. [consulté le 22 juillet 2024].

CARTWRIGHT, Mark. L'Impact de la Conquête Normande de l'Angleterre. *World History Encyclopedia*. 2019. [en ligne]. Disponible sur : <https://www.worldhistory.org/trans/fr/2-1323/limpact-de-la-conquete-normande-de-langleterre>. [consulté le 21 juillet 2024].

COSTĂCHESCU, Adriana. Comment créer une terminologie ? (Lexique de l'informatique). Synergies Royaume-Uni et Irlande. en ligne. 2012, no. 5, p. 141-155. Faculté des Lettres, Université de Craiova (Roumanie). Disponible sur: <https://gerflint.fr/Base/RU-Irlande5/adriana.pdf>. [consulté le 21 juillet 2024].

FRANCETERME. en ligne. Disponible sur: <https://www.culture.fr/franceterme>. [consulté le 22 juillet 2024].

FUNK, J., WEILAND, V. and WENZ, K. Purisme linguistique et Sprachkritik en français. HESO. en ligne. 2018, no. 3, p. 111-117. Disponible sur: <https://heiu.uni-heidelberg.de/journals/heso/article/view/23892>. [consulté le 21 juillet 2024].

GRAND DICTIONNAIRE TERMINOLOGIQUE. en ligne. Disponible sur: <https://vitrinelinguistique.oqlf.gouv.qc.ca/>. [consulté le 22 juillet 2024].

GRIFFITHS, Ieuan. The Scramble for Africa: Inherited Political Boundaries. *The Geographical Journal*. 1986, vol. 152, no. 2, p. 204-216. The Royal Geographical Society (with the Institute of British Geographers). [en ligne]. Disponible sur : <https://www.jstor.org/stable/634762>. [consulté le 21 juillet 2024].

HAIGH, Thomas. The History of Information Technology. *Annual Review of Information Science and Technology*. en ligne. 2011, vol. 45, p. 1-55. University of Wisconsin. Disponible sur: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.2011.1440450116>. [consulté le 22 juillet 2024].

HIGONNET, Patrice Louis-René. The Origins of the Seven Years' War. *The Journal of Modern History*. 1968, vol. 40, no. 1, p. 57-90. The University of Chicago Press. [en ligne]. Disponible sur : <https://www.jstor.org/stable/1877721>. [consulté le 21 juillet 2024].

HODKOVÁ, Zuzana. Už to nejde zastavit. Expert popisuje, co s lidmi udělá umělá inteligence. *Seznam Zprávy*. en ligne. 2024. Disponible sur: <https://www.seznamzpravy.cz/clanek/porady-inside-talks-uz-to-nejde-zastavit-expert-popisuje-co-s-lidmi-udela-umela-inteligence-251049>. [consulté le 22 juillet 2024].

HOUEBINE, Anne-Marie. Le centralisme linguistique. Brève histoire d'une norme prescriptive. *La Linguistique*. en ligne. 2016, vol. 52, no. 1, p. 35-54. Disponible sur: <https://www.jstor.org/stable/24886342>. [consulté le 21 juillet 2024].

JEWELL, Catherine. L'intelligence artificielle, une invention aussi révolutionnaire que l'électricité. *OMPI Magazine*. en ligne. 2019. Disponible sur: https://www.wipo.int/wipo_magazine/fr/2019/03/article_0001.html. [consulté le 22 juillet 2024].

KVAPILÍK, Martin. Kybernetika po více než šedesáti letech. *Katolický týdeník*. en ligne. 2010. Disponible sur: <https://katyd.cirkev.cz/clanky/kybernetika-po-vice-nez-sedesati-letech.html>. [consulté le 22 juillet 2024].

LAROUSSE. en ligne. Disponible sur : <https://www.larousse.fr/>. [consulté le 2 juin 2024].

LAWLOR, Benjamin R. *The Age of Globalization: Impact of Information Technology on Global Business Strategies*. Bryant University. en ligne. 2008. Disponible sur: https://digitalcommons.bryant.edu/cgi/viewcontent.cgi?article=1000&context=honors_cis. [consulté le 22 juillet 2024].

LE PETIT ROBERT .en ligne. Disponible sur : <https://dictionnaire.lerobert.com>. [consulté le 2 juin 2024].

MERRIAM-WEBSTER. en ligne .Disponible sur: <https://www.merriam-webster.com/>. [consulté le 22 juillet 2024].

MERRIAM-WEBSTER. We Added 690 New Words to the Dictionary for September 2023: Ngl, this update is a big one. en ligne.2023. Disponible sur: <https://www.merriam-webster.com/wordplay/new-words-in-the-dictionary>. [consulté le 22 juillet 2024].

MUSEÉ INFORMATIQUE. Comment s'est-elle développée l'informatique en France ? en ligne .2022. Disponible sur: <https://www.museeinformatique.fr/comment-s'est-elle-developpe-l'informatique-en-france/#:~:text=l'informatique%20existe%20!-,Les%20tous%20premiers%20pas%20de%20l'informatique%20en%20France,num%C3%A9rique%20%C3%A9lectronique%20con%C3%A7ue%20en%20France>. [consulté le 22 juillet 2024].

NADZIEJKA, David E. *Technical Communication. Special Issue: Research in Technical Communication*. en ligne. 1992, vol. 39, no. 4, p. 663-664. Society for Technical Communication. Disponible sur: <https://www.jstor.org/stable/43090185>. [consulté le 21 juillet 2024].

OBSERVATOIRE DE LA LANGUE FRANÇAISE. L'Organisation internationale de la francophonie. Estimation du nombre de francophones. en ligne .2018. Disponible sur : <http://observatoire.francophonie.org/wp-content/uploads/2018/09/Francophones-Statistiques-par-pays.pdf> . [consulté le 21 juillet 2024].

OPUS. en ligne. Disponible sur: <https://opus.nlpl.eu/>. [consulté le 22 juillet 2024].

PARACRAWL. en ligne. Disponible sur: <https://paracrawl.eu/>. [consulté le 22 juillet 2024].

PHILLIPSON, Robert. *Linguistic Imperialism*. 2018. ISBN 9781405194730. doi:10.1002/9781405198431.wbeal0718.pub2.

SCOTT, Hamish. The Seven Years War and Europe's "Ancien Régime". *War in History*. 2011, vol. 18, no. 4, p. 419-455. Sage Publications, Ltd. [en ligne] Disponible sur : <https://www.jstor.org/stable/26098283>. [consulté le 21 juillet 2024].

STATISTIQUE CANADA. *Histoire des ordinateurs*. en ligne. 2021. Disponible sur: [https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch4/history-histoire/5214792-fra.htm#:~:text=1946,%C2%AB%20v%C3%A9ritable%20ordinateur%20%C2%BB%20\(%20c..](https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch4/history-histoire/5214792-fra.htm#:~:text=1946,%C2%AB%20v%C3%A9ritable%20ordinateur%20%C2%BB%20(%20c..) [consulté le 22 juillet 2024].

TASSAVA, Christopher J. *The American Economy during World War II*. Economic History Association. en ligne .2023. Disponible sur: <https://eh.net/encyclopedia/the-american-economy-during-world-war-ii/>. [consulté le 22 juillet 2024].

TECHOPEDIA. en ligne. Disponible sur: <https://www.techopedia.com/>. [consulté le 22 juillet 2024].

WATKINS, Jay. Operation Paperclip: The Secret Intelligence Program to Bring Nazi Scientists to America. *Studies in Intelligence*. en ligne. 2014, vol. 58, no. 3, p. 210-218. Disponible sur: <https://www.cia.gov/resources/csi/studies-in-intelligence/volume-58-no-3/operation-paperclip-the-secret-intelligence-program-to-bring-nazi-scientists-to-america-pdf-210-8kb/>. [consulté le 22 juillet 2024].

WILSON, Richard. English And French in England 1100–1300. *History*. 1943, no. 28, p. 37-60. [en ligne]. Disponible sur : <https://www.jstor.org/stable/24401821>. [consulté le 21 juillet 2024]

Liste des graphiques

Le graphique 1 L'analyse du terme « téléverser »	29
Le graphique 2 L'analyse du terme « pare-feu »	30
Le graphique 3 L'analyse du terme « logiciel »	31
Le graphique 4 L'analyse du terme « matériel »	32
Le graphique 5 L'analyse du terme « module d'extension »	34
Le graphique 6 L'analyse du terme « mémoire vive »	35
Le graphique 7 L'analyse du terme « serveur mandataire »	36
Le graphique 8 L'analyse du terme « témoins de connexion »	37
Le graphique 9 L'analyse du terme « code à octets »	39
Le graphique 10 L'analyse du terme « IUG »	40
Le graphique 11 L'analyse du terme « informatique en nuage »	42
Le graphique 12 L'analyse du terme « boucle de programmation »	44
Le graphique 13 L'analyse du terme « cadre d'application »	45
Le graphique 14 L'analyse du terme « entrepôt de données »	46
Le graphique 15 L'analyse du terme « métadonnées »	49
Le graphique 16 L'analyse du terme « informatique décisionnelle »	50

Liste des tableaux

Le tableau 1 L'exemple de tableau résultant	25
Le tableau 2 La liste des termes sélectionnés	27
Le tableau 3 L'analyse du terme « téléverser »	28
Le tableau 4 L'analyse du terme « pare-feu »	30
Le tableau 5 L'analyse du terme « logiciel »	31
Le tableau 6 L'analyse du terme « matériel »	32
Le tableau 7 L'analyse du terme « bogue »	33
Le tableau 8 L'analyse du terme « dialogueur »	33
Le tableau 9 L'analyse du terme « module d'extension »	34
Le tableau 10 L'analyse du terme « mémoire vive »	35
Le tableau 11 L'analyse du terme « serveur mandataire »	36
Le tableau 12 L'analyse du terme « témoins de connexion »	37
Le tableau 13 L'analyse du terme « code à octets »	38
Le tableau 14 L'analyse du terme « IUG »	40
Le tableau 15 L'analyse du terme « informatique en nuage »	41
Le tableau 16 L'analyse du terme « application dorsale »	43
Le tableau 17 L'analyse du terme « application frontale »	43
Le tableau 18 L'analyse du terme « boucle de programmation »	44
Le tableau 19 L'analyse du terme « cadre d'application »	45
Le tableau 20 L'analyse du terme « entrepôt de données »	46
Le tableau 21 L'analyse du terme « outil d'alimentation »	47
Le tableau 22 L'analyse du terme « logiciel local »	48
Le tableau 23 L'analyse du terme « métadonnées »	48
Le tableau 24 L'analyse du terme « informatique décisionnelle »	50
Le tableau 25 Les résultats de l'analyse	52

Liste des images

L'image 1 CWB.....	22
L'image 2 Code HTML	22
L'image 3 Code python	23
L'image 4 Code python 2	24
L'image 5 Sortie du programme	25
L'image 6 Code python 3	26