

Université de Tartu

Institut des langues et des cultures étrangères

Département d'études romanes

Artjom Romanov

ANGLICISMES DANS LE DISCOURS DES DÉVELOPPEURS
FRANCOPHONES : USAGES, ADAPTATION MORPHOLOGIQUE
ET POSITIONS IDÉOLOGIQUES

Sous la direction de Marge Käsper

Mémoire de licence

Tartu 2026

Table des matières

Introduction	4
1. Cadre théorique	5
1.1. Définition et typologie des anglicismes	5
1.2. Le processus d'adaptation des anglicismes	7
1.2.1. Intégration morphologique des anglicismes	7
1.2.2. Variation orthographique	8
1.3. Enjeux sociolinguistiques	8
2. Méthodologie	10
2.1. Construction du corpus	10
2.1.1. Le choix de la plateforme Reddit et du corpus thématique défini	13
2.2. Critères de sélection et de délimitation	14
2.3. Cadre analytique	15
2.4. Procédure analytique	15
3. Analyse	17
3.1. Aperçu quantitatif	17
3.1.1. Intégration des anglicismes : emprunts bruts ou adaptés	17
3.1.2. Formes les plus fréquentes	19
3.1.3. Première discussion sur les résultats	21
3.2. Identification et classification	21
3.2.1. Emprunt lexical direct	22
3.2.2. Emprunts morphologiquement adaptés	22
3.2.3. Emprunts sémantiques et calques	23
3.3. Le cas particulier de <i>dev</i>	24
3.4. Concurrence terminologique et stabilisation des formes	26
3.4.1. Position anti-anglicismes	26
3.4.2. Position pro-anglicismes	28
3.4.3. Acceptation mixte	29
Conclusion	31
Bibliographie	33
Resümee	37
Annexes	38
Annexe 1. Fonctionnement de Reddit	38
Annexe 2. Configuration PRAW et script Python de collecte du corpus	40
2.1. Client PRAW	40

2.2. Script Python	40
Annexe 3. Exemple d'utilisation de ChatGPT	43
Annexe 4. Le tableau chiffré des anglicismes relevés	45

Introduction

L'usage des technologies numériques est un phénomène mondial. Dans le secteur informatique, l'anglais occupe une position de langue de référence pour la documentation, les outils et l'innovation terminologique. Cette dominance a conduit à une large diffusion d'anglicismes dans la langue française, en particulier dans les domaines professionnels et technologiques.

L'objectif de ce travail est d'analyser les anglicismes du français de l'informatique à partir d'un corpus de commentaires Reddit issus du forum francophone r/developpeurs. L'étude adopte une démarche de linguistique de corpus à visée descriptive : elle décrit la distribution des anglicismes (terminologie technique vs lexique socio-professionnel), les types d'emprunt et les mécanismes d'intégration.

L'analyse répondra à quatre questions de recherche :

- Comment se répartissent les anglicismes du français de l'informatique entre les termes techniques et le vocabulaire socio-professionnel du secteur ?
- Quels mécanismes d'intégration morphologique dominant : verbalisation, flexion, dérivation, et quelles bases deviennent productives ?
- Dans quelle mesure la variation orthographique indique-t-elle une intégration non stabilisée plutôt qu'une simple variation accidentelle ?
- Quels types d'emprunts structurent le corpus et comment se combinent-ils avec l'intégration morphologique ?

Le corpus est constitué d'un espace de communication informel au sein de la communauté francophone, en rapport étroit avec l'informatique. Les anglicismes ont été recensés en fonction de leur niveau d'intégration dans la langue française et sont répartis en différentes catégories.

1. Cadre théorique

1.1. Définition et typologie des anglicismes

L'emprunt lexical est un phénomène qui se produit depuis des siècles. Dans un article fondateur devenu un classique dans les études sur l'emprunt, Haugen (1950 : 211-212) définit l'emprunt comme un mécanisme d'adaptation linguistique qui entraîne l'évolution du langage au contact d'autres systèmes.

Le cas français est également lié à une tradition de norme et de réglementation. Dès 1635, l'Académie française a pour mission d'établir des règles pour la langue, de la rendre pure, éloquente et capable de manier les arts et les sciences. La préface de la première édition du Dictionnaire de l'Académie française (1694) reflète cette orientation prescriptive : l'Académie revendique une posture prudente à l'égard des innovations, en indiquant qu'elle ne voulait pas prendre en charge plusieurs mots nouvellement inventés ni certaines tournures introduites par « la Licence » ou « le Caprice de la Mode ».

Dans le français contemporain, l'anglais occupe une position particulièrement dominante dans les domaines scientifiques et techniques, ce qui explique la forte fréquence des anglicismes dans le vocabulaire spécialisé. Ce phénomène recoupe des questions historiques et sociolinguistiques plus larges, où les préoccupations linguistiques rencontrent des préoccupations symboliques. Le terme *franglais*, popularisé par le pamphlet de René Etienne « *Parlez-vous le franglais ?* (1964) », résume ces débats.

Comme l'affirme Zanola (2008 : 88-90), la notion de *franglais* ne correspond pas à un système linguistique réel ou stable, mais désigne plutôt un ensemble d'emprunts ordinaires accompagnés d'une forte idéologie. Pour les Français, le franglais est perçu comme le danger d'une contamination linguistique qui appauvrit leur langue, le risque de perdre une valeur partagée, un lieu d'identité nationale, au-delà des visions idéologiques et des enjeux politiques. Tout au long de la seconde moitié du XX^e siècle, les institutions françaises et les autorités de politique linguistique ont tenté de réguler ces emprunts en proposant des équivalents officiels et en promouvant un certain purisme linguistique. Pourtant, comme le montre Zanola, la crainte d'une langue

prétendument « hybride » oublie la nature dynamique et adaptative du français, qui a toujours intégré des éléments étrangers. Comme le montre Walter (1997), la langue française s'est façonnée au fil des siècles grâce à des emprunts provenant d'une grande variété de langues - non seulement l'anglais moderne, mais aussi le grec, le latin, les dialectes régionaux et d'autres langues européennes. Loin de porter atteinte à l'intégrité du français, ces emprunts ont historiquement contribué à sa capacité d'adaptation. Ainsi, la présence actuelle d'anglicismes technologiques ne doit pas être perçue comme une dégradation, mais comme la continuité d'une longue tradition de progression lexicale.

Selon la classification proposée par Sablayrolles et Pruvost (2016), les emprunts en français se répartissent en plusieurs catégories :

- Les emprunts lexicaux – le mot étranger (simple ou composé) est intégré directement au français. Dans ce cas, le sens peut rester identique (*camping* de l'anglais, *spaghetti* de l'italien ou *manga* du japonais).
- Les emprunts sémantiques – un mot déjà présent en français acquiert un nouveau sens sous l'influence d'une autre langue, en général, de l'anglais. Ainsi, le verbe réaliser a pris le sens de 'comprendre' (*realize* de l'anglais).
- Les calques sémantiques – il s'agit de traductions littérales d'expressions ou de composés étrangers. Par exemple, *lune de miel* traduit en français l'expression de l'anglais *honeymoon*. Ces calques, lorsqu'ils respectent les structures françaises, sont en général bien acceptés.
- Les calques syntaxiques – ce sont des constructions grammaticales françaises qui reproduisent mot à mot une structure étrangère. La majorité des calques syntaxiques actuels provient de l'anglais. La locution *faire sens* signifie 'avoir du sens' (sur *to make sense*). Ces calques remplacent souvent des tournures françaises préexistantes, ce qui explique qu'ils soient régulièrement critiqués.

Dans le français de l'informatique, la domination des emprunts lexicaux directs s'explique par plusieurs facteurs :

- l'innovation terminologique apparaît d'abord en anglais
- la documentation, les interfaces et les ressources techniques circulent majoritairement en anglais

- les équivalents français, lorsqu'ils existent, peuvent être concurrencés par une forme anglaise déjà stabilisée dans l'usage
- les intitulés de postes et certaines pratiques professionnelles (par exemple *junior/senior, lead, remote*) sont souvent empruntés comme terminologie de secteur

1.2. Le processus d'adaptation des anglicismes

L'intégration d'un emprunt dans une langue ne se limite pas à sa présence lexicale. L'acte d'emprunter est, comme l'explique Guilbert (1975 : 89, 92), un acte de créativité et de création linguistique plutôt qu'un simple « vol » d'une autre langue, ce qui souligne la vitalité et l'expansion du français.

1.2.1. Intégration morphologique des anglicismes

Les travaux de Sablayrolles (2019) et d'Ashrafova (2024 : 101-102) montrent que, pour que les emprunts puissent fonctionner dans la langue, ils doivent être adaptés de manière à s'inscrire dans son système grammatical et morphologique, ce qui garantit leur conformité aux schémas morphologiques du français. À cet égard, les suffixes français, notamment -er pour l'infinitif et -é pour le participe passé, permettent de réinterpréter des formes anglaises comme des formes françaises. Une fois adaptés, ces emprunts peuvent être :

- conjugués (*je checke, nous avons uploadé*), en suivant le régulier infinitive des verbes en -er
- dérivés, en produisant des formes telles que *uploadé, marketing* ou d'autres dérivés
- flexionnels, comme dans *les weekends* ou *des emails*, qui suivent les règles françaises de pluralisation

Comme le souligne Sablayrolles (2019), les néologismes, qu'ils soient empruntés ou formés de manière interne, ne deviennent pleinement opérationnels que lorsqu'ils s'alignent sur les schémas morphologiques existants. Les adaptations observées dans les emprunts d'origine anglaise confirment que la morphologie du français n'importe pas simplement des formes anglaises : elle les réinterprète et les restructure afin

qu'elles puissent participer aux processus natifs de conjugaison, de dérivation et de flexion.

Une autre preuve apparaît dans une étude effectuée par Cartier (2019). Cartier (2019 : 19) montre que les mots anglais liés aux technologies donnent fréquemment naissance à de vastes familles de dérivés :

- facebook → facebooker, facebookien(ne), anti-facebook
- twitter → tweeter, twitto(s), retweeter, tweeteur(euse), twitterisation
- snapchat → snapchater, snapchat(s), snapchateur/snapchateuse
- instagram → instagrammeur/euse, instagrameur, instagrammer

Cela démontre que les anglicismes techniques non seulement entrent en français, mais qu'ils deviennent rapidement des bases lexicales pleinement intégrées, capables de supporter les affixations françaises telles que -eur, -age, -iser ou re-.

1.2.2. Variation orthographique

L'intégration graphique n'est pas encore stabilisée. La coexistence de plusieurs graphies pour un même lexème n'est pas un détail superficiel mais un problème structurel qui affecte la manière dont les mots sont intégrés dans un système linguistique (Janssen 2006 : 168-169). Ceci s'applique directement aux anglicismes technologiques en français tels que *debugger vs debugger vs debuguer*.

La variation orthographique révèle ainsi non seulement l'hésitation des usagers quant aux normes à suivre, mais aussi la difficulté d'adapter des formes d'origine anglaise aux schémas morphologiques français.

1.3. Enjeux sociolinguistiques

Les anglicismes ne sont pas seulement des formes lexicales mais également des marqueurs sociaux. Selon Bourdieu (1982), la langue fonctionne comme un capital symbolique permettant aux locuteurs d'exprimer une appartenance à un groupe. Dans les environnements technologiques, la terminologie anglaise joue un rôle majeur dans la construction identitaire.

Dans cette perspective sociolinguistique, plusieurs études montrent que l'usage des anglicismes techniques répond à des dynamiques sociales plus larges que la simple précision terminologique. L'anglais est fréquemment employé comme un marqueur d'appartenance à une culture mondialisée, permettant aux locuteurs de s'identifier à des pratiques internationales et de renvoyer une image socialement plus moderne (Martin 2019). Dans les environnements technologiques, ce phénomène est encore plus marqué : utiliser la terminologie anglaise devient une manière d'exprimer son intégration dans des communautés professionnelles globalisées, tout en signalant une familiarité avec les outils, les plateformes et les innovations qui proviennent majoritairement du monde anglophone. En outre, l'emploi d'anglicismes permet aux spécialistes de se distinguer des non-experts, puisque la maîtrise du lexique technique anglais sert souvent d'indicateur implicite de compétence. Ces dynamiques sociales contribuent à expliquer pourquoi les formes anglaises demeurent solidement ancrées dans l'usage, même lorsque des équivalents français existent, comme *pare-feu* face à *firewall*.

L'Académie française recommande de limiter les anglicismes, se référant aux principes de la loi Toubon (1994), qui encourage l'utilisation d'équivalents français, y compris dans le français technique, afin de réduire l'emploi excessif des emprunts. La nouvelle terminologie est d'abord créée et circule au sein de domaines spécialisés, où elle acquiert son sens à l'intérieur de pratiques professionnelles spécifiques (Depecker 2005 : 6). Sa diffusion au-delà de ces domaines dépend ensuite de l'usage réel qu'en font les locuteurs, plutôt que d'une imposition abstraite, puisque c'est l'ancrage dans la pratique qui détermine son adoption. Ainsi, on pourrait dire que la terminologie ne s'impose pas par décret, mais qu'elle se diffuse par l'usage (Rousseau 2005 : 96-100). Dans les domaines techniques, l'usage prévaut sur les normes pour trois raisons principales :

- la rapidité de l'innovation terminologique
- la circulation mondiale des pratiques professionnelles
- la centralité de l'anglais dans la documentation et les outils

2. Méthodologie

2.1. Construction du corpus

Le corpus utilisé pour cette étude est constitué de commentaires publiés sur le forum Reddit r/developpeurs, une communauté francophone en ligne dédiée au développement de logiciels et aux technologies numériques (pour une présentation de l'interface de ce forum voir l'Annexe 1). Cette plateforme a été choisie parce qu'elle représente un espace d'interaction informel mais hautement spécialisé, où les participants discutent régulièrement de questions techniques, d'outils, de pratiques professionnelles et de choix linguistiques liés à l'informatique. Les commentaires Reddit offrent un discours spontané et non édité, ce qui les rend particulièrement adaptés à l'observation de l'utilisation réelle du langage plutôt qu'aux normes prescriptives. Parallèlement, la communauté rassemble des utilisateurs ayant des degrés d'expertise variables, des débutants aux professionnels, permettant l'émergence d'une diversité de pratiques linguistiques et d'attitudes métalinguistiques.

Le corpus a été compilé à l'aide de Python, via la bibliothèque Python Reddit API Wrapper (PRAW s.d.), collectant une année de données datant du 18.10.2024 au 15.10.2025. Les données ont été collectées à l'aide d'un script de Python (pour lire le script, voir l'Annexe 2).

Les données ont été traitées avec l'aide de ChatGPT (OpenAI 2022), avec pour tâche de rechercher des anglicismes nécessaires pour constituer le corpus d'étude. L'aide de ChatGPT (OpenAI 2022) a également été utilisée lors de la rédaction finale des commentaires (un exemple de la manière dont il a été utilisé pour la compilation finale des commentaires est disponible dans l'Annexe 3).

L'ensemble de données a été compilé à partir du corpus r/developpeurs et comprend du matériel tiré de 228 threads, comprenant 22 149 commentaires écrits par 5 721 auteurs distincts. ChatGPT (OpenAI 2022) a été utilisé comme outil d'appui analytique : des instructions lui ont été fournies afin de filtrer les données et d'identifier des anglicismes attestés dans le corpus. Les étapes :

1. Le corpus a été importé depuis le fichier `r_developpeurs_corpus.csv` compilé à l'aide du code composé dans Python (22 149 commentaires).
2. Chaque commentaire a été traité comme une unité et on calcule, pour une forme donnée, le nombre de commentaires dans lesquels celle-ci apparaît au moins une fois. Cette mesure permet d'estimer la répartition de l'anglicisme dans la communauté. On calcule aussi le nombre total d'occurrences de la forme dans tout le corpus, en comptant chaque apparition. Cette mesure rend compte de l'intensité d'usage.

Par exemple : dev apparaît dans 3010 commentaires, mais dev totalise 4 335 occurrences.

3. Une normalisation légère a été appliquée pour faciliter les recherches :
 - mise en minuscules
 - harmonisation des apostrophes
 - suppression des accents pour le groupement (*dév* - *dev*)
4. La notion d'anglicisme appliquée comprenait aussi bien les emprunts lexicaux que les emprunts lexicaux adaptés
5. Les exclusions qui ont été explicitement prises en compte :
 - noms propres employés comme Windows, GitHub, LinkedIn, etc. non comptabilisés comme anglicismes
 - mais si un nom propre a été recatégorisé dans l'usage (minuscule, déterminant, pluriel) ou servait de base à une dérivation (*googler*), il a été traité séparément dans une sous-catégorie « dérivés de noms propres ».
6. Une série d'anglicismes a ainsi été établie à partir du corpus, en listant les formes pertinentes pour les besoins de l'étude (*dev*, *junior*, *senior*, *framework*, *backend*, etc.), ainsi que leurs variantes.
7. Chaque entrée a été annotée (catégorie grammaticale, type d'emprunt : brut / adapté) et accompagnée d'exemples.
8. Pour éviter de disperser artificiellement les fréquences, les variantes ont été regroupées sous une même entrée canonique :
 - variantes de casse : *Dev/DEV/dev*
 - variantes accentuées : *dév/dévs*

- pluriels : *dev/devs, junior/juniors*
 - variantes graphiques : *back-end/backend, full-stack/fullstack*
 - variantes morphologiques dans le cas des adaptés (base anglaise + morphologie française)
9. Le groupement a été appliqué : une entrée = un ensemble de variantes listées
10. Deux types de comptage ont été distingués :
- Occurrences (*token frequency*) - nombre total de fois où une forme ou ses variantes apparaissent dans tout le corpus.
 - Présence par commentaire (*document frequency*) - nombre de commentaires contenant au moins une occurrence de la forme.

(Exemple : *dev* apparaît dans 3 010 commentaires ≠ *dev* totalise 4 355 occurrences)

11. Pour certains items, des règles spécifiques ont été ajoutées afin d'éviter des faux positifs :
- *boss* : comptage du nom *boss* uniquement (exclusion de *bossier* et d'autres formes verbales, dont le sens et l'étymologie sont différents (le mot familier pour 'travailler' *bossier* vient du mot dialectal français signifiant 'se courber' (Larousse)).
 - *manager* : selon l'objectif, possibilité de compter seulement certains contextes, par exemple nom avec déterminant, mais cette règle doit être explicitée si elle est retenue.
 - il peut s'agir d'un nom emprunté / d'une étiquette de travail (*un manager, les managers*), il est donc perçu comme un anglicisme compté dans l'analyse du corpus
 - mais il est également utilisé comme un verbe entièrement français (*manager une équipe*), ce qui signifie qu'il ne s'agit pas d'un anglicisme et n'a donc pas été compté
12. À partir de l'inventaire et les règles définies, les résultats ont été générées :
- un tableau qui montre les fréquences totales des formes, en leur forme canonique et en variantes principales
 - et, si besoin, un tableau du nombre de commentaires contenant :
 - distribution des résultats
 - diagnostic de contexte et d'intégration

- variation orthographique
- cooccurrences ou collocations

Une attention particulière a été portée aux discussions où les utilisateurs commentent le choix de la langue elle-même, par exemple en justifiant l'utilisation de termes anglais, en critiquant les anglicismes ou en proposant des équivalents français. Ces discussions linguistiques en ligne - soit relevant de la « linguistique populaire » (Preston 2008) - sont particulièrement pertinentes pour comprendre non seulement comment les anglicismes sont utilisés, mais aussi comment ils sont évalués et négociés au sein de la communauté.

2.1.1. Le choix de la plateforme Reddit et du corpus thématique défini

Reddit est une plateforme en ligne organisée en communautés thématiques appelées subreddits. Chaque subreddit fonctionne comme un forum dédié à un intérêt spécifique : les utilisateurs publient, et d'autres répondent dans un fil de commentaires. Les discussions sont structurées hiérarchiquement : un commentaire peut recevoir des réponses, et ces réponses peuvent être répondues à leur tour, produisant des échanges longs et détaillés. Le système de vote positif/négatif affecte la visibilité du contenu dans un fil.

Comparable à Twitter / X, la participation de Reddit est relativement ouverte, rassemblant divers utilisateurs. Cette diversité est particulièrement précieuse pour étudier les anglicismes au sein du corpus : plutôt que d'observer un discours unique et homogène (par exemple, un média ou un groupe étroitement défini), il devient possible de saisir une gamme de registres et d'attitudes. La recherche sur le forum de développeurs a ainsi donné une liste où il y a des contributions hautement techniques aux côtés de réponses humoristiques, de positions normatives ou de débats sur la terminologie.

Ce type de données permet également d'observer le vocabulaire en contexte, au sein d'échanges spontanés avec tous les termes techniques thématiques (outils, documentation, routines professionnelles, etc.). Les anglicismes apparaissent donc non seulement comme des éléments lexicaux mais comme des choix situés, qui peuvent passer inaperçus ou, au contraire, déclencher des réactions : reformulation,

propositions d'équivalents français, justifications pragmatiques, ironie ou critique. En ce sens, Reddit, ou plus précisément le corpus choisi, n'est pas simplement un « réservoir de mots », mais un lieu où l'usage devient visible à travers l'interaction, la négociation et parfois la tension autour des normes.

Enfin, le format rend la micro-dynamique particulièrement facile à observer : un terme spécifique peut être repris et discuté, corrigé, adapté en français ou défendu au nom de la précision, de l'efficacité ou de l'habitude professionnelle. Cette combinaison d'un ton de conversation libre, de la diversité des participants et d'une interaction structurée fait de Reddit un site de terrain pertinent pour examiner comment les anglicismes circulent et comment ils sont reçus dans leur contexte.

2.2. Critères de sélection et de délimitation

Les données sélectionnées incluent des commentaires qui contiennent :

- emprunts lexicaux directs à l'anglais ou formes morphologiquement adaptées, relevant soit de la terminologie technique, soit du lexique socio-professionnel du secteur informatique
- références explicites à l'anglais comme langue
- déclarations concernant les anglicismes (acceptation, rejet, justification ou critique)
- une comparaison entre la terminologie française et anglaise
- traitement des noms propres : exclus des comptages, sauf lorsqu'ils sont morphologiquement intégrés (nom communisé / dérivation), analysés séparément.

Seuls les commentaires textuels rédigés en français ont été retenus. Les extraits de code, les fichiers de configuration et les sorties purement techniques ont été exclus, car ils ne représentent pas des choix en langage naturel. Les commentaires composés uniquement de liens ou d'émojis ont également été exclus. Une attention particulière a été accordée aux discussions dans lesquelles la langue elle-même devient un sujet de discussion, en particulier lorsque les participants commentent explicitement l'utilisation de termes anglais ou français.

Le corpus est de nature qualitative plutôt que quantitative. Cette étude adopte une démarche mixte à visée descriptive : les comptages et fréquences présentés servent à décrire la distribution des formes dans le corpus, sans prétendre à une généralisation statistique ni à des tests d'inférence. L'analyse qualitative vise ensuite à interpréter ces tendances à partir d'exemples contextualisés. Son objectif n'est pas de mesurer statistiquement la fréquence, mais d'identifier des schémas récurrents d'utilisation, d'adaptation et de positionnement idéologique liés aux anglicismes.

2.3. Cadre analytique

L'analyse combine une approche linguistique descriptive et une perspective terminologique, fondées sur les données du corpus.

D'une part, les anglicismes sont classés selon leurs propriétés linguistiques, en s'appuyant sur des typologies établies (emprunts lexicaux, emprunts sémantiques, calques, etc.). Une attention particulière est portée à :

- intégration morphologique (verbalisation, pluralisation, dérivation)
- variation orthographique
- critères d'intégration
- productivité des bases empruntées et schémas de dérivation en français

D'autre part, l'étude examine les aspects sociaux des anglicismes. Cette double perspective permet de relier la forme linguistique au sens social : non seulement quels types d'anglicismes sont utilisés, mais pourquoi ils sont défendus, critiqués ou négociés.

2.4. Procédure analytique

L'analyse s'est déroulée en plusieurs étapes :

1. Identification et classification des anglicismes

Toutes les formes dérivées de l'anglais ont été identifiées et regroupées selon leur catégorie grammaticale (nom, verbe, adjectif) et le type d'emprunt. Chaque élément a été classé dans l'une des options suivantes :

- emprunts lexicaux directs
- emprunts morphologiquement adaptée
- emprunt sémantique
- calque

2. Observation de la variation formelle

Des variantes orthographiques et morphologiques ont été enregistrées afin d'identifier l'instabilité ou la compétition entre les formes.

3. Analyse terminologique

Les commentaires contenant des comparaisons explicites (anglais vs équivalents français) ont été examinés afin d'illustrer la concurrence terminologique.

4. Interprétation linguistique et terminologique

Les résultats ont été interprétés à partir des catégories d'emprunt, des indices d'intégration et des phénomènes de variation et de stabilisation terminologique.

3. Analyse

3.1. Aperçu quantitatif

Cette section propose un aperçu descriptif des anglicismes du français de l'informatique. Les fréquences rapportées ici sont descriptives et propres au corpus. Elles ne doivent pas forcément constituer une inférence statistique sur l'ensemble du français de l'informatique. Les résultats sont présentés en distinguant, lorsque ceci est pertinent, la terminologie technique et le lexique socio-professionnel du secteur. La figure 1 présente la répartition entre anglicismes bruts et anglicismes adaptés. La figure 2 répertorie les formes d'anglicisme les plus fréquentes.

3.1.1. Intégration des anglicismes : emprunts bruts ou adaptés

Dans l'ensemble du corpus, les anglicismes représentent environ 17 265 occurrences sur environ 1 280 058 tokens. Cela correspond à une estimation de 1,349 % de l'ensemble des tokens, ce qui indique que contrairement à ce qu'on pourrait peut-être s'attendre, le matériau lexical d'origine anglaise ne domine pas sur un tel forum, il ne constitue pas au fond une part faible mais mesurable du discours écrit des utilisateurs de r/developpeurs. Comme la notion d'anglicisme ne forme pas une catégorie objectivement close, et puisque les comptages dépendent fortement de choix relatifs à la tokenisation, aux variantes orthographiques, aux expressions polylexicales et aux recouvrements entre formes, cette estimation doit toujours être considérée comme approximative plutôt que définitive.

Aux fins de la présente étude, il est donc plus informatif de restreindre la portée aux anglicismes qui sont directement pertinents pour les questions de recherche. Le nombre d'anglicismes bruts est de 11666 et adapté est de 582, ce qui signifie que le nombre total est de 12248 (0,96% de l'ensemble des tokens).

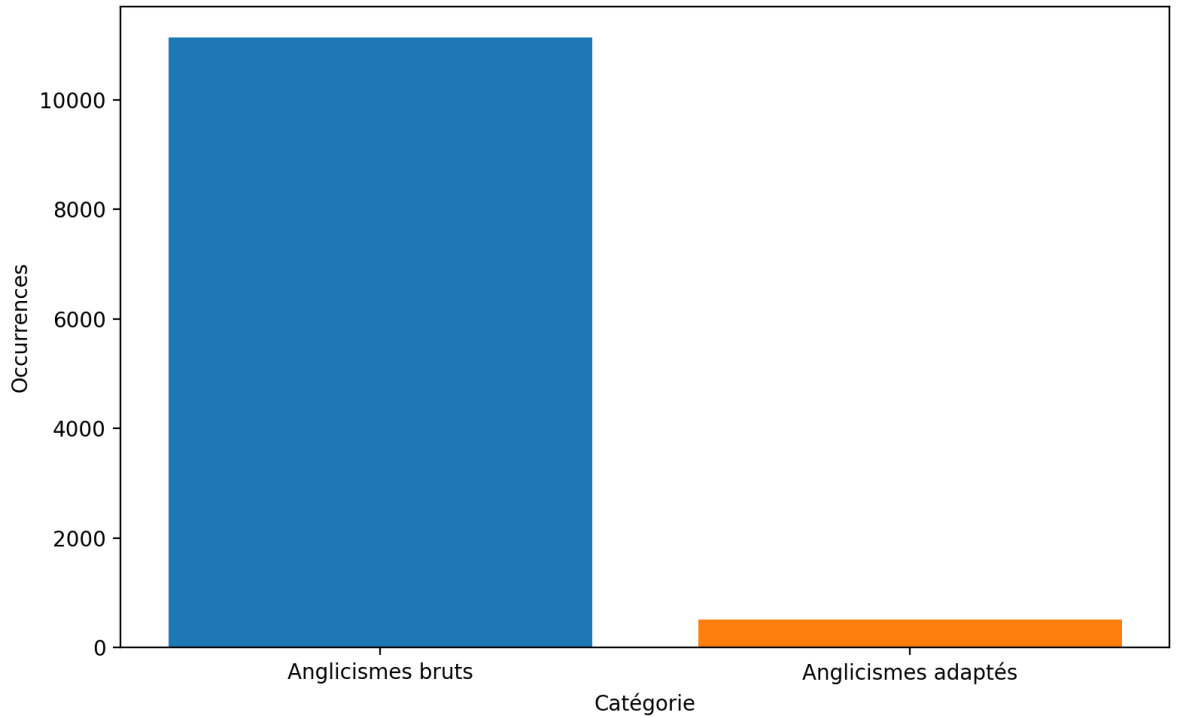


Figure 1. Répartition entre les anglicismes bruts et adaptés dans le corpus.

La figure 1 présente la distribution des occurrences d’anglicisme en deux catégories.

Les anglicismes bruts sont des mots qui apparaissent dans une graphie proche de l’anglais (par ex. *framework, backend, commit, ticket, debug, patch, mais aussi junior, senior, lead, remote*), y compris des abréviations (*dev*¹).

Les anglicismes adaptés sont des mots où une base anglaise est intégrée à la morphologie française. Par exemple, la verbalisation française, les éléments dérivationnels français ou les formations hybrides combinant du matériel anglais et français (comme *debug/ger, refactor/er, manage/ment*).

La figure 1 montre un fort déséquilibre : les anglicismes bruts sont nettement plus nombreux que les anglicismes adaptés. Pour ce corpus, la prédominance des formes brutes suggère que l’emprunt direct est la norme, compatible avec le fait que le génie logiciel s’appuie fortement sur la terminologie anglaise, tant dans la documentation que dans les interactions professionnelles quotidiennes. Les formes adaptées restent

¹ Pour les spécificités de cette forme particulière, voir 3.3 ci-dessous

néanmoins importantes analytiquement, car elles offrent les indices les plus clairs des mécanismes d'intégration, qui seront examinés dans les sections suivantes.

3.1.2. Formes les plus fréquentes

Alors que la figure 1 décrivait comment les anglicismes tendent à apparaître, la figure 2 met en évidence les éléments les plus fréquents.

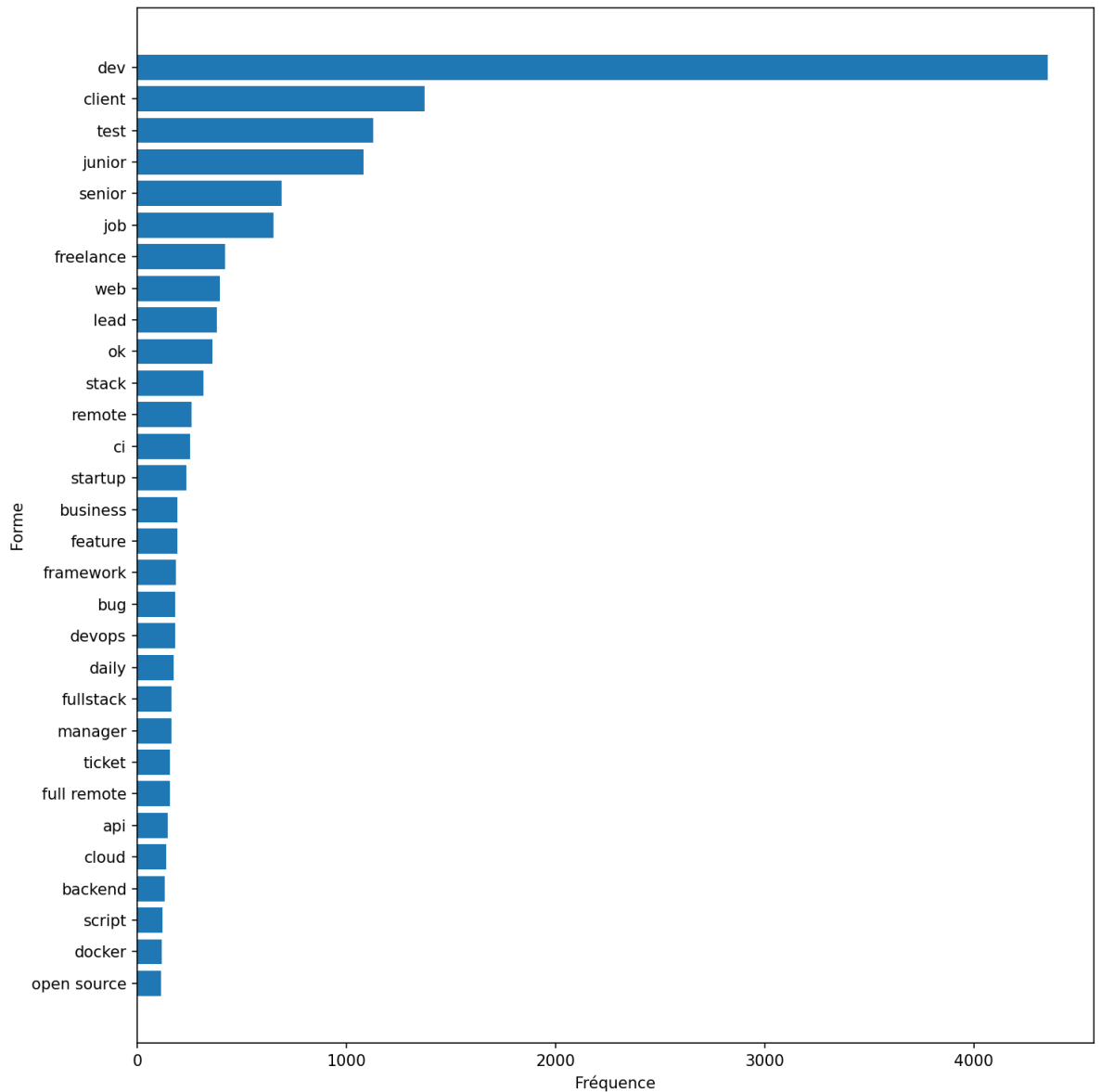


Figure 2. Les anglicismes les plus fréquents.

La forme la plus fréquente est de loin *dev*, ce qui souligne le statut particulier de ce mot dans la communauté. Dans ce corpus, *dev* fonctionne à la fois comme un « développeur » et une activité de « développement ».

Au-delà de cet élément supérieur, les formes les plus fréquents se répartissent en deux groupes thématiques :

- Vocabulaire de travail et de carrière : *junior, senior, emploi, manager, lead, freelance, remote / full remote*.
 - Cet ensemble indique que les anglicismes dans r/développeurs ne se limitent pas à des termes techniques : ils sont également fortement liés aux réalités du travail et au positionnement professionnel. En d'autres termes, le corpus reflète le fait que les développeurs discutent non seulement du code et des outils, mais aussi de l'embauche, de la hiérarchie et des conditions de travail.
- Flux de travail technique et infrastructure : *stack, framework, feature, bug, devops, backend, cloud, docker, ticket, open source*.
 - Ces formulaires correspondent à des concepts du génie logiciel, y compris l'outillage, les processus de production et les environnements de développement. Leur récurrence confirme que le vocabulaire anglais sert de système de référence partagé pour discuter efficacement des pratiques techniques.

Pour une vue détaillée des fréquences des anglicismes attestés dans le corpus, un tableau séparé a été créé et présenté en Annexe 4. Ce tableau inclut tous les anglicismes dont la fréquence était supérieure à 100 occurrences et dont l'usage nous allons analyser ci-dessous.

En ce qui concerne l'ensemble des formes relevées, notons enfin qu'il y avait aussi quelques formes moins fréquentes dans le corpus. Cette liste permet de mettre en évidence des usages qui, malgré leur faible occurrence, contribuent à la diversité du corpus. Les formes rares sont présentées ci-dessous :

- unit test (3) et unittest (1)
- scalable (5)
- revue de code (8)

- livraison continue (4)
- pair programming (16)
- checker (19)
- patch (8), patches (3), patcher (2), patching (1)
- brancher (6), branch (3), branching (3)
- hotfix (4), hotfixs (1), hotfixer (1)
- switch (15), switcher (8), switching (2) etc.

3.1.3. Première discussion sur les résultats

Ensemble, les figures 1 et 2 soutiennent plusieurs implications qui orientent l'analyse qualitative.

Premièrement, la prédominance des formes brutes indique que les étiquettes anglaises circulent comme des unités « prêtes à l'emploi », probablement parce qu'elles correspondent à la langue des outils, de la documentation et des normes professionnelles internationales. Deuxièmement, la présence de formes adaptées montre que le matériel anglais peut devenir productif dans l'usage français. Les sections suivantes se concentrent donc sur les mécanismes d'adaptation et sur les contextes interactionnels où les choix de forme sont négociés, justifiés ou critiqués.

3.2. Identification et classification

Dans un processus d'identification et de regroupement, précisant leur analyse quantitative, les anglicismes peuvent être analysés de manière plus systématique. Comme cela a été mentionné dans l'extrait de procédure analytique, chaque item a été classé soit en emprunt lexical direct, en emprunt morphologiquement adapté, en emprunt sémantique ou en calque, suivant des typologies d'emprunt lexical largement utilisées en linguistique de contact, notamment celles présentées par Guilbert (1975). Cette classification permet d'examiner chaque occurrence séparément du point de vue morphologique et orthographique.

3.2.1. Emprunt lexical direct

Les emprunts lexicaux directs sont des mots anglais qui sont importés en français sans modification formelle. Ils sont intégrés syntaxiquement (articles, pluralisation) mais conservent leur forme anglaise. Ce type d'intégration est bien documenté dans les études sur la réception du lexique anglais dans le discours écrit français (Biedermann-Pasques & Humbley 1995 : 57-59, 63). Le corpus révèle une forte présence de tels emprunts liés au développement logiciel et aux pratiques professionnelles, par exemple *job*, *dev*, *manager*, *stack*, *startup*, *freelance*, *lead*, *web*, *remote* et *junior/senior*. Ces éléments sont utilisés comme noms ou adjectifs français ordinaires (*un job*, *des devs*, *un junior*), et leurs significations restent très proches de leurs originaux anglais. Certains éléments tels que *junior*, *senior*, *lead* ou *freelance* peuvent fonctionner à la fois comme noms et adjectifs, ce qui montre la flexibilité de la langue anglaise et l'importe en français. La dérivation nominale est également fréquente. Des suffixes tels que *-eur*, *-age*, *-isation* ou *-able* sont utilisés pour former de nouveaux éléments lexicaux à partir de bases anglaises.

Bon nombre de ces termes font référence à des concepts pour lesquels il n'existe aucun équivalent français largement accepté ou dont les équivalents français sont perçus comme artificiels ou peu pratiques. Les études sur l'implantation terminologique montrent que l'usage dans les communautés spécialisées prévaut fréquemment sur les recommandations institutionnelles, en particulier dans les domaines techniques structurés par des pratiques internationales (Rousseau 2005 : 96, 100).

3.2.2. Emprunts morphologiquement adaptés

Les emprunts morphologiquement adaptés sont constitués de bases lexicales anglaises adaptées à la morphologie française par des terminaisons verbales du premier groupe en *-er*, comme dans *debugger*, *checker*, *scaler*, *refactorer* (je debugge, nous scalons). Ces verbes sont conjugués comme des verbes natifs français et participent pleinement à la syntaxe verbale française. Les locuteurs ne changent pas de langue mais traitent plutôt les racines anglaises comme une base pour le système grammatical français. Ce mécanisme est décrit pour le français contemporain dans les travaux sur les emprunts verbaux, qui montrent l'existence d'un ensemble important de verbes créés sur base

anglaise et majoritairement fléchis en -er, en particulier dans les usages techniques (Zayed 2021 : 116).

Cependant, certaines normes ne sont pas établies, comme le mot *debugger*. Les ressources terminologiques institutionnelles telles que celles publiées par le Ministère de la Culture enregistrent explicitement des formes comme *déboguer* et *débogage* comme équivalents français établis de *to debug*, illustrant la normalisation de ces formes adaptées dans le domaine des technologies de l'information. D'autres variantes existent comme *debuguer* et *debuger*, mais elles sont moins fréquentes. Une telle fluctuation orthographique peut être interprétée comme une phase de transition dans l'intégration lexicale. Les utilisateurs négocient entre la proximité de la forme source anglaise et la conformité aux conventions orthographiques françaises. Ce phénomène corrobore l'observation de Janssen (2006 : 168-169) selon laquelle l'instabilité graphique est une caractéristique structurelle des processus d'emprunt plutôt qu'une erreur marginale.

Les dérivés de noms (verbes issus de marques) ont été mentionnés comme cas particuliers d'intégration, mais ils n'ont pas été intégrés aux statistiques générales afin de ne pas confondre dénomination et emprunt lexical.

3.2.3. Emprunts sémantiques et calques

Les emprunts sémantiques et les calques sont moins fréquents, mais ils sont linguistiquement significatifs. Comme par exemple, de *support* utilisé avec le sens anglais d'*assistance/service* plutôt que de support physique. Dans le discours médiatique, les anglicismes peuvent être attestés « avec un sens différent », ce qui signale des ajustements sémantiques et une possible concurrence de valeurs au sein d'une même forme (Biedermann-Pasques & Humbley 1995 : 57)

Les calques sémantiques tels que *revue de code* ou *livraison* continuent de traduire des expressions anglaises tout en respectant pleinement les structures syntaxiques françaises. Dans la terminologie technoscientifique, le calque, notamment le calque morphologique, est souvent présenté comme un procédé néologique pertinent et avantageux pour la création de termes français, tout en conservant une transparence formelle (Di Spaldro, Auger & Ladouceur 2010). En revanche, les calques syntaxiques

comme *faire sens*, calqués sur *to make sense*, reproduisent une construction anglaise qui entre directement en concurrence avec une expression française existante (*avoir du sens*).

3.3. Le cas particulier de *dev*

Parce que *dev* est à la fois extrêmement fréquent et structurellement ambigu (développeur / développement), je le traite séparément de la liste générale des fréquences d'anglicismes.

Comme cette forme correspond également en français au mots *développeur*, *développement*, etc, et que la troncation comme moyen d'abréviation est en général beaucoup utilisée en français, l'on pourrait penser qu'il s'agirait d'une forme autochtone française. Les dictionnaires anglais, cependant, mentionnent explicitement *dev* comme abréviation de *development* et/ou *developer* (Collins Dictionary). Au-delà, par exemple, des ressources éducatives/professionnelles estoniennes présentent explicitement *dev* comme un raccourci de *development* (« Dev on lühend sõnast Development ») (Tallinna Polütehnikum), et la documentation informatique d'organismes publics estoniens explique de manière similaire Dev dans des composés comme DevOps au sens de « développement logiciel » (Riigi infosüsteemi amet 2021 : 8). Cela rend plausible l'analyse de *dev* non comme une abréviation strictement interne à une langue, mais comme un raccourci professionnel partagé, que les francophones adoptent puis intègrent au fonctionnement morphosyntaxique du français.

Cette forme fonctionne donc en français principalement comme une abréviation technolocale dont le sens reste ancré dans le domaine du développement logiciel, mais dont la valeur précise se détermine selon le contexte. En pratique, *dev* est polyfonctionnel : il peut renvoyer à une personne (un développeur / des développeurs) comme à l'activité (le développement). Cette double valeur est cohérente avec la manière dont les sources lexicographiques en anglais décrivent *dev* comme pouvant renvoyer à la fois à *developer* et à *development*, y compris dans des emplois adjectivaux tels que « *dev team* ».

Regroupée à travers ses principales variantes écrites (*dev*, *devs*, *Dev*, *DEV*, *dév*, *dévs*), les formes de la famille *dev* totalisent 4355 occurrences et apparaissent dans 3010 commentaires ($\approx 13,6\%$ de tous les commentaires). En comparaison, *développeur(s)* totalise 816 occurrences, et *développement/developpement* totalise 289 occurrences dans le même corpus, ce qui signifie que *dev* est environ 5,34x plus fréquent que *développeur(s)* dans ce corpus.

Le pluriel révèle une concurrence entre un pluriel marqué en -s et un pluriel invariable de type abréviation. Avec des déterminants pluriels, les formes marquées (*les/des devs*) totalisent 656 instances, tandis que les invariantes (*les/des dev*) en totalisent 204. Ce profil est compatible avec l'idée que l'intégration peut être avancée tout en permettant une micro-variation des formes courtes ou des items technolèctaux (Ashrafova 2024 : 108).

Le genre est majoritairement masculin dans des contextes déterminants (*un dev* 409, *le dev* 375, *une dev* 3, *la dev* 3). Pour l'analyse, l'enjeu principal est que *dev* n'est pas syntaxiquement « étranger » : il entre dans les systèmes d'accords français et montre une affectation par défaut stable, qui est l'un des diagnostics clés de l'intégration grammaticale dans les emprunts (Ashrafova 2024 : 105).

Le corpus montre *dev* comme tête nominale dans le schéma dominant N + spécification postposée, par exemple *dev senior/junior/web/backend/fullstack*. Les fréquences relatives indiquent que l'ordre compatible avec le français est majoritaire :

- développeur senior (68) vs senior développeur (15)
- développeur web (92) vs web développeur (5)
- développeur junior (30) vs junior développeur (4)

Cette accommodation au schéma nominal du français correspond à l'idée que les emprunts tendent à s'insérer dans les cadres syntaxiques de la langue réceptrice plutôt qu'à maintenir systématiquement l'ordre de la langue source (Ashrafova 2024 : 109).

Enfin, *dev* n'est pas limité à la référence à une personne : le corpus inclut aussi *en dev* (116), supportant une lecture secondaire plus proche de « en développement ».

3.4. Concurrence terminologique et stabilisation des formes

Dans le corpus, certains échanges rendent visibles des processus de concurrence terminologique : les locuteurs comparent des formes anglaises et des équivalents français (traductions, calques, formes adaptées), discutent leur précision et leur acceptabilité technique, et signalent parfois des usages jugés stabilisés ou, au contraire, variables.

Trois positions émergent du corpus : les anti-anglicismes, les pro-anglicismes et l'acceptation mixte ou conditionnelle. Les exemples, cherchés au moyen des mots-clés *anglicisme(s)*, *franglais*, *anglais / en anglais*, *français*, *équivalent(s)*, *traduction/traduire*, *terme(s)*, proviennent directement du corpus analysé. Par ailleurs, la structure de Reddit facilite l'observation de ces positionnements : tout utilisateur peut commenter, et les réponses s'organisent en fils (« threads ») hiérarchisés. Cela rend les discussions publiques facilement accessibles.

3.4.1. Position anti-anglicismes

Dans ce type de commentaires, les locuteurs proposent ou valorisent des équivalents français, en avançant des critères de lisibilité, de cohérence terminologique ou de conformité aux recommandations. Principalement, les locuteurs se positionnent comme défendant le français.

- Instruction explicite :
 - « Est-ce que tu pourrais réécrire ton poste sans anglicisme ? »
- Plainte de mauvaise lisibilité :
 - « ...ne pas aimer lire des phrases à moitié en anglais et à moitié en français. »
- Norme sociale (« ça n'a pas sa place ici ») :
 - « c'est chelou de dire fuck dans un pays francophone ... »
- « Liste franglais » comme affichage négatif :
 - « 'Thread', 'CTO', 'networking', 'calls', 'founders'... Non vraiment les anglicismes ça devient insupportable. »
- Équivalents français proposés comme solution évidente :
 - « Manager -> Responsable »

Cette position anti s'inscrit dans la logique de l'« enrichissement » du français : en particulier, le *Vocabulaire des techniques de l'information et de la communication* publié dans le cadre du dispositif d'enrichissement du Ministère de la Culture présente le travail terminologique comme produisant un lexique « de qualité » qui suit les règles de formation des mots français et vise des formes « facilement compréhensibles » et utilisables dans des secteurs spécialisés (Ministère de la Culture 2017). La position anti du corpus peut donc être présentée comme s'alignant sur une idéologie existante de « l'adéquation du français » dans les domaines techniques modernes : si le français peut nommer le concept, l'anglais est présenté comme évitable.

Cette position anti est également ancrée historiquement et sociolinguistiquement. Les travaux sur la réception des mots anglais dans le discours public français (notamment la presse) montrent que l'intégration de ces formes fait l'objet d'ajustements et de débats pratiques, en particulier autour de leur attestation lexicographique et de leur variation graphique/orthographique, ce qui conduit à formuler des principes d'harmonisation pour stabiliser leur écriture et leur insertion dans l'usage (Biedermann-Pasques & Humbley 1995 : 57). Les commentaires normatifs du corpus peuvent donc être interprétés non comme des préférences individuelles isolées, mais comme l'expression locale en ligne d'une tradition plus ancienne d'évaluation des emprunts à l'anglais. De manière connexe, Elchacar documente les anglicismes comme un site durable de préoccupation normative dans les commentaires linguistiques francophones (dans son cas, le Québec), renforçant l'idée que les « anglicismes » fonctionnent fréquemment comme un déclencheur de discours normatif plutôt que comme une simple catégorisation descriptive (Elchacar 2022 : 1, 13).

Cette position « anti-anglicismes » ne doit cependant pas être considérée comme une défense du français standard ni comme un rejet des pratiques non standard. L'un des exemples ci-dessus montre que les locuteurs peuvent utiliser le verlan (*chelou* - verlan de *louche* (*bizarre, étrange*')) (Wiktionnaire, « chelou ») ou le français familier tout en qualifiant les emprunts à l'anglais de socialement inacceptables : le commentaire « *c'est chelou de dire fuck dans un pays francophone ...* » recourt à une ressource familière française (verlan : *chelou*) sans doute notamment afin de créer un contraste avec l'emprunt anglais (*fuck*) présenté comme déplacé dans un contexte francophone.

Enfin, les ressources terminologiques institutionnelles comme FranceTerme, Académie française ou Ministère de la culture opérationnalisent concrètement la logique « éviter la forme anglaise » en associant les termes français recommandés à des équivalents étrangers explicitement marqués (par exemple, l'entrée *bogue* de FranceTerme répertorie *bug* comme équivalent anglais). Ce type de structuration soutient les passages du corpus qui sont d'avis d' « utiliser le terme français » comme une norme actionnable.

3.4.2. Position pro-anglicismes

Dans ce type de commentaires, la forme anglaise est justifiée par sa stabilité d'usage, sa présence dans la documentation, servant comme les outils, ou plus précisément comme vocabulaire qui divise les concepts techniques en distinctions plus fines et renvoie plus directement à un sens spécifique.

- Argument de précision :
 - « Les termes en informatique c'est précis, ça renvoie à un sens précis. »
- L'écosystème est d'abord en anglais :
 - « Il y a-t-il une seule techno dont la doc soit majoritairement francophone ? ... quel serait l'intérêt ? »
- L'anglais comme condition normale du travail :
 - « le taff est en anglais dans les bureaux français ? »
- Le pro-anglais présenté comme évitant l'isolement :
 - « ...un séparatisme francophone... ne fera que faire grossir ce handicap. »

Bien que cette position puisse sembler contredire les objectifs institutionnels d' « enrichissement du français », les sources institutionnelles (Ministère de la Culture 2017 ; FranceTerme) aident en fait à expliquer pourquoi des arguments pro-anglais surgissent : de nombreux concepts techniques circulent avec une étiquette anglaise comme référence de base, et même les ressources officielles doivent tenir compte de ce fait. Elles enregistrent aussi systématiquement la contrepartie anglaise comme « équivalent étranger », ce qui soutient indirectement la logique du corpus selon laquelle l'anglais reste l'étiquette de référence que les usagers doivent comprendre. Un point

essentiel est que les publications terminologiques du Ministère de la Culture et FranceTerme enregistrent systématiquement l'équivalent étranger à côté de la recommandation française. Par exemple, *bogue* y est répertorié avec l'équivalent anglais *bug* et relié à une famille de dérivés (*débogage*, *déboguer*, *débogueur*).

Du point de vue linguistique, ces commentaires fournissent surtout des indices sur la stabilisation terminologique : ils permettent d'illustrer quelles formes sont perçues comme standard et comment s'articule la concurrence entre formes anglaises, calques et formes adaptées. En d'autres termes, la position pro présente à plusieurs reprises l'anglais comme le chemin le plus court vers la précision et l'interopérabilité. Cette logique apparaît en effet clairement dans un commentaire recensé du corpus qui décrit l'anglais comme l'environnement de travail le plus habituel même dans des échanges entre francophones dans certains domaines comme le jeu vidéo, où les termes anglais sont très spécialisés comme *spawn*, *respawn*, *headshot*, *ragequit* – « Tous les langages sont en anglais, il y a bien plus d'exposition en terme de réseaux sociaux, influenceurs devs anglophones que français », comme le décrit un commentateur.

3.4.3. Acceptation mixte

Ici, les locuteurs tracent une ligne médiane négociée, explicitement visible dans le corpus à travers des formulations comme : « A est acceptable, mais B ne l'est pas ».

- « 'Backend', 'Frontend' c'est acceptable puisqu'il n'y a pas vraiment d'équivalent. Pourquoi utiliser un terme anglais quand il y en a un français ? »
- « 'US', 'Singapore', 'Thread', 'DM', ça n'a pas de rapport avec la tech. Il y a des termes français qui existent, utiliser l'anglais, pour ça, n'a pas de sens »
- « Autant je suis pour défendre la langue française ... mais ... dans la tech ça n'a absolument aucun sens. Choisis mieux tes batailles »

Cette position mixte correspond précisément à ce que prédisent les chercheurs lorsque les emprunts sont traités comme socialement négociables plutôt que mécaniquement importés. La discussion de Beaud sur les « appropriations néologiques » présente l'emprunt à l'anglais non comme une importation passive, mais comme un processus actif d'adaptation et de négociation sociale, ce que réalise précisément la position mixte : elle classe certains emprunts comme intégrés/fonctionnels tout en interprétant

d'autres comme socialement marqués ou idéologiquement suspects (Beaud 2023 : 259-260).

L'analyse des anglicismes verbaux par Zayed (2021 : 115) fournit un appui à ce type de raisonnement, en modélisant l'emprunt comme une compétition sémantique, où le sens des verbes français à base anglaise (par exemple, *liker*, *zoomer*) est comparé à la fois au verbe source anglais et à d'éventuels concurrents français (*aimer*, *tricher*, etc.). Dans le corpus, l'intégration verbale est attestée par des verbes à base anglaise insérés dans des cadres syntaxiques français, par exemple *debugguer* (« Oui, on a 50000 lignes de COBOL des années 70 à debugguer. Tu viens ? ») ou *refactorer* (« ... tu te rendras vite compte que le refactorer est impossible et qu'il faut le faire vivre. »). Celles-ci montrent une verbalisation productive et contextualisée, compatible avec l'idée d'une concurrence sémantique entre formes empruntées et ressources françaises.

La position mixte gagne également en profondeur explicative lorsqu'elle est connectée aux perspectives du discours numérique. Les travaux de Paveau (2017) sur le discours numérique et la stabilisation lexicale des formes liées aux plateformes (par exemple, *tweeter/retweeter/twittosphère*) illustrent comment les environnements numériques produisent une innovation lexicale et une conventionnalisation rapide, qui peuvent normaliser rapidement certains emprunts tout en laissant d'autres contestés (Paveau, 2017). Un cas particulièrement transparent de dérivation à partir d'un nom propre apparaît également : « Poser une question sans googler avant c'est un bon anti-pattern », ce qui constitue un indice d'intégration morphologique du référent (*Google - googler*) et rejoint les observations sur la conventionnalisation rapide de formes liées aux environnements numériques. Cela aide à expliquer pourquoi les participants au corpus peuvent tolérer certaines étiquettes techniques bien installées tout en rejetant d'autres éléments anglais comme relevant d'un style optionnel.

Dans l'ensemble, le corpus suggère que les éléments présentant une intégration morpho-syntaxique plus forte (par exemple, *googler*, *debugguer*, *refactoriser*) sont plus faciles à conventionner, tandis que les éléments anglais moins intégrés sont plus souvent encadrés comme un style facultatif et donc contestés.

Conclusion

Ce mémoire de licence a examiné l'usage de la langue anglaise dans le corpus des développeurs francophones sur la plateforme de communication à grande échelle Reddit. La première recherche, qui combine des questions quantitatives, s'est intéressée à ce comment les anglicismes se répartissent entre le vocabulaire technique et l'usage socioprofessionnel de la langue à ce sujet. Dans le corpus, les anglicismes sont répandus et ne se limitent pas à la dénomination technique ; les termes sur les flux de travail et les infrastructures sont également denses avec un vocabulaire axé sur la carrière et le lieu de travail. Les anglicismes de cette communauté servent d'étiquettes aux outils, aux processus et, enfin, aux rôles et à l'ancienneté. Cela indique que l'influence de l'anglais structure la discussion sur l'identité professionnelle et les réalités du travail dans le secteur des technologies.

La deuxième question portait sur les mécanismes linguistiques de l'intégration morphologique. L'analyse quantitative révèle une tendance claire où la plupart des anglicismes circulent sous forme d'étiquettes toutes faites, tandis qu'une plus petite quantité est adaptée selon la morphologie française. Dans le corpus, les anglicismes bruts dominent massivement. Cependant, la minorité adaptée est cruciale : chaque fois que les locuteurs ont besoin qu'une forme se comporte comme celles du français, l'intégration devient productive, généralement à travers la morphologie verbale française et les modèles associés.

La troisième question était de savoir si la variation orthographique reflète une instabilité dans l'intégration ou de simples fautes de frappe. Les preuves indiquent le premier. Les mots comme *debugger*, *debouger* et *debuger* reflètent différentes stratégies – conserver une forme anglaise, simplifier ou redéfinir la forme. Plus généralement, cette variation se concentre précisément là où les locuteurs eux-mêmes semblent hésiter sur la manière de classer ces unités, les traitant tantôt comme une étiquette anglaise à conserver, tantôt comme un mot français à « normaliser » et à intégrer. Cela conforte l'interprétation selon laquelle la communauté n'est pas encore parvenue à un consensus sur une forme d'écriture acceptée.

La quatrième question était de savoir quels types d'emprunts existent et comment ils se combinent avec l'intégration. Dans l'ensemble, les types d'emprunt ne se font pas

concurrence dans un simple cadrage « anglais vs français ». Il forme un continuum où les besoins d'usage (comme précision, rapidité, références partagées) interagissent avec la grammaire française.

Le mémoire a ainsi atteint ses objectifs en offrant un compte rendu fondé sur des données empiriques analysées selon les questions de recherche définies. Il a relevé les anglicismes les plus significatifs dans un espace de développement francophone et a montré comment la morphologie française incorpore sélectivement les bases anglaises, en prêtant aussi attention aux variations orthographiques et les positionnements idéologiques qui façonnent les choix lexicaux. L'utilité de cette recherche est d'avoir analysé une série importante des cas d'usage où règne la pragmatique de la vie quotidienne d'un domaine concret. Il fournit ainsi une vision réelle et empirique du changement linguistique du point de vue d'une linguistique spécialisée, montrant comment la pragmatique, structuration et innovation linguistiques coexistent dans ce domaine. En termes de terminologie et de pédagogie, il révèle à quel point les équivalents français sont fonctionnels et comment les mots ont un poids social ou idéologique. Il montre comment les désaccords sur les anglicismes traduisent des problèmes plus profonds, tels que la précision, l'inclusion, l'accessibilité, la légitimité professionnelle, qui peuvent réorienter les pratiques de documentation, la formation et la modération communautaire.

Bibliographie

ACADÉMIE FRANÇAISE. 1694. « Préface de la première édition ». Le Dictionnaire : les neuf préfaces. En ligne : <https://www.academie-francaise.fr/le-dictionnaire-les-neuf-prefaces/preface-de-la-premiere-edition-1694>. Consulté le 11 janvier 2026.

ACADÉMIE FRANÇAISE. s.d. « Les missions ». En ligne : <https://www.academie-francaise.fr/institution/les-missions>. Consulté le 11 janvier 2026.

ACADÉMIE FRANÇAISE. s.d. « Néologismes & anglicismes », *Dire, ne pas dire*. En ligne : <https://www.academie-francaise.fr/dire-ne-pas-dire/neologismes-anglicismes>. Consulté le 11 janvier 2026.

ASHRAFOVA, I. 2024. « The Syntactic Adaptation of English Borrowings in Modern French », *Eurasian Geographic and Linguistic Journal of Languages and Education (EGJLLE)*, 1(1).

BEAUD, L. 2023. « Emprunts à l'anglais et appropriations néologiques. Note de lecture », *Vocabulaire de l'éducation et de la recherche* (Ministère de la Culture, 2022).

BIEDERMANN-PASQUES, L., HUMBLEY, J. 1995. « La réception de mots anglais dans les journaux français : l'application de quelques principes d'harmonisation graphique », *Langue française* 108, p. 57–65.

BOURDIEU, P. 1982. *Ce que parler veut dire : l'économie des échanges linguistiques*. Paris : Fayard.

CARTIER, E. 2019. « Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille », dans KACPRZAK, A., MUDROCHOVÁ, R., SABLAYROLLES, J.-F. (dir.), *L'emprunt en question(s). Conceptions, réceptions, traitements lexicographiques*. Limoges : Lambert-Lucas, p. 145–186.

COLLINS DICTIONARY. s.d. « dev », *Collins English Dictionary*. En ligne : <https://www.collinsdictionary.com/dictionary/english/dev>. Consulté le 11 janvier 2026.

DEPECKER, L. 2005. « Contribution de la terminologie à la linguistique », *Langages* 157, p. 6–13.

DI SPALDRO, J. ; AUGER, P. ; LADOUCEUR, J. 2010. « *Le calque technoscientifique : un procédé néologique avantageux pour la terminologie française ?* », *Neologica*, 4, p. 163–184.

ELCHACAR, M. 2022. « Toujours un problème, les anglicismes ? Les commentaires linguistiques sur les anglicismes dans les chroniques de langage québécoises contemporaines », *Actes du congrès annuel de l'Association canadienne de linguistique (ACL/CLA)*.

ÉTIEMBLE, R. 1964. *Parlez-vous français ?* Paris : Gallimard.

FRANCETERME. s.d. « bogue », entrée terminologique. En ligne : <https://www.culture.fr/franceterme/terme/INFO89>. Consulté le 11 janvier 2026.

GUILBERT, L. 1975. *La créativité lexicale*. Paris : Larousse.

HAUGEN, E. 1950. « The Analysis of Linguistic Borrowing », *Language* 26(2), p. 210–231.

JANSSEN, M. 2006. « Orthographic Variation in Lexical Databases », in *Proceedings of the XII EURALEX International Congress (EURALEX 2006)*, p. 167–172.

LAROUSSE. s.d. « bosser ». *Dictionnaire de français Larousse*. En ligne : <https://www.larousse.fr/dictionnaires/francais/bosser/10311>. Consulté le 12 janvier 2026.

LOI n° 94-665 du 4 août 1994 relative à l'emploi de la langue française (dite « loi Toubon »). 1994. En ligne : <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000005616341>. Consulté le 11 janvier 2026.

MARTIN, É. 2019. « Français : la marque de fabrique de la presse à scandale française », *Corela* 17-1. DOI : 10.4000/corela.7943.

MINISTÈRE DE LA CULTURE. 2017. *Vocabulaire des techniques de l'information et de la communication*. (PDF). En ligne : <https://www.culture.gouv.fr/content/download/169264/1885116>. Consulté le 11 janvier 2026.

OPENAI. 2022. ChatGPT, grand modèle de langage. En ligne : <https://chat.openai.com/>. Consulté le 16 janvier 2026.

PAVEAU, M.-A. 2017. *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*. Paris : Hermann.

PRAW. s.d. « Quick Start ». PRAW documentation. En ligne : https://praw.readthedocs.io/en/stable/getting_started/quick_start.html. Consulté le 15 octobre 2025.

PRESTON, D. R. 2008. « *Qu'est-ce que la linguistique populaire ? Une question d'importance* ». *Pratiques*, 139-140. DOI : <https://doi.org/10.4000/pratiques.1176>.

RIIGI INFOSÜSTEEMI AMET. 2021. « Projekti « Ettevõtjate ühtse kontaktpunkti infotehnoloogilise arhitektuuri analüüs lõpparuanne » (PDF), 23.07.2021. En ligne : <https://www.ria.ee/sites/default/files/documents/2022-11/Ettevotjate-uhitse-kontaktpunkti-infotehnoloogilise-arhitektuuri-analuus-23.07.2021.pdf>. Consulté le 12 janvier 2026.

ROUSSEAU, L.-J. 2005. « Terminologie et aménagement des langues », *Langages* 157, p. 93-102.

SABLAYROLLES, J.-F. ; PRUVOST, J. 2016. *Les néologismes*. Paris : Presses Universitaires de France, coll. « Que sais-je ? ».

SABLAYROLLES, J.-F. 2019. *Comprendre la néologie. Conceptions, analyses, emplois*. Limoges : Lambert-Lucas.

TALLINNA POLÜTEHNIKUM. s.d. « Tarkvaraarendaja ». En ligne : <https://tptlive.ee/et/tarkvaraarendaja>. Consulté le 12 janvier 2026.

WALTER, H. 1997. *L'aventure des mots français venus d'ailleurs*. Paris : Robert Laffont.

WIKTIONNAIRE. s.d. *chelou*. Dans *Wiktionnaire, le dictionnaire libre*. En ligne: [https://fr.wiktionary.org/wiki/chelou#:~:text=\(verlan\)%20\(fin%20XX%20e%20si%C3%A8cle\)%20Verlan%20de,de%20bizarre%2C%20par%20inversion%20des%20d eux%20syllabes](https://fr.wiktionary.org/wiki/chelou#:~:text=(verlan)%20(fin%20XX%20e%20si%C3%A8cle)%20Verlan%20de,de%20bizarre%2C%20par%20inversion%20des%20d eux%20syllabes). Consulté le 12 janvier 2026.

ZANOLA, M. 2008. « De quelques raisons du français », *Synergies Italie* 4.

ZAYED, A. 2021. « Les emprunts verbaux du français à l'anglais : une analyse sémantique. » *Travaux de linguistique*, 83, 115–145.

Resüme

“Anglitsismid frankofoonsete arendajate diskursuses: kasutusala, morfoloogiline kohandamine ja ideoloogilised seisukohad.”

Käesolevas bakalaureusetöös on analüüsitud inglise laenude kasutust frankofoonsete arendajate veebidiskursuses. Empiiriliseks aluseks on avalikud arutelud Redditi kogukonnas *r/developpeurs*, mis koondab peamiselt prantsuskeelseid IT-valdkonna kasutajaid. Korpus koostati Pythoniga ning see hõlmab kommentaare ajavahemikust 18.10.2024-15.10.2025.

Antud töö eesmärgiks oli kirjeldada, kuidas inglise päritolu keeleüksused ja -vormid jaotuvad tehnilise terminoloogia ning sotsiaal-professionaalse sõnavara vahel, millised on domineerivad morfoloogilise kohandamise mehhanismid ning kuidas ortograafiline varieeruvus viitab vormide stabiliseerumisele või vastupidi. Samuti on käsitletud, kuidas kasutajad võrdlevad ingliskeelseid vorme ja prantsuskeelseid vasteid, vaagivad nende täpsust ning arutavad, millised nimetused on “normaalsed”, “standardiseerunud” või vastupidi.

Tulemused näitavad, et anglitsismid esinevad korpuses regulaarselt ning ei piirdu üksnes kitsalt tehniliste mõistetega: sagedased on ka töökorralduse ja karjääriga seotud nimetused. Valdav osa anglitsismidest on kohandamata vormid, kuid nähtav on ka inglise keelest laenatud sõnade produktiivne kohandamine prantsuse keele morfoloogiasse. Ortograafiline ja vormiline varieeruvus osutab, et osa üksusi on kasutuses veel ebastabiilsed, paljud aga toimivad vabalt aktiivses kasutuses.

Aruteludest eristub kolm hoiakut: anglitsismidele vastandumine, anglitsismide pooldamine ning segatud/tingimuslik aktsepteerimine. Töö järeldus on, et *r/developpeurs*'i arutelud kajastavad korraga nii tehnilise suhtluse praktilisi vajadusi kui ka keeleideoloogilist normiloomet, kus inglise ja prantsuse vorme pidevalt läbi räägitakse.

Annexes

Annexe 1. Fonctionnement de Reddit

Reddit est une plateforme de discussion organisée en communautés thématiques appelées *subreddits* (préfixe *r/*). Chaque subreddit regroupe des publications et des fils de commentaires centrés sur un thème, avec des règles de modération propres à la communauté.

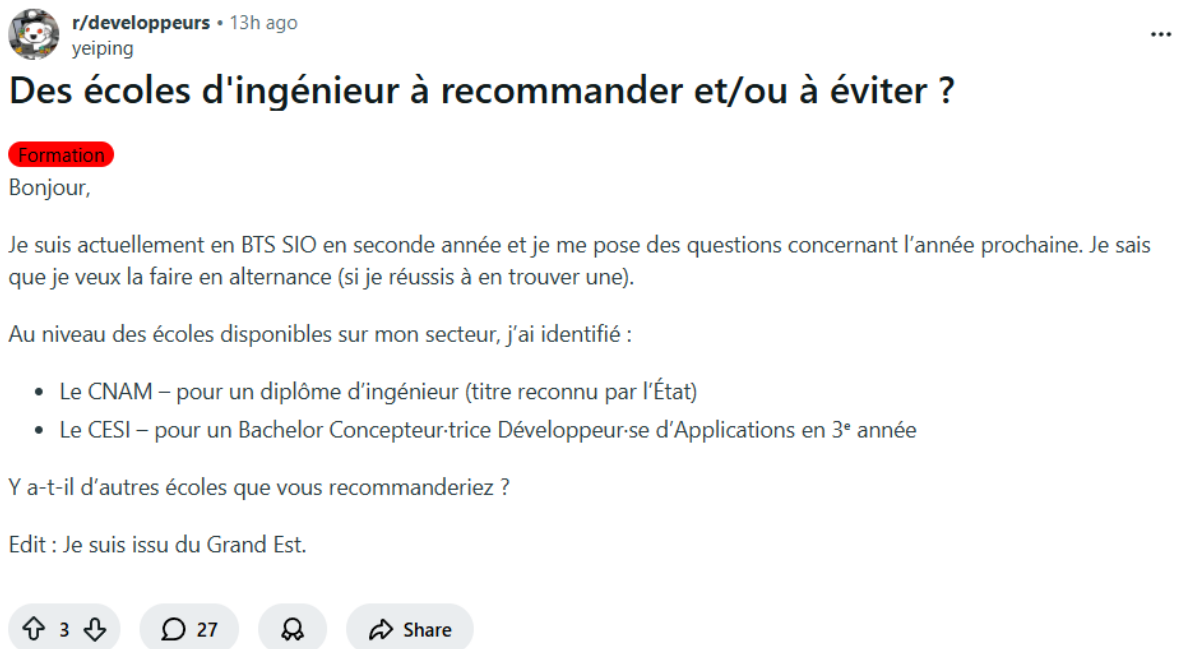


Figure 3. Exemple d'un post dans *r/developpeurs*

Figure 3 montre l'interface d'un post dans le subreddit *r/developpeurs*. En haut figurent le nom de la communauté, le temps écoulé depuis la publication et le pseudonyme de l'auteur. Le post comporte un titre (ici : « Des écoles d'ingénieur à recommander et/ou à éviter ? ») et un contenu textuel. En bas du post, on observe les principales interactions : le vote positif/ négatif (flèches haut/bas) qui attribue un score au post, le nombre de commentaires (icône bulle, ici 27), des actions de diffusion comme Share. Les votes influencent la mise en avant des contenus à l'intérieur du subreddit.



Figure 4. Exemple d'un fil de commentaires Reddit

Figure 4 illustre le fonctionnement des commentaires, organisés sous forme de fil arborescent. Chaque commentaire affiche un pseudonyme, un temps de publication, un score et des actions : Reply (répondre), Award (attribuer une récompense) et Share. Les réponses à un commentaire sont visuellement indentées afin de montrer la structure conversationnelle (qui répond à qui). Certains comptes peuvent porter des badges, signalant un statut ou une reconnaissance au sein de la communauté.

Annexe 2. Configuration PRAW et script Python de collecte du corpus

Cette annexe documente la configuration PRAW et le script Python utilisé pour extraire les publications et commentaires. Le script a été compilé le 15 octobre 2025, une nouvelle vérification pourrait donc entraîner l'ajout de données.

2.1. Client PRAW

Exemple d'un client PRAW. Les valeurs remplacées par des espaces réservés.

```
reddit = praw.Reddit(  
    client_id="X",  
    client_secret="Y",  
    user_agent="Z",  
)
```

2.2. Script Python

Script Python complet utilisé pour la collecte (développé par l'auteur à partir des exemples PRAW).

```
import os  
import time  
import praw  
import pandas as pd  
  
def get_reddit():  
    return praw.Reddit(  
        client_id="X",  
        client_secret="Y",  
        user_agent="Z",  
    )  
  
def save_partial(rows, path):
```

```

if rows:
    pd.DataFrame(rows).to_csv(path, index=False)

def main():
    subreddit_name = "developpeurs"
    time_filter = "year"
    scan_limit = 1000
    keep_threads = 500
    sleep_secs = 3
    checkpoint_every = 50
    final_csv = "r_developpeurs_corpus.csv"
    partial_csv = "r_developpeurs_partial.csv"

    reddit = get_reddit()
    reddit.read_only = True

    subreddit = reddit.subreddit(subreddit_name)
    posts = list(subreddit.top(time_filter=time_filter,
limit=scan_limit)[:keep_threads])

    rows = []

    for i, submission in enumerate(posts, start=1):
        try:
            submission.comments.replace_more(limit=0)
            for comment in submission.comments.list():
                rows.append({
                    "thread_id": submission.id,
                    "thread_title": submission.title,
                    "thread_permalink":
f"https://www.reddit.com{submission.permalink}",
                    "thread_num_comments": submission.num_comments,
                    "thread_score": submission.score,

```

```

        "thread_flair": submission.link_flair_text,
        "comment": comment.body,
        "comment_author": str(comment.author),
        "comment_score": comment.score,
        "comment_created_utc": pd.to_datetime(comment.created_utc,
unit="s", utc=True),
    })
except Exception:
    pass

if i % checkpoint_every == 0:
    save_partial(rows, partial_csv)

time.sleep(sleep_secs)

df = pd.DataFrame(rows)
if not df.empty:
    df = df[df["comment"].notna() & ~df["comment"].isin(["[deleted]",
"[removed]"])]

df.to_csv(final_csv, index=False)
save_partial(rows, partial_csv)

if __name__ == "__main__":
    main()

```

Annexe 3. Exemple d'utilisation de ChatGPT

J'ai utilisé ChatGPT pour la correction de textes lors de la préparation de ce travail.

Les requêtes suivantes ont été saisies dans le chatbot : « Make the text more coherent and academic :

Dans ce corpus, dev n'est pas simplement un autre mot d'apparence anglaise. C'est une forme abrégée très courante utilisée par les gens de la technologie, et elle peut être comprise de deux manières :

- cela peut venir de l'anglais comme développeur ou développement
- ou il peut s'agir d'une forme abrégée française de développeur / développement

Donc dev est un cas limite, ça a l'air anglais, mais ça fonctionne aussi comme un mot normal en français.

Ceci est important pour la méthode car dev apparaît si souvent qu'il peut masquer d'autres mots dans les résultats de fréquence. C'est pourquoi je parle de dev séparément et l'utilise principalement pour montrer comment une forme fait partie du français (orthographe, pluriel, genre et comment elle s'intègre dans les phrases), au lieu de la traiter comme un anglicisme normal tel que *framework* ou *backend*. »

Le résultat obtenu était : « Dans ce corpus, **dev** n'est pas simplement un autre élément fréquent qui a l'air anglais : c'est une abréviation technolocale à très haute fréquence dont le statut est structurellement ambigu. Il peut être interprété soit comme (i) une abréviation d'origine anglaise (**dev** < *developer/development*) circulant à l'échelle internationale dans les communautés techniques, soit comme (ii) une troncation française de *développeur / développement* qui se trouve converger avec l'abréviation anglaise. **Dev** est donc mieux traité comme une unité liminaire : il se comporte comme un élément lexicalisé en français, mais il n'est pas aisément classable comme un « anglicisme normal » au même titre que *framework* ou *backend*.

Cela a une importance méthodologique, car **dev** est si fréquent qu'il peut masquer la distribution d'autres formes s'il est conservé dans les classements généraux de fréquence. C'est pourquoi la présente étude présente **dev** dans une sous-section dédiée et l'utilise principalement comme un indicateur de schémas d'intégration

(orthographe, genre/nombre et comportement morphosyntaxique) plutôt que comme un simple token dans la liste des fréquences d'anglicismes ».

J'ai modifié le résultat comme suit : Parce que *dev* est à la fois extrêmement fréquent et structurellement ambigu (développeur / développement), je le traite séparément de la liste générale des fréquences d'anglicismes.

Comme cette forme correspond également en français au mots *développeur*, *développement*, etc, et que la troncation comme moyen d'abréviation est en général beaucoup utilisée en français, l'on pourrait penser qu'il s'agirait d'une forme autochtone française. [...] Cela rend plausible l'analyse de *dev* non comme une abréviation strictement interne à une langue, mais comme un raccourci professionnel partagé, que les francophones adoptent puis intègrent au fonctionnement morphosyntaxique du français.

Cette forme fonctionne donc en français principalement comme une abréviation technolocale dont le sens reste ancré dans le domaine du développement logiciel, mais dont la valeur précise se détermine selon le contexte. En pratique, *dev* est polyfonctionnel : il peut renvoyer à une personne (un développeur / des développeurs) comme à l'activité (le développement). Cette double valeur est cohérente avec la manière dont les sources lexicographiques en anglais décrivent *dev* comme pouvant renvoyer à la fois à *develop* et à *development*, y compris dans des emplois adjectivaux tels que « *dev team* ».

Pour ne pas rendre mon texte inutilement lourd ou trop artificiellement académique, je n'ai pas utilisé toutes les propositions de Chat GPT. En plus, pour lier le passage à ce qui précédait et suivait, j'ai reformulé encore le texte.

Annexe 4. Le tableau chiffré des anglicismes relevés

Le tableau 1 ci-dessous complète le graphique présenté dans la partie 3.1.2 en Figure 2 en détaillant le nombre étendu d'éléments analysés. Il indique le nombre exact de fréquences pour chaque forme attestée, leur présentation étant triées par ordre décroissant afin que les termes les plus fréquents apparaissent en premier.

Pour garder le tableau lisible et éviter une liste trop longue, seuls les termes avec une fréquence supérieure à 100 sont inclus.

Phrase et variants	Fréquence
dev, devs, dév, dévs	4355
client, clients	1372
test, tests, tester, testing	1129
junior, juniors	1082
senior, seniors	689
job, jobs	650
freelance, freelances	419
web, webs	395
lead, leads, leader	379
ok, okay	374
stack, stacks	317
remote, remotes	258
ci, CI (continuous integration)	250
startup, startups	235
business	193
feature, features	191
framework, frameworks	184
bug, bugs, buger	182
devops	182

daily, dailys	172
fullstack, full-stack	163
manager (nom), managers	161
tickets, ticket, ticketing	156
full remote	155
api, apis, API, APIs	144
cloud, clouds	139
backend, back-end, backends, back-ends	132
script, scripts, scripter, scripting	120
docker	115
open source	112
commit, commits, commiter	108
support	106
boss	102
pr, pull request	100

Tableau 1. Les anglicismes les plus fréquents dans le corpus.

Abstract

Ce mémoire examine l'usage des anglicismes dans le discours de développeurs francophones sur la base d'un ensemble de commentaires publiés sur Reddit (r/developpeurs). Les anglicismes classés de la terminologie technique et du vocabulaire socioprofessionnel ont été étudiés sur la base des mécanismes de leur intégration dans la langue française. Les résultats montrent une nette prédominance des anglicismes bruts, ce qui est cohérent avec le rôle central de l'anglais dans la documentation et la pratique professionnelle, mais aussi l'existence d'une intégration productive lorsque les formes doivent fonctionner comme des unités françaises. L'analyse de l'échange met en évidence trois positions idéologiques récurrentes (anti, pro et mixte), montrant que le choix des formes est non seulement terminologique, mais également socialement cohérent au sein de la communauté.

Märksõnad: prantsuse keel, anglitsismid, Reddit, digidiskursus, arvutikeel, korpuslingvistika.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Artjom Romanov,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Anglicismes dans le discours des développeurs francophones : usages, adaptation morphologique et positions idéologiques“, mille juhendaja on Marge Käsper, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Artjom Romanov

16.01.2026