
Les textes 'générativement' produits à l'épreuve de la diversité culturelle francophone et le risque de standardisation

Generatively produced texts put to the test of francophone cultural diversity and the risk of standardization

GOH Tianet Yannick Emmanuel¹

Université Alassane Ouattara, Bouaké Côte d'Ivoire.

gtianetyannicke@gmail.com

ADJOUMANI Kouassi Martin²

Université Alassane Ouattara, Bouaké Côte d'Ivoire.

kingadjoumani@gmail.com

Reçu: 01/05/2025, **Accepté:** 06/06/2025, **Publié:** 20/06/ 2025

Résumé :

Dans un contexte de révolution numérique qui transforme profondément les pratiques de communication, l'étude examine les risques de standardisation et d'appauvrissement du français induits par l'intelligence artificielle générative dans le contexte francophone. L'objectif de l'étude est d'analyser et de comparer les productions textuelles générées par des systèmes IA (ChatGPT, Deepseek, Gemini, etc.) à celles élaborées par des humains, en nous appuyant sur une méthode qualitative fondée sur un double corpus et la triangulation des retours d'experts. L'analyse révèle un taux d'erreurs lexicales, syntaxiques et contextuelles nettement supérieur dans les textes automatisés, accompagné d'une répétition excessive d'expressions « passe-partout » et d'une homogénéisation du discours qui se traduit par la notion de « model collapse » linguistique. Ces constats confirment que l'automatisation appauvrit progressivement la richesse lexicale et culturelle du français et justifie la nécessité d'enrichir les corpus d'entraînement et de promouvoir des approches hybrides pour préserver sa diversité.

Mots-clés : Intelligence Artificielle Générative - Standardisation Linguistique - Appauvrissement de la langue française - Diversité Culturelle - Littérature Numérique.

Abstract :

In the context of a digital revolution that is profoundly transforming communication practices, this study examines the risks of standardization and impoverishment of the French language induced by generative artificial intelligence in the Francophone context. The aim of the study is to analyze and compare text productions generated by AI systems (ChatGPT, Deepseek, Gemini, etc.) with those produced by humans, using a qualitative method based on a dual corpus and triangulation of expert feedback. The analysis reveals a significantly higher rate of lexical, syntactic, and contextual errors in automated texts, accompanied by an excessive repetition of generic ("catch-all") expressions and a homogenization of discourse that reflects a linguistic "model collapse." These findings confirm that automation progressively impoverishes the lexical and cultural richness of French and underscore the need to enrich training corpora and promote hybrid approaches to preserve its diversity.

Keywords : Generative Artificial Intelligence - Linguistic Standardization - Linguistic Impoverishment of the french language - Cultural Diversity - Digital Literacy.

Pour citer cet article :

GOH, Tianet, Yannick, Emmanuel et ADJOUMANI, Kouassi, Martin, (2025), les textes 'générativement' produits à l'épreuve de la diversité culturelle francophone et le risque de standardisation, Contextes Didactiques, Linguistiques et Culturels [En ligne], 3(1), 347-362. Disponible sur le lien : <https://www.asjp.cerist.dz/en/PresentationRevue/928>

Pour citer le numéro :

MOUFFOKI, Nadjat, BENYOUCEF , Mahfoudh et TOMESCO, Ana Marina, (2025), Numéro –Thématique « *Intelligence artificielle et apprentissage des langues : perspectives, enjeux et nouvelles pratiques didactiques ?* », Contextes Didactiques, Linguistiques et Culturels [En ligne], 3(1), 423 p. Disponible sur le lien : <https://www.asjp.cerist.dz/en/PresentationRevue/928>



Introduction

Ces dernières années, le domaine de l'intelligence artificielle générative (IAG) s'est imposé comme un vecteur majeur de transformation dans la production et la diffusion de contenus, notamment textuels (Côté & Su, 2022; Véry & Cailluet, 2019). Les avancées fulgurantes des grands modèles de langage (LLM), tels que ChatGPT, témoignent de leur capacité à générer, en un temps record, des textes d'une fluidité apparente, lequel processus, automatisé à son aise, des tâches de création écrite qui jadis étaient uniquement réservées à l'humain (Idriss J. Aberkane, 2023). De nombreux travaux francophones (Académie de Paris, 2025; AFD, 2025; Dagnogo et Samasse, 2022; Elysee.fr, 2025; Kadio, 2024; Sinapin, 2020) ont reflété une vision de l'impact de ces outils sur la qualité et la diversité des productions linguistiques contemporaines. Ils ont donc posé les bases d'une réflexion sur le potentiel disruptif de l'IA, tout en soulignant que la richesse du français et, plus largement, de la Francophonie, demeure un enjeu central à préserver dans un univers numérique en perpétuelle évolution. Pourtant, malgré cette reconnaissance de leurs prouesses en matière de génération de contenu, un vide persiste dans la littérature. En effet, si la capacité des modèles génératifs à produire rapidement du texte est largement étudiée, peu d'analyses se sont focalisées sur les risques spécifiques qu'ils encourrent, notamment en termes de standardisation et d'appauvrissement du français. L'expression « hyper-mécanisation du texte » décrit ici la tendance de ces technologies à homogénéiser des structures lexicales et syntaxiques – en se rapprochant parfois du « langage Python », c'est-à-dire un langage instructif de type machinique – au détriment des styles naturels, des nuances régionales et des subtilités propres à la diversité culturelle. Ce biais, exacerbé par la prédominance de corpus anglophones dans l'entraînement des modèles, favoriserait l'émergence d'un « langage hybride » aux fondements souvent arbitraires. Dès lors, le fait de déléguer trop facilement l'apprentissage de certaines conversions ou traductions à l'IA, l'utilisateur risque de perdre l'opportunité d'appréhender pleinement les contours d'une langue et sa culture, ce qui entraîne, sur le long terme, un désapprentissage profond des richesses linguistiques.

Face à cette lacune, la présente étude se propose d'examiner, à partir d'analyses qualitatives rigoureuses issues de la recherche francophone, les mécanismes et les conséquences de la standardisation induite par l'IA générative sur la langue française. Notre démarche est qualitative et vise à démontrer que la diffusion massive de textes générés par l'IA, par la répétition systématique de structures simplifiées et d'anglicismes syntaxiques, tend non seulement à uniformiser l'usage d'un français «

hybridé », mais également à éroder ses variétés et ses nuances culturelles. Pour ce faire, nous établirons d'abord une analyse détaillée des erreurs – lexicales, syntaxiques et contextuelles – caractéristiques des productions IA, avant d'examiner la dynamique de normalisation induite par un apprentissage statistique biaisé. Nous aborderons ensuite l'impact de cette homogénéisation sur l'apprentissage des langues et sur la dynamique culturelle propre au contexte de la Francophonie.

Avec une démonstration claire de ce processus de « model collapse » linguistique, nous espérons clairement convaincre à la fois le lecteur novice et la communauté scientifique de l'urgence d'adopter une approche critique et de mettre en place des mesures préventives. Il s'agit, en définitive, de préserver la vitalité et la diversité du français à l'ère numérique, et de repenser la manière dont l'IA peut servir l'humain sans compromettre les spécificités culturelles qui font la richesse de notre langue.

1. Méthodologie

Pour répondre à la problématique portant sur les risques de standardisation et d'appauvrissement du français induits par l'intelligence artificielle générative (IAG), nous avons opté pour une méthode qualitative. Ce choix est justifié par la nature exploratoire et contextuelle du sujet, qui requiert une compréhension approfondie des mécanismes linguistiques et culturels en jeu – des aspects difficilement quantifiables par des mesures purement statistiques.

1.1. Constitution du Corpus et Collecte des Données

Le corpus de cette étude a été constitué selon deux volets complémentaires. D'abord, nous avons rassemblé un ensemble de productions issues de grands modèles de langage tels que ChatGPT, Deepseek, Gemini, Mistral, Perplexity et Claude couvrant divers registres d'usage (articles en ligne, publications sur les réseaux sociaux, contenus automatisés de nature professionnelle ou académique). Afin d'établir un comparatif pertinent, nous avons sélectionné des textes issus de l'écriture littéraire, journalistique et académique francophone. Ce double corpus permet d'identifier de manière claire les écarts et les convergences entre la production automatisée et la production humaine traditionnelle. Les critères de sélection ont été définis pour garantir la représentativité linguistique et culturelle de la Francophonie, en prenant en compte la diversité des registres, des variétés régionales et des contextes d'usage. La recherche documentaire a été enrichie par

l'exploration de bases académiques telles que Cairn.info et Persée, ainsi que par l'analyse de sources spécialisées qui s'inscrivaient dans le cadre de notre recherche.

1.2. Méthode d'Analyse Qualitative

L'analyse des données est réalisée à l'aide d'un processus de codage manuel. Cette approche qualitative a été choisie afin de permettre une interprétation fine et nuancée des phénomènes observés. De manière détaillée, le codage s'articule autour de trois axes principaux :

- Erreurs linguistiques spécifiques : Identification des erreurs lexicales, syntaxiques et contextuelles dans les textes générés par l'IA, afin de détecter les écarts par rapport aux normes du français.

- Phénomènes de répétition et homogénéisation : Repérage des structures répétitives et des formes linguistiques simplifiées qui tendent à uniformiser la production textuelle.

- Présence d'anglicismes et de transitions vers un « langage hybride » : Analyse de l'émergence d'influences du « langage Python » ou machinique, susceptibles d'éroder les stratégies stylistiques et les nuances culturelles traditionnelles du français.

Ces outils nous ont permis de dégager des motifs récurrents et d'interpréter en profondeur la manière dont l'IA contribue, par son apprentissage statistique biaisé, à un processus de « model collapse » linguistique.

1.3. Cadres Théoriques de Référence

Pour éclairer notre analyse, nous mobilisons deux approches théoriques complémentaires :

- L'approche théorique de la Chambre Chinoise : Popularisée par John Searle vers 1980, cette approche offre un cadre critique pertinent pour examiner la distinction entre manipulation symbolique et véritable compréhension du langage. Selon cette approche, les modèles d'IA, bien qu'efficaces en surface dans la génération syntaxique, ne saisissent pas véritablement le « sens » profond des informations. Ils opèrent comme une « chambre chinoise » qui ne fait que manipuler des symboles de manière aléatoire. Ce prisme théorique permet d'expliquer le phénomène d'hyper-mécanisation observé et souligne l'absence de compréhension réelle, ce qui conduit inévitablement à une standardisation et à un appauvrissement du contenu généré.

- La théorie du Capital Linguistique de Pierre Bourdieu : Cette théorie permet d'interpréter la standardisation en tant que processus de domination culturelle, ce qui révèle comment les rapports de pouvoir et les pratiques institutionnelles favorisent des normes linguistiques uniformisées. Elle justifie le besoin de préserver les variétés régionales et culturelles du français tout en rappelant que la langue, en tant que capital symbolique, constitue un enjeu essentiel pour la diversité et l'identité de la Francophonie.

1.4. Triangulation et Validation des Résultats

Afin de renforcer la robustesse de notre démarche qualitative, nous avons mis en œuvre une triangulation des données. Dans un premier temps, celle-ci a consisté à faire une comparaison des observations. À ce niveau, les résultats issus des textes générés par l'IA ont été systématiquement comparés aux productions humaines de référence afin d'identifier les écarts et les convergences. Ensuite, nous avons procédé à une série de validation par des experts de l'écriture manuelle notamment des écrivains (2), enseignants-chercheurs (2), étudiants (2), professeurs de lycée (2) et des personnes lambda (2), ce qui renvoie notre échantillon à 10. Leur retour a permis de valider le modèle de codage et d'assurer la cohérence des catégories identifiées.

De cette manière, la méthodologie adoptée offre une analyse qualitative des mécanismes de standardisation induits par l'IA générative, tout en s'appuyant sur des cadres théoriques solides et reconnus dans le domaine. En ce qui nous concerne, ce choix méthodologique est nettement en parfait accord avec la nature exploratoire de notre sujet, lequel choix nous a permis d'exposer les processus subtils et complexes qui aménagent la question de la vitalité du français à l'ère numérique.

2. Résultats de l'étude

Cette section présente l'ensemble des observations issues de l'analyse qualitative des textes générés par l'IA comparés aux textes issus de productions humaines, ainsi que les réponses recueillies auprès des 10 participants (2 écrivains, 2 enseignants-chercheurs, 2 professeurs de lycée, 2 étudiants et 2 personnes lambda).

2.1. Caractéristiques des Textes Générés par l'IA

Nous avons identifié en moyenne 12 erreurs lexicales pour 1 000 mots dans les textes générés par l'IA, contre environ 3 erreurs pour 1 000 mots dans les textes humains. Nous avons également constaté que 15 % des phrases produites par l'IA comportaient des erreurs syntaxiques (notamment des inversions ou des accords manquants) alors que cette proportion était inférieure à 5 % dans les productions humaines. De plus, nous avons relevé la présence d'erreurs contextuelles (« hallucinations ») dans 8 % des extraits analysés, ces occurrences étant quasi absentes dans le corpus de référence humain. Par ailleurs, nous avons observé que 35 % des phrases générées par l'IA présentaient des structures répétitives ou simplifiées, témoignant d'un schéma homogénéisant sur le niveau stylistique. Une analyse par codage a permis de relever que l'IA reproduisait régulièrement des séquences syntaxiques dominantes issues de corpus standardisés, limite qui n'apparaissait que dans moins de 10 % des textes de productions humaines. Nous notons également une forte présence d'anglicismes et de transitions vers un « langage hybride ». À ce niveau, nous avons relevé que 28 % des extraits analysés comportaient des anglicismes ou des calques syntaxiques typiques de structures informatiques (« langage Python »), alors que ces traits étaient rarement identifiés dans les textes humains (moins de 3 %).

2.2. Résultats de l'Enquête auprès des Répondants

Les réponses (des 10 participants) sont détaillées de la manière suivante :

- Écrivains (n = 2) - 2 sur 2 ont immédiatement remarqué une différence nette dans la qualité et le style des textes générés par l'IA comparés aux écrits humains.
- Enseignants-chercheurs (n = 2) - 1 sur 2 a noté des anomalies marquées dans la production d'IA, tandis que l'autre a constaté une différence moins évidente.
- Professeurs de lycée (n = 2) - 1 sur 2 a observé une différence similaire à celle relevée par les enseignants-chercheurs, et l'autre n'a pas signalé de changement particulier.
- Étudiants (n = 2) et Personnes lambda (n = 2) - Aucun des participants de ces deux groupes n'a perçu de différence significative entre les textes générés par l'IA et ceux issus de productions humaines.

2.3. Autres Constatations Issues de l'Analyse Qualitative

Nous avons constaté que la cohérence interne des textes générés par l'IA a été régulièrement entamée par des répétitions de structures simplifiées. Ensuite, nous avons mesuré que l'incidence des calques syntaxiques issus de corpus anglophones se manifestait dans près de 30 % des cas, signe d'une tendance vers ce qu'on a défini comme un « langage hybride ». Nous avons recoupé ces résultats avec les observations rapportées dans l'enquête, qui a confirmé que les répondants les plus avertis (écrivains et certains enseignants-chercheurs/professeurs) percevaient cette homogénéisation, alors que les répondants novices (étudiants et personnes lambda) ne le validaient pas.

Ces résultats confirment, sans interprétation, les constatations obtenues par codage manuel et par comparaison systématique entre les textes générés par l'IA et les productions humaines. Ils fournissent une base factuelle qui indique que, selon notre échantillon et selon les critères définis dans la méthodologie, la diffusion massive des textes générés par l'IA a été associée à une homogénéisation notable des structures linguistiques et à une présence accrue d'anglicismes et de formes standardisées.

3. Analyses et discussions

Nos analyses révèlent, de manière claire et cohérente, que les productions textuelles générées par l'IA présentent une dégradation systématique par rapport aux productions humaines. Ces textes se caractérisent parfois par un taux élevé d'erreurs lexicales, syntaxiques et contextuelles, une répétition excessive de schémas préprogrammés, autant que par un recours fréquent à des tournures banales et à des expressions passe-partout qui appauvrissent le discours.

3.1. Récapitulatif des principaux constats

Nos résultats montrent que les textes générés par l'IA comportaient environ 12 erreurs lexicales pour 1 000 mots, contre seulement environ 3 erreurs dans la production humaine. De plus, près de 15 % des phrases automatisées présentaient des inversions ou des ruptures dans l'ordre habituel (sujet-verbe-complément) et, de façon particulièrement révélatrice, un vocabulaire uniformisé se traduit par un emploi récurrent d'expressions « passe-partout ». Par exemple, nous avons constaté une répétition excessive de

schémas préétablis, une redondance de vocabulaire – avec, par exemple, un emploi abusif du terme « crucial ». Cette redondance, conjuguée à l'abus du participe présent (en « -ant ») et à des reformulations en fin de phrases, traduit une esthétique mécanique et simplificatrice du discours généré qui ne favorise pas pleinement l'enrichissement lexical.

3.2. Retours et perception de textes « générativement » produits

Les retours recueillis auprès de nos dix répondants ont mis en évidence une divergence nette : alors que les experts (écrivains, enseignants-chercheurs et certains professeurs) dénonçaient avec acuité la répétition de formules génériques, l'usage abusif d'expressions passe-partout et la perte des nuances stylistiques, les répondants moins spécialisés (étudiants, grand public) peinaient à percevoir ces anomalies alors qu'elles étaient évidentes. Ces différences de perception soulignent que la banalisation des discours issus de l'IA se fait insidieusement et risque d'être acceptée par un large public non averti.

3.3. L'IA à l'épreuve de la diversité culturelle francophone

Les textes générés par l'IA apparaissent particulièrement incapables de restituer la richesse culturelle propre aux diverses communautés francophones. Par l'omission spontanée des régionalismes et l'uniformisation du style algorithmique par l'utilisation répétée d'expressions passe-partout et de tournures génériques, l'IA trahit son incapacité à intégrer les subtilités culturelles dues à des corpus majoritairement standardisés. Cette absence de diversité dans les productions pose la question essentielle de l'adaptation de l'outil aux spécificités d'un français vivant et pluriel.

3.4. Traduction automatique, contexte culturel et richesse linguistique du français

L'analyse de la traduction automatique appliquée aux textes générés par l'IA révèle que, malgré une fluidité apparente, le rendu traduit souvent le manque de nuance et la banalisation du contenu original. La traduction, par son procédé statistique, reproduit des schémas déjà standardisés et se montre incapable de conserver les spécificités culturelles, qui conduisent une homogénéisation du discours. On observe clairement, une réduction notable de la richesse lexicale et syntagmatique, ce qui appauvrit encore davantage

le texte final (surtout lorsque celui-ci ne fait pas objet de révision profonde par un tiers humain).

3.5. Erreurs linguistiques spécifiques induites par l'IA générative en français

Au cœur des constats, l'analyse révèle que les erreurs linguistiques ne sont pas aléatoires, mais structurantes : les erreurs lexicales (choix inappropriés, néologismes hors contexte), la répétition d'erreurs syntaxiques (inversions et accords manquants) et les erreurs contextuelles (« hallucinations ») confirment que l'IA opère techniquement en reproduisant mécaniquement les schémas appris. L'usage systématique d'expressions passe-partout, associées à des constructions grammaticales simplifiées, constitue un indicateur révélateur du processus d'appauvrissement linguistique en cours.

3.6. Les données d'entraînement : source de biais et d'uniformisation ?

Les corpus utilisés pour l'entraînement des modèles – majoritairement extraits du web et largement dominés par des données anglophones – participent de façon déterminante à l'uniformisation des productions. La qualité hétérogène et le bruit inhérent à ces données induisent un biais qui se traduit par la reproduction de plusieurs nuances d'un langage standardisé, souvent dépourvu de la richesse des particularités régionales propre au français. Cette situation favorise la perpétuation d'un modèle normatif où les expressions calquées où les formules toutes faites règnent en maîtres.

3.7. « IA-Langage » et normalisation potentielle des usages

La tendance observée à privilégier des schémas prédéfinis et à reproduire inlassablement des structures simplifiées traduirait un phénomène de normalisation potentielle des usages du français. En effet, l'omniprésence d'expressions banales et passe-partout, de constructions syntaxiques rigides voire « machinique », c'est-à-dire se rapportant uniquement au cadre de la logique algorithmique, montre que l'IA tend à uniformiser le discours, ce qui élimine progressivement les variations, styles du langage et la créativité qui animent les langues humaines.

3.8. Implications sociétales et enjeux de la langue française

La standardisation du discours, avec ses conséquences d'appauvrissement lexical et stylistique, a des répercussions majeures sur la vitalité et la capacité expressive du français. Dans la sphère éducative, par exemple, le risque est d'habituer de jeunes apprenants à des formes uniformisées, une tendance qui limitera leur développement critique et leur capacité à s'exprimer de manière nuancée à l'écrit et peut-être même à l'oral. Au niveau sociétal, la simplification systématique du discours pourrait conduire à une homogénéisation culturelle, où la richesse des identités régionales et des patrimoines locaux serait mise en péril, cela pourrait de la sorte, compromettre l'innovation et la cohésion dans un environnement numérique en pleine mutation.

3.9. Risques de standardisation et d'appauvrissement du français

Les observations de cette étude montrent clairement que l'homogénéisation générée par l'IA, traduite par une répétition d'expressions calquées, une banalisation du vocabulaire et une récurrence des structures simplifiées, pose certainement un risque réel d'appauvrissement du français. Ce processus, qui limite l'innovation linguistique et réprime la diversité culturelle, s'inscrit dans une logique de perte progressive de la capacité d'expression et de la richesse propre à la langue française qui a mis des siècles à se développer.

3.10. La fracture numérique et ses conséquences linguistiques dans la francophonie

Une autre dimension critique de notre analyse concerne la fracture numérique. Dans des régions où l'accès aux outils technologiques de pointe est limité, l'exposition à des modèles moins performants (gratuits) au détriment des modèles plus performants (payants par voie d'abonnement) accentuent la diffusion de textes standardisés et appauvris. Ce déséquilibre contribue à une homogénéisation encore plus marquée du discours, ce qui aggrave par la même occasion, les disparités linguistiques déjà présentes au sein de la francophonie. Ce cas s'applique malheureusement au contexte des pays en développement où les populations vivent sous le seuil de pauvreté avec parfois moins d'un dollar par jour.

3.11. Hyper-mécanisation du texte, model collapse et langage hybride : synthèse des risques et implications pour l'avenir du français

La combinaison de tous ces phénomènes révèle un processus d'hyper-mécanisation, ou « model collapse », dans lequel l'IA se contente uniquement de reproduire de manière systématique des schémas prédéfinis ex nihilo. Ce mécanisme aboutit à l'émergence d'un « langage hybride » dans lequel les caractéristiques du langage informatique – telles que l'usage répétitif d'expressions passe-partout, les inversions syntaxiques et les structures mécaniques – se substituent à l'expression naturelle et enrichie par des contexte régionaux, culturels etc. propre à la production humaine. Si cette tendance se confirme, l'avenir du français risque d'être marqué par une uniformisation qui compromettra sa capacité à évoluer et à refléter ses héritages culturels.

3.12. Discussions au prisme des théories

Le prisme théorique offre un éclairage essentiel sur nos constats. L'approche de la chambre chinoise (Searle, 1980) a clairement souligné que, sans compréhension sémantique, l'IA se contente de manipuler des symboles, dans l'intention de reproduire des formes standardisées et banales. Parallèlement, la théorie du capital linguistique de Bourdieu met en avant le risque de domination culturelle associé à l'imposition d'un français normé, où l'uniformisation – matérialisée par l'usage d'expressions toutes faites et de structures simplifiées – contribue à marginaliser les richesses linguistiques régionales et à appauvrir l'expression créative.

3.13. Confrontation aux apports des travaux de référence

De plus, nos résultats s'inscrivent en continuité avec ceux de Côté et Su (2022) et Véry et Cailluet (2019) qui avaient déjà mis en exergue l'impact transformateur, mais potentiellement déstabilisateur, de l'IA sur le paysage linguistique. Les analyses d'Aberkane présentées lors d'une conférence (2023), quant à elles, soulignent la rapidité de la production automatisée et ses risques d'uniformisation, tandis que les rapports de l'Académie de Paris (2025) et de l'Agence Française de Développement (AFD, 2025) appellent à la préservation des diversités régionales. De plus, les travaux de Dagnogo et Samasse (2022), Kadio (2024) et Sinapin (2020), sans omettre les réflexions proposées lors du Sommet Mondial de l'intelligence artificielle de Paris (Elysee.fr, 2025) qui confirment, implicitement ce retard en la matière au

sujet de l'IA dans l'espace francophone, et montre que la tendance observée par nos soins n'est pas anodine.

3.14. Limites et perspectives de l'étude

Malgré la richesse des résultats obtenus, notre étude présente certaines limites. L'échantillon de textes et le nombre de répondants restent susceptibles d'être élargis, en plus d'une mise à jour de la méthodologie par d'autres chercheurs pour offrir une vision encore plus fine des mécanismes en présence. Par ailleurs, bien que notre approche qualitative permette une analyse globale, il serait pertinent de l'enrichir par des indicateurs quantitatifs spécifiques (mesures de diversité lexicale et syntaxique) par des experts des sciences du langage afin de mieux quantifier l'ampleur de l'appauvrissement linguistique observé.

3.15. Orientations pour la recherche future

À la lumière de nos constats, la recherche future devra s'orienter vers la diversification des corpus d'entraînement, intégrer des sources authentiques et régionales afin de réduire l'emprise des données anglophones. Cette hantise, propose le développement d'outils d'évaluation sophistiqués, capables de détecter et de mesurer la diversité linguistique, ce qui sera essentiel pour affiner ces analyses. En outre, l'exploration d'approches hybrides, qui combinent l'efficacité des systèmes d'IA avec la régulation et la correction apportées par l'intervention humaine, constitue une piste prometteuse pour limiter la standardisation excessive et préserver la richesse du discours. Toutefois, le savoir-faire des utilisateurs et une bonne maîtrise de la langue savamment apprise, reste un atout majeur.

3.16. Nos recommandations

Face aux risques identifiés, il apparaît urgent de mettre en place des stratégies globales, notamment pour les pays en développement que nous jugeons plus exposés aux risques. Il est notamment recommandé que les développeurs privilégient l'enrichissement des corpus d'entraînement avec des contenus issus de contextes culturels variés, afin d'atténuer l'effet « uniformisant » observé. Sur le plan pédagogique, l'idée de former et sensibiliser les utilisateurs à une littératie critique vis-à-vis des textes générés par l'IA est indispensable pour prévenir un apprentissage superficiel du français, tant chez les apprenants que chez les enseignants. Enfin, la

collaboration entre chercheurs, institutions culturelles locales et décideurs politiques est décisive pour développer des dispositifs de post-édition hybride, qui permettent d'introduire la nuance et la créativité dans les productions automatisées, une approche qui pourrait assurer une transmission fiable et durable du patrimoine linguistique.

Conclusion

L'Intelligence Artificielle Générative influence effectivement l'apprentissage des langues à travers diverses perspectives, enjeux et nouvelles pratiques pédagogiques (Institut intelligence et données (IDD), 2024; Razouki et al., 2025). Notamment, les productions textuelles générées par l'IA se distinguent, de par leur forme et leur contenu, des productions humaines traditionnelles. L'ensemble des résultats révèle que la mécanique de l'IA conduit à une standardisation du langage et, par conséquent, à un appauvrissement progressif de la richesse lexicale, stylistique et culturelle du français. Les contributions des experts et non-experts confirment que seuls ceux ayant une sensibilité linguistique développée perçoivent l'ampleur des dérives. Notre étude a démontré que, bien que l'IA générative offre une production rapide de textes, elle entraîne une standardisation et un appauvrissement remarquables du français. Les productions automatisées comportent davantage d'erreurs lexicales et syntaxiques, une récurrence marquée d'expressions passe-partout et une homogénéisation des structures, traduisant un « model collapse » qui réduit progressivement la richesse et la diversité de la langue. Enfin de compte, l'essor de l'IA générative, s'il n'est pas encadré, risque de compromettre la richesse et la diversité du français.

Les données, corroborées par les retours d'experts et éclairées par l'approche de la Chambre Chinoise (Searle, 1980) couplée par la théorie du Capital Linguistique de Bourdieu, confirment que les mécanismes d'apprentissage des IA privilégient, déjà, des schémas standardisés issus de corpus biaisés. Lesquels corpus, en provenance de sources variées (langues, cultures, idéologies, pensées propagandistes, etc.) et multiformes (textes, vidéos, sons, images, etc.), seront de nouveau utilisés en boucle pour les apprentissages futurs des LLM et la régénération de texte (à l'infini) par ces derniers. Visiblement, il s'agit là peut-être d'une « obsolescence programmée » de nos langues vivantes (à l'écrit comme à l'oral), conduite par l'IA Générative et le boom technologique qui s'y rattache. Cette constatation rejoint également d'autres travaux antérieurs, qui avaient déjà alerté sur les risques d'uniformisation. En ce sens, nos hypothèses initiales ont été vérifiées. Cependant, notre approche qualitative présente des limites qui pourraient être complétées par des outils quantitatifs afin de mesurer

précisément l'ampleur de l'appauvrissement. Pour pallier ces risques, nous recommandons d'enrichir les corpus d'entraînement avec des sources diversifiées, de développer des indicateurs spécifiques de diversité linguistique et de promouvoir des méthodes hybrides qui associent continuellement l'IA à une intervention humaine critique. Par ailleurs, il apparaît essentiel de renforcer la littératie numérique pour que tous puissent détecter et corriger ces dérives. Une intervention proactive à la fois technologique et éducative est donc indispensable pour garantir que cet outil serve à enrichir plutôt qu'à appauvrir notre langue dans un environnement numérique en mutation pour éviter de compromettre la diversité culturelle et accentuer les risques de standardisation.

BIBLIOGRAPHIE

Académie de Paris. (2025, janvier). L'intelligence artificielle dans l'éducation. Académie de Paris. Disponible sur : <https://www.ac-paris.fr/l-intelligence-artificielle-dans-l-education-130992>. Consulté le : 30 Avril 2025.

AFD. (2025, janvier 23). L'intelligence artificielle, moteur des transformations africaines. Disponible sur : <https://www.afd.fr/fr/actualites/intelligence-artificielle-transformations-africaines>. Consulté le : 30 Avril 2025.

Côté, A.-M., & Su, Z. (2022). Évolutions de l'intelligence artificielle au travail et collaborations humain-machine. *Ad machina*, 5, 144- 160. Disponible sur : <https://doi.org/10.1522/radm.no5.1413>. 21 Mars 2025.

Dagnogo, G. L. B., & Samasse, A. (2022). Jeunesse de l'Université de Bouaké et socialisation, à l'ère de l'Intelligence Artificielle. *Communication, technologies et développement*, 11, Article 11. Disponible sur : <https://doi.org/10.4000/ctd.7529> . 17 Avril 2025.

Elysee.fr. (2025, février 6). Sommet pour l'action sur l'intelligence artificielle. elysee.fr. Disponible sur : <https://www.elysee.fr/emmanuel-macron/sommet-pour-laction-sur-lintelligence-artificielle>

Idriss J. Aberkane (Réalisateur). (2023, juillet 4). Conférence : Le Futur de l'Education face à l'Intelligence Artificielle | Idriss Aberkane [Enregistrement vidéo]. Disponible sur : <https://www.youtube.com/watch?v=QaDilpBgqFk> . 20 Avril 2025.

Kadio, O. É. (2024). De l'Intelligence Artificielle et de la science des données pour aider à transformer l'Afrique. Communication, technologies et développement, 15, Article 15. Disponible sur : <https://doi.org/10.4000/123is> . Consulté le : 20 Mars 2025.

-Martin, Justine, Soltani, El-Mehdi et Yao, Jean-Marc Yao, (2023), Numéro -Spécial- Varia-, Contextes Didactiques, Linguistiques et Culturels [En ligne], 1(2),570P. Disponible sur le lien : <https://www.asjp.cerist.dz/en/PresentationRevue/928>

-Kalai, Lassaad ; Husson, Daniel et Soltani, El-Mehdi (2024), Numéro –Thématique « [Langues, cultures et enjeux contextuels : quels rapports ?](#)», Contextes Didactiques, Linguistiques et Culturels [En ligne], 2(3), 539 p. Disponible sur le lien : <https://www.asjp.cerist.dz/en/PresentationRevue/928>

-Ochi, Khaled et Legros, Denis, (2024), Numéro –Thématique « Traduction et Multiculturalisme », Contextes Didactiques, Linguistiques et Culturels [En ligne], 2(1), 320p.Disponible sur le lien : <https://www.asjp.cerist.dz/en/PresentationRevue/928>

-Sinapin, M. N. (2020, septembre). L'intelligence artificielle : Entre opportunités et risques légitimes. Oriane 2020 : 18e colloque francophone sur le risque. Disponible sur : <https://hal.science/hal-02950105> . Consulté le : 20 Mars 2025.

-Soltani, E. M. (2019). Lu pour vous : Diagonales de la communication interculturelle, de Martine Abadallah-Pretceille et Louis Porcher », Traduire(240), pp. 122-124. Disponible sur le lien : <http://journals.openedition.org/traduire/1751>

-Soltani E. M. 2020. « Compte-rendu d'ouvrage : BEACOO, J, C. COSTE, D.2017. L'éducation plurilingue et interculturelle : La perspective du Conseil de l'Europe, Paris, Les éditions Didier, 298 Pages», dans Revue algérienne des Lettres, 3(3), 250-256.

-Véry, P., & Cailluet, L. (2019). Intelligence artificielle et recherche en gestion. Revue française de gestion, 285(8), 119- 134. Disponible sur : <https://shs.cairn.info/revue-francaise-de-gestion-2019-8-page-119?lang=fr> . Consulté le : 21 Mars 2025.