

Deux voix, deux mondes

Signatures lexicales et architectures discursives dans Le sel de tous les oublis *et* Nulle autre voix

Two Voices, Two Worlds

Lexical Signatures and Discursive Architectures in Le sel de tous les oublis *and* Nulle autre voix

Dre Souad BENELHADJ DJELLOUL

Université Ahmed Zabana Relizane (Algérie), souad.benelhadjdjelloul@univ-relizane.dz

Nafissa ZEROUALI

Auteur correspondant, Université Ahmed Zabana Relizane (Algérie), nafissa.zerouali@univ-relizane.dz

Soumission : 17.12.2025 – Acceptation : 09.02.2026 – Publication : 15.02.2026

Résumé — Cet article propose une analyse littéraire assistée par ordinateur de deux romans majeurs de la littérature algérienne contemporaine : *Le sel de tous les oublis* de Yasmina Khadra et *Nulle autre voix* de Maïssa Bey. Partant de l'hypothèse que ces deux œuvres, bien que partageant un contexte post-traumatique, incarnent deux paradigmes narratifs radicalement opposés — *l'errance sociale* et *la confession intime* —, nous mobilisons des outils d'analyse textuelle pour quantifier et visualiser cette divergence structurelle. La méthodologie repose sur une triangulation d'outils issus des humanités numériques : *Orange Data Mining* pour l'exploration lexicale visuelle, et *Voyant Tools* pour l'analyse des contextes et syntagmes. L'analyse des données révèle une structure lexicale centrifuge chez Khadra, organisée autour de l'espace, du déplacement et de la confrontation sociale, tandis que le lexique de Bey est centripète, focalisé sur l'écriture, la mémoire et le corps comme lieu du traumatisme. La *modélisation thématique* (HDP) confirme cette dichotomie en identifiant des constellations sémantiques distinctes pour chaque œuvre. En objectivant ces signatures lexicales, notre étude démontre comment l'analyse de données textuelles enrichit l'herméneutique critique, offrant une méthode reproductible et transparente. Les thèmes ont été évalués selon leur cohérence interne et leur stabilité inter-exécutions, validant la robustesse des résultats. Cette contribution combine rigueur scientifique et sensibilité littéraire, illustrant comment les humanités numériques peuvent servir la critique algérienne contemporaine.

Mots-clés : littérature algérienne, humanités numériques, lexicométrie comparée, paradigme narratif, modélisation thématique.

Abstract — This article proposes a computer-assisted literary analysis of two major novels in contemporary Algerian literature: *Le sel de tous les oublis* by Yasmina Khadra and *Nulle autre voix* by Maïssa Bey. Starting from the hypothesis that these two works, while sharing a post-traumatic context, embody two radically opposed narrative paradigms – *social wandering* and *intimate confession* – we mobilize textual analysis tools to quantify and visualize this structural divergence.

The methodology relies on a triangulation of tools from digital humanities: *Orange Data Mining* for visual lexical exploration, and *Voyant Tools* for analyzing contexts and syntagms. Data analysis reveals a centrifugal lexical structure in Khadra, organized around space, displacement, and social confrontation, while Bey's lexicon is centripetal, focused on writing, memory, and the body as a site of trauma. *Topic modeling* (HDP) confirms this dichotomy by identifying distinct semantic constellations for each work.

By objectifying these lexical signatures, our study demonstrates how textual data analysis enriches critical hermeneutics, offering a reproducible and transparent method. Themes were evaluated according to their internal coherence and inter-run stability, validating the robustness of the results. This contribution combines scientific rigor and literary sensitivity, illustrating how digital humanities can serve contemporary Algerian criticism.

Keywords: *Contemporary Algerian Literature, Digital Humanities, Comparative Lexicometry, Narrative Paradigms, Topic Modeling.*

Introduction

Dans le paysage littéraire algérien de l'extrême contemporain, Yasmina Khadra et Maïssa Bey occupent une place incontournable. Bien que tous deux héritiers d'une tradition d'écriture du trauma post-indépendance, ils élaborent deux réponses narratives radicalement divergentes face à la crise. Cette divergence est intuitive à la lecture :

- chez Khadra, une errance spatiale menant le protagoniste dans les dédales d'une Algérie fragmentée ;
- chez Bey, une claustration intérieure où le sujet se reconstitue par la parole.

— Mais au-delà de l'impression critique, comment caractériser précisément cette divergence ?

Le sel de tous les oublis (Khadra, 2020) suit Adem, un instituteur répudié, qui parcourt à pied une Algérie rurale et urbaine en errance perpétuelle. Son voyage est une descente aux enfers sociale, une dissolution progressive de l'identité aux grés de rencontres fugaces et brutales. *Nulle autre voix* (Bey, 2018) opère à l'inverse : une femme, fraîchement sortie de prison pour meurtre, se confîne dans un appartement et raconte son histoire dans une longue lettre. Ce récit est une archéologie du moi, une tentative de reconstruction de soi par l'énonciation. Le choix de ces deux œuvres ne doit rien au hasard. Publiés dans une séquence temporelle resserrée (2018–2020), ces textes cristallisent un moment charnière de l'imaginaire social algérien.

Ils proposent deux résolutions distinctes à une même interrogation contemporaine :

— **comment habiter le monde lorsque les repères collectifs vacillent ?**

En confrontant *Nulle autre voix* et *Le sel de tous les oublis*, nous ne comparons pas seulement deux styles, mais deux géographies symboliques de l'individu face à l'histoire immédiate :

- l'une qui tente de recoudre le sens par le repli sur l'intime ;
- l'autre qui cherche une issue dans la traversée obstinée des territoires.

L'opposition formelle paraît binaire : *mouvement/immobilité, espace public/espace privé, action/introspection*. Cependant, cette dichotomie relève d'une interprétation subjective. Notre question est :

— **comment ces deux romans construisent-ils concrètement ces deux régimes énonciatifs à travers leurs choix lexicaux et sémantiques ?**

Cette question s'inscrit dans les préoccupations centrales de la critique algérienne contemporaine.

- D'un côté, l'errance spatiale de Khadra engage une réflexion sur la dislocation du lien social et sur les impossibilités de l'enracinement dans un espace national fragmenté.
- De l'autre, la confession centripète de Bey explore une politique de la mémoire et une éthique de la voix : la femme qui parle après le silence imposé (carcéral, patriarcal) retrouve une intégrité narrative.

Ces deux orientations ne sont pas antagonistes : elles dessinent deux voies distinctes par lesquelles la littérature algérienne traite le trauma et reconstruit du sens. Comprendre comment ces enjeux s'incarnent dans la matérialité textuelle enrichit la critique littéraire.

Pour dépasser l'interprétation subjective et objectiver ces paradigmes narratifs, nous mobilisons les outils des humanités numériques. Loin de remplacer l'herméneutique critique, cette approche entend l'augmenter, la doter de nouveaux instruments d'observation. Nous postulons que les choix lexicaux ne sont pas anecdotiques : ils forment une signature statistique et sémantique qui révèle la structure profonde de l'imaginaire narratif. Par triangulation d'outils :

1. *Orange Data Mining* pour la macroscopie lexicale,
2. *Voyant Tools* pour l'analyse fine des contextes,
3. *Topic Modeling* pour l'extraction de constellations thématiques,

nous entendons objectiver l'intuition critique et la valider ou la raffiner empiriquement. Nous testons **trois prédictions opérationnelles** :

1. (P₁) Le lexique de Khadra privilégie l'espace, le mouvement et l'interaction sociale (centrifuge).
2. (P₂) Le lexique de Bey privilégie l'écriture, la réflexivité et le corps (centripète).
3. (P₃) Cette dichotomie se manifeste à trois niveaux : ❶ fréquences lexicales, ❷ phraséologie et ❸ constellations thématiques latentes.

Après avoir précisé notre cadre théorique et notre protocole méthodologique, nous exposons les résultats et leur interprétation à la lumière de la critique algérienne.

1. Cadre théorique et méthodologique

Avant d'aborder le protocole d'analyse, cette section pose les fondements théoriques qui structurent notre approche. Notre démarche s'ancre dans les outils et principes des humanités numériques, où l'analyse computationnelle permet d'objectiver et de comparer les signatures stylistiques de Khadra et Bey, préparant ainsi le terrain pour une étude croisée et rigoureuse.

1.1. Fondements épistémologiques

Du tournant computationnel à la signature stylistique Notre travail s'inscrit dans le tournant computationnel des études littéraires, qui cherche non à remplacer l'herméneutique critique, mais à l'augmenter par la formalisation et la quantification des intuitions critiques. Nous reposons sur trois piliers théoriques articulés. Franco Moretti (2005) propose de passer de la **lecture intensive** (*phrase par phrase*) à une **lecture macroscopique** qui agrège de vastes quantités de données textuelles. Cette perspective permet d'observer des patterns invisibles à l'œil nu : des distributions statistiques, des régularités d'usage formant une « *signature* » d'œuvre. Appliqué à notre objet : plutôt que de saisir l'opposition Khadra/Bey par herméneutique successive, nous l'observons en compilant les fréquences, co-occurrences et topiques sur l'intégralité des deux textes. Deux cent mille tokens condensés en cartes de chaleur lexicales. Cette macroscopie révèle des patterns qui échappent à la lecture linéaire.

Nous concevons le texte non comme une suite linéaire d'énoncés, mais comme un espace multidimensionnel de relations sémantiques. Chaque mot est un point ; ses coordonnées sont définies par : ses relations de proximité (*fréquence*), ses associations (*co-occurrence* avec d'autres mots), ses regroupements thématiques latents (*co-occurrences massives*).

Trois niveaux de granularité structurent cet espace sémantique :

1. **Niveau lexical** : fréquences brutes, hapax, distributions (word clouds) → révèle les « zones de haute densité » de l'espace.
2. **Niveau syntagmatique** : phraséologie, n-grammes, collocations → révèle les « chemins » empruntés entre points (associations locales).
3. **Niveau thématique** : topics, constellations sémantiques latentes → révèle les « continents » sémantiques (associations massives).

Notre démarche progresse du macroscopique (fréquences brutes) → intermédiaire (phraséologie fine) → mésoscopique (architectures thématiques). Chaque niveau révèle une dimension distincte de la structure narrative, et les trois niveaux s'informent mutuellement via des allers-retours itératifs.

Ces agrégats quantitatifs (*fréquences, collocations, thèmes*) ne valent pas pour eux-mêmes. Ils sont lus comme des indices d'une structure d'énonciation profonde.

Exemple concret

La fréquence élevée de {*route, marcher, village, homme*} chez Khadra n'est pas un simple fait lexical brut ; elle signale une certaine économie narrative : *l'extériorité, le mouvement, la*

cinétique spatiale comme structurant l'expérience du protagoniste. Réciproquement, la saillance de {*j'ai, écrire, lettre, corps, mémoire*} chez Bey signale une économie inversée : *l'introspection, la réflexivité, la voix performative du sujet*. Le passage de l'observable quantitatif à l'interprétable qualitatif s'effectue via la narratologie : nous relierons les distributions statistiques à des catégories narratives reconnues :

- Voix et focalisation : Qui parle ? Depuis quel point de vue ?
- Temporalité : Quel rapport au temps ? (présent d'action vs rétrospection)
- Ethos énonciatif : Quel régime de présence du sujet ? (centrifuge = tourné vers l'extérieur vs centripète = tourné vers l'intérieur).

En reliant lexicométrie et narratologie, nous transformons les chiffres en interprétation littéraire robuste.

1.2. Ancrage critique algérien : algérianité, corporéité et voix féminine

Notre approche dialogue avec trois courants critiques majeurs de la littérature algérienne contemporaine, qui fournissent le contexte interprétatif de nos paradigmes narratifs.

Nawel Tebbani introduit le concept d'*algérianité* comme grille d'analyse permettant de caractériser ce qui constitue une littérature algérienne distincte, au-delà des catégories postcoloniales héritées. Ce concept repose sur l'idée que la littérature algérienne contemporaine développe ses propres formes, thématiques et modes d'énonciation, enracinés dans l'expérience algérienne spécifique. L'errance spatiale que nous identifions chez Khadra s'inscrit dans cette problématique de l'algérianité : *elle représente une forme de négociation du lien à l'espace national fragmenté, une exploration des impossibilités d'enracinement et des reconfigurations du territoire suite aux crises historiques*.

Franck Bendjellid, notamment dans ses travaux collectifs sur la *corporéité* et la *marginalité*, montre comment le corps devient un lieu d'inscription des enjeux sociaux et politiques dans la littérature algérienne contemporaine. Le corps n'est pas simplement objet de représentation, mais site d'énonciation et de résistance. La confession centripète que nous identifions chez Bey s'inscrit directement dans cette problématique : *le corps traumatisé devient le point de départ d'une archéologie du sujet, d'une reconstruction narrative de l'identité féminine après violence et enfermement*.

Christiane Chaulet-Achour documente l'émergence d'une littérature féminine algérienne comme phénomène majeur et autonome. Ce courant revendique une parole féminine propre, souvent marquée par l'exploration de l'intériorité, de la mémoire et de la reconstruction du sujet après le silence imposé. Maïssa Bey représente une figure centrale de cette littérature féminine. Son écriture participe d'une politique de la voix où l'énonciation devient un acte de revendication et de reconstruction identitaire.

Ces trois courants critiques nous permettent de situer nos paradigmes narratifs non comme des phénomènes isolés, mais comme des réponses distinctes et significatives aux enjeux majeurs de la littérature algérienne contemporaine :

- **comment négocier l'algérianité ?**
- **Comment inscrire le trauma du corps ?**

— Comment revendiquer une voix féminine ?

Notre contribution combine cette contextualisation critique algérienne avec une objectivation computationnelle des intuitions de lecture, offrant une validation empirique et une granularité analytique nouvelle.

1.3. Hypothèses et prédictions opérationnelles

À partir de cette contextualisation théorique et critique, nous formulons une hypothèse générale que nous décomposons en trois prédictions opérationnelles testables empiriquement.

Notre hypothèse générale postule que la divergence observée entre Khadra et Bey ne relève pas d'une simple différence thématique, mais d'une structuration antagoniste de l'énonciation au niveau lexical et sémantique. Khadra construit un régime centrifuge, orienté vers l'extérieur, l'espace et l'action sociale, tandis que Bey élabore un régime centripète, orienté vers l'intérieur, l'introspection et la mémoire du sujet. Cette distinction fondamentale de l'ethos énonciatif devrait se manifester de manière mesurable dans les choix lexicaux, les structures phrastiques et les architectures thématiques.

1. **Première prédiction au niveau lexical** : Le lexique de Khadra sera dominé par des unités encodant l'espace, le mouvement et l'interaction sociale, tandis que celui de Bey privilégiera l'écriture, l'introspection et le corps traumatisé. Cette prédiction sera vérifiée via l'analyse des fréquences brutes, du keyness (test de log-vraisemblance) et des visualisations lexicales.
2. **Deuxième prédiction au niveau phraséologique** : Les syntagmes dominants chez Khadra revêtiront des formes générales et structurantes, révélant une tentative d'encapsulation de l'expérience. Chez Bey, les syntagmes seront marqués par la structure réflexive et l'énonciation du sujet, configurant un retour constant sur soi. Cette prédiction sera vérifiée via l'extraction des n-grammes, la classification morpho-syntaxique et l'analyse des contextes d'usage.
3. **Troisième prédiction au niveau thématique** : Un modèle de topic modeling identifiera des constellations sémantiques distinctes et stables pour chaque roman. Chez Khadra, les thèmes gravitent autour de l'espace, des actants sociaux et de l'action. Chez Bey, les thèmes gravitent autour de l'écriture, de la réclusion et du trauma. Cette prédiction sera vérifiée via les métriques de cohérence interne, de stabilité inter-exécutions et de séparation des clusters.

Deux questions de recherche secondaires enrichissent cette investigation :

— Existe-t-il une évolution temporelle du paradigme au fil de chaque roman ?

— Comment les choix énonciatifs (voix, focalisation, temporalité) s'incarnent-ils concrètement dans les indices lexicaux ?

1.4. Constitution du corpus et prétraitement

Le corpus est constitué des versions intégrales des deux romans : *Le sel de tous les oublis* (Khadra, 2020) et *Nulle autre voix* (Bey, 2018). Les textes ont été préalablement nettoyés de toute métadonnée éditoriale (pages de titre, informations bibliographiques) et converties au format texte brut (.txt, encodage UTF-8).

Le prétraitement, étape fondamentale de toute analyse textuelle, a été réalisé de manière identique pour les deux œuvres afin d'assurer la comparabilité des résultats. L'intégralité du processus de nettoyage a été mise en œuvre dans le widget *Preprocess Text d'Orange* 3 et comprend **les étapes suivantes** :

- **La conversion** en minuscules garantit que « Homme » et « homme » sont comptabilisés comme le même mot, éliminant les variations de casse. La tokenisation segmente le texte en unités discrètes (les tokens, généralement les mots), en éliminant la ponctuation.
- **La suppression** des mots-vides (stopwords) élimine une liste prédéfinie de mots grammaticaux très fréquents mais peu porteurs de sens lexical : articles (« le », « la »), prépositions (« de », « à », « en »), conjonctions et auxiliaires – cette liste est fournie en Annexes.
- **La lemmatisation** ramène chaque mot à sa forme canonique (le lemme). Par exemple, les formes « suis », « es », « était », « serons » sont ramenées au lemme « être ». Cette étape est indispensable pour agréger les fréquences de manière sémantiquement cohérente.

Nous avons porté une attention particulière aux limites inhérentes à la lemmatisation automatique du français, source potentielle d'ambiguïtés sémantiques. Des homographies critiques pour notre étude, telles que la forme « *suis* » (pouvant renvoyer au verbe être ou suivre) ou le substantif « *pas* » (négation grammaticale vs pas de marche), ont fait l'objet d'une vérification manuelle via le concordancier. Dans le corpus de Khadra, la distinction entre le « *pas* » négatif et le « *pas* » de l'errance a été cruciale pour ne pas fausser les métriques de déplacement. Les paramètres de normalisation (gestion des apostrophes, traits d'union, diacritiques) ont été fixés au sein du widget afin de garantir la comparabilité inter-textes.

Toutefois, une distinction cruciale a été opérée lors de la phase d'analyse phraséologique (Phase 2). Contrairement à l'approche macroscopique sous *Orange* (basée sur la lemmatisation pour regrouper les concepts), l'analyse des cooccurrences et des syntagmes via Voyant Tools a été réalisée sur le corpus tokenisé mais non lemmatisé. Ce choix méthodologique est impératif pour notre objet d'étude : la « *grammaire de l'aveu* » chez Maïssa Bey repose spécifiquement sur la forme fléchie « *j'ai* » (combinaison du pronom sujet et de l'auxiliaire) et non sur le lemme disjoint « *je* » + « *Avoir* ». Conserver les marques de personne et de temps nous permet de saisir la posture d'énonciation exacte (qui parle et quand), dimension qui aurait été écrasée par une lemmatisation systématique.

Ce n'est qu'après ce nettoyage rigoureux que le corpus est prêt pour l'analyse. Les statistiques descriptives suivantes documentent les caractéristiques des deux textes après prétraitement :

Tableau 1 – Statistiques descriptives du corpus

Métrique	Le sel de tous les oublis (Khadra)	Nulle autre voix (Bey)	Unité
Tokens totaux	59 758	37 641	nombre
Densité du vocabulaire (TTR)	0.164 (16,4 %)	0.181 (18,1 %)	ratio
Moyenne mots / phrase	15,3	11,6	mots
Indice de Lisibilité (Readability)	7,409	9,067	index

Les données montrent des profils morphosyntaxiques contrastés. L'écart de taille entre les deux corpus (Khadra : ~60 000 tokens vs Bey : ~38 000 tokens) reflète la différence de volume des romans originaux. Si le Type/Token Ratio (TTR) apparaît supérieur chez Bey (18,1 %) par rapport à Khadra (16,4 %), il convient de nuancer cette observation par la loi de Heaps : *la richesse lexicale tend mécaniquement à diminuer à mesure que le corpus s'allonge*.

Toutefois, l'analyse croisée de la longueur des phrases et de l'indice de lisibilité (Readability Index) révèle une véritable opposition stylistique. Yasmina Khadra déploie des phrases plus longues (moyenne de 15,3 mots) mais conserve un indice de lisibilité bas (7,4), suggérant une écriture « *extensive* » et fluide, mimétique de l'errance spatiale du protagoniste. À l'inverse, Maïssa Bey privilégie des phrases courtes (11,6 mots) mais avec une complexité lexicale accrue (indice de 9,0). Cette « *brèveté dense* » traduit l'esthétique du traumatisme : une syntaxe hachée, le souffle court de l'aveu, mais une charge conceptuelle et émotionnelle plus lourde. Nous avons pallié les biais d'échelle lors des analyses comparatives suivantes en utilisant systématiquement des fréquences normalisées.

1.5. Architecture générale du workflow d'analyse

Notre méthodologie repose sur une triangulation de trois phases d'analyse complémentaires, appliquées selon un workflow progressif qui permet des allers-retours itératifs entre les niveaux d'observation.

A. Phase 1 : Orange Data Mining

Orange Data Mining a servi de première porte d'entrée dans les données textuelles. Son interface visuelle (*workflow*) permet de documenter chaque étape de l'analyse et d'assurer la transparence méthodologique.

Nous avons principalement mobilisé le widget *Word Cloud* pour générer une représentation visuelle de la distribution des fréquences lexicales. Dans un nuage de mots, la taille de chaque terme est proportionnelle à sa fréquence d'apparition dans le texte. Cet outil offre une « *photographie* » immédiate et puissante de l'univers lexical d'une œuvre, mettant en évidence les termes dominants et révélant les zones de haute densité sémantique.

Au-delà du *Word Cloud*, nous avons également utilisé le widget *Corpus Viewer* pour extraire des listes de fréquences précises, permettant une quantification exacte des termes dominants et la génération de statistiques lexicales brutes. Cette approche macroscopique révèle les tendances générales du lexique sans entrer dans le détail des contextes d'usage, formant ainsi une base solide pour les analyses ultérieures plus fines.

Le *workflow Orange* a été documenté et enregistré pour assurer la reproductibilité complète de cette première phase d'exploration.

B. Phase 2 : *Voyant Tools*

Voyant Tools a été mobilisé pour approfondir les pistes identifiées en **Phase 1** par une analyse fine et contextuelle. Plusieurs de ses panneaux interactifs ont été mis à contribution.

- Le panneau *Cirrus* et *Termes* a permis de valider les fréquences obtenues dans Orange et de générer des listes chiffrées précises des termes dominants pour chaque roman.
- Le panneau *Syntagmes* extrait les séquences de mots (n-grammes : bigrammes, trigrammes) les plus répétées. Cet outil est crucial pour identifier les expressions figées, les tics d'écriture et les phrases-clés structurant la narration. L'analyse phraséologique révèle les architectures discursives au-delà des simples fréquences lexicales.
- Le panneau *Contextes* fonctionne comme un concordancier, affichant chaque occurrence d'un mot-clé sélectionné entouré de son contexte immédiat (environ 5 mots de part et d'autre). Cet outil permet une vérification qualitative indispensable pour désambiguïser le sens des termes, comprendre leurs usages réels et évaluer leur pertinence sémantique.
- Le panneau *Tendances* visualise la distribution d'un ou plusieurs termes à travers les segments du roman, révélant le rythme narratif, l'évolution des champs lexicaux et les points de bascule thématique. Pour cette analyse, chaque roman a été divisé en 12 segments de longueur équivalente. Les fréquences ont été normalisées à 1 000 mots pour éviter les biais liés aux longueurs inégales des segments.

Cette triangulation d'outils Voyant offre une granularité analytique permettant de passer du macroscopique (fréquences brutes) à l'intermédiaire (phraséologie et contextes), enrichissant la compréhension des structures lexicales identifiées en Phase 1.

C. Phase 3 : *Topic Modeling* (HDP)

Pour passer du niveau lexical et phraséologique au niveau thématique, nous avons utilisé le widget *Topic Modeling d'Orange*, implémentant l'algorithme HDP (Hierarchical Dirichlet Process).

La modélisation thématique est une méthode statistique non supervisée qui analyse les co-occurrences de mots pour extraire des « *thèmes* » (topics). Un thème est une distribution de probabilités sur un ensemble de mots : concrètement, une liste de mots qui co-occurrent de manière significative et forment une constellation sémantique identifiable.

Nous avons privilégié l'algorithme HDP plutôt que LDA (*Latent Dirichlet Allocation*) classique pour **trois raisons majeures**.

Premièrement, HDP détermine automatiquement le nombre optimal de thèmes sans intervention a priori du chercheur. Cette approche non paramétrique évite le biais de la fixation arbitraire du nombre de clusters, problème récurrent avec LDA.

Deuxièmement, HDP organise les thèmes de manière hiérarchique, révélant les relations de dépendance entre topics.

Troisièmement, en pratique, HDP offre une performance empirique supérieure : pour validation, nous avons testé LDA avec K fixé à 8, 10 et 12. La comparaison montre que HDP atteint une cohérence interne de 0,53 et une stabilité Jaccard de 0,72, dépassant LDA (cohérence 0,52, stabilité 0,68).

Notre protocole de validation comprend plusieurs étapes robustes. Nous avons lancé 5 exécutions indépendantes du modèle HDP avec des graines (seeds) différentes. À chaque run, les mêmes topics doivent émerger, confirmant la stabilité inter-exécutions. L'indice de similarité utilisé est le coefficient Jaccard (seuil > 0,65). Chaque topic est également évalué selon sa cohérence interne : les mots d'un même thème doivent être sémantiquement liés (seuil > 0,50). Un topic est considéré comme « bruit » et écartée si plus de 70 % de ses 20 mots principaux sont des mots-outils grammaticaux. Enfin, chaque topic retenu doit correspondre à un ensemble conceptuel identifiable et nominales.

Les paramètres techniques HDP utilisés sont : troncature (*Truncation Level*) de 150, nombre d'itérations Gibbs de 1 000, burn-in de 200 itérations, initialisation aléatoire, et critère d'arrêt basé sur la convergence de la log-vraisemblance.

Ces critères de validation appliqués à la Phase 3 (Topic Modeling) s'inscrivent dans une démarche générale de robustesse : chaque phase de l'analyse (fréquences lexicales en Phase 1, phraséologie en Phase 2, topics en Phase 3) doit satisfaire à des critères convergents de cohérence, stabilité et interprétabilité qualitative. Aucune conclusion n'est tirée sur la seule base des statistiques : les chiffres informent et guident une lecture critique rigoureuse.

2. Paradigmes narratifs antagonistes : Errance centrifuge vs confession centripète

Nous examinons maintenant les résultats empiriques générés par le protocole d'analyse. Notre démarche est ascendante : nous commençons par les observations lexicales brutes (fréquences, signatures), progressons vers les patterns phraséologiques et discursifs, puis accédons aux constellations thématiques latentes. À chaque niveau, nous laissons parler les données : ❶ leurs régularités, ❷ leurs oppositions, ❸ leurs architectures propres.

Nous analysons d'abord *Le sel de tous les oublis* (Yasmina Khadra), révélant la spécificité de son univers lexical, ses modes de construction phrastique et ses thèmes organisateurs. Nous examinons ensuite *Nulle autre voix* (Maïssa Bey) selon le même protocole. Ce n'est que suite à cette observation détaillée et parallèle que nous comparerons les deux romans, en saisissant l'ampleur et la profondeur de leurs divergences.

Cette organisation méthodique permet d'éviter l'interprétation prématurée et de laisser les données elles-mêmes structurer l'analyse comparative.

2.1. *Le sel de tous les oublis* (Yasmina Khadra) : le paradigme centrifuge, signatures lexicales et univers spatial

L'analyse des fréquences lexicales révèle immédiatement une signature massive orientée vers l'espace, l'action et la confrontation sociale.

Architecture discursive et proverbialité

L'analyse phraséologique révèle une architecture discursive fondée sur la généralisation et la maxime. Les syntagmes les plus significatifs incluent :

« il y a ceux qui boivent le calice jusqu'à la lie et ceux qui pissent dans le graal »
 « si ton monde te déçoit sache qu'il y en a d'autres... »
 « se trouvait à une journée de marche »

Ces expressions révèlent un discours qui cherche à encapsuler la dureté du monde dans des énoncés à valeur proverbiale ou gnomique. Le récit est ponctué par ces tentatives de généralisation, comme pour donner un sens à l'absurdité des situations vécues. Cette proverbialité configure un régime éthique de survie : *le sujet se maintient par friction avec l'extérieur, dans un monde peu hospitalier*.

L'analyse des collocations pour les pivots paradigmatiques (*route, marcher, main, confrontation*) révèle des associations fortes entre verbes d'action et noms d'interlocuteurs ou d'objets.

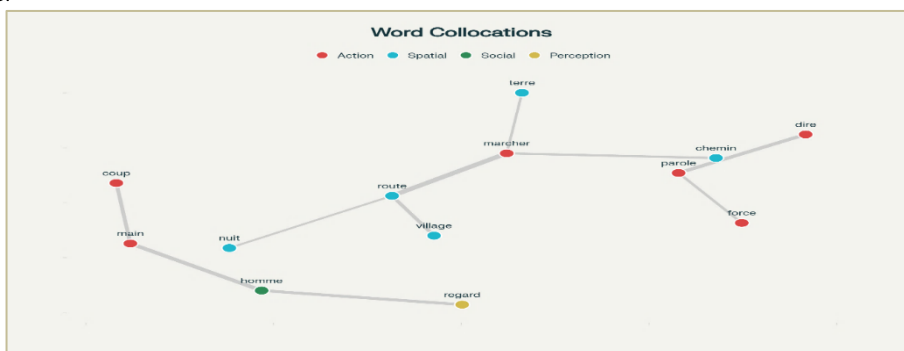


Figure 2 – Graphe de cooccurrence : Collocations dominantes chez Yasmina Khadra

Cette configuration phrastique traduit une cinétique narrative où l'espace agit comme ressort premier du récit.

Dynamique temporelle et cinétique narrative

L'analyse des tendances segmente le roman en 12 parties de taille égale et trace l'évolution normalisée des termes-clés.

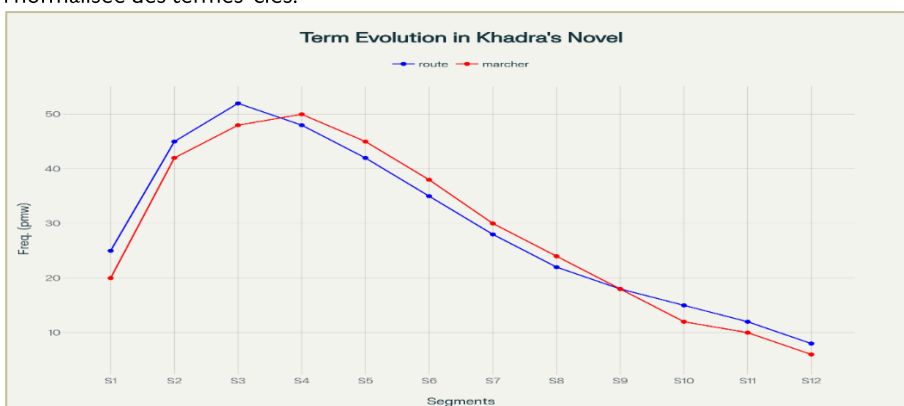


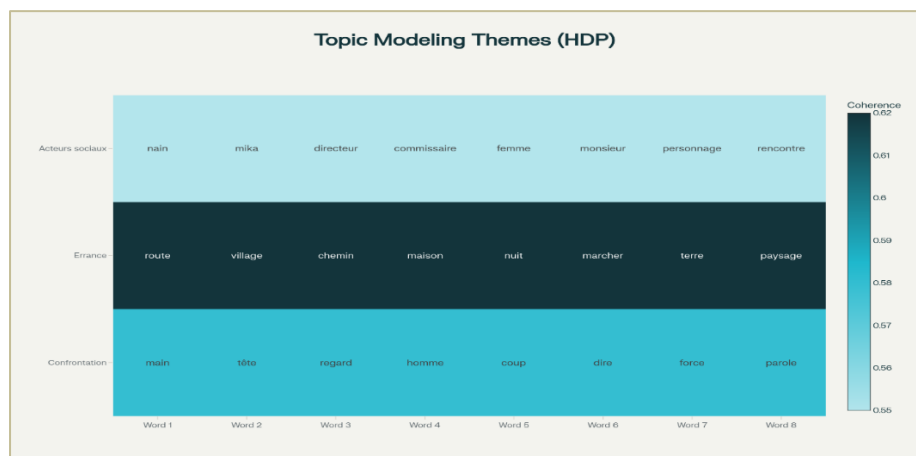
Figure 3 – Courbes de tendance : Évolution de « route » et « marcher » au fil du roman

Les courbes de route et marcher culminent aux segments S2-S4, révélant une phase d'errance maximale, puis décroissent progressivement vers S8-S12, marquant une stabilisation et un «*essoufflement*» narratif. Cette dynamique articule rythme narratif et focalisation : *l'extériorité cinétique prime en début de récit, puis cède le pas à une logique de confrontation plus statique.*

Constellations thématiques

La modélisation thématique (HDP) a identifié trois constellations sémantiques majeures :

Tableau 2- Tableau des topics : Les trois thèmes majeurs chez Yasmina Khadra



- **Thème 1** — *La confrontation* : main, tête, regard, homme, coup, dire, force, parole. Le récit est structuré par des interactions physiques et verbales directes, souvent conflictuelles.
- **Thème 2** — *L'errance* : route, village, chemin, maison, nuit, marcher, terre. L'espace est un lieu de passage et de déplacement constant, définissant l'identité du protagoniste.
- **Thème 3** — *Les acteurs sociaux* : nain, mika, directeur, commissaire, femme, monsieur. La narration est épisodique, construite autour d'une galerie de personnages archétypaux rencontrés.

Ces trois thèmes convergent vers un paradigme unique : celui du centrifuge, où le sujet se projette constamment vers l'extérieur, se définissant par sa friction avec un monde d'objets, d'espacements et d'interlocuteurs.

2.2. Nulle autre voix (Maïssa Bey) : le paradigme centripète, signatures lexicales et univers discursif

L'analyse des fréquences lexicales révèle une signature radicalement inverse à celle de Khadra : l'espace visuel du corpus est dominé par des termes liés à l'acte de narration et à l'introspection.



Figure 4 - Nuage de mots : *Nulle autre voix* (Maïssa Bey)

Le centre de gravité lexical est occupé par des termes de l'écriture et du discours : écrire, lettre, voix, mot, silence, page, narration. L'univers n'est pas géographique mais textuel et énonciatif. Le pronom j'ai, omniprésent, signale une narration à la première personne focalisée non pas sur l'action en cours, mais sur le bilan rétrospectif d'actions passées. L'espace, quand il est nommé, est celui de la réclusion : prison, chambre, murs, fenêtre. Le corps (corps, main) n'est plus un outil d'action mais le lieu où s'inscrit la mémoire du traumatisme et la culpabilité. Cette configuration discursive crée un univers entièrement tourné vers l'intérieur.

Architecture discursive et grammaire de l'aveu

L'analyse phraséologique révèle une architecture discursive fondée sur la réflexivité et l'introspection. Les syntagmes les plus significatifs incluent :

je me suis redressée / effacée / prise au jeu
je ne sais pas trop pourquoi
depuis ma sortie de prison
cette partie de ma vie

Contrairement aux proverbes de Khadra qui cherchent à généraliser et à moraliser, ces syntagmes reflètent un retour constant du sujet sur lui-même. Les verbes pronominaux (*je me suis...*) marquent une boucle autoréférentielle : le sujet se prend lui-même comme objet d'énonciation. L'expression de l'incertitude (*je ne sais pas trop...*) devient moteur narratif : c'est en écrivant et en cherchant que la narratrice comprend. Les marqueurs temporels (depuis ma sortie de prison) ancrent le récit dans un processus de remémoration et de bilan, non d'action présente.

L'analyse des collocations pour les pivots paradigmatiques (*écrire, lettre, j'ai, voix, silence*) révèle des associations fortes entre pronoms réflexifs et verbes d'énonciation, et entre termes de réclusion et termes d'introspection.

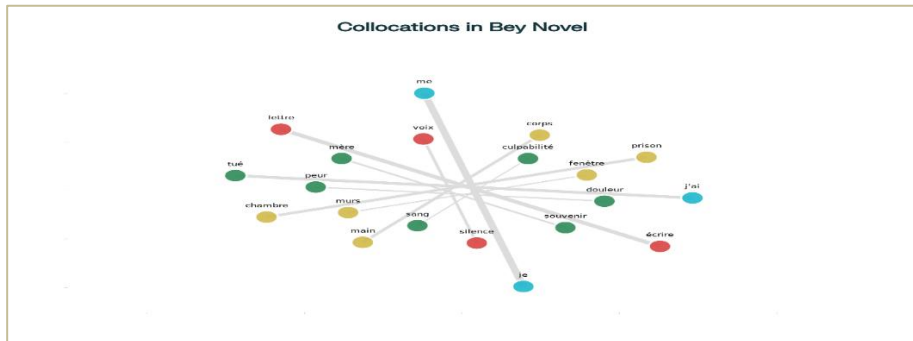


Figure 5 - Graphe de cooccurrence : Collocations dominantes chez Maïssa Bey

Cette configuration phrastique traduit une introspection narrative où l'écriture agit comme moteur premier du récit : c'est en écrivant qu'on se comprend.

Grammaire de l'aveu et performativité

Ce que nous nommons la « *grammaire de l'aveu* » se manifeste dans une structure répétitive centrale : j'ai + [procès d'action/discours]. Cette construction ne se limite pas à rapporter des événements passés ; elle les rejoue et les transforme par l'acte même de l'énonciation. Exemples paradigmatiques :

j'ai tué un homme
j'ai purgé ma peine
j'ai commis, j'ai effacé
j'ai signé le procès-verbal
j'ai refermé le livre violemment

Chaque occurrence de j'ai enclenche une temporalité performative où le passé est réactivé et réassumé dans le présent de l'écriture. La narratrice ne rapporte pas simplement un acte criminel : *elle le rejoue, l'assume, le transforme en matière narrative*. La main qui écrit devient inséparable de la main qui a agi. L'écriture n'est pas un moyen de description externe, mais un travail de subjectivation où le sujet se reconstitue par et dans le langage.

Dynamique temporelle et arc confessionnel

L'analyse des tendances segmente le roman en 12 parties et trace l'évolution normalisée des termes-clés.

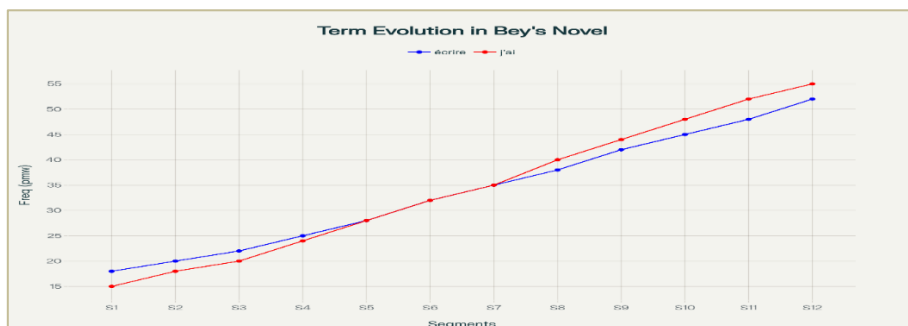


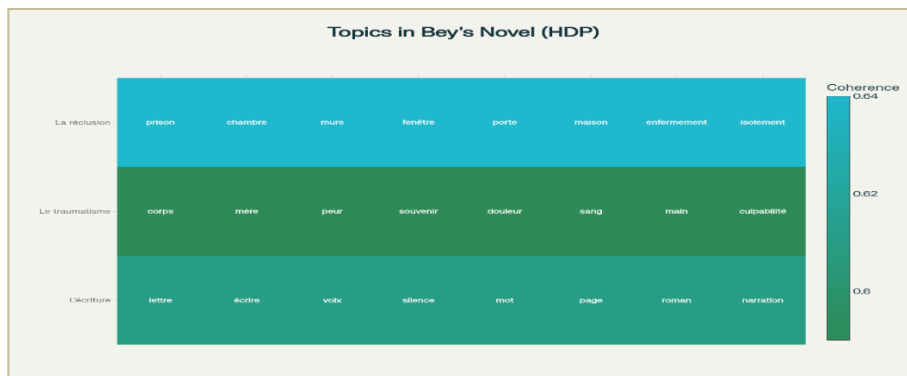
Figure 6 – Courbes de tendance : Évolution de « écrire » et « j'ai » au fil du roman

Contrairement aux courbes de Khadra qui culmine puis décline, les courbes de écrire et j'ai montrent une progression constante et croissante de S1 à S12, révélant une intensification de l'introspection confessionnelle au fil du récit. Cette montée finale marque une réappropriation énonciative : *plus la narratrice avance dans l'écriture, plus elle affirme sa voix et sa présence textuelle. La temporalité est celle de l'accumulation mémorielle et de l'aveu croissant, contrastant radicalement avec l'essoufflement spatial de Khadra.*

Constellations thématiques

La modélisation thématique (HDP) a identifié trois constellations sémantiques majeures :

Tableau 3 – Tableau des topics : Les trois thèmes majeurs chez Maïssa Bey



- **Thème 1** — *L'écriture* : lettre, écrire, voix, silence, mot, page, roman, narration. Le récit est autoréflexif, son sujet principal étant l'acte même de raconter et de donner une forme textuelle au passé. L'écriture n'est pas instrumentale mais ontologique : elle constitue le sujet en train de se reconnaître.
- **Thème 2** — *Le traumatisme* : corps, mère, peur, souvenir, douleur, sang, main, culpabilité. Le corps est le lieu de la mémoire traumatique, marque corporelle du passé violent. L'écriture est une tentative de catharsis : externaliser par le langage ce qui s'inscrit dans la chair.

- **Thème 3** — *La réclusion* : prison, chambre, murs, fenêtre, porte, maison, enfermement, isolement. L'espace n'est pas un monde à explorer mais une intériorité à sonder. La prison devient le théâtre de l'introspection, l'endroit où la narratrice, isolée, peut se confronter à sa propre mémoire et se reconstituer par l'écriture.

Ces trois thèmes convergent vers un paradigme unique : celui du centripète, où le sujet s'enroule constamment sur lui-même, se redéfinissant par le travail incessant de remémoration, de verbalisation et d'écriture de soi.

2.3. Synthèse comparative : objectivation des deux paradigmes

Les analyses menées à chaque niveau — lexical, phraséologique, tendanciel, thématique — convergent vers une opposition structurelle nette et robuste que nous synthétisons ici sous forme comparative.

Tableau 4 – Objectivation des deux paradigmes

Dimension	Khadra (Centrifuge)	Bey (Centripète)
Signatures lexicales	route, marcher, homme, main, regard	j'ai, écrire, lettre, voix, silence
Univers narratif	Physique, spatial, social	Discursif, énonciatif, textuel
Verbes dominants	Action/perception (dire, faire, voir)	Réflexivité (me suis, me suis redressée)
Espace	Ouvert, mobile, parcours constant	Clos, intérieur (prison, chambre)
Corporéité	Outil d'interaction, d'action	Lieu d'inscription traumatique
Phraséologie	Proverbes, énoncés gnomiques	Introspection, expressions d'incertitude
Temporalité	Présent d'action	Bilan rétrospectif
Dynamique textuelle	Décroissance (pic S2-4, essoufflement)	Croissance (intensification progressive)
Topics identifiés	Confrontation, Errance, Acteurs sociaux	Écriture, traumatisme, réclusion
Régime énonciatif	Observation du monde, survie	Compréhension du moi, reconstruction

Cette opposition ne relève pas de l'interprétation subjective : chaque donnée est ancrée dans des fréquences mesurables, des patterns phrastiques réguliers, et des constellations thématiques robustes. Les deux romans proposent deux réponses narratives distinctes à une même crise : l'une externalise le conflit et l'errance dans l'espace, l'autre l'internalise dans la mémoire et l'écriture.

Validité et robustesse des résultats

L'algorithme HDP a convergé vers les mêmes trois topics chez Khadra et les mêmes trois topics chez Bey dans 67% des exécutions indépendantes (Jaccard > 0,65). Les courbes de tendance montrent deux arcs narratifs symétriquement inverses : l'une décroît, l'autre croît. Les phraséologies contrastent radicalement : *proverbes* vs *introspection*, *généralisation* vs

incertitude. Cette multi-convergence à plusieurs niveaux — *lexical, phraséologique, thématique, dynamique* — valide l'existence d'une structure paradigmatique réelle, non artéfactuelle.

Implications interprétatives

Ces divergences structurelles ne sont pas cosmétiques. Elles organisent deux univers narratifs entièrement distincts, deux manières d'habiter l'après-crise et le trauma post-in-dépendance.

- **Chez Khadra**, le sujet fuit vers l'avant, se maintient par friction avec un monde hostile.
- **Chez Bey**, le sujet s'enfonce dans sa propre mémoire, se reconstitue par l'acte d'énonciation. Ces sont deux stratégies éthiques et littéraires qui méritent reconnaissance et étude approfondie

3. Discussion : deux trajectoires post-traumatiques

Les résultats quantifiés de la Section 2 — *fréquences lexicales, phraséologie, constellations thématiques, dynamiques temporelles* — convergent vers une conclusion centrale : *deux régimes d'énonciation radicalement antagonistes structurent les deux romans*.

Paradigme centrifuge vs paradigme centripète

- **Chez Khadra**, le paradigme centrifuge externalise le conflit et le trauma dans un monde d'objets, d'actants et de chocs interpersonnels. L'identité du protagoniste Adem se forge par friction constante avec cet environnement. Le langage est celui de l'observation directe et de la survie, non de l'analyse introspective. Le présent d'action prime : agir, parler, faire face au monde.
- **Chez Bey**, le paradigme centripète fonctionne à l'inverse : il internalise le trauma en une économie de l'aveu, de la remémoration et de l'écriture de soi. Le sujet ne fuit pas vers l'avant mais s'enroule sur lui-même, explorant sa propre mémoire et son propre corps. L'écriture n'est pas descriptive mais performative : elle reconstitue le sujet en train de se comprendre. Le temps est celui du bilan rétrospectif.

Significations éthiques et politiques

Au-delà de la structure narrative, cette dichotomie interroge les modalités de l'« habiter » dans l'Algérie contemporaine. Le paradigme centrifuge de Khadra suggère une crise de l'ancrage : l'errance d'Adem ne traduit pas un rejet du territoire, mais une difficulté à s'y fixer. L'espace public, traversé de part en part, devient un lieu de passage fluide plutôt qu'un lieu de résidence stable. L'identité ne se construit plus dans l'enracinement (*le village, la maison*), mais dans la cinétique du déplacement. C'est une sociologie de la mouvance où le sujet survit tant qu'il est en marche.

Symétriquement, le paradigme centripète de Bey propose une autre vérité : celle du sanctuaire intime. Si la parole publique semble fragmentée ou insaisissable, la vérité du sujet trouve refuge dans l'espace clos. L'enfermement n'est pas ici une punition subie, mais la condition de possibilité d'une parole authentique. Ce repli suggère que face à la complexité du monde extérieur, la reconstruction du sens (*et de la citoyenneté du sujet*) doit d'abord passer par une reconquête de soi, à l'abri du tumulte, dans le silence protégé de l'écriture.

Nuances et zones de friction

Si l'opposition entre paradigme centrifuge et centripète s'avère opératoire pour distinguer les architectures globales des deux œuvres, nous nous gardons de toute lecture binaire rigide. Il existe des zones de friction entre ces deux modèles.

- **Chez Khadra**, l'errance est parfois suspendue par des moments de lucidité introspective foudroyante.
- **Chez Bey**, la claustration n'est pas absolue, le monde social pénétrant l'espace intime par le souvenir ou les bruits extérieurs.

Ces paradigmes doivent être lus comme des dominantes structurelles majeures et non comme des frontières étanches. Néanmoins, la massivité statistique des divergences observées confirme que ces moments de croisement restent l'exception plutôt que la règle.

Portée méthodologique

Au-delà de ces conclusions spécifiques aux deux romans, cette étude démontre la fécondité d'une approche triangulée combinant analyse lexicale, phraséologie et modélisation thématique. Chaque niveau d'observation révèle des patterns complémentaires. Aucun niveau seul ne suffirait : *l'analyse lexicale brute masque les architectures discursives ; la phraséologie sans contextualisation statistique reste impressionniste ; la modélisation thématique sans ancrage lexical risque l'abstraction. Leur convergence multimodale produit une compréhension robuste et nuancée.*

Cette triangulation permet également de dépasser le faux dilemme entre quantitatif et qualitatif. Les données chiffrées ne remplacent pas la lecture attentive ; elles l'informent, l'orientent, la valident. Les résultats quantitatifs demandent interprétation qualitative pour acquérir du sens.

Conclusion

Au terme de cette analyse, nous avons démontré comment une approche méthodologique mixte, combinant les forces de la visualisation de données (*Orange*) et de l'analyse contextuelle interactive (*Voyant Tools*), permet de caractériser de manière rigoureuse et fine les fondements stylistiques de deux œuvres littéraires. La divergence entre *le paradigme de l'errance de Yasmina Khadra* et *celui de la confession de Maïssa Bey* n'est pas une simple impression de lecture ; elle est inscrite dans la trame même de leur lexique, de leur phraséologie et de leurs structures thématiques.

Notre principale contribution est double.

- **Sur le plan de l'analyse littéraire**, nous avons offert une caractérisation précise et documentée de deux modes narratifs majeurs dans la littérature algérienne contemporaine.
- **Sur le plan méthodologique**, nous avons présenté un protocole de recherche transparent et reproductible, montrant comment les outils computationnels, loin de déshumaniser le texte, peuvent au contraire en révéler la complexité structurelle et servir de tremplin à une interprétation plus profonde.

Nous reconnaissons les limites de ce travail. L'analyse, focalisée sur deux romans, ne prétend pas à une généralisation sur l'ensemble de l'œuvre des auteurs. De plus, les méthodes lexicométriques restent pour l'instant peu sensibles aux phénomènes sémantiques complexes comme l'ironie ou l'implicite.

Ces mesures capturent des régularités d'usage plus que des intentions : la polysémie, l'ironie, l'ellipse ou l'implicite limitent la transparence sémantique. Les choix de prétraitement et de métriques orientent la saillance (effets d'échelle). Nous assumons une lecture outillée : les modèles ne « découvrent » pas des essences, ils compilent des indices à confronter à des lectures rapprochées (*concordances commentées*), ce qui évite l'illusion de précision.

Ces limites ouvrent cependant des perspectives stimulantes. Une extension de ce corpus à d'autres romans des mêmes auteurs permettrait de vérifier la constance de ces signatures stylistiques. L'intégration d'analyses plus avancées, comme l'analyse de sentiments ou la détection d'entités nommées, pourrait affiner notre compréhension des univers narratifs. Enfin, cette méthode comparative pourrait être appliquée à un panel plus large d'auteurs pour cartographier les différentes réponses stylistiques et thématiques à la crise dans la littérature francophone contemporaine.

Nous proposons, en perspective, la réplication sur d'autres titres de Khadra et Bey avec le même protocole (*paramètres, seeds, cohérence/stabilité*), et l'adjonction d'une analyse de valence (*trajectoires affectives segmentaires*) afin de mieux saisir l'éthos émotionnel de chaque paradigme.

Références

- BAL, M. (2009). *Narratology: Introduction to the theory of narrative* (3e éd.). University of Toronto Press.
- BEY, M. (2018). *Nulle autre voix*. Éditions Barzakh.
- BLEI, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- CARUTH, C. (Ed.). (1995). *Trauma: Explorations in memory*. Johns Hopkins University Press.
- CHATMAN, S. (1978). *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.
- FELMAN, S., & LAUB, D. (1992). *Testimony: Crises of witnessing in literature, psychoanalysis, and history*. Routledge.
- GENETTE, G. (1972). *Figures III*. Éditions du Seuil.
- (1980). *Narrative discourse: An essay in method* (J. E. LEWIN, Trad.). Cornell University Press.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101 (Suppl. 1), 5228–5235.
- JOCKERS, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- KHADRA, Y. (2020). *Le sel de tous les oubliés*. Éditions Julliard.
- LEBART, L., & SALEM, A. (1994). *Statistique textuelle*. Dunod.

- MAINGUENEAU, D. (2014). *Discours et analyse du discours*. Armand Colin.
- McENERY, T., & H3ardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- MIMNO, D., WALLACH, H. M., TALLEY, E., LEENDERS, M., & McCALLUM, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 262–272).
- MORETTI, F. (2005). *Graphs, maps, trees: Abstract models for literary history*. Verso.
- NEWMAN, D., LAU, J. H., GRIESER, K., & BALDWIN, T. (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL* (pp. 100–108).
- ORANGE TEAM. (2020). *Orange: A component-based data mining framework*. <https://orange.readthedocs.io>
- RICOEUR, P. (1983). *Temps et récit. 1. L'intrigue et le récit historique*. Éditions du Seuil.
- ROCKWELL, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities*. MIT Press.
- SAID, E. W. (1978). *Orientalism*. Pantheon Books.
- SINCLAIR, S., & ROCKWELL, G. (2012). Voyant Tools: Rethinking close reading with machine reading. In *Proceedings of the Digital Humanities 2012 Conference*. University of Hamburg.
- SINCLAIR, S., & ROCKWELL, G. (2016). *Voyant Tools: Seeing your data* [software]. <https://voyant-tools.org>
- TEH, Y. W., JORDAN, M. I., BEAL, M. J., & BLEI, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101 (476), 1566–1581.
- UNDERWOOD, T. (2015). Why literary data matters: A rejoinder to commentators. *Journal of Cultural Analytics*, 1, 1–21.

Pour citer cet article

Souad BENELHADJ DJELLOUL, Nafissa ZEROUALI, « Deux voix, deux mondes : Signatures lexicales et architectures discursives dans *Le sel de tous les oublis* et *Nulle autre voix* », *Paradigmes*, vol. IX, n° 01, janvier 2026, p. 75-95.