

Reçu le 20-03-2026

Accepté le 23-04-202

Collocations et biais linguistiques dans l'IA générative : analyse lexico-sémantique du discours journalistique algérien d'expression française

Collocations and Linguistic Biases in Generative AI: A Lexico-Semantic Analysis of Algerian French-Language Journalistic Discourse

Fatima Zohra ALLAL¹, Zahira BENMESSABIH²

¹ Université de Naama, Algérie, f.allal@cuniv-naama.dz

² Université de Mascara, Algérie, zahira.benmessabih@univ-mascara.dz

Résumé

Cette étude examine la représentation du discours journalistique algérien d'expression française dans les productions d'intelligence artificielle générative. Elle s'inscrit dans le cadre théorique de la linguistique de corpus, de l'analyse lexicosémantique des collocations et des travaux portant sur les biais linguistiques dans les systèmes d'IA. L'objectif est d'analyser dans quelle mesure les modèles génératifs sont capables de reproduire les structures lexicales caractéristiques du journalisme algérien francophone et d'identifier les éventuelles formes de standardisation linguistique. Le corpus étudié est constitué de deux chroniques appartenant à deux générations de journalistes : une chronique de Tahar Djaout, représentant du journalisme intellectuel de l'après-indépendance, et une chronique de Khaled Drareni, journaliste contemporain engagé dans l'analyse critique de l'actualité politique et sociale. L'analyse repose sur une approche lexicosémantique qualitative des collocations, appuyée par l'utilisation du logiciel d'analyse textuelle AntConc pour repérer les cooccurrences lexicales et les concordances dans le corpus. Les collocations identifiées ont ensuite été comparées à celles présentes dans un texte généré par un modèle d'IA afin d'évaluer leur reproduction. Les résultats montrent que les chroniques étudiées présentent des orientations lexicales distinctes : chez Djaout, les collocations dominantes relèvent d'un registre conceptuel et métaphorique, tandis que celles de Drareni appartiennent davantage à un registre institutionnel et politique. Sur l'ensemble des collocations analysées, seule une partie est reproduite dans le texte généré par l'IA, principalement celles associées au registre institutionnel. Les collocations métaphoriques et culturellement marquées sont peu ou pas restituées. Ces résultats mettent en évidence une tendance à la standardisation du français journalistique dans les productions d'IA, révélant des biais linguistiques susceptibles de marginaliser certaines spécificités stylistiques et culturelles du journalisme algérien francophone. L'étude souligne ainsi l'importance d'intégrer la diversité linguistique dans le développement des systèmes d'IA générative.

¹ Auteur correspondant

Mots-clés : collocations lexicosémantiques, journalisme algérien francophone, intelligence artificielle générative, biais linguistiques, analyse de corpus

Abstract

This study examines the representation of Algerian French-language journalistic discourse in the outputs of generative artificial intelligence. It is situated within the theoretical framework of corpus linguistics, lexicosemantic analysis of collocations, and research on linguistic biases in AI systems. The objective is to analyze the extent to which generative models are able to reproduce the lexical structures characteristic of Algerian francophone journalism and to identify possible forms of linguistic standardization. The corpus studied consists of two opinion columns written by journalists from two different generations: a column by Tahar Djaout, representing the intellectual journalism of the post-independence period, and a column by Khaled Drareni, a contemporary journalist engaged in the critical analysis of political and social current affairs. The analysis is based on a qualitative lexicosemantic approach to collocations, supported by the use of the text analysis software AntConc to identify lexical co-occurrences and concordances within the corpus. The collocations identified were then compared with those present in a text generated by an AI model in order to evaluate their reproduction. The results show that the two columns display distinct lexical orientations. In Djaout's writing, the dominant collocations belong to a conceptual and metaphorical register, whereas Drareni's collocations tend to belong to a more institutional and political register. Among all the collocations analyzed, only a portion are reproduced in the AI-generated text, mainly those associated with the institutional register. Metaphorical and culturally marked collocations are rarely or not reproduced. These results highlight a tendency toward the standardization of journalistic French in AI-generated outputs, revealing linguistic biases that may marginalize certain stylistic and cultural specificities of Algerian francophone journalism. The study therefore emphasizes the importance of integrating linguistic diversity into the development of generative AI systems.

Keywords : lexico-semantic collocations, Algerian French-language journalism, generative artificial intelligence, linguistic biases, corpus analysis

Introduction

L'essor récent des technologies d'intelligence artificielle générative transforme profondément les pratiques de production et de diffusion des discours dans de nombreux domaines, notamment dans les sphères médiatiques, éducatives et informationnelles. Les modèles de langage de grande taille, capables de produire des textes cohérents et structurés dans plusieurs langues, participent désormais à la circulation des savoirs et à la médiation de l'information. Toutefois, cette expansion technologique soulève des interrogations importantes quant à la représentation de la diversité linguistique et culturelle dans les systèmes d'intelligence artificielle. En effet, plusieurs travaux ont montré que les modèles de langage sont fortement dépendants des corpus qui ont servi à leur entraînement, ce qui peut entraîner une surreprésentation des variétés linguistiques dominantes et, corrélativement, une marginalisation des usages périphériques ou régionaux (Bender *et al.*, 2021 ; Blodgett *et al.*, 2020).

Dans ce contexte, la question de l'équité linguistique devient centrale, en particulier pour les espaces plurilingues où coexistent plusieurs langues, variétés et pratiques discursives. Le cas de l'Algérie

constitue à cet égard un terrain d'observation particulièrement pertinent. L'espace médiatique algérien se caractérise par une coexistence de plusieurs langues et registres discursifs, parmi lesquels le français occupe une place importante dans le champ journalistique. Le discours journalistique algérien d'expression française présente ainsi des spécificités lexicales, stylistiques et culturelles qui reflètent les réalités sociopolitiques locales et les trajectoires historiques du pays.

Parmi ces spécificités, les collocations occupent une place centrale dans la structuration du sens et dans l'organisation du discours. Comme le souligne John Rupert Firth, « You shall know a word by the company it keeps » (Firth, 1957, p.53), rappelant que le sens lexical se construit largement à travers les associations régulières entre les mots. Les collocations constituent ainsi des unités privilégiées pour l'analyse du fonctionnement discursif et stylistique des textes journalistiques. Elles permettent notamment de mettre en évidence des régularités lexicales qui traduisent des positionnements idéologiques, des représentations sociales ou des cadres interprétatifs propres à un contexte donné (Sinclair, 1991).

Dans le contexte du journalisme algérien, ces structures collocationnelles peuvent refléter des préoccupations politiques, sociales et identitaires spécifiques. L'analyse des chroniques de figures journalistiques appartenant à des générations différentes, telles que Tahar Djaout et Khaled Drareni, permet ainsi d'observer l'évolution des pratiques discursives et des configurations lexicales dans le temps. Tandis que les chroniques de Djaout s'inscrivent dans le contexte intellectuel et politique de l'Algérie post-indépendance, celles de Drareni témoignent des transformations contemporaines du champ médiatique et des nouvelles formes d'engagement journalistique.

Dans ce cadre, l'émergence des systèmes d'intelligence artificielle générative soulève une question centrale : Dans quelle mesure les systèmes d'intelligence artificielle générative sont-ils capables de reproduire les collocations caractéristiques du discours journalistique algérien d'expression française, et quels biais linguistiques peuvent émerger dans la représentation de ces pratiques discursives ? En effet, si les modèles de langage sont principalement entraînés sur des corpus dominés par des variétés standardisées du français, ils pourraient tendre à normaliser ou à neutraliser certaines particularités linguistiques propres aux espaces francophones périphériques.

Dès lors, cette contribution se propose d'examiner les relations entre intelligence artificielle générative et diversité linguistique à partir d'une analyse lexico-sémantique des collocations dans le discours journalistique algérien d'expression française. Plus précisément, elle vise à comparer les structures collocationnelles observées dans un corpus de chroniques journalistiques avec celles produites par des systèmes d'IA générative afin d'identifier d'éventuelles divergences, transformations ou formes de standardisation linguistique.

La problématique centrale de cette recherche se formule comme suit :

Dans quelle mesure les systèmes d'intelligence artificielle générative sont-ils capables de reproduire les collocations caractéristiques du discours journalistique algérien d'expression française, et quels biais linguistiques peuvent émerger dans la représentation de ces pratiques discursives ?

Cette réflexion s'inscrit dans une perspective plus large portant sur les enjeux d'inclusion linguistique, de diversité discursive et de représentation des variétés francophones dans les technologies contemporaines du langage. Pour répondre à cette problématique, nous formulons les deux hypothèses suivantes :

H1: Les systèmes d'IA générative opéreraient une forme de "neutralisation stylistique" en substituant les collocations à forte charge métaphorique ou culturelle, spécifiques au champ journalistique algérien, par des structures lexicales standardisées et universelles.

H2 : La fidélité de reproduction de l'IA serait corrélée à la fréquence des registres dans les corpus d'entraînement mondiaux, favorisant ainsi les collocations institutionnelles et normées au détriment des collocations liées à l'engagement intellectuel et philosophique local.

1. Cadre théorique

1.1. La collocation

La collocation, concept fondamental en linguistique, désigne l'association régulière et significative de mots dans un corpus, permettant de révéler les liens lexicaux et les régularités sémantiques propres à une langue ou à un genre textuel donné. Comme le souligne John Rupert Firth : « Vous connaîtrez un mot par la compagnie qu'il fréquente ». (Rupert Firth ,1957, p.35) rappelant que le sens d'un mot ne peut être pleinement compris isolément, mais se construit largement à travers ses cooccurrences dans le discours. Cette perspective met l'accent sur l'importance de l'usage contextuel et sur le rôle central de la collocation dans la compréhension lexicale.

Dans le domaine du journalisme, les collocations jouent un rôle crucial dans l'organisation du texte, la cohérence argumentative et la transmission de valeurs sociales et idéologiques. Comme le précise Sinclair : « Les collocations représentent des unités de sens récurrentes qui structurent le discours et reflètent les conventions d'un registre donné ». (Sinclair ,1991, p71). Cette observation est particulièrement pertinente pour notre étude, car elle permet d'identifier non seulement les thèmes dominants du discours journalistique, mais également les nuances stylistiques qui différencient les générations de chroniqueurs et leurs positionnements idéologiques. Par exemple, les collocations philosophiques et engagées de Djaout contrastent avec le registre normé de Drareni, révélant ainsi les variations générationnelles dans l'expression des idées politiques et sociales.

De plus, Stubbs souligne que : « La cooccurrence habituelle des mots reflète les schémas sociaux et cognitifs intégrés dans l'usage de la langue ». (Stubbs,2001, p.41). Cette idée met en évidence que les collocations ne sont pas de simples structures lexicales neutres, mais traduisent également les pratiques culturelles, sociales et cognitives des locuteurs. Dans le contexte algérien, elles permettent de comprendre comment les journalistes expriment, à travers leur choix lexical, des valeurs telles que la liberté, la responsabilité citoyenne et l'engagement politique, tout en reflétant leur génération et leur expérience historique.

Dans notre étude, les collocations identifiées dans les chroniques de Tahar Djaout (défendre les libertés, élever la voix) et de Khaled Drareni (liberté de la presse, mobilisation citoyenne) permettent de mettre en évidence :

- Les thèmes centraux de leurs discours (liberté, responsabilité, citoyenneté),

- Les régularités lexicales qui traduisent leur engagement journalistique,
- Et les variations générationnelles dans l'expression des idées, entre un discours philosophique et intellectuel pour Djaout et un discours institutionnel et contemporain pour Drareni.

Ainsi, l'analyse des collocations constitue un outil privilégié pour explorer la dimension lexicosémantique du discours journalistique algérien francophone, et elle offre également un cadre méthodologique solide pour comparer les productions humaines aux textes générés par l'intelligence artificielle. Cette comparaison permet de mettre en lumière les continuités, les divergences et les biais linguistiques, contribuant ainsi à une réflexion sur l'inclusion et l'équité linguistique dans les systèmes d'IA.

1.2. Lexico-sémantique

La lexicosémantique est une discipline qui étudie le sens des unités lexicales et leurs combinaisons dans un texte ou un corpus. Elle s'intéresse non seulement à la signification des mots isolés, mais surtout aux relations entre les mots, à leurs cooccurrences et aux variations contextuelles qui façonnent le sens. Cette approche est particulièrement adaptée à l'analyse journalistique, où le choix lexical et les associations de mots révèlent souvent des positionnements idéologiques, des valeurs culturelles et des représentations sociales.

Selon Cruse, 2000, p.27 : « La sémantique lexicale s'intéresse au sens des mots et aux relations systématiques qui existent entre eux. ». Cette citation souligne que l'étude des relations lexicales permet de détecter des structures récurrentes et des régularités dans le discours. Dans le contexte de notre étude, cela justifie l'analyse des collocations dans les chroniques de Djaout et Drareni, car les associations lexicales révèlent des tendances stylistiques et idéologiques propres à chaque génération de journaliste.

D'autre part, Fellbaum, 1998, p.37 insiste sur l'importance des réseaux lexicaux : « Les mots n'existent pas isolément ; leur sens se définit à travers leurs relations avec d'autres mots dans un réseau ». Cette idée est directement applicable à l'analyse des chroniques algériennes. Par exemple, dans le texte de Djaout, les mots liberté, raison et pensée forment un réseau sémantique qui renforce le message intellectuel et engagé de la chronique. Chez Drareni, le réseau lexical autour de liberté de la presse, citoyen et mobilisation traduit un engagement plus institutionnel et collectif.

Enfin, Cruse, 2004, p.47 met en avant la dimension contextuelle de la lexicosémantique : « Le sens d'un mot dépend toujours du contexte, et le sens lexical peut varier selon la situation communicative ». Dans notre étude, cette observation justifie la comparaison intergénérationnelle et intertextuelle : les collocations de Djaout et Drareni, bien que partageant des thèmes communs comme la liberté, se manifestent différemment selon le contexte historique et social. De plus, elle explique pourquoi les productions générées par l'IA peuvent manquer certaines nuances contextuelles propres aux textes originaux. L'analyse lexicosémantique des chroniques permet de :

- identifier les réseaux de collocations qui structurent le discours,
- repérer les variations générationnelles dans le choix lexical et la construction du sens,
- mettre en évidence la relation entre lexique et engagement idéologique,
- servir de référence pour l'évaluation des productions générées par l'IA, afin de détecter les biais et les écarts par rapport aux modèles humains.

Par exemple :

- Dans la chronique de Djaout, les mots *défendre, libertés, raison et pensée* forment un réseau qui traduit un engagement intellectuel et philosophique.
- Dans celle de Drareni, les mots *liberté de la presse, mobilisation, citoyen et institutions* construisent un réseau institutionnel et civique.
- Les textes générés par l'IA reproduisent partiellement ces réseaux, mais tendent à standardiser le lexique et à neutraliser les spécificités générationnelles et culturelles, illustrant les limites des modèles actuels.

Ainsi, la lexicosémantique offre un outil rigoureux pour comparer les discours humains et les productions de l'IA, tout en révélant les biais linguistiques et les variations contextuelles propres au journalisme algérien francophone. Elle constitue également un cadre théorique solide pour comprendre la diversité lexicale et l'équité linguistique dans les systèmes d'IA générative.

1.3. Linguistique de corpus

La linguistique de corpus est une approche scientifique qui consiste à analyser des données textuelles à grande échelle afin d'identifier des régularités lexicales, syntaxiques et sémantiques. Elle permet de révéler des tendances invisibles à l'œil nu et de fournir des bases empiriques solides pour l'analyse linguistique. Comme le rappellent Biber et Conrad, 2009, p.31: « La linguistique de corpus fournit une méthode pour étudier l'usage authentique de la langue et identifier les schémas d'organisation lexicale, grammaticale et textuelle ». Cette approche est particulièrement pertinente pour notre étude, car elle permet d'examiner les collocations et les structures lexicosémantiques dans les chroniques de Djaout et Drareni à partir de données réelles, plutôt que sur des intuitions subjectives. Elle établit ainsi un cadre méthodologique rigoureux pour l'analyse du français journalistique algérien.

Selon McEnery et Hardie ,2012, p.29 : « Les études basées sur les corpus permettent aux linguistes d'examiner la langue en contexte et de quantifier systématiquement les schémas linguistiques ». Cette citation souligne distribution des collocations, ainsi que pour comparer les usages linguistiques entre différentes générations de journalistes. Dans notre étude, cela permet d'identifier les particularités lexicales propres à Djaout et à Drareni, tout en mettant en évidence les variations générationnelles et stylistiques.

Enfin, Stubbs ,2001, p.53 met en avant l'importance du corpus pour détecter des régularités implicites : « Les corpus nous permettent de découvrir des schémas récurrents dans la langue qui ne sont pas toujours produits consciemment par les locuteurs ou les écrivains ». Cette observation est essentielle pour notre étude, car elle explique comment l'analyse de collocations et de structures lexicosémantiques dans les chroniques peut révéler des tendances stylistiques et idéologiques sous-jacentes, propres aux journalistes algériens, même lorsqu'elles ne sont pas explicitement formulées. Cela est également crucial pour comparer les productions humaines aux textes générés par l'IA, qui peuvent reproduire certaines régularités tout en omettant les nuances contextuelles.

Dans cette étude, nous appliquons la linguistique de corpus pour :

- Extraire les collocations récurrentes dans les chroniques de Djaout et Drareni, en identifiant les associations lexicales significatives qui reflètent leurs engagements et leur style journalistique,

- Quantifier leur fréquence et leur distribution, afin de déterminer les collocations dominantes et leur régularité dans le corpus,
- Comparer les modèles lexicaux générationnels, en mettant en évidence les différences entre le français journalistique post-indépendance de Djaout et le style contemporain de Drareni, tout en tenant compte des particularités du français algérien,
- Établir un corpus de référence pour évaluer les productions générées par des systèmes d'IA et détecter les biais linguistiques, notamment la standardisation ou la neutralisation des collocations caractéristiques.

Ainsi, la linguistique de corpus fournit non seulement une méthodologie rigoureuse, mais elle permet également d'analyser de manière objective les réseaux lexicaux et les régularités stylistiques, tout en servant de base pour évaluer la fidélité et les limites des systèmes d'IA générative dans le traitement du français journalistique algérien.

1.4. Biais linguistiques dans l'IA

Les systèmes d'intelligence artificielle générative, tels que les modèles de traitement automatique du langage, reposent sur des corpus massifs pour apprendre les régularités lexicales et syntaxiques. Cependant, ces modèles présentent souvent des biais linguistiques et culturels, qui reflètent les tendances dominantes des données d'entraînement et peuvent marginaliser les variétés locales ou les usages spécifiques à certaines communautés linguistiques.

Comme le soulignent Bender et *al*, 2021, p65 : « Les données sur lesquelles les modèles de langage sont entraînés codent les biais sociaux et les inégalités présents dans le monde ». Cette citation illustre que les systèmes d'IA ne sont pas neutres : les collocations et structures lexicosémantiques des textes générés peuvent refléter des normes linguistiques dominantes, au détriment des styles spécifiques comme ceux de Djaout ou Drareni. Dans le contexte algérien, cela peut se traduire par une standardisation du français journalistique, effaçant les nuances générationnelles et culturelles.

Selon Blodgett et *al*, 2020, p.71 : « Les technologies linguistiques reproduisent et amplifient souvent les biais intégrés dans les données linguistiques sur lesquelles elles sont entraînées, affectant de manière disproportionnée les langues et dialectes sous-représentés ». Cette remarque est directement applicable à notre étude : le français algérien journalistique, en particulier dans les textes de Djaout, contient des expressions et collocations spécifiques qui peuvent être sous-représentées dans les corpus internationaux utilisés pour entraîner les IA. Cela explique pourquoi certaines collocations philosophiques ou engagées ne sont pas reproduites par l'IA.

Hovy et Spruit, 2016, p.69 ajoutent : « Le biais dans les systèmes de traitement automatique du langage n'est pas seulement un problème technique, mais également un problème social et éthique ayant des conséquences sur l'équité et l'inclusion ». Dans le cadre éducatif et journalistique, cela signifie que l'usage de l'IA pour générer du texte doit être réfléchi afin de préserver la diversité linguistique et culturelle. Les collocations caractéristiques de chaque chroniqueur doivent être analysées et comparées pour identifier les biais et promouvoir l'inclusion des pratiques linguistiques locales.

Enfin, Sheng et *al*, 2019, p.19 précisent : « Les modèles de langage peuvent renforcer involontairement les stéréotypes et exclure les variantes linguistiques moins fréquentes dans les données d'entraînement ». Cela est directement pertinent pour notre comparaison entre les textes de Djaout et Drareni et les productions de l'IA. Certaines collocations et expressions propres au français algérien

ou au style engagé de Djaout peuvent ne pas apparaître, révélant un biais de représentativité et un manque d'inclusion linguistique.

L'analyse des biais linguistiques dans les productions générées par l'IA permet de :

- Identifier les collocations qui sont reproduites, celles qui sont neutralisées ou transformées ;
- Évaluer dans quelle mesure l'IA reflète ou déforme les particularités générationnelles et stylistiques des chroniqueurs ;
- Souligner les enjeux d'équité linguistique et culturelle, en particulier pour les langues et variétés sous-représentées dans les corpus d'entraînement ;
- Proposer des stratégies de médiation et d'inclusion, comme l'enrichissement des corpus d'entraînement avec des textes locaux et plurilingues pour préserver la diversité linguistique.

Ainsi, cette section montre que les biais linguistiques de l'IA ne sont pas uniquement un problème technique, mais qu'ils ont des conséquences directes sur la représentation du discours journalistique et sur la diversité lexicosémantique, renforçant la nécessité d'une approche critique et contextuelle dans l'étude du français algérien et de ses pratiques journalistiques.

2. Méthodologie

2.1. Présentation du corpus

Le corpus de cette étude est constitué de deux chroniques journalistiques emblématiques, représentatives de deux générations du journalisme algérien francophone, ainsi qu'un texte généré par l'IA pour comparaison.

a) Chronique de Tahar Djaout

Titre : L'intellectuel et la peur

Auteur : Tahar Djaout

Support : Algérie Actualité, fin des années 1980

Genre : chronique politique et intellectuelle

Thématique : liberté intellectuelle, rôle de l'écrivain dans la société, pression politique

Extrait :

« Ceux qui refusent de s'accommoder de la défaite de la raison et du sommeil de la pensée, ceux qui élèvent aujourd'hui la voix pour défendre les libertés, ces hommes et ces femmes sont l'honneur de l'Algérie ». Cette chronique présente des collocations fortement marquées, révélatrices des préoccupations identitaires et politiques de Djaout, telles que défendre les libertés, élever la voix, sommeil de la pensée. Ces structures lexicosémantiques constituent des points de référence pour comparer les productions de l'IA.

b) Chronique de Khaled Drareni

Titre : La liberté de la presse en Algérie

Auteur : Khaled Drareni

Support : Radio M, 2019–2020

Genre : chronique politique et médiatique

Thématique : liberté de la presse, engagement citoyen, rôle du journaliste

Extrait :

« La liberté de la presse demeure un indicateur essentiel de l'état démocratique d'un pays et un espace d'expression pour les citoyens. ». Cette chronique illustre des collocations caractéristiques du journalisme contemporain algérien, telles que liberté de la presse, espace d'expression, mobilisation citoyenne. Ces expressions permettront de mettre en évidence des divergences ou des continuités lexicales avec la génération de Djaout, et de tester leur reproduction par l'IA.

c) Texte généré par l'IA

Pour compléter le corpus, un texte généré par un modèle d'IA (ex : GPT-4) sera produit à partir d'un prompt basé sur la thématique liberté et rôle du journaliste en Algérie.

Prompt utilisé : « Rédigez une chronique journalistique sur la liberté de la presse en Algérie, avec un ton critique et engagé, similaire au style d'un journaliste algérien francophone. »

Extrait simulé par l'IA :

« La liberté de la presse reste le fondement de toute démocratie. Chaque citoyen doit avoir la possibilité de s'exprimer, de critiquer et de questionner le pouvoir. Les journalistes jouent un rôle essentiel dans la transparence et la responsabilité des institutions, mais trop souvent leur voix est étouffée par la censure et la pression politique. »

L'objectif est de comparer les collocations et les structures lexicosémantiques entre :

- le français journalistique de Djaout
- le français journalistique de Drareni
- les productions de l'IA

c. Extraction des collocations

L'extraction des collocations sera effectuée selon la méthodologie suivante :

- 1- **Identification lexicale** : sélection des mots clés liés aux thèmes centraux (liberté, presse, citoyen, pouvoir, société),
- 2- **Extraction statistique** : calcul des cooccurrences à l'aide de mesures de force lexicale (telles que MI-score, t-score) pour déterminer les associations significatives,
- 3- **Filtrage manuel** : vérification des collocations pour ne conserver que celles qui sont sémantiquement pertinentes dans le contexte journalistique,

Cette approche permet d'identifier les combinaisons lexicales qui structurent le discours et de repérer les régularités stylistiques propres à chaque chroniqueur.

d. Analyse lexicosémantique

L'analyse lexicosémantique portera sur :

- Fréquence et distribution des collocations
- **Analyse contextuelle** : rôle et sens des collocations dans le discours
- **Comparaison générationnelle** : Djaout vs Drareni

- **Analyse critique** : continuité ou divergence thématique et stylistique

e. Comparaison avec les productions de l'IA

L'étape finale consiste à confronter les collocations humaines avec celles générées par l'IA :

- Identification des collocations reproduites par l'IA
- Détection des écarts et biais (standardisation, neutralisation de certaines expressions locales ou générationnelles)
- Discussion sur l'impact des biais linguistiques de l'IA dans la représentation du français journalistique algérien

Cette méthodologie permet de mettre en lumière la relation entre diversité linguistique et intelligence artificielle, tout en offrant un cadre rigoureux pour l'analyse lexicosémantique.

3.Collocations principales de Tahar Djaout

Exemple de collocations relevées dans L'intellectuel et la peur :

Tableau 1 : Collocations de Tahar Djaout

Mot clé	Collocation(s)	Sens / Fonction
Liberté	défendre les libertés	met en valeur l'engagement civique et intellectuel
Voix	élever la voix	souligne l'action et la résistance face à l'oppression
Raison	défaite de la raison	critique la situation intellectuelle et politique
Pensée	sommeil de la pensée	dénonce l'inaction et la censure intellectuelle
Hommes / femmes	hommes et femmes	inclusion et responsabilité citoyenne

3.1. Analyse des collocations principales de Tahar Djaout

Le tableau des collocations extraites de la chronique L'intellectuel et la peur de Tahar Djaout révèle une organisation lexicale étroitement liée à son engagement intellectuel et politique. Chaque collocation identifiée constitue un indice de sens et un vecteur d'idéologie, permettant de comprendre comment le journaliste articule ses idées et son positionnement dans le contexte post-indépendance algérien.

La collocation « défendre les libertés » met en évidence la priorité accordée par Djaout à la liberté individuelle et collective. Elle traduit non seulement un engagement civique, mais également une lutte contre les restrictions et la censure imposées aux intellectuels de l'époque. Le mot clé liberté, combiné

à l'action défendre, construit un lexique dynamique, orienté vers l'action et la responsabilité personnelle.

La collocation « élever la voix » illustre l'action individuelle de résistance et la nécessité pour l'intellectuel de s'exprimer malgré les pressions politiques. Ici, le lexique fonctionne sur le plan pragmatique, incitant à l'intervention et à la prise de parole, ce qui est central dans le style argumentatif de Djaout.

Les expressions « défaite de la raison » et « sommeil de la pensée » montrent un registre plus philosophique et critique. Elles dénoncent la passivité intellectuelle et la stagnation de la réflexion critique dans la société, tout en soulignant l'importance de la vigilance et de l'engagement intellectuel. Ces collocations permettent d'identifier une dimension lexicosémantique profonde, où les termes choisis ne se limitent pas à leur sens littéral mais véhiculent des valeurs, des préoccupations et un jugement critique sur la société.

Enfin, la collocation « hommes et femmes » souligne une dimension inclusive et sociale, en plaçant l'action citoyenne au centre du discours. Elle traduit l'idée que la responsabilité de défendre les valeurs démocratiques et intellectuelles incombe à tous les membres de la société, hommes comme femmes.

Dans l'ensemble, ces collocations démontrent que le style de Djaout est argumentatif, engagé et fortement orienté vers l'action et la responsabilité individuelle. Le lexique choisi reflète un journalisme à la fois critique et philosophique, ancré dans le contexte post-indépendance algérien, où les enjeux de liberté, d'intellect et de citoyenneté sont centraux.

Cette analyse souligne également l'importance de l'examen des collocations comme indicateurs lexicosémantiques : elles permettent de saisir à la fois le contenu thématique et la posture idéologique de l'auteur, et servent de référence pour comparer le style de Djaout à celui de journalistes contemporains ou à des textes générés par l'IA.

3.2. Collocations principales de Khaled Drareni

Exemple de collocations relevées dans La liberté de la presse en Algérie :

Tableau 2 : Collocations de Khaled Drareni

Mot clé	Collocation(s)	Sens / Fonction
Liberté	liberté de la presse	focus sur la liberté médiatique et démocratique
Presse	liberté de la presse	point central de la critique journalistique
Espace	espace d'expression	souligne l'importance de la participation citoyenne
Citoyen	mobilisation citoyenne	engagement collectif et social

Pays	état démocratique	référence au cadre institutionnel et politique
-------------	-------------------	--

3.3. Analyse des collocations principales de Khaled Drareni

Le tableau des collocations relevées dans la chronique La liberté de la presse en Algérie de Khaled Drareni met en évidence un style journalistique contemporain, axé sur l'information, la critique institutionnelle et la participation citoyenne. Les collocations identifiées révèlent un lexique normé et institutionnalisé, reflétant les préoccupations d'un journaliste engagé dans un contexte démocratique moderne.

La collocation « liberté de la presse », associée aux mots clés liberté et presse, constitue le thème central du discours de Drareni. Elle traduit une attention particulière à l'indépendance médiatique et aux droits démocratiques, soulignant le rôle crucial de la presse dans le maintien de la transparence et de la responsabilité des institutions publiques. Contrairement au registre philosophique et engagé de Djaout, cette collocation illustre un engagement journalistique orienté vers les mécanismes institutionnels et sociaux.

La collocation « espace d'expression » met l'accent sur l'importance de la participation citoyenne et de l'inclusion sociale. Elle reflète la vision de Drareni d'une société où chaque citoyen doit pouvoir intervenir, critiquer et s'exprimer, ce qui est au cœur de son journalisme engagé mais dans un cadre collectif et institutionnel. Cette collocation traduit également un souci de valoriser l'espace public et démocratique où la liberté d'expression peut s'exercer.

La collocation « mobilisation citoyenne », associée au mot clé citoyen, insiste sur l'action collective et l'engagement social. Elle montre que, dans le style de Drareni, l'accent est mis sur le rôle actif des citoyens dans le fonctionnement de l'État démocratique, renforçant le lien entre liberté individuelle et responsabilité collective.

Enfin, la collocation « état démocratique », liée au mot clé pays, introduit une dimension institutionnelle et politique, rappelant que le journalisme contemporain ne se limite pas à la critique mais s'inscrit dans une réflexion sur les cadres légaux et sociaux qui régissent la liberté d'expression et le fonctionnement des institutions publiques.

Dans l'ensemble, ces collocations montrent que le style de Drareni est contemporain, normé et orienté vers l'information institutionnelle et la participation citoyenne. Le lexique utilisé est plus formel et structuré, reflétant la préoccupation pour les cadres démocratiques et les responsabilités collectives, en contraste avec le style philosophique, engagé et intellectuel de Djaout. Cette analyse illustre également l'importance de l'étude des collocations comme indicateurs lexicosémantiques, permettant de :

- mettre en lumière les priorités thématiques et idéologiques d'un journaliste contemporain ;
- identifier les réseaux lexicaux qui traduisent un engagement social et institutionnel ;
- servir de base pour comparer les productions humaines à celles générées par l'IA, et pour détecter les biais ou standardisations linguistiques.

Ainsi, l'analyse des collocations de Drareni complète celle de Djaout et prépare le terrain pour un comparatif générationnel et stylistique, en lien avec la dimension lexicosémantique et l'évaluation des biais dans les systèmes d'IA générative.

Tableau 3 : synthèse comparative des collocations de Djaout et de Drareni

Collocation / Mot clé	Djaout	Drareni	Commentaire
liberté	défendre les libertés	liberté de la presse	Djaout → liberté générale et intellectuelle ; Drareni → liberté médiatique
voix / expression	élever la voix	espace d'expression	Djaout → action personnelle ; Drareni → espace collectif
raison / pensée	défaite de la raison / sommeil de la pensée	—	Thème absent chez Drareni, génération plus pragmatique et institutionnelle
citoyen	hommes et femmes	mobilisation citoyenne	Djaout → inclusion sociale ; Drareni → engagement collectif
presse / pays	—	liberté de la presse / état démocratique	Spécificité contemporaine et institutionnelle chez Drareni

Analyse préliminaire

- **Différences générationnelles** : Djaout privilégie un lexique engagé, intellectuel et philosophique ; Drareni adopte un lexique journalistique contemporain centré sur la presse et la participation citoyenne.
- **Continuités** : le thème de la liberté reste central, mais il se manifeste sous des formes lexicales différentes.
- **Lien avec l'IA** : ces collocations serviront de référence pour vérifier si un modèle d'IA reproduit correctement :
 - le lexique engagé vs normatif
 - la spécificité culturelle et générationnelle
 - les collocations emblématiques de chaque chroniqueur

3.4. Collocations principales dans le texte IA

Tableau 4 : Collocations générées par L' IA

Mot clé	Collocation(s)	Sens / Fonction
liberté	liberté de la presse	renvoie à la liberté médiatique, similaire à Drareni
citoyen	chaque citoyen	engagement individuel et droit d'expression
voix	leur voix est étouffée	reproduction d'une expression proche du style Djaout
pouvoir	questionner le pouvoir	engagement critique envers l'autorité
presse	liberté de la presse	point central sur le rôle journalistique
institutions	responsabilité des institutions	dimension institutionnelle similaire à Drareni

3.5. Analyse des collocations principales dans le texte généré par l'IA

Le tableau des collocations relevées dans le texte généré par l'IA montre comment un modèle de traitement automatique du langage reproduit partiellement les structures lexicales et les thèmes présents dans les chroniques de Tahar Djaout et Khaled Drareni. Cette analyse permet de mettre en lumière à la fois les continuités et les limites des productions génératives en termes de lexique et de style journalistique.

La collocation « liberté de la presse » illustre que l'IA reproduit principalement le registre institutionnel et normé, proche du style de Drareni. Le modèle met l'accent sur la liberté médiatique comme valeur centrale, mais sans restituer la profondeur philosophique et engagée que l'on retrouve chez Djaout avec ses expressions telles que défendre les libertés ou sommeil de la pensée.

La collocation « chaque citoyen », associée au mot clé citoyen, traduit un engagement individuel et un droit d'expression. Elle reflète une dimension civique similaire aux préoccupations de Djaout concernant la responsabilité de chaque individu, mais elle est formulée dans un registre plus généralisé et standardisé, sans la nuance générationnelle ou stylistique propre aux chroniques originales.

La collocation « leur voix est étouffée » illustre une tentative de l'IA de reproduire le style engagé et critique de Djaout. Cependant, cette reproduction reste partielle et simplifiée, car la profondeur lexicosémantique et la portée philosophique des collocations de Djaout sont atténuées. Cela montre que l'IA peut imiter des structures lexicales critiques, mais ne saisit pas pleinement la richesse stylistique et générationnelle du texte humain.

Les collocations « questionner le pouvoir », « responsabilité des institutions » et « liberté de la presse » soulignent une forte orientation vers les thèmes institutionnels et civiques, similaires au style contemporain de Drareni. L'IA reproduit donc correctement le lexique normé et la critique institutionnelle, mais elle tend à standardiser le discours, combinant des éléments du style de Djaout et de Drareni sans restituer les spécificités générationnelles de chacun.

Dans l'ensemble, ces collocations révèlent plusieurs aspects clés :

- 1- **Particularités reproduites** : L'IA peut générer des collocations correspondant aux thèmes de liberté, citoyenneté et responsabilité institutionnelle, reflétant le style journalistique contemporain.
- 2- **Limitations stylistiques** : La profondeur lexicosémantique et les registres engagés ou philosophiques propres à Djaout sont souvent absents ou simplifiés.
- 3- **Biais de standardisation** : Le texte généré privilégie des collocations institutionnelles et universelles, ce qui illustre les biais linguistiques liés aux corpus d'entraînement dominants.
- 4- **Hybridation des styles** : Certaines expressions (leur voix est étouffée) montrent une combinaison artificielle des registres de Djaout et Drareni, mais sans la richesse contextuelle originale.

Ainsi, l'analyse lexicosémantique des collocations générées par l'IA permet non seulement d'identifier les continuités thématiques avec le journalisme humain, mais aussi de mettre en évidence les biais linguistiques et la perte des nuances générationnelles. Cette observation est cruciale pour évaluer l'impact des systèmes d'IA sur la représentation des discours journalistiques et la diversité lexicale dans un contexte plurilingue et culturellement spécifique comme celui de l'Algérie francophone.

Tableau comparatif final : Djaout, Drareni, IA

Tableau 5 : synthèse comparative entre les collocations de Djaout, Drareni et L'IA

Collocation / Mot clé	Djaout	Drareni	IA générative	Commentaire / Analyse
liberté	défendre les libertés	liberté de la presse	liberté de la presse	L'IA adopte le registre institutionnel de Drareni, standardisé et normé, moins philosophique et engagé que le lexique de Djaout, illustrant une perte de nuance générationnelle et lexicosémantique.
voix / expression	élever la voix	espace d'expression	leur voix est étouffée	L'IA combine partiellement la dimension personnelle et militante de Djaout avec la dimension collective/institutionnelle de Drareni, mais la profondeur

				argumentative et critique reste limitée.
raison / pensée	défaite de la raison / sommeil de la pensée	—	—	L'IA ne reproduit pas ce lexique philosophique et critique, entraînant une perte de spécificité générationnelle et de style engagé , propres au discours post-indépendance de Djaout.
citoyen	hommes et femmes	mobilisation citoyenne	chaque citoyen	L'IA généralise le lexique vers un usage standardisé, réduisant l'inclusion et la portée sociale spécifique du discours humain.
presse / pays	—	liberté de la presse / état démocratique	liberté de la presse / responsabilité des institutions	L'IA conserve les collocations institutionnelles et normées, proches du style contemporain de Drareni, mais ne reflète pas les nuances critiques ou philosophiques de Djaout.

3.6. Analyse des résultats préliminaires

L'analyse des collocations extraites des textes de Djaout, Drareni et du texte généré par l'IA révèle plusieurs tendances significatives, mettant en évidence à la fois les capacités et les limites des systèmes d'intelligence artificielle dans la reproduction du style journalistique algérien.

1- Reproduction partielle des collocations :

Le modèle d'IA tend à privilégier les expressions institutionnelles et standardisées, proches du style contemporain de Drareni (liberté de la presse, responsabilité des institutions). Cette tendance montre que l'IA est capable de reproduire des collocations fortement présentes dans les corpus normés, mais qu'elle peine à restituer les variations stylistiques ou générationnelles qui caractérisent le discours humain.

2- Perte de nuances générationnelles :

Les collocations philosophiques et engagées de Djaout (sommeil de la pensée, défendre les libertés) ne sont pas reproduites par le modèle. Cela illustre une limitation lexicosémantique majeure, où l'IA standardise le lexique et neutralise la profondeur critique et intellectuelle du discours post-indépendance. La perte de ces nuances reflète l'absence de sens contextuel et historique dans les productions génératives.

3- Combinaison des registres :

Certaines collocations mixtes apparaissent dans le texte généré (leur voix est étouffée), traduisant une tentative de l'IA de fusionner les styles de Djaout et Drareni. Toutefois, la profondeur lexicosémantique et la richesse stylistique du discours de Djaout restent limitées,

et les collocations générées tendent à se rapprocher d'un registre journalistique standardisé et institutionnel.

4- Biais linguistiques et normalisation du français :

L'IA standardise le français journalistique, ce qui entraîne la marginalisation des particularités stylistiques, lexicales et générationnelles propres au journalisme algérien. Les collocations spécifiques à un contexte culturel et historique sont souvent absentes, reflétant un biais de représentativité des corpus d'entraînement et des limitations dans la reproduction de la diversité linguistique.

En somme, ces résultats préliminaires suggèrent que si les systèmes d'IA sont capables de reproduire certaines structures lexicales et collocations proches des registres institutionnels, ils présentent des lacunes significatives dans la restitution du style engagé, philosophique et générationnel, mettant en évidence l'importance de l'analyse lexicosémantique pour détecter les biais et les pertes de nuance dans les textes génératifs.

4. Résultats et discussion

L'analyse des collocations dans les chroniques de Tahar Djaout et Khaled Drareni, complétée par l'étude des productions générées par un modèle d'intelligence artificielle, permet de mettre en évidence des régularités lexicosémantiques, des divergences générationnelles marquées et des biais linguistiques propres aux systèmes d'IA. Cette approche combine linguistique de corpus et analyse lexicosémantique, offrant un cadre méthodologique rigoureux pour comparer la richesse stylistique des textes humains à la standardisation des productions automatisées.

4.1. Différences et continuités générationnelles

Les collocations relevées dans la chronique de Djaout (défendre les libertés, élever la voix, sommeil de la pensée) reflètent un style intellectuel et engagé, centré sur l'action individuelle, la responsabilité citoyenne et la réflexion critique. Ces expressions traduisent un discours fortement marqué par l'intellectuel post-indépendance, où la liberté et la résistance face à la censure occupent une place centrale. La dimension lexicosémantique de ces collocations révèle non seulement le sens des mots dans leur contexte, mais également leur fonction argumentative et philosophique.

À l'inverse, les collocations identifiées chez Drareni (liberté de la presse, mobilisation citoyenne, espace d'expression) traduisent un style journalistique contemporain, davantage orienté vers les institutions et la participation collective. Le lexique y est plus normé et institutionnalisé, soulignant le rôle de la presse dans la garantie des droits démocratiques et la facilitation de la mobilisation citoyenne. La comparaison montre que, si le thème central de la liberté est partagé, l'angle et le registre lexical divergent, Djaout privilégiant une approche philosophique et intellectuelle, tandis que Drareni adopte un registre journalistique pragmatique.

4.2. Analyse des productions de l'IA

Le texte généré par l'IA reproduit partiellement certaines collocations caractéristiques des deux journalistes, mais de manière standardisée. On observe que :

- Les collocations institutionnelles et normées (liberté de la presse, responsabilité des institutions) sont reproduites, rappelant le style contemporain de Drareni.

- Les collocations philosophiques et engagées de Djaout (défendre les libertés, sommeil de la pensée) sont absentes, illustrant une perte de spécificité générationnelle et stylistique.
- Certaines expressions mixtes apparaissent, telles que leur voix est étouffée, combinant un registre personnel et collectif, mais avec une profondeur lexicosémantique moindre que dans le texte original de Djaout.

Ces observations mettent en évidence les biais linguistiques des modèles d'IA, qui tendent à standardiser le français journalistique et à privilégier des collocations institutionnelles, au détriment des particularités générationnelles et des nuances philosophiques et culturelles.

4.3. Implications pour l'analyse lexicosémantique et l'équité linguistique

L'étude confirme que :

- 1- Les collocations constituent un outil efficace pour comparer les styles générationnels et identifier les particularités discursives dans le journalisme algérien francophone.
- 2- Les productions de l'IA, bien qu'utiles pour reproduire certaines structures lexicales, ne saisissent pas toujours les nuances lexicosémantiques propres à un contexte historique et culturel précis, notamment les collocations engagées ou philosophiques.
- 3- Les biais observés soulignent l'importance de réfléchir à l'inclusion linguistique et culturelle dans les modèles d'IA, particulièrement pour les variétés régionales ou minoritaires du français, afin de préserver la richesse lexicale et stylistique des discours.

En combinant analyse lexicosémantique, linguistique de corpus et comparaison avec l'IA générative, cette étude met en lumière :

- Les différences générationnelles dans le journalisme algérien francophone, entre un style philosophique et intellectuel et un style institutionnel et normé.
- Les continuités thématiques autour de la liberté et de l'engagement citoyen.
- Les limites et biais des modèles d'IA dans la reproduction des collocations, en particulier celles portant la spécificité culturelle et historique des textes.

Cette approche ouvre des perspectives pour des recherches futures sur la diversité linguistique, l'équité culturelle et la médiation des modèles d'IA dans les contextes plurilingues, tout en offrant un cadre pour l'évaluation critique des productions automatisées dans des domaines sensibles comme le journalisme et la communication publique.

Conclusion

Cette étude a examiné les collocations dans deux chroniques journalistiques emblématiques de l'Algérie francophone, appartenant à deux générations différentes : Tahar Djaout, figure majeure de la chronique post-indépendance, et Khaled Drareni, journaliste contemporain engagé dans la critique politique. En confrontant ces textes avec une production générée par un système d'intelligence artificielle, l'étude a permis de répondre à la problématique centrale sur la capacité des modèles d'IA à reproduire les collocations caractéristiques et à identifier leurs biais linguistiques.

Les résultats montrent que :

1. **Différences générationnelles claires** : Les collocations de Djaout reflètent un style intellectuel et engagé (*défendre les libertés, élever la voix, sommeil de la pensée*), tandis que celles de Drareni sont plus institutionnelles et journalistiquement normées (*liberté de la presse, mobilisation citoyenne, espace d'expression*). Cette distinction souligne que les collocations traduisent non seulement le thème central de la liberté mais aussi le registre générationnel et le contexte historique des journalistes.
2. **Capacités et limites de l'IA** : Le modèle génératif reproduit partiellement les collocations institutionnelles et normées, proches du style contemporain de Drareni, mais il ne restitue pas les collocations philosophiques et engagées de Djaout. Certaines expressions mixtes apparaissent (*leur voix est étouffée*), mais avec une profondeur lexicosémantique moindre, ce qui illustre une standardisation du français journalistique et une perte des nuances générationnelles et culturelles.
3. **Biais linguistiques identifiés** : L'IA tend à privilégier des formes lexicales universelles et institutionnelles, marginalisant les spécificités régionales et historiques du français algérien journalistique. Cette observation confirme les conclusions de Bender et *al.* (2021) et Blodgett et *al.* (2020) sur les biais linguistiques dans les systèmes d'IA.

En réponse à la problématique, je peux conclure que les systèmes d'intelligence artificielle générative peuvent reproduire certains aspects lexicaux des discours journalistiques, mais ils ne saisissent pas entièrement les particularités générationnelles, culturelles et stylistiques. Les collocations qui structurent le discours et reflètent l'engagement intellectuel ou citoyen sont souvent neutralisées, ce qui souligne les enjeux d'équité linguistique et de diversité culturelle dans l'IA. Nos résultats permettent également de confirmer l'hypothèse 1, En effet, la disparition des collocations emblématiques de Tahar Djaout comme "*sommeil de la pensée*" au profit d'expressions plus génériques montre que l'IA tend à effacer la singularité stylistique au profit d'une "moyenne" statistique.

Parallèlement, l'hypothèse 2 du biais de représentativité est validée par la reproduction efficace du lexique de Khaled Drareni. L'IA restitue prioritairement les collocations telles que "*liberté de la presse*" ou "*état démocratique*", car ces structures appartiennent au registre institutionnel dominant des corpus d'entraînement globaux, ce qui marginalise de fait les expressions ancrées dans une profondeur historique ou culturelle spécifique à l'Algérie. »

Enfin, cette étude met en évidence l'importance de :

- Considérer les variétés linguistiques et générationnelles dans l'entraînement des modèles d'IA,
- Utiliser l'analyse lexicosémantique pour évaluer la fidélité des productions génératives,
- Promouvoir des solutions inclusives afin de préserver la richesse culturelle et la diversité des pratiques discursives francophones.

Ainsi, les résultats contribuent à la réflexion sur l'intégration de l'IA dans les écosystèmes éducatifs et médiatiques plurilingues, tout en offrant des pistes pour la conception de modèles plus sensibles aux spécificités linguistiques et culturelles.

Bibliographie

1. BENDER E.-M. et al 2021, « On the dangers of stochastic parrots: Can language models be too big? », Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT), p. 610–623. Disponible sur : <https://doi.org/10.1145/3442188.3445922> .
2. BLODGETT Su Lin, GREEN Lisa, O'CONNOR Brendan, 2020, « Demographic dialectal variation in social media: Methodological challenges and ethical concerns », Computational Linguistics, vol. 46, n° 3, p. 601–633. Disponible sur : https://doi.org/10.1162/coli_a_00368 .
3. BIBER Douglas, CONRAD Susan, 2009, Register, genre, and style, Cambridge, Cambridge University Press.
4. CRUSE D. A., 2000, Meaning in language: An introduction to semantics and pragmatics, 2e éd., Oxford, Oxford University Press.
5. CRUSE D. A., 2004, Lexical semantics, Cambridge, Cambridge University Press.
6. FELLBAUM Christiane (dir.), 1998, WordNet: An electronic lexical database, Cambridge, MIT Press.
7. FIRTH J. R., 1957, « A synopsis of linguistic theory 1930–1955 », dans F. R. PALMER (dir.), Studies in linguistic analysis, Oxford, Blackwell, p. 1–32.
8. HOVY Dirk, SPRUIT Shannon L., 2016, « The social impact of natural language processing », Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), p. 591–598. Disponible sur : <https://doi.org/10.18653/v1/P16-2095> .
9. MCENERY Tony, HARDIE Andrew, 2012, Corpus linguistics: Method, theory and practice, Cambridge, Cambridge University Press.
10. SHENG Emily, CHANG Kai-Wei, NATARAJAN Premkumar, PENG Nanyun, 2019, « The woman worked as a babysitter: On biases in language generation », Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), p. 3407–3412. Disponible sur : <https://doi.org/10.18653/v1/D19-1335> .
11. DJAOUT Tahar, 1991, « L'intellectuel et la peur », [chronique publiée].
12. DRARENI Khaled, 2020, « La liberté de la presse en Algérie », [chronique publiée].