CZU: 811.133.1`373.45=111 https://doi.org/10.59295/drgl2026.22

LE LEXIQUE DE L'IA EN FRANÇAIS : LE RETOUR DES ANGLICISMES

THE AI GLOSSARY IN FRENCH: THE RETURN OF ANGLICISMS

Felicia DUMAS
Habilitate Doctor, University Professor
Alexandru Ioan Cuza University of Iaşi, Romania
ORCID: https://orcid.org/0000-0001-5053-2337

Abstract: ***We offer a lexical analysis of the various words, phrases, and acronyms used in French to refer to components and aspects of AI (artificial intelligence) use. Most of them are Anglicisms, whose use has not really been stopped by the various ministerial terminology commissions, which are supposed to defend French as the national language. We will also analyze several discursive contexts, taken from the French press, in which these lexemes are used, in order to highlight their integration into the language. Cases of lexical derivation will also be studied, with the aim of highlighting the same process of lexico-semantic integration. We will analyze words such as chatbot, chatter, and overfitting; phrases such as feature engineering, AI governance, algorithmic audit, and digital sovereignty; and acronyms such as AI and GDPR. This brand-new, rich, and complex lexicon marks a new period of borrowing from English into French, with the massive return of Anglicisms into contemporary French.***

Keywords*: AI, Anglicisms, loanwords, lexical analysis, derivation, acronyms, lexical integration.*

Introduction

L'IA – l'intelligence artificielle – a fait irruption dans nos vies d'un seul coup et sans prévenir, avec toutes ses conséquences anthropologiques et lexicales, nous obligeant à nous remettre en question en tant qu'êtres humains.

Nous proposons par la suite une analyse lexicale des différents mots, syntagmes et sigles employés en langue française pour désigner ses composantes et différents aspects de son utilisation. La plupart d'entre eux sont des anglicismes, dont l'emploi n'a pas vraiment été stoppé par les différentes commissions ministérielles de terminologie, censées défendre le français en tant que langue nationale, de la République Française. Nous analyserons aussi plusieurs contextes discursifs, tirés de la presse française, d'emploi de ces lexèmes, afin de mettre en évidence leur intégration dans la langue. Des cas de dérivation lexicale seront aussi étudiés, dans le but de souligner le même processus, d'intégration lexico-sémantique. Des mots tels *chatbot, chatter, overfitting*, des syntagmes tels *feature engineering, gouvernance de l'IA, audit algorithmique, souveraineté numérique*, et des sigles, comme *IA* ou *RGPD* font partie de ce lexique, et feront l'objet de notre analyse. Il s'agit d'un lexique tout nouveau, riche et complexe, qui marque une nouvelle période d'emprunts faits par le français à l'anglais, le retour massif des anglicismes en langue française contemporaine.

Les anglicismes de l'IA en langue française

Définis de façon générale comme des emprunts à l'anglais intégrés en langue française, les anglicismes ont été toujours présents dans le lexique du français, surtout aux XXème et XXIème siècles et notamment dans les domaines de l'informatique et des sciences technologiques de pointe, mais pas seulement. En ce début du XXIème siècle, c'est le cas de l'IA, qui dépasse largement l'informatique, se prolongeant dans la vie de tous les jours.

Comme pour la grande majorité des emprunts, on peut remarquer le fait que la plupart des anglicismes sont des substantifs, car ceux-ci peuvent être francisés assez facilement à l'aide des déterminants propres à la langue française. Comme nous le disions ailleurs [1], le signe de la francisation, définie comme la réussite du processus d'intégration de ces emprunts dans la langue française, est l'adoption définitive des marques françaises à tous les niveaux linguistiques : phonétique, morpho-syntaxique et sémantico-lexical [2].

Le lexème *chatbot* est un nom masculin issu d'un mot-valise, composé de la fusion du signifiant complet du mot anglais : *chat* „bavardage", et de *bot*, l'aphérèse du mot anglais *robot*. Il désigne en langue française un « programme informatique basé sur l'intelligence artificielle, capable de répondre en temps réel aux questions d'un internaute, faisant ainsi office de conseiller virtuel. Recommandation officielle : *dialogueur*. Synonyme : agent conversationnel. Il est notamment utilisé pour la vente, le service après-vente et le marketing sur les sites de commerce et de services en ligne » [3]. C'est le « robot vivant » avec lequel on discute lorsqu'on veut parler à différents agents de distribution de services ou de commerce, en milieu virtuel. C'est un nom très employé en langue française de nos jours, comme le prouve le contexte suivant, tiré d'un texte numérique :

> « En 2025, les chatbots IA transforment les modes d'interaction entre les entreprises et les clients. L'IA générative a fait un bond en avant en révolutionnant les chatbots qui offrent désormais un support client de haut niveau via des canaux de communication de service alimentés par l'IA, ne nécessitant aucune interaction humaine. Cette nouvelle vague de chatbots IA gagne en popularité auprès des représentants du support client en particulier » [4].

Par exemple, le chatbot de la SNCF renseigne les voyageurs sur leurs horaires [5]. En Roumanie, il remplace les opérateurs téléphoniques.

Il est également employé dans des textes scientifiques, comme dans le contexte suivant, extrait d'un article de traductologie, où il prend la marque du pluriel français :

> « Depuis 2019-2020, plusieurs cours du programme ULiège intègrent progressivement la traduction automatique (TA), la post-édition (PE) et, plus récemment, les chatbots, non comme fin en soi, mais comme objet d'analyse et d'aide à la pratique » [6].

Le verbe *chatter* est dérivé, quant à lui de l'anglicisme *chat*, dont il reprend le noyau sémantique de base, avec le suffixe verbal français *-er* ; il signifie donc „discuter", en ligne notamment. Il est utilisé dans des contextes faisant référence à des rencontres amoureuses sur Internet. Il a donné naissance, par dérivation lexicale avec le suffixe agentif *-eur* au nom *chatteur*, utilisé dans le même type de contextes.

Le substantif *overfitting* est un anglicisme technique, employé dans les finances appelées prédictives et en statistique. C'est un « phénomène où le modèle „apprend par cœur" son jeu de données et échoue sur de nouvelles données. C'est un problème fréquent en finance prédictive » [7]. Son équivalent français est le nom « surapprentissage », utilisé

surtout en statistique, en alternance avec lui. Voici un contexte d'emploi de cet anglicisme extrait d'un texte numérique également :

> « À l'inverse de l'overfitting, l'underfitting survient quand les données d'entraînement sont hautement biaisées. Par exemple, si un problème est excessivement simplifié, le modèle ne fonctionne pas correctement sur les données d'entraînement. Il peut être causé par des données contenant du bruit ou des valeurs erronées. Le modèle ne sera donc pas capable de dégager de patterns à partir du jeu de données » [8].

Le syntagme *feature engineering* est un autre anglicisme technique qui désigne « un processus qui consiste à sélectionner, transformer et créer des variables pertinentes afin d'optimiser les performances d'un modèle d'apprentissage automatique. Par exemple, dans l'immobilier, créer un indicateur « distance au métro » améliore les prédictions de prix » [9]. Son équivalent français serait l'expression « ingénierie des caractéristiques ». Cet anglicisme est assez présent dans des textes numériques, hébergés par la toile, dont voici un exemple :

> « Le Feature Engineering est un processus qui consiste à transformer les données brutes en caractéristiques représentant plus précisément le problème sous-jacent au modèle prédictif. Pour faire simple, il s'agit d'appliquer une connaissance du domaine pour extraire des représentations analytiques à partir des données brutes et de les préparer pour le Machine Learning » [10].

Un autre syntagme d'origine anglaise employé dans le domaine de l'IA est celui de *Deep Learning*, emprunté en même temps qu'un autre, celui de *Machine Learning*. Les deux font référence à des modèles d'apprentissage ; le premier est défini comme une « méthode de l'IA utilisant des réseaux neuronaux artificiels pour reconnaître des modèles complexes » [11], et le deuxième comme « une composante de l'IA permettant aux ordinateurs d'apprendre sans programmation explicite » [12]. Voici un contexte numérique d'emploi des deux syntagmes, à visée d'initiation lexico-sémantique :

> « Le *deep learning* ou apprentissage profond est un type d'intelligence artificielle dérivé du *machine learning* (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées » [13].

Quelques syntagmes et expressions français sont utilisés aussi dans le domaine de l'IA, à côtés des anglicismes. L'une des expressions est celle de *gouvernance de l'IA,* en relation avec des pratiques et de l'éthique de son utilisation dans différents domaines de nos jours. Elle désigne le « cadre de politiques, de normes et de pratiques permettant de superviser, contrôler et encadrer l'usage de l'intelligence artificielle au sein d'une organisation. Les grandes banques déploient des comités de gouvernance IA » [14].

Dans le même domaine, on rencontre également deux autres syntagmes, que nous aimerions mentionner ici : *audit algorithmique* et *souveraineté numérique*. L'audit algorithmique est défini comme « un processus d'évaluation et de vérification des modèles de l'IA dans le but d'analyser leurs performances, leur conformité réglementaire et leur absence de biais. La CNIL (Commission nationale de l'informatique et des libertés) recommande son usage en France » [15]. Il est employé dans de nombreux textes virtuels, dont voici un exemple de contexte (numérique) :

> « L'audit algorithmique consiste à s'assurer qu'un algorithme d'apprentissage n'est pas biaisé. "*Comment ?*", me demandez-vous. Eh bien, l'une des façons de faire, c'est de créer des profils

fictifs de clients qu'un enquêteur envoie ensuite à l'algorithme qu'il souhaite tester. L'enquêteur récolte toutes les données de l'algorithme et en déduit les biais potentiels dans le processus de prise de décision. Est-ce que les femmes ont les mêmes offres d'emploi que les hommes ? » [16].

Le syntagme *souveraineté numérique* fait référence à la « capacité d'un État ou d'un ensemble régional, comme l'Union européenne, à contrôler et développer ses propres infrastructures technologiques et solutions de l'IA pour limiter sa dépendance extérieure » [17]. Il apparaît souvent dans des textes numériques, comme on peut le voir de l'exemple suivant :

> « La réflexion sur la souveraineté numérique naît d'une préoccupation : le refus de voir les peuples, les communautés d'utilisateurs, les États, les individus perdre le contrôle de leur destin au profit d'entités mal identifiées, non légitimes, et dont l'objectif n'est pas la promotion de l'intérêt général » [18].

Le lexique de l'IA comprend aussi de nombreux sigles en langue française, dont la plupart abrègent des signifiants d'origine anglaise, donc des anglicismes. C'est le cas des sigles tels *IA, RGPD, LLM* ou *GAN*. Le premier, qui traduit le sigle anglais *AI*, fait référence au domaine même, celui d'une forme d'intelligence non humaine, artificielle, robotique. Il s'agit d'un « domaine scientifique qui vise à créer des machines capables de simuler certaines fonctions cognitives humaines (raisonnement, vision, langage). Par exemple, les assistants virtuels comme Siri ou Alexa sont des applications concrètes d'IA (IBM) » [19]. Ce sigle est très fréquent dans les textes numériques, ainsi que dans la presse écrite, de nos jours :

> « L'intelligence artificielle (IA) a fait des avancées significatives au fil des années à tel point qu'elle est devenue omniprésente dans la vie de tous les jours, même parfois sans que vous en ayez conscience. Découvrez avec Futura comment l'IA intervient dans différents aspects de votre quotidien, en présentant 6 utilisations essentielles de cette technologie » [20].

> « Pourquoi la justice ne doit pas être soumise à l'IA ? Déjà sous-financée et cible de graves critiques, la justice risque demain d'être en partie automatisée, en s'appuyant sur l'intelligence artificielle. Un rêve que caressent les entrepreneurs du numérique et que dénonce Manuela Cadelli, juge à Namur, dans son ouvrage *Justice et IA : une enquête critique* » [21].

Le RGPD est un sigle français désignant « le règlement européen qui encadre l'utilisation des données personnelles afin de protéger la vie privée des individus et d'imposer des obligations de transparence et de sécurité » [22]. Il connaît une grande fréquence d'emploi dans la presse française contemporaine. Voici un titre tiré du journal *Le Monde* du 11 juillet 2025 : « Les caméras augmentées dans les bureaux de tabac pour évaluer l'âge des clients sont contraires au RGPD » [23].

LLM et *GAN* sont deux sigles d'origine anglaise, gardés tels quels par le français. Le premier est le résultat de la siglaison du syntagme *Large Language Model* (« Grands Modèles de langage »), qui veut dire « un modèle d'IA entraîné sur d'énormes quantités de textes pour comprendre et générer du langage humain » [24]. Le deuxième abrège la construction phrastique *Generative Adversarial Networks* et désigne « un système composé de deux réseaux neuronaux (générateur et discriminateur) travaillant en opposition pour créer du contenu réaliste » [25].

> « Les LLM impressionnent par leur puissance, mais les SLM gagnent du terrain grâce à leur agilité, leur rapidité et leur déploiement local. L'enjeu n'est pas de choisir, mais de combiner les deux pour répondre au mieux aux usages. Découvrez leurs différences en video. » [26],

lit-on sans aucune explication supplémentaire dans le même quotidien français, *Le Monde*.

Le *Generative Adversarial Networks* (*GAN*) est un anglicisme technique et spécialisé, qui se trouve en situation d'homophonie avec le nom d'un assureur français, plus largement connu par les locuteurs français moyens : Gan assurances : https://www.gan.fr/page-daccueil/particuliers/.

Pour conclure

Tous ces anglicismes cohabitent avec des noms, des sigles et des syntagmes français qui désignent différents aspects de l'IA (des plus généraux aux plus techniques et ultraspécialisés), dans l'usage et la mémoire des sujets parlants. Sous la pression des commissions ministérielles de terminologie, qui surveillent la pénétration excessive des mots d'origine anglaise dans le lexique du français (ou ce que la presse française appelle « la manie collective des anglicismes » [27]), certains d'entre eux sont utilisés aussi en simultanéité avec leurs correspondants français, proposés par ces commissions ; d'autres ont été complétement traduits, comme l'*IA*, par exemple (de *AI*), ou bien, d'autres encore résistent au niveau de l'usage, sous leur forme d'origine, tout en étant bien intégrés du point de vue lexical en langue française [28] (ni *dialogueur,* ni *agent conversationnel* n'ont pu remplacer l'anglicisme *chatbot*). Leur utilisation massive à l'époque actuelle, du développement fulgurant de l'IA, marque une nouvelle période dans l'histoire du lexique de la langue française (qui évoque un changement de paradigme), à savoir le retour massif des anglicismes, dans le sillage de leur présence importante déjà, dans le même domaine, de l'informatique, les dernières décennies du siècle dernier.

Références bibliographiques

1. DUMAS, Felicia (2008) – *Lexicologie française*, Iaşi, Casa editorială Demiurg, 210 p.
2. GUIRAUD, Pierre (1965) – *Les mots étrangers*, Paris, P.U.F., 123p.
3. *Larousse* en ligne, consulté à l'adresse URL suivante: www.larousse.fr.
4. https://www.liveagent.fr/blog/benefits-of-chatbots/?fh_id=a701213d6c5a4b6c#, consulté en ligne le 10 février 2026.
5. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 10 février 2026.
6. BADA, Valérie, SCHUMACHER, Perrine (2026) – « L'intégration de l'IA dans la formation des traducteur.ices: évolution sous double contraintes? Le cas de l'Université de Liège (Belgique) », *Al-Kimiya*, no 27, 2026, p. 28. Consulté en ligne le 30 mars 2026 : https://e-journals.usj.edu.lb/alkimiya/vol27/iss27/3/.
7. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 10 février 2026.
8. https://liora.io/overfitting-tout-savoir, consulté en ligne le 11 février 2026.
9. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 10 février 2026.

10. https://liora.io/feature-engineering, consulté en ligne le 11 février 2026.
11. https://www.roboto.fr/blog/52-termes-d-ia-generative-a-connaitre-en-2025-lexique-complet, consulté en ligne le 12 février 2026.
12. https://www.roboto.fr/blog/52-termes-d-ia-generative-a-connaitre-en-2025-lexique-complet, consulté en ligne le 12 février 2026.
13. https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262/, consulté en ligne le 12 février 2026.
14. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 12 février 2026.
15. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 12 février 2026.
16. https://project.inria.fr/matheseunesacreehistoire/audit-biais-algorithmiques/, consulté en ligne le 12 février 2026.
17. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 12 février 2026.
18. https://shs.cairn.info/magazine-cahiers-francais-2020-2-page-18?lang=fr, consulté en ligne le 12 février 2026.
19. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 12 février 2026.
20. https://www.futura-sciences.com/tech/questions-reponses/intelligence-artificielle-intelligence-artificielle-six-usages-revolutionnaires-quotidien-11341/, consulté en ligne le 13 février 2026.
21. https://www.lemonde.fr/idees/article/2026/02/09/pourquoi-la-justice-ne-doit-pas-etre-soumise-a-l-ia_6666045_3232.html?search-type=classic&ise_click_rank=1, consulté en ligne le 13 février 2026.
22. https://bigmedia.bpifrance.fr/nos-dossiers/lexique-ia-40-mots-pour-comprendre-lintelligence-artificielle, consulté en ligne le 12 février 2026.
23. https://www.lemonde.fr/pixels/article/2025/07/11/les-cameras-augmentees-dans-les-bureaux-de-tabac-sont-contraires-au-rgpd-juge-la-cnil_6620704_4408996.html?search-type=classic&ise_click_rank=4, consulté en ligne le 13 février 2026.
24. https://www.roboto.fr/blog/52-termes-d-ia-generative-a-connaitre-en-2025-lexique-complet, consulté en ligne le 13 février 2026.
25. https://www.roboto.fr/blog/52-termes-d-ia-generative-a-connaitre-en-2025-lexique-complet, consulté en ligne le 13 février 2026.
26. https://www.lemonde.fr/intelligences-numeriques/article/2025/12/08/ia-entre-petits-modeles-slm-et-grands-modeles-de-langage-llm_6656455_521.html?search-type=classic&ise_click_rank=1, consulté en ligne le 13 février 2026.

27. https://www.lexpress.fr/culture/lutte-contre-les-anglicismes-les-coups-de-maitre-des-commissions-de-terminologie-RX7KM64JLRBD7EJSJZP6JVHZGE/, consulté en ligne le 13 février 2026.
28. PERGNIER, Maurice (1989) – *Les anglicismes. Danger ou enrichissement pour la langue française?* Paris, P.U.F., 242 p.