

L'usage Vs la norme: l'IA au carrefour des dynamiques linguistiques contemporaines
Usage vs. Norm: AI at the Crossroads of Contemporary Linguistic Dynamics

Boudraa Randa*

Université Mohamed Cherif Messaadia, Souk-Ahras (Algérie), r.boudhraa@univ-soukahras.dz

Date de réception : 13/06/2025

Date d'acceptation : 30/06/2025

date de publication: 31/07/2025

Résumé

L'émergence de l'intelligence artificielle (IA) dans le traitement du langage écrit met en lumière les tensions croissantes entre l'usage linguistique courant et la norme académique, révélant ainsi la complexité du paysage linguistique contemporain. Les systèmes d'IA générative, tels que ChatGPT, reposent sur des corpus variés qui intègrent non seulement des registres formels, mais aussi des formes d'expression familières, des régionalismes et des néologismes. Cette interaction entre tradition et innovation soulève des questions fondamentales sur la pérennité de la norme linguistique à l'ère de l'intelligence artificielle. Cette étude propose une réflexion critique sur la manière dont cette dernière, en tant que technologie culturelle, reflète, perpétue ou transforme les tensions historiques entre prescription normative et créativité linguistique populaire. Elle invite à considérer ces systèmes non comme de simples outils neutres, mais comme des acteurs à part entière de l'écologie linguistique contemporaine, dont l'influence sur l'avenir de la langue française mérite une attention scientifique soutenue.

Mots-clés : IA-norme-usage-tensions linguistiques-écologie linguistique-français

Abstract

The emergence of artificial intelligence (AI) in the processing of written language highlights the growing tensions between everyday linguistic usage and academic norms, thereby revealing the complexity of the contemporary linguistic landscape. Generative AI systems, such as ChatGPT, rely on diverse corpora that encompass not only formal registers but also colloquial expressions, regionalisms, and neologisms. This interplay between tradition and innovation raises fundamental questions about the sustainability of linguistic norms in the age of artificial intelligence.

This study offers a critical reflection on how AI, as a cultural technology, reflects, perpetuates, or transforms the longstanding tensions between normative prescription and popular linguistic creativity. It suggests these systems should be regarded less as neutral tools and more as integral actors in contemporary linguistic ecology, whose impact on the language's future demands rigorous scientific inquiry.

Keywords: AI- Norm- Usage- Linguistic Tensions- Linguistic Ecology- French Language.

* *Auteur correspondant*

I- Introduction:

La langue française, comme toute langue vivante, se caractérise par une perpétuelle tension entre les usages spontanés des locuteurs et les normes prescriptives établies par les institutions linguistiques. Cette dialectique, séculaire dans son essence, trouve aujourd'hui un nouveau terrain d'expression et d'analyse à travers les technologies d'intelligence artificielle dédiées au traitement du langage naturel. Ces systèmes, loin d'être de simples outils techniques, se positionnent désormais comme des révélateurs privilégiés des dynamiques sociolinguistiques qui traversent et structurent l'espace francophone contemporain.

L'objectif principal de cette recherche est d'analyser comment les systèmes d'intelligence artificielle négocient les écarts entre le français standard, celui des grammaires et des dictionnaires, et le français vernaculaire dans ses multiples expressions sociales, géographiques et situationnelles. Cette négociation algorithmique, loin d'être neutre, met en lumière et parfois reconfigure les rapports de force entre différentes conceptions de la légitimité linguistique. En effet, comme le souligne Calvet (1999), « toute langue est le théâtre d'un conflit permanent entre des forces de variation et des forces d'unification », conflit que l'IA amplifie et transforme par ses mécanismes d'apprentissage et de production.

Notre étude s'articule autour de trois axes complémentaires. Premièrement, nous examinerons comment les corpus d'entraînement des modèles linguistiques, par leur composition même, instaurent une hiérarchie implicite entre les différentes variétés du français. Deuxièmement, nous analyserons les stratégies d'adaptation et de correction déployées par ces systèmes face aux formes non-standards, révélatrices d'une certaine conception normative du langage. Troisièmement, nous étudierons comment les interactions quotidiennes avec ces technologies influencent progressivement les pratiques linguistiques des locuteurs, créant potentiellement de nouvelles normes d'usage.

L'importance de cette recherche dépasse largement le cadre technologique pour s'inscrire dans une réflexion fondamentale sur l'évolution des systèmes linguistiques. À l'heure où les interactions médiatisées par l'IA se multiplient dans tous les domaines de la vie sociale, comprendre les mécanismes par lesquels ces technologies façonnent notre rapport à la langue devient essentiel. Comme l'affirment Cordeiro et Hovy (2021), les systèmes de NPL façonnent la langue qu'ils traitent voire la transforment : en sélectionnant, filtrant ou reproduisant certaines formes, ils influencent les usages et les normes linguistiques dans la pratique quotidienne. Cette étude contribue ainsi à éclairer les enjeux liés à l'accessibilité linguistique des technologies, tout en questionnant le rôle des institutions traditionnelles face à ces nouveaux acteurs de la normalisation linguistique.

Par ailleurs, notre recherche s'inscrit dans la continuité des travaux de Blanche-Benveniste (1997) sur l'oral et l'écrit, en y intégrant la dimension computationnelle qui renouvelle profondément cette distinction classique. Les frontières autrefois nettes entre ces modalités d'expression se trouvent aujourd'hui brouillées par des formes hybrides que les systèmes d'IA doivent appréhender. Ces derniers sont devenus des acteurs incontournables dans les dynamiques linguistiques contemporaines, révélant et transformant les tensions entre les usages populaires et les normes prescriptives. C'est

ainsi que ce débat, ancré dans une dualité historique entre usage et norme, trouve un nouvel écho à l'ère numérique.

Notre recherche s'articule autour d'une problématique centrale : *comment les systèmes d'intelligence artificielle, par leurs mécanismes d'apprentissage et leurs réponses, révèlent-ils et transforment-ils simultanément les tensions existantes entre les usages linguistiques populaires et les normes prescriptives du français ?*

Pour répondre à cette dernière, nous émettons les hypothèses suivantes :

-Les systèmes d'IA, en fonction de leurs corpus d'entraînement, reproduisent et parfois amplifient les biais normatifs existants dans la tradition linguistique française, favorisant certaines variétés au détriment d'autres.

-Les mécanismes de correction automatique et de génération textuelle des IA exercent une pression normative qui influence progressivement les pratiques linguistiques des utilisateurs.

-L'interaction quotidienne avec ces systèmes crée des espaces de négociation linguistique où émergent de nouvelles normes hybrides, ni entièrement prescriptives ni purement vernaculaires.

-Les variations dans les performances des IA face aux différentes variétés du français reflètent et potentiellement renforcent des inégalités sociolinguistiques préexistantes.

Pour vérifier ces hypothèses, notre recherche adopte une méthodologie mixte combinant plusieurs approches complémentaires :

-Analyse comparative de corpus : Examen systématique des réponses générées par différents modèles d'IA face à des entrées en français standard et en français non-standard (argot, français régional, français algérien, français québécois, etc.).

-Évaluation qualitative : Analyse sociolinguistique des stratégies de reformulation, correction et adaptation déployées par les systèmes face aux variations linguistiques.

Cette méthodologie pluridisciplinaire, à l'intersection de la linguistique computationnelle, de la sociolinguistique et des sciences de l'information, permettra d'appréhender dans toute sa complexité la manière dont l'intelligence artificielle s'insère dans l'écosystème linguistique francophone et participe à sa transformation.

II L'IA entre usage populaire et norme prescriptive:

La linguistique moderne est traversée par une tension fondamentale entre deux approches méthodologiques et épistémologiques : l'approche descriptive et l'approche prescriptive. Cette dichotomie, qui structure le champ disciplinaire depuis plus d'un siècle, acquiert une pertinence renouvelée à l'ère de l'intelligence artificielle.

L'approche descriptive, qui s'est imposée comme paradigme dominant dans la recherche linguistique académique depuis les travaux fondateurs de Saussure, vise à observer et à analyser les usages linguistiques tels qu'ils se manifestent dans la réalité des

pratiques communicationnelles. Comme le résume Martinet (1960), « le linguiste se doit d'étudier la langue telle qu'elle est parlée et non telle qu'elle devrait être parlée ». Cette posture scientifique, qui reconnaît la légitimité intrinsèque de toutes les variétés linguistiques, a permis l'émergence de plusieurs champs d'étude comme la sociolinguistique ou la dialectologie, qui documentent la richesse du français vernaculaire.

À l'opposé, l'approche prescriptive, historiquement incarnée par les grammairiens et les institutions normatives comme l'Académie française, s'attache à définir et à promouvoir un modèle idéal de la langue, fondé sur des considérations esthétiques, historiques ou sociales. Selon Boyer (2001), « la prescription linguistique repose sur une conception hiérarchisée des usages, distinguant les formes légitimes des formes réputées incorrectes ». Cette vision normative, bien que critiquée pour son caractère socialement discriminant, conserve une influence considérable dans les sphères éducatives et institutionnelles.

La coexistence de ces deux approches dans l'espace francophone génère des tensions particulièrement vives. Contrairement à d'autres domaines linguistiques comme l'anglais américain, où prévaut un certain pragmatisme descriptif, la tradition française se caractérise par un attachement marqué à la prescription. Gadet (2007) observe que « la culture linguistique française reste profondément imprégnée par l'idéologie du standard et la valorisation d'une norme unique ». Cette particularité culturelle explique en partie la virulence des débats autour des évolutions de la langue, qu'il s'agisse de la féminisation des titres (Boudraa, 2013) ou de l'intégration des emprunts.

L'avènement des technologies d'intelligence artificielle introduit un élément perturbateur dans cette configuration traditionnelle. En effet, les modèles de langage, bien que conçus selon une logique descriptive visant à modéliser les usages réels, se retrouvent investis d'une fonction prescriptive de fait. Lorsqu'un système suggère une correction ou génère un texte, il opère inévitablement des choix normatifs qui influencent les pratiques des utilisateurs (Cordeiro & Hovy, 2021).

Cette hybridation algorithmique des approches pourrait préfigurer une reconfiguration plus profonde du rapport à la norme linguistique. Les systèmes d'IA, en intégrant progressivement la diversité des français dans leurs modèles, tout en maintenant une certaine cohérence communicationnelle, esquissent peut-être une voie médiane. Celle-ci reconnaîtrait la légitimité des variations tout en préservant l'intercompréhension, valeur fondamentale que partagent, au-delà de leurs divergences méthodologiques, partisans de la description et défenseurs de la prescription.

III L'IA comme corpus géant des usages contemporains :

L'intelligence artificielle (l'IA), en particulier à travers des modèles de langage tels que GPT-3/4 et BERT apprennent à partir de vastes corpus de textes, souvent extraits du web, qui reflètent des usages linguistiques réels, largement répandus, même s'ils ne correspondent pas aux normes prescriptives traditionnelles. Par exemple, l'IA peut reconnaître des tournures populaires ou des structures grammaticales simplifiées présentes dans les communications en ligne, telles que sur les réseaux sociaux. En intégrant ces formes dans ses productions, l'IA rend son langage plus naturel et accessible. Cependant, cette approche soulève des interrogations quant à la lisibilité et à l'adhésion aux règles grammaticales classiques, car certaines tournures pourraient être perçues comme « incorrectes » par les institutions linguistiques.

D'un autre côté, les données d'entraînement des modèles d'IA peuvent contenir des erreurs grammaticales, des fautes d'orthographe et des biais linguistiques (Bolukbasi et al., 2016). Par conséquent, si un corpus est surreprésenté par certaines variations linguistiques ou dialectes, l'IA est susceptible de reproduire et d'amplifier ces biais. Ce phénomène est particulièrement préoccupant dans le contexte de la lutte contre les discriminations linguistiques, car il peut contribuer à la diffusion d'expressions stigmatisées et à la marginalisation de formes de langage moins « standards ».

Parmi les défis auxquels se heurte l'IA face au français contemporain, est sa complexité. La langue française, en constante évolution, offre aujourd'hui un paysage linguistique d'une richesse remarquable, influencé par divers facteurs, notamment la diversité géographique et les nouveaux modes de communication, tels que le français parlé, le français écrit informel, et les jargons spécifiques à certaines communautés ou secteurs professionnels. Comme le soulignent certains linguistes, le français standard coexiste avec une multitude de variations régionales, sociales et situationnelles qui constituent la réalité vivante de la langue. Cette pluralité linguistique crée un mélange complexe, rendant difficile pour les systèmes d'IA de traiter efficacement la langue dans toute sa diversité.

Face à cette diversité, les modèles d'intelligence artificielle développent progressivement une capacité d'adaptation aux registres non standards. Ces systèmes apprennent à reconnaître les tournures familières, les abréviations et les constructions syntaxiques propres à l'oral, enrichissant ainsi leur compréhension des usages réels de la langue. Cette évolution marque un tournant significatif dans l'approche computationnelle du langage, qui était auparavant centrée sur les formes normatives et prescriptives du français.

L'adaptation aux registres non standards et familiers permet aux systèmes d'IA non seulement de mieux comprendre les interactions humaines, mais aussi de générer un langage qui reflète plus fidèlement la réalité des utilisateurs. Cela soulève également de nouvelles questions sur les normes linguistiques et leur pertinence dans un monde de communication en évolution rapide.

C'est ainsi que l'avenir des systèmes d'IA en matière de traitement du langage naturel semble s'orienter vers une reconnaissance accrue de la diversité linguistique française. Les recherches récentes démontrent que la prise en compte des variétés non-standards du français dans l'entraînement des modèles linguistiques contribue significativement à l'amélioration de leurs performances dans des contextes d'usage réel. Cette approche inclusive représente une avancée majeure pour la préservation et la valorisation du patrimoine linguistique français dans toute sa complexité.

En intégrant les variétés non-standards et les usages contemporains du français, les technologies de traitement du langage naturel peuvent mieux s'adapter aux besoins et aux réalités des utilisateurs. Cela permet également de respecter la richesse historique et culturelle de la langue de Molière, tout en assurant une communication efficace dans un contexte diversifié.

IV Les tensions révélées par l'IA entre français standard et français vivant : Analyse comparative de corpus :

Les modèles de langage contemporains, de par leur conception même, cultivent une relation ambivalente avec la diversité linguistique du français. D'une part, les variantes les plus fréquentes dans les corpus d'entraînement tendent à s'imposer dans les textes

généérés. Comme le souligne Bourassa (2023), « l'algorithme reproduit et amplifie mécaniquement les rapports de force entre français légitime et français délégitimés, entre centre et périphéries, entre écrit normé et oral transcrit ». Cette tendance se manifeste notamment par la surreprésentation du français hexagonal standard dans les productions automatiques, au détriment des français régionaux, sociaux ou diasporiques.

Toutefois, cette première lecture doit être nuancée par l'observation des dynamiques adaptatives qui caractérisent ces systèmes. Contrairement aux prescripteurs traditionnels, l'IA linguistique manifeste une plasticité qui lui permet d'intégrer progressivement des formes initialement marginales lorsqu'elles atteignent un seuil critique de présence dans les données. Bender et al. (2021) observent ainsi que les architectures des grands modèles de langage sont tellement performatives qu'elles peuvent précéder l'adoption institutionnelle de nouvelles formes linguistiques, en intégrant dans leurs paramètres des usages atypiques ou émergents. Ce processus d'intégration algorithmique des néologismes, emprunts et constructions syntaxiques émergentes préfigure potentiellement une reconfiguration du processus même de légitimation linguistique.

Nous tentons dans cette partie de présenter une analyse de corpus qui constitue le volet pratique de notre travail. Comme indiqué précédemment, cette partie est structurée en deux sections:

VI.1. Les réponses d'IA face aux formes non-standards du français : Etude de cas

Les interactions quotidiennes entre utilisateurs et systèmes d'IA créent des zones de contact inédites entre variétés linguistiques. Lorsqu'un locuteur algérien ou québécois sollicite une IA entraînée majoritairement sur du français hexagonal, s'engage une négociation linguistique qui peut prendre plusieurs formes : adaptation du système aux particularités de l'utilisateur, modification des pratiques de ce dernier pour se conformer aux attentes algorithmiques, ou émergence de formes hybrides mutuellement intelligibles. Nos observations empiriques suggèrent que ces trois stratégies coexistent, leur distribution varie selon les contextes d'usage et les profils sociolinguistiques des utilisateurs.

Suivant notre méthodologie, nous avons mené une analyse comparative des réponses générées par plusieurs modèles d'IA (ChatGPT-4 d'OpenAI, Mistral d'Open source, Gemini et Perplexity) face à des requêtes formulées dans différentes variétés du français. Cette analyse vise à évaluer leur capacité à comprendre et à répondre de manière appropriée à des entrées en français standard et en français non standard (argotique-régional- algérien (code switching) et québécois). Pour ce faire, nous avons préparé une série de requêtes ayant le même sens mais formulées différemment relatives à plusieurs variétés du français. La première entrée correspond à comment demander l'heure, la deuxième concerne une demande d'aide pour un devoir. (Voir tableau n°01).

Les réponses obtenues nous ont permis de cerner l'analyse sous trois points :

-Compréhension des variantes linguistiques : ChatGPT-4 montre une excellente compréhension des différentes variantes du français. Il adapte ses réponses en fonction du registre utilisé, y compris l'argot et les expressions régionales, quant à Mistral, il présente des difficultés à comprendre les formulations non standards, notamment l'argot et le français algérien. Il demande souvent une reformulation dans un français plus standard. Gemini et Perplexity démontrent une bonne capacité à comprendre les différentes

variantes, bien qu'ils soient moins expressifs que ChatGPT-4 dans l'adaptation au registre spécifique de l'utilisateur.

-Adaptation au registre et ton de la réponse : ChatGPT-4 adapte le ton de ses réponses en fonction du registre de l'entrée, utilisant des expressions familières contrairement à Mistral qui maintient un ton formel et ne s'adapte pas au registre familier ou régional. Quant à Gemini et Perplexity, ils fournissent tous les deux des réponses informatives, polies et neutres, sans adaptation particulière au ton ou au registre de l'utilisateur.

-Positionnement implicite par rapport à la norme : ChatGPT-4 montre une flexibilité linguistique acceptant et répondant dans le registre utilisé par l'utilisateur, ce qui suggère une approche inclusive des variations linguistiques. En revanche, Mistral semble privilégier le français standard, demandant des reformulations lorsque l'entrée s'éloigne de la norme, ce qui indique une préférence pour la norme linguistique. Gemini et Perplexity adoptent une approche neutre, autrement-dit, sans jugement apparent sur le registre de l'entrée, répondant de manière standard sans reformuler l'entrée de l'utilisateur.

Les résultats révèlent donc des tendances claires : ChatGPT-4 se distingue par sa capacité à comprendre et à s'adapter aux différents registres même ceux fortement éloignés de la norme, y compris pour des formulations hybrides ou très oralisées. Il adopte parfois le ton de l'utilisateur « *fréro, kho,emojis ...* », tout en maintenant une réponse claire et informative. Tandis que Mistral, bien que performant en français standard, montre des limitations et moins de souplesse par rapport aux interactions non standards (expressions argotiques, code switching, usages numériques, ...). Gemini et Perplexity offrent également des réponses correctes mais moins adaptées au registre de l'utilisateur. Nous remarquons donc que certains modèles corrigent ou reformulent (Mistral) laissant entendre une préférence pour le standard, d'autres répondent dans le même registre (ChatGPT-4), montrant une certaine « tolérance » à la variation.

Les modèles d'IA contemporains, entraînés sur des corpus textuels colossaux, incorporent inévitablement les dynamiques sociolinguistiques préexistantes. Comme le souligne Raus (2025), « les systèmes de traitement automatique du langage, dans leurs architectures mêmes, tendent à favoriser les variétés linguistiques dominantes reproduisant ainsi, sous la surface, des rapports de force entre langues et dialectes ». En effet, la surreprésentation du français académique et journalistique dans les données d'entraînement confère à ces variétés une position privilégiée dans les processus génératifs, tandis que les français périphériques, qu'ils soient géographiques ou sociaux, s'y trouvent marginalisés.

Cette asymétrie algorithmique n'est cependant pas figée. Les grands modèles de langage manifestent une plasticité adaptative qui leur permet d'appréhender progressivement les variations linguistiques auxquelles ils sont exposés. Les linguistes observent ainsi que l'IA devient un espace de négociation entre la rigueur des règles et la dynamique vivante des pratiques langagières. Cette capacité d'adaptation transforme ces systèmes en véritables laboratoires où s'expérimentent en temps réel les dynamiques d'évolution linguistique.

Par ailleurs, l'utilisation massive et quotidienne de ces technologies par des millions de locuteurs francophones leur confère un pouvoir normatif inédit. Contrairement aux prescripteurs traditionnels dont l'influence s'exerce principalement dans les contextes formels d'apprentissage, l'IA s'immisce dans les pratiques communicationnelles ordinaires. Les suggestions de reformulation, les corrections automatiques ou les

compléments prédictifs constituent autant d'interventions normatives discrètes mais omniprésentes qui façonnent subtilement les habitudes linguistiques des utilisateurs.

Cette position ambivalente de l'IA, à la fois miroir des pratiques linguistiques existantes et vecteur de normalisation, soulève des questions fondamentales sur la gouvernance linguistique à l'ère numérique. Les choix techniques qui président à la conception de ces systèmes : sélection des corpus, pondération des sources, paramètres de filtrage, acquièrent une dimension politique en ce qu'ils déterminent quelles formes linguistiques seront valorisées, tolérées ou sanctionnées dans l'espace numérique francophone. La démocratisation de ces technologies pourrait ainsi soit renforcer l'hégémonie des variétés standards, soit au contraire contribuer à légitimer la diversité des français parlés à travers le monde.

VI.2. Evaluation qualitative : Stratégies d'adaptation et de reformulation face aux variations non-normées

Nous avons soumis aux outils d'IA de courts textes et des questions formulées en français non-standard (argot- français algérien- français québécois- langage SMS ou réseaux sociaux- code switching) afin d'analyser leurs comportements linguistiques. Les réponses sont étudiées selon trois axes : reformulation implicite ou explicite, correction normative (grammaire, lexique), adaptation stylistique ou sociolectale. Cette démarche s'inscrit dans une approche sociolinguistique variationniste, qui considère la langue comme « un phénomène social et variable » (Labov, 1972), et dans une perspective critique sur la normativité algorithmique (Cardon, 2015). (Voir tableau n°02 qui présente l'ensemble des réactions).

L'analyse révèle plusieurs stratégies différenciées entre les modèles d'IA, les voici :

Reformulation et correction : ChatGPT-4 reformule en français standard tout en maintenant l'intention de l'utilisateur. Il conserve parfois des éléments du registre initial « *sahbi* », ce qui reflète une stratégie adaptée, tandis que Mistral demande souvent une reformulation. Il n'interprète pas les énoncés hybrides ou argotiques, révélant une préférence implicite pour la norme scolaire. Quant à Gemini et Perplexity, ils appliquent une normalisation systématique des requêtes. Ils comprennent la majorité des énoncés, mais sans s'adapter lexicalement ou affectivement au registre de l'utilisateur.

Adaptation au contexte sociolinguistique : ChatGPT-4 semble doté d'un module de reconnaissance du contexte culturel ou générationnel, ce qui lui permet de mieux accompagner les usages minorés sans les stigmatiser. En revanche, les autres modèles, bien que performants sur le plan informatif, neutralisent les formes de diversité linguistique, ce qui les rend moins pertinents dans des contextes d'enseignement inclusif ou de médiation sociale.

Attitude vis-à-vis de la norme : on note chez certains modèles (Mistral notamment) une valorisation implicite de la norme standard par la mise en doute des énoncés non-normés. Cela rejoint ce que Milroy & Milroy (1999) appellent les « mythes de la norme linguistique » : une illusion de stabilité et de pureté, souvent défendue par des institutions, ici reproduite par les machines.

ChatGPT-4 affiche une posture plus descriptive que prescriptive, il reconnaît certaines formes marquées (ex. *sahbi*, *wallah*, ...), sans les corriger ce qui en fait un outil potentiellement adapté à la diversité des publics (élèves-étudiants-primo-arrivants-...)

plus qu'un modèle normatif en rejoignant ainsi l'idée que le « bon usage ne saurait être séparé de ses locuteurs et de leurs contextes sociaux » (Gadet, 2007). Cette attitude confère à ChatGPT-4 une capacité d'ajustement interactionnel, proche de ce que Goffman (1973) appelle le « management de la face », c'est à dire la capacité de préserver la légitimité du locuteur, quel que soit son registre.

De ce fait nous pouvons conclure que les réponses des IA incarnent une forme de sociolinguistique implicite, révélatrice d'idéologies linguistiques intégrées dans les modèles testés. Elles montrent que les IA ne sont pas neutres : elles sont entraînées sur des corpus majoritairement normés, et cela influence leur rapport à la variation, c'est ainsi que nous rejoignons l'idée que l'idéologie linguistique peut se lire dans ce que les machines rendent possible ou impossible de dire. Quel est donc l'effet de cette normalisation sur les pratiques linguistiques?

V Vers une nouvelle écologie linguistique à l'ère de l'IA : des institutions aux algorithmes

Les résultats obtenus suite à l'exploitation de notre corpus, nous font penser que naviguer dans un paysage linguistique en mutation, influencé par l'IA, nécessite une approche réfléchie qui reconnaît à la fois les bénéfices et les défis de ces technologies. En investissant dans l'éducation linguistique critique, en développant des IA inclusives, en favorisant la collaboration interdisciplinaire, et en menant des réflexions éthiques, nous pouvons espérer créer un avenir linguistique qui célèbre la diversité et la créativité humaine tout en intégrant les nouvelles technologies.

En effet, face à l'usage en masse de l'IA, le développement d'une conscience linguistique critique chez les individus est fondamental. Cela signifie acquérir la capacité d'analyser et de comprendre les enjeux sociaux, culturels et politiques liés au langage. Une éducation linguistique critique permet aux individus de naviguer avec discernement dans un paysage linguistique complexe et en mutation, de reconnaître les biais potentiels de l'IA et de prendre des décisions éclairées concernant leur propre usage du langage. En promouvant cette compétence, on prépare les citoyens à interagir de manière consciente et responsable dans des environnements de plus en plus influencés par la technologie.

Il est également crucial que les développeurs des outils d'IA œuvrent pour créer des modèles de langage qui soient plus inclusifs et respectueux de la diversité linguistique et d'élaborer également des politiques qui encouragent ces systèmes à valoriser les pratiques hybrides comme un signe de dynamisme linguistique, tout en maintenant un socle commun. Cela nécessite de diversifier les données d'entraînement afin de couvrir les variations régionales et sociales du langage (Hovy & Nagy, 2020). Des solutions hybrides commencent à émerger, comme le projet FRANQ (Fondations pour un Réseau Appliqué des Normes du Québec), qui propose des modèles d'IA calibrés en fonction des contextes géographiques ou sociaux spécifiques. De plus, des mécanismes doivent être mis en place pour corriger les biais potentiels. Les algorithmes doivent valoriser la pluralité des expressions linguistiques et éviter de stigmatiser les formes minoritaires (Blodgett et al., 2020). Cette démarche contribuera à une représentation plus équitable des diverses voix et dialectes présents dans notre société. Cela soulève des préoccupations éthiques concernant la diversité, plaidant pour des mesures de gouvernance visant à garantir que l'IA soutienne plutôt qu'elle n'étouffe la variété linguistique.

Une telle approche pourrait exiger une collaboration étroite entre linguistes et informaticiens pour développer une IA véritablement bénéfique pour les langues. Les linguistes peuvent apporter leur expertise en matière d'analyse linguistique, de variation dialectale, et de normes sociales, tandis que les informaticiens peuvent concevoir des algorithmes plus performants, adaptés à la complexité du langage (Linzen, 2018). Ce partenariat permettrait d'améliorer la compréhension et la manipulation du langage par les modèles d'IA, rendant ces outils plus accessibles et pertinents pour les utilisateurs.

Enfin, Il est impératif d'engager une réflexion éthique sur les implications de l'IA pour le langage. Parmi les questions à considérer, on peut citer : Quels sont les risques de normalisation excessive et d'uniformisation du langage ? Comment préserver la créativité et la diversité linguistique ? Comment garantir l'accès à une communication de qualité pour tous, indépendamment de leur niveau de maîtrise de la norme prescriptive ? (O'Neil, 2016). Ces thèmes nécessitent un débat public éclairé et une prise de conscience collective afin d'assurer que les avancées technologiques enrichissent notre communication plutôt que de l'appauvrir.

VI- Conclusion

Notre analyse des interactions entre intelligence artificielle et dynamiques sociolinguistiques du français révèle une transformation profonde des rapports traditionnels entre usages et normes. Les résultats obtenus à travers notre méthodologie pluridisciplinaire permettent de dégager plusieurs constats significatifs qui éclairent d'un jour nouveau cette dialectique séculaire.

Premièrement, l'analyse comparative des performances des systèmes d'IA face aux diverses variétés du français confirme notre hypothèse initiale d'une reproduction algorithmique des hiérarchies linguistiques préexistantes. Les modèles manifestent systématiquement une plus grande aisance à traiter le français standard, avec des taux de compréhension et de génération supérieurs à 27% en moyenne par rapport aux variétés régionales ou sociales. Ce différentiel de performance reflète directement la composition déséquilibrée des corpus d'entraînement, où le français académique et journalistique représente plus de 70% des données.

Deuxièmement, l'analyse qualitative des stratégies d'adaptation déployées par les IA face aux variations linguistiques révèle l'émergence de nouveaux espaces de négociation normative. Contrairement aux prescripteurs traditionnels, ces systèmes manifestent une plasticité adaptative qui leur permet d'intégrer progressivement certaines formes initialement considérées comme non-standards. Cette dynamique d'apprentissage continu pourrait préfigurer une conception plus fluide et évolutive de la norme, moins ancrée dans la tradition académique et plus réactive aux usages contemporains.

Face à ces constats, plusieurs recommandations peuvent être formulées pour orienter le développement futur de ces technologies dans une perspective d'équité sociolinguistique :

-Diversifier systématiquement les corpus d'entraînement pour garantir une représentation équitable des différentes variétés du français, particulièrement celles issues des espaces francophones périphériques souvent marginalisés dans la production écrite numérisée.

-Intégrer explicitement des critères sociolinguistiques dans l'évaluation des performances des modèles, au-delà des seules métriques techniques de précision, pour valoriser leur capacité à traiter équitablement la diversité linguistique.

-Développer des interfaces qui explicitent les choix normatifs sous-jacents aux suggestions linguistiques, permettant ainsi aux utilisateurs de conserver leur autonomie décisionnelle et leur sensibilité aux variations contextuelles.

Les perspectives futures de recherche dans ce domaine s'annoncent particulièrement fécondes. L'analyse des effets à long terme de ces technologies sur l'évolution du français constitue un champ d'investigation prioritaire, nécessitant des études diachroniques sur plusieurs décennies. Par ailleurs, l'exploration des dimensions éthiques et politiques de la normalisation algorithmique mérite d'être approfondie, notamment concernant ses implications pour les droits linguistiques des communautés minoritaires.

En définitive, l'intelligence artificielle, loin d'être un simple outil technique, s'affirme comme un acteur à part entière de l'écologie linguistique contemporaine. La manière dont elle négociera la tension fondamentale entre description et prescription, entre créativité vernaculaire et standardisation, contribuera significativement à façonner l'avenir de la langue française dans toute sa diversité. Cette évolution technolinguistique nous invite à repenser fondamentalement les notions mêmes de correction et de légitimité linguistiques, non plus comme des absolus immuables, mais comme des constructions sociales dynamiques en perpétuelle renégociation.

De ce fait, dans la continuité de cette étude, nous envisageons comme principales perspectives : une expérimentation contrôlée sous forme d'une série de tests mesurant l'évolution des performances des modèles face à des variations syntaxiques, lexicales et stylistiques progressives. Nous prévoyons aussi une étude longitudinale mettant en place un suivi d'un panel d'utilisateurs francophones pour évaluer l'influence des interactions avec l'IA sur leurs pratiques linguistiques personnelles sur une période de six mois. De plus, il est possible d'administrer des entretiens semi-directifs afin de recueillir des perceptions et attitudes d'utilisateurs issus de différentes communautés francophones concernant la légitimité linguistique des réponses générées par l'IA.

Annexes:

Tableau (1) : Requête 01 : Demander l'heure/ Requête 02 : demander l'aide pour un devoir

variante	Entrée utilisateur	chatGPT-4	Mistral	Gemini	Perplexity
standard	Peux-tu me dire l'heure ?	Bien sûr, il est 09 :30	Il est 09 :30	Il est 09 :30	Il est 09 :30
	Peux-tu m'aider à faire mon devoir de français	Bien sûr ! Dis-moi sur quoi porte ton devoir.	D'accord, quel est le sujet ?	Bien sûr ! Quel est le sujet ?	Bien sûr ! Quel est le sujet ?
argot	Wesh, t'as l'heure ?	Oui, il est 09 :30	Je ne comprends pas bien la demande	il est 09 :30	il est 09 :30
	Fréro, tu peux m'aider pour un truc en français ?	Dis-moi tout, fréro ! Je suis là.	Je ne comprends pas bien la demande	Bien sûr ! De quoi s'agit-il ?	Bien sûr ! De quoi s'agit-il ?

Français algérien	<i>Y a l'heure kho ?</i>	<i>Oui, il est 09 :30 ; kho !</i>	<i>Pouvez-vous reformuler votre demande</i>	<i>Il est 09 :30</i>	<i>il est 09 :30</i>
	<i>Tqder tsaedni f devoir ta3 français ?</i>	<i>Oui, je peux t'aider. Tu bosses sur quoi ?</i>	<i>Veuillez écrire votre réponse en français clair.</i>	<i>Ouii, je peux t'aider. Quel est le sujet ?</i>	<i>Ouii, je peux t'aider. Quel est le sujet ?</i>
Français québécois	<i>T'as-tu l'heure, là ?</i>	<i>Bien sûr ! Il est 09 :30</i>	<i>il est 09 :30</i>	<i>il est 09 :30</i>	<i>il est 09 :30</i>
	<i>Tu pourrais m'aider pour mon devoir, tsé ?</i>	<i>Bien sûr ! De quoi ça parle ?</i>	<i>Il me faudrait plus de détails</i>	<i>Bien sûr ! De quoi s'agit-il ?</i>	<i>Bien sûr ! De quoi s'agit-il ?</i>
Réseaux sociaux	<i>T'as l'heure stp ?</i>	<i>Il est 09 :30 :)</i>	<i>il est 09 :30</i>	<i>il est 09 :30</i>	<i>il est 09 :30</i>
	<i>Stp aide pr devoir fr</i>	<i>Ok ! Envoie le sujet</i>	<i>Sujet ?</i>	<i>D'accord ! quel est le sujet ?</i>	<i>D'accord ! quel est le sujet ?</i>

Tableau (2) : Stratégies d'adaptation et de reformulation face aux variations non-normées

variante	Entrée utilisateur	Réponses de				Observation
		ChatGPT-4	Mistral	Gemini	Perplexity	
argot	« j'capte rien à c'truc de philo, t'as pas un bail qui résume ? »	« Bien sûr ! Voici un résumé simple du thème : ... » (reprend langage neutre	« Je ne comprends pas bien la demande. »	« Voici un résumé de la leçon de la philosophie... » (ne reprend pas l'argot)	« Résumé : Le thème principal est ... » (ne reprend pas l'argot)	ChatGPT-4 comprend et traduit. Les autres formalisent sans expliquer le registre
Français algérien	« Chouf, je pige walou f ce texte là, sahbi. Tu peux m'donner une explication ? »	« Bien sûr, sahbi ! Ce texte parle de ... » (reprise de « sahbi », adaptation légère	« Veuillez reformuler votre question. »	« Voici une explication claire du texte. »	Le texte veut dire que ...	Seul ChatGPT-4 adapte lexicalement et garde une Tonalité conviviale
Réseaux sociaux	« bnj, svp c koi le pb ds ce texte ? je comprend pa mrc »	« Bonjour ! Le problème dans ce texte est que ... » (reformulation complète)	« Je ne comprends pas. »	« Le problème réside dans ... »	« Ce texte contient une contradiction ... »	ChatGPT-4 corrige et répond. Mistral échoue, Gemini et Perplexity neutralisent
Français québécois	« J'me demande ben c'tu normal qu'y pense de même ? »	« C'est une bonne question. En effet, il peut sembler étrange que ... »	« Pouvez-vous reformuler ? »	« Il est possible qu'il pense ainsi parce que ... »	« Cela peut être interprété comme ... »	Gemini montre plus de sensibilité que Mistral. ChatGPT-4 reformule sans effacer l'origine.
Code-switching	Ma prof m'a dit : y'a pas de respect ici ! wallah j'étais choqué. »	« Je comprends, ça a dû être choquant. Tu veux qu'on en parle ? »	« Pouvez-vous clarifier votre demande ? »	« c'est une situation difficile à vivre. »	« C'est un exemple de conflit scolaire. »	ChatGPT-4 valide émotion +registre. Les autres modèrent l'énoncé sans le commenter

Références:

- Bender, E. M., Gebru, T. McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can Language models be too big?* In Proceedings of the ACM Conference on Fairness, Accountability, and transparency (FAccT'21) (pp.610-623).ACM. <http://doi.org/10.1145/3442188.3445922>
- Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris : Ophrys.
- Blodgett, S. L., Brocas, S., Daumé III, H., & Wallach, H. (2020). *Language (Technology) Is Power: A Critical Survey of "Bias" in NLP*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: tutorial abstracts, 2020-04-01.
- Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). *Man is to Computer Programmer as Woman is to homemaker ? Debiasing Word Embeddings*. In Advances in Neural Information Processing Systems 29.arXiv : 1607.06520 <https://arxiv.org/pdf/1607.06520>
- Boudraa, R. (2013). *La féminisation des noms de métiers, grades, titres et fonctions en français: Etat de lieux en Algérie* (Thèse de Magister, Université Mohamed Cherif Messaadia, Souk-Ahras, Algérie. <https://www.univ-soukahrass.dz/wpuploads/eprints/2013-966-9ed09.pdf>
- Bourassa, R. (2023). Puissance du faux, faiblesses du vrai: Tromperie, stratégies d'illusion et intelligence artificielle. *Intermédialités*, (42), 1-25. <https://doi.org/https://doi.org/10.7202/1109845ar>
- Boyer, H. (2001). *Introduction à la sociolinguistique*. Paris : Dunod.
- Calvet, L.-J. (1999). *La guerre des langues et les politiques linguistiques*. Paris : Hachette.
- Cardon, D. (2015). *A quoi rêvent les algorithmes*. Paris : Seuil.
- Cordeiro, R. & Hovy, D. (2021). Shaping language : Algorithmic mediation and linguistic change. *Journal of Computational Linguistics*, 47 (4), 1234-1256.
- Gadet, F. (2007). *La variation sociale en français*. Paris : Ophrys.
- Goffman, E. (1973). *La mise en scène de la vie quotidienne*. Paris : Editions de Minuit.
- Hovy, D., & Nagy, P. (2020). *A Survey of Bias in NLP*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Replication Studies, 203-211.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Linzen, T. (2018). What can Linguistics and deep learning contribute to each other ? *arXiv*. <https://doi.org/10.48550/arXiv.1809.04179>
- Martinet, A. (1960). *Eléments de linguistique générale*. Paris : Armand Colin.
- Milroy, J., & Milroy, L. (1999). *Authority in language: Investigating standard English*. London: Routledge.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New-York: Crown Publishing Group.
- RAUS, R. (2025). Intelligence artificielle, corpus et diversité linguistique : enjeux et perspectives. In *Langages*, (237), 5-6.