Revue Internationale de la Recherche Scientifique et de l'Innovation (Revue-IRSI)
ISSN: 2960-2823
Vol. 4, No. 5, Septembre 2026

Analyse linguistique des commentaires des touristes sur la plateforme numérique Booking : étude des facteurs de satisfaction au Sénégal

Auteurs :
Pape Mactar DIAW [1]
Docteur en sociologie du tourisme,
Laboratoire de recherche en sciences économiques et sociales (LARSES),
Université Assane Seck de Ziguinchor.

Assane NDIAYE [2]
Doctorant en linguistique,
École doctorale pluridisciplinaire,
Université Numérique Cheikh Hamidou Kane de Dakar (Sénégal)

Résumé

Cette recherche examine, à partir d'une analyse linguistique outillée, les facteurs de satisfaction et d'insatisfaction exprimés par les touristes ayant séjourné dans des établissements d'hébergement sénégalais référencés sur Booking.com. À partir d'un plan d'échantillonnage stratifié par pôle touristique (Dakar, Thiès-Diourbel, Casamance, Sine Saloum, Nord et Sénégal oriental), un corpus de 372 commentaires exploitables a été constitué et soumis à une chaîne de traitement combinant statistique textuelle, traitement automatique du langage naturel (tokenisation, lemmatisation, suppression des mots vides), analyse lexicométrique (fréquences, TF-IDF, n-grammes, champs lexicaux, réseau de cooccurrences), modélisation thématique (Latent Dirichlet Allocation) et analyse des sentiments globale et par aspects (Aspect-Based Sentiment Analysis, ABSA), cette dernière reposant sur un dictionnaire lexical spécialisé dans le domaine hôtelier en raison de l'indisponibilité, dans l'environnement de traitement, des modèles neuronaux pré-entraînés de type CamemBERT ou XLM-RoBERTa. Les résultats indiquent une tonalité globalement positive du corpus (73,4 % des commentaires), portée principalement par la qualité de l'accueil et du personnel, la propreté et le calme des lieux, tandis que le confort des chambres, le rapport qualité-prix et, dans une moindre mesure, la connexion internet concentrent les motifs d'insatisfaction.

Mots-clés : *UGC ; e-WOM ; Booking.com ; analyse linguistique ; satisfaction touristique ; Sénégal.*

Digital Object Identifier (DOI): https://doi.org/10.5281/zenodo.**22533577**

Ceci est un article en accès libre sous la licence CC BY-NC-ND.

1. Introduction

Dans un monde où les services touristiques sont de plus en plus dématérialisés, les sources de recherche d'informations, d'évaluation et de partage d'expériences évoluent également. Par exemple, sur des sites de réservation en ligne comme Booking.com, les voyageurs laissent des milliers d'avis sur leurs expériences. Ces contenus générés par les utilisateurs (User Generated Content) sont une source d'information particulièrement intéressante pour mieux comprendre les facteurs déterminant la satisfaction touristique. Ces données, plus connues sous l'appélation *electronic Word-of-Mouth* (e-WOM) par les chercheurs, avec un impact supérieur au WOM ou bouche-à-oreille car la portée du message est internationale, et la source est traçable. Les recherches sur l'e-WOM ont mis en évidence que les messages en ligne des consommateurs ont un impact sur les intentions et comportements d'achat. Dans le secteur du tourisme, les services sont intangibles et le futur consommateur se base donc sur l'expérience d'autres consommateurs pour réduire l'incertitude qui peut entourer la décision d'achat. Par exemple, les travaux de Jalilvand et Samiei (2012) montrent que les avis en ligne ont une conséquence sur l'image et l'intention de visiter une destination.

Comme le suggèrent les travaux de Litvin, Goldsmith et Pan (2008), l'argument du bouche-à-oreille numérique est devenu un élément stratégique de compétitivité pour les entreprises touristiques. Le développement de plateformes digitales a favorisé l'émergence d'une méthodologie innovante visant à automatiser l'analyse des avis des utilisateurs par le biais de techniques d'exploration de textes (text mining) et d'analyse des émotions (sentiment analysis). Cette méthode d'analyse des sentiments vise à reconnaître automatiquement les émotions, opinions et jugements exprimés dans des textes. Selon Liu (2012), cette analyse permet d'identifier si un texte exprime un sentiment positif, négatif ou neutre, afin de mieux comprendre la perception des consommateurs. Cette méthode s'est popularisée et est désormais largement utilisée pour traiter les vastes volumes de données générés par les plateformes digitales. Dans le secteur du tourisme, plusieurs études montrent que l'analyse linguistique des avis aide à identifier précisément les facteurs influençant la satisfaction des touristes. À partir d'une analyse de plus de 40 000 commentaires recueillis sur Booking.com et TripAdvisor, Rita, Ramos, Borges-Tiago et Rodrigues (2022) notent que les points souvent associés aux bonnes notes portent sur la qualité du personnel, l'accueil chaleureux et la localisation des hôtels. En revanche, les avis négatifs portent principalement sur les tarifs, la facturation et le rapport qualité-prix. Les experts montrent également que le style utilisé par les voyageurs varie en fonction de la plateforme utilisée et de leur nationalité, ce qui souligne l'importance d'une étude linguistique approfondie des commentaires en ligne.

Les recherches récentes tendent aujourd'hui vers une analyse par aspects (Aspect-Based Sentiment Analysis), allant au-delà de l'évaluation de la polarité des commentaires. Cette méthode ne cherche plus seulement à déterminer si un commentaire est positif ou négatif, elle identifie également les aspects spécifiques liés à ces émotions, tels que la propreté, le confort, la restauration, le service ou l'accessibilité. Chu, Chen, Yang et Wang (2022) montrent que les modèles fondés sur BERT sont désormais capables d'extraire simultanément les caractéristiques mentionnées et l'impact émotionnel qui leur est attribué avec une précision nettement supérieure aux techniques statistiques traditionnelles. Leurs recherches montrent que les aspects relatifs à la nourriture, aux prix, à la propreté et à la fréquentation des sites touristiques sont les principaux déclencheurs des émotions négatives exprimées dans les avis en ligne. Ces progrès s'inscrivent dans un mouvement plus large d'évolution des techniques de traitement automatique des langues. Auparavant, les recherches se faisaient principalement à partir de dictionnaires et de techniques statistiques. Les études récentes, par contre, font usage

de modèles de *Deep Learning* et peuvent ainsi intégrer le contexte langagier des mots. Cette mise à jour accroît considérablement la précision du classement des émotions, notamment sur de longs textes où les subtilités, l'ironie ou les motivations sous-jacentes amoindrissent l'efficacité des méthodes classiques. Les revues de littérature centrées sur l'analyse des données textuelles dans le domaine touristique confirment cette montée en puissance des modèles neuronaux, devenus aujourd'hui la référence des études en intelligence artificielle dans le tourisme. Les recherches sur l'usage des avis en ligne montrent également que les données textuelles apportent une plus-value importante par rapport aux simples notes numériques attribuées aux hôtels. Si l'évaluation globale ne donne qu'une idée générale de la satisfaction, les détails linguistiques des commentaires permettent de déterminer les motifs exacts qui sous-tendent cette note. L'étude de García-Cumbreras et coll. (2014) montre que l'analyse des commentaires par aspects fournit aux gestionnaires d'hôtels des informations nettement plus utiles pour improviser leurs services. Grâce à leur méthode, il est possible, entre autres, d'identifier les points à travailler en priorité pour améliorer l'expérience des visiteurs.

En dépit de l'existence d'une vaste littérature internationale, les études centrées sur le continent africain restent relativement peu nombreuses. La majorité des recherches se concentrent sur des destinations en Europe, en Asie ou en Amérique du Nord, où des quantités considérables de commentaires sont produites sur les plateformes numériques. Dans le contexte de l'Afrique subsaharienne, et particulièrement au Sénégal, rares sont les recherches qui utilisent des méthodes d'analyse linguistique pour déterminer les facteurs de la satisfaction touristique à partir des commentaires postés sur Booking.com. Ce déficit d'autant plus montre que le tourisme est crucial pour l'économie du Sénégal et que les hôtels sont actuellement largement représentés sur les plateformes numériques mondiales. Cette étude vise justement à pallier ce manque dans la recherche scientifique en examinant les avis déposés sur Booking.com à propos des lieux touristiques au Sénégal. Ainsi, dans quelle analyse l'examen linguistique des critiques laissées par les touristes sur Booking.com peut-elle aider à déterminer les éléments majeurs de satisfaction et de mécontentement des visiteurs au sein des logements au Sénégal ? L'objectif de cette recherche est de cerner à la fois les facteurs clés de satisfaction et de mécontentement des visiteurs dans le contexte sénégalais, mais aussi de souligner les particularités linguistiques des commentaires écrits dans ce contexte africain. Cette approche contribuera à approfondir la compréhension du comportement des touristes tout en offrant aux acteurs du domaine des indicateurs précieux pour rehausser la qualité de leurs services.

2. Matériel et méthodes

Cette recherche s'inscrit dans une démarche mixte alliant méthode quantitative et qualitative dominante fondées sur l'exploitation de données statistiques et textuelles issues des plateforme Booking.com. Elle mobilise les méthodes de *Text Mining*, de Traitement Automatique du Langage Naturel (Natural Language Processing) et d'analyse des perceptions afin d'examiner les commentaires publiés par les touristes. Cette approche permet d'extraire automatiquement les informations contenues dans les avis des utilisateurs, d'identifier les principaux thèmes évoqués et de mesurer les sentiments associés aux différentes dimensions de l'expérience touristique. Elle porte sur les établissements d'hébergement touristique localisés au Sénégal et référencés sur Booking.com. Le choix de cette plateforme se justifie par son importance dans le secteur de l'hébergement touristique, son audience internationale et le volume considérable de commentaires qu'elle met à la disposition des utilisateurs. Elle est la principale plateforme mondiale de réservation d'hébergements en ligne, opérant dans plus de 220 pays et territoires avec environ 31 millions d'annonces, dont 4,4 millions d'établissements. En 2025, Booking Holdings a enregistré 186,1 milliards de dollars de réservations brutes et plus de 1,235 milliard

de nuitées réservées, grâce à une base de plusieurs centaines de millions de clients. La plateforme s'appuie sur un puissant référencement numérique et un algorithme de classement fondé sur la pertinence, les performances commerciales et les interactions des utilisateurs. Son système d'avis vérifiés, notés de 1 à 10, constitue un indicateur déterminant de la réputation des établissements et influence fortement les décisions de réservation (Booking Holdings, 2026 ; Investors Business Daily, 2025). Dans ce contexte, les établissements retenus pour cette étude sont répartis dans les principales destinations touristiques du pays (cf. Figure 1**)**, notamment en raison de leur forte rétention des établissements et services touristique comme :

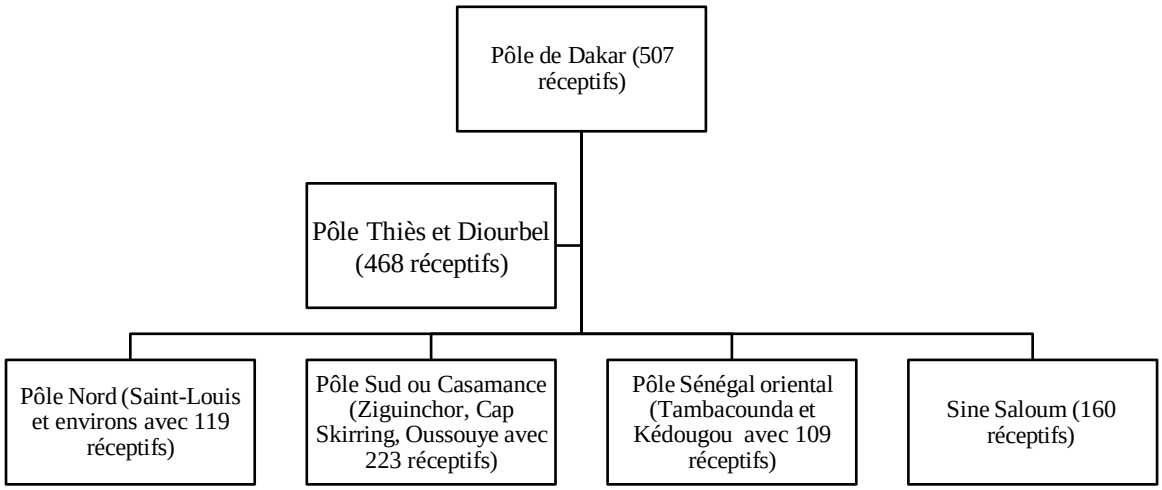


Figure 1. répartition des établissements touristiques dans les principaux pôles touristiques.
Source : selon les données de la Direction de la réglementation touristique du Sénégal, DRT, 2025.

Le choix de ces pôles répond à plusieurs considérations scientifiques. Premièrement, ils concentrent une part importante de l'offre nationale d'hébergement touristique, permettant ainsi de disposer d'un volume suffisant de commentaires publiés sur Booking.com. Deuxièmement, ils présentent une diversité de formes de tourisme (balnéaire, culturel, patrimonial, écologique et d'affaires), ce qui favorise une analyse comparative des facteurs de satisfaction exprimés par les touristes. Enfin, cette répartition géographique permet de couvrir les principales destinations fréquentées par les visiteurs nationaux et internationaux, améliorant ainsi la représentativité du corpus étudié. Les données analysées proviennent exclusivement des commentaires publiés sur Booking.com par les visiteurs ayant séjourné dans les établissements sélectionnés. Le corpus comprend les avis textuels, les notes globales attribuées aux établissements, la date du séjour, la langue du commentaire ainsi que les informations non nominatives disponibles concernant le profil du voyageur. Le corpus est constitué à partir des commentaires répondant aux critères suivants :

- être publiés pour un établissement situé au Sénégal ;
- contenir un texte exploitable ;
- rédigé en français, en anglais, en allemand ou en espagnol.

Selon la DRT (2025) 1 586 établissements touristiques sont identifiés et répartis comme suit : 327auberges, 228 campements, 423 hôtels et 608 résidences. Sur la base de cet effectif, nous partons sur un échantillonnage de 25 %, soit 396,5 arrondi à 397 avis/commentaires collectés. La répartition proportionnelle des établissements et avis est décrite dans le tableau 1, selon la

méthode de stratification : l'extraction brute des avis ne restitue pas nativement une structure exploitable directement : établissement, note, nationalité du voyageur et texte de l'avis apparaissent regroupés de façon hétérogène selon les pôles. Un script de reconstitution structurée (identification des marqueurs de numérotation des avis, des ruptures d'établissement et des mentions de nationalité) a donc été développé afin de reconstituer, pour chaque avis, ces quatre champs. Cette étape de structuration conduit à un corpus effectivement mobilisé pour les analyses de 372 commentaires, soit 93,7 % de l'échantillon visé (397) ; l'écart correspond aux avis dont la segmentation, insuffisamment fiable dans l'export brut, ne permettait pas une attribution non ambiguë à un établissement ou à un auteur, et qui ont été exclus par précaution méthodologique plutôt que reconstitués par estimation. De la même façon, la note Booking.com de l'établissement, bien que prévue comme variable de cadrage, n'a pu être extraite de façon exploitable que pour une partie des établissements identifiés (25 sur 54, concentrés sur trois des six pôles) ; cette limite de couverture est assumée et rapportée telle quelle dans les résultats plutôt que comblée artificiellement.

Tableau 1. répartition de la taille de l'échantillon par pôle

Pôle touristique	**Régions couvertes**	**Nombre de réceptifs**	**Poids du pôle (%)**	**Échantillon à enquêter**	**Corpus effectivement analysé**
Dakar	Dakar	507	31,97 %	127	113
Thiès-Diourbel	Thiès, Diourbel	468	29,51 %	117	112
Casamance	Ziguinchor, Kolda, Sédhiou	230	14,50 %	58	58
Sine Saloum	Fatick, Kaolack, Kaffrine	153	9,65 %	38	37
Pôle Nord	Saint-Louis, Louga, Matam	119	7,50 %	30	30
Sénégal oriental	Tambacounda, Kédougou	109	6,87 %	27	22
Total		**1 586**	**100 %**	**397**	372

Source : auteurs, 2026 et données DRT, 2025.

Pour rappel, la collecte des données s'est effectuée entrele mois de Juillet et Août 2026. Avant leur analyse, les commentaires font l'objet d'un processus de prétraitement destiné à améliorer la qualité des données textuelles. Cette étape comprend la suppression des caractères spéciaux, des balises HTML, des espaces superflus et des mots vides (*stop words*), ainsi que la conversion des textes en minuscules. Les commentaires vides, dupliqués ou ne comportant qu'une

appréciation sans contenu descriptif sont exclus de l'analyse. Cette procédure vise à constituer une base de données homogène et suffisamment représentative des expériences touristiques exprimées sur la plateforme. La variable dépendante est la satisfaction du touriste, appréhendée à travers la note globale attribuée à l'établissement ainsi que le sentiment exprimé dans le commentaire. Les variables explicatives correspondent aux principales dimensions de la qualité de service identifiées dans la littérature sur le tourisme et l'hôtellerie : la qualité de l'accueil, le professionnalisme du personnel, la propreté, le confort des chambres, les équipements, la restauration, la connexion Internet, l'accessibilité, la sécurité, le climat ou encore le rapport qualité-prix.

L'analyse des données est conduite en plusieurs étapes complémentaires : une analyse descriptive permet de caractériser le corpus à travers la distribution des commentaires, la fréquence des notes, la répartition linguistique des avis ainsi que les mots et expressions les plus fréquemment employés ; une analyse lexicale est réalisée afin d'identifier les unités lexicales dominantes, les cooccurrences de mots et les principaux champs sémantiques présents dans les commentaires. Cette étape vise à mettre en évidence les thèmes les plus récurrents dans le discours des touristes. Une analyse des sentiments est ensuite conduite afin de déterminer automatiquement la polarité des commentaires (positive, négative ou neutre) ; mais aussi une analyse des sentiments par aspects (Aspect-Based Sentiment Analysis, ABSA) est réalisée afin d'associer chaque opinion exprimée aux différentes dimensions de la qualité de service (propreté, accueil, localisation, restauration, confort, etc.). Cette approche permet d'identifier avec précision les facteurs qui contribuent à la satisfaction ou à l'insatisfaction des touristes. Les analyses statistiques et textuelles sont réalisées sous Python.

Sur le plan technique, la tokenisation, l'étiquetage morphosyntaxique et la lemmatisation du corpus ont été réalisés à l'aide du modèle français de la bibliothèque spaCy, en conservant uniquement les catégories grammaticales porteuses de sens (noms, verbes, adjectifs, adverbes et noms propres) et en écartant les mots vides ainsi que les lemmes de moins de trois caractères. Les indicateurs de fréquence lexicale (TF) et de pondération TF-IDF ont été calculés à l'aide du module scikit-learn (CountVectorizer), chaque commentaire étant traité comme un document distinct pour le calcul du TF-IDF. Le réseau de cooccurrences lexicales a été construit et représenté à l'aide de la bibliothèque network.

3. Résultats

Les résultats sont présentés en suivant la logique méthodologique exposée précédemment : caractérisation statistique du corpus, profils des répondants et des établissements, analyses lexicale et sémantique, modélisation thématique, analyse des sentiments globale et par aspects, comparaison inter-pôles, puis hiérarchisation des facteurs de satisfaction et d'insatisfaction. L'ensemble des indicateurs présentés est calculé directement à partir du corpus collecté ; lorsque l'information n'a pu être extraite de façon fiable pour l'ensemble des observations (par exemple la note Booking.com de certains établissements), la limite est explicitement signalée plutôt que comblée par estimation.

3.1 Statistiques descriptives du corpus

Le corpus final rassemble 372 commentaires exploitables sur les 397 avis visés par le plan d'échantillonnage stratifié (soit un taux de réalisation de 93,7 %), l'écart s'expliquant par la présence, dans l'export brut, d'un nombre résiduel de commentaires tronqués ou dont la segmentation ne permettait pas une attribution fiable à un avis individuel ; ces observations ont

été exclues par précaution méthodologique plutôt que reconstituées. Le tableau II présente les indicateurs de longueur du corpus.

Tableau 2. Statistiques descriptives du corpus (n = 372 commentaires)

Indicateur	Valeur
Nombre de commentaires	372
Nombre total de mots	19069
Nombre de mots uniques	4019
Longueur moyenne (mots)	51,26
Médiane (mots)	36,00
Écart-type (mots)	54,47
Minimum (mots)	1
Maximum (mots)	415

Source : auteurs, 2026.

Le corpus totalise 19 069 mots (occurrences), pour 4 019 formes lexicales distinctes, ce qui correspond à un ratio type/token (type-token ratio) de 21,1 %, révélateur d'une diversité lexicale modérée typique des corpus d'avis en ligne, où les touristes recourent à un vocabulaire relativement stable pour décrire des expériences de service comparables (accueil, chambre, restauration). La longueur moyenne d'un commentaire est de 51,3 mots, mais la médiane, nettement inférieure (36 mots), et l'écart-type élevé (54,5) indiquent une distribution fortement asymétrique à droite : la majorité des avis sont courts et synthétiques, tandis qu'une minorité de commentaires très détaillés, pouvant atteindre 415 mots, tire la moyenne vers le haut. Cette hétérogénéité rédactionnelle, déjà documentée par Xiang et al. (2017) dans leur comparaison des plateformes d'avis, invite à ne pas limiter l'analyse à la fréquence brute des termes et justifie le recours complémentaire à la pondération TF-IDF ainsi qu'à l'analyse par aspects, plus robustes face à la variabilité de longueur des textes. La section suivante situe ce corpus du point de vue du profil des touristes qui en sont les auteurs.

3.2 Profil sociodémographique des touristes

Les commentaires du corpus renseignent, pour 361 avis sur 372 (97,0 %), la nationalité déclarée du voyageur, permettant de reconstituer le profil géographique de la clientèle des établissements sénégalais sur Booking.com. 37 nationalités distinctes sont représentées. La figure 2 présentent les douze nationalités les plus fréquentes.

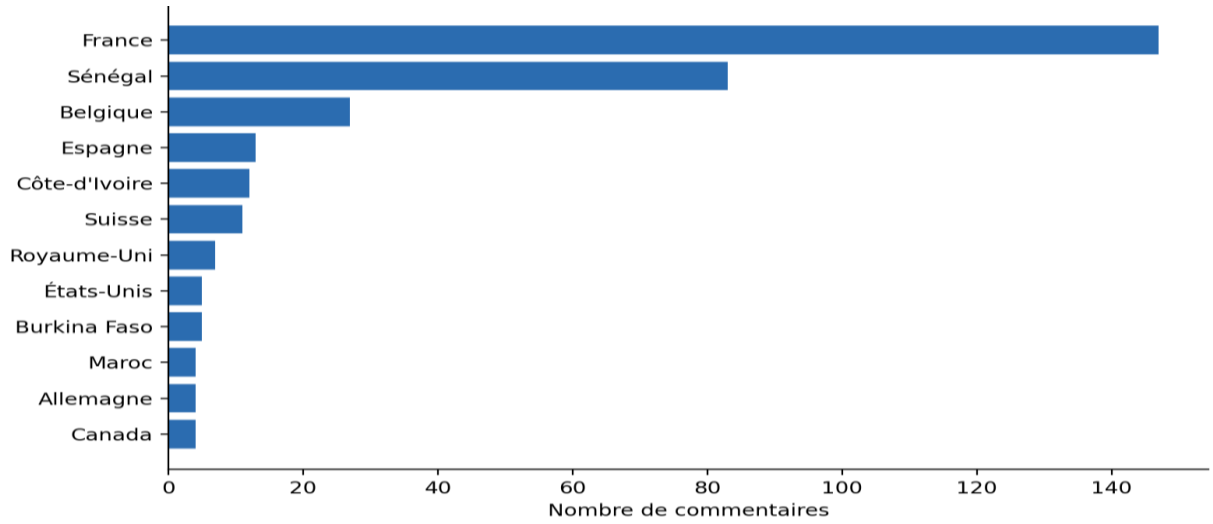


Figure 2. nationalités des touristes les plus représentées (top 12). **Source :** auteurs, 2026.

La clientèle française domine très largement le corpus (39,5 %), conformément au poids historique et actuel de la France dans les flux touristiques à destination du Sénégal, suivie par la clientèle sénégalaise (22,3 %), ce qui traduit l'importance de la diaspora et du tourisme domestique ou d'affaires dans la fréquentation des établissements référencés sur Booking.com. Les touristes belges (7,3 %), espagnols (3,5 %) et ivoiriens (3,2 %) complètent le noyau des nationalités les plus représentées, cependant qu'une longue traîne de 32 autres nationalités (Suisse, Royaume-Uni, États-Unis, pays d'Afrique de l'Ouest et centrale, Maghreb, Europe du Nord et de l'Est, Asie, Amériques) atteste d'une fréquentation internationale diversifiée, cohérente avec le statut de Booking.com comme plateforme mondiale de réservation (Booking Holdings, 2026). Cette diversité linguistique et culturelle de l'échantillon, également relevée par Rita et al. (2022) dans leur comparaison de plateformes d'avis, a une incidence directe sur l'analyse lexicale : le corpus comporte, à côté de la majorité de commentaires rédigés en français, des avis en anglais et en espagnol, dont la présence a été prise en compte dans l'interprétation des résultats lexicaux et de sentiment. Le dépouillement du corpus permet d'identifier 54 établissements distincts couverts par les 372 commentaires analysés, répartis dans les six pôles touristiques retenus par l'étude. La figure 3 croise cette répartition par pôle et par catégorie d'hébergement (hôtels, résidences, campements, auberges et autres formes d'hébergement de type villa ou maison d'hôtes).

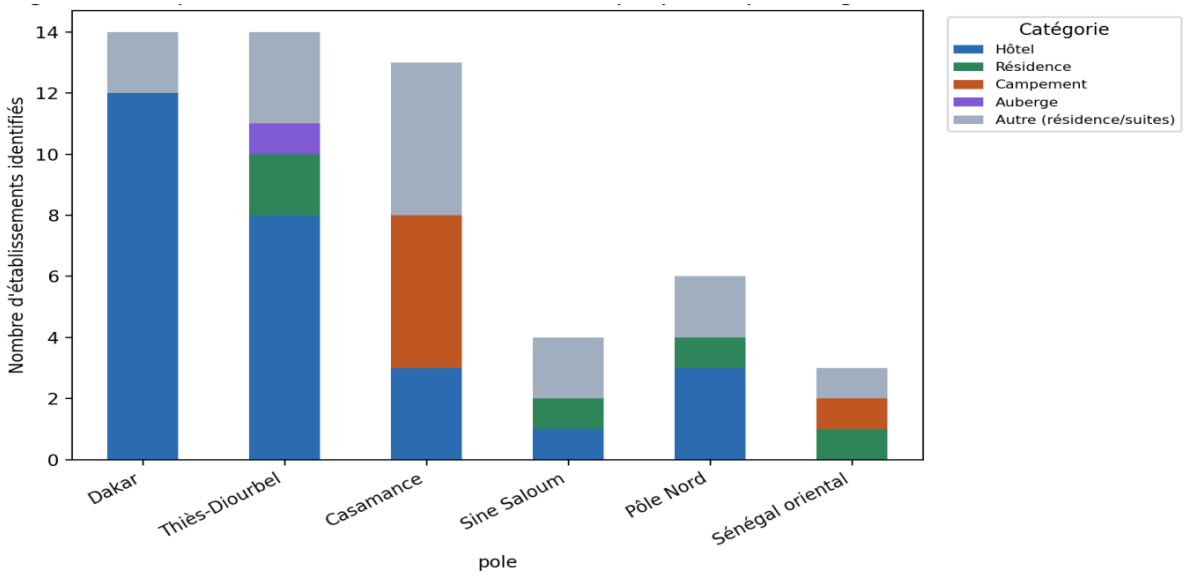


Figure 3. Répartition des établissements par pôle et par catégorie. **Source :** auteurs, 2026.

Les hôtels constituent la catégorie dominante (27 établissements, 50,0 %), suivis par les hébergements de type villa, maison d'hôtes ou suites, regroupés sous la catégorie « Autre » (15 établissements, 27,8 %), puis par les campements (6, 11,1 %), les résidences (5, 9,3 %) et les auberges (1, 1,9 %). Cette structure reflète la diversité des formes d'hébergement touristique attendue dans des pôles aux vocations différenciées : Dakar et Thiès-Diourbel concentrent l'essentiel de l'offre hôtelière balnéaire et urbaine, tandis que la Casamance se distingue par une part plus importante de campements et d'hébergements alternatifs, cohérente avec son positionnement de tourisme rural, communautaire et écologique déjà documenté dans la littérature sur le tourisme sénégalais. La note Booking.com moyenne des établissements, quant à elle, n'a pu être extraite de façon fiable et systématique pour l'ensemble des 54 établissements : elle figurait explicitement dans le texte collecté pour 25 d'entre eux (46,3 %), concentrés dans les pôles Thiès-Diourbel (n = 8, note moyenne de 8,15/10), Casamance (n = 12, note moyenne de 8,88/10) et Nord (n = 5, note moyenne de 8,20/10), avec une moyenne générale de 8,51/10 sur cette sous-population documentée. Cette information n'était en revanche pas disponible de façon exploitable dans les extractions relatives aux pôles Dakar, Sine Saloum et Sénégal oriental, où la note ne figurait pas systématiquement à proximité du nom de l'établissement dans l'export brut ; plutôt que d'imputer une valeur estimée. Cette hétérogénéité de couverture n'affecte cependant pas les analyses linguistiques et de sentiment qui suivent, celles-ci portant sur l'intégralité des 372 commentaires indépendamment de la disponibilité de la note chiffrée. L'analyse se porte à présent sur le contenu lexical des commentaires eux-mêmes.

3.3 Analyse lexicale : fréquences et TF-IDF

Le corpus a fait l'objet d'un prétraitement standard en trois étapes : tokenisation, lemmatisation et suppression des mots vides, réalisées à l'aide du modèle français de spaCy (fr_core_news_sm), en conservant uniquement les catégories grammaticales porteuses de sens (noms, verbes, adjectifs, adverbes, noms propres). Ce traitement réduit le corpus à 8 688 occurrences lexicales significatives, correspondant à 2 800 lemmes distincts. Pour chaque lemme, la fréquence brute (TF) et la pondération TF-IDF moyenne, calculée en traitant chaque commentaire comme un document distinct, ont été calculées. Le tableau 3 présente les 30

premiers rangs de ce classement (le tableau complet des 100 premiers rangs est disponible en annexe).

Tableau 3. Top 30 des mots les plus fréquents du corpus lemmatisé (fréquence et TF-IDF)

Rang	Mot	Fréquence	TF (%)	TF-IDF moyen
1	personnel	185	2,13	0,0520
2	chambre	178	2,05	0,0544
3	petit	130	1,50	0,0449
4	hôtel	123	1,42	0,0374
5	piscine	84	0,97	0,0309
6	accueil	79	0,91	0,0341
7	bon	77	0,89	0,0324
8	déjeuner	57	0,66	0,0279
9	agréable	54	0,62	0,0238
10	calme	53	0,61	0,0249
11	gentillesse	50	0,58	0,0211
12	emplacement	50	0,58	0,0271
13	propre	45	0,52	0,0200
14	accueillir	42	0,48	0,0177
15	confortable	40	0,46	0,0153
16	qualité	40	0,46	0,0185
17	service	39	0,45	0,0186
18	plage	39	0,45	0,0165
19	restaurant	39	0,45	0,0161
20	endroit	37	0,43	0,0166
21	prix	36	0,41	0,0159
22	eau	34	0,39	0,0153
23	vue	34	0,39	0,0168
24	heure	34	0,39	0,0138
25	beaucoup	32	0,37	0,0126
26	bel	31	0,36	0,0168
27	salle	30	0,34	0,0136
28	hôtel	30	0,34	0,0159
29	excellent	30	0,34	0,0161
30	super	29	0,33	0,0187

Source : auteurs, 2026.

Le lemme « personnel » arrive en tête du classement (185 occurrences, TF = 2,13 %), immédiatement suivi par « chambre » (178), « petit » (130, très majoritairement dans l'expression « petit-déjeuner »), « hôtel » (123) et « piscine » (84). Cette hiérarchie lexicale, largement dominée par le champ humain du service (personnel, accueil, gentillesse) et par les composantes matérielles de l'hébergement (chambre, piscine, petit-déjeuner), converge avec les résultats de Rita et al. (2022), pour qui la qualité du personnel et l'accueil figurent parmi les dimensions les plus fréquemment associées aux avis positifs sur Booking.com et TripAdvisor. La pondération TF-IDF, qui atténue le poids des termes ubiquitaires au profit de ceux qui discriminent davantage les commentaires entre eux, conserve globalement ce classement tout

en le nuançant : des termes comme « rapport », « qualité » ou « emplacement », moins fréquents en valeur absolue, obtiennent un poids TF-IDF relativement élevé, signe qu'ils interviennent dans des sous-ensembles de commentaires plus spécifiques, souvent liés à l'expression d'un jugement comparatif ou évaluatif plutôt que descriptif. L'analyse lexicale isolée ne permettant pas de restituer la structure syntagmatique du discours, elle est complétée par l'étude des associations de mots contigus.

3.4 Analyse des n-grammes

L'examen des séquences de deux (bigrammes) et de trois (trigrammes) lemmes contigus permet de restituer des expressions figées ou des associations sémantiques que l'analyse lexicale isolée ne peut révéler. La figure 4 illustrant les bigrammes dominants.

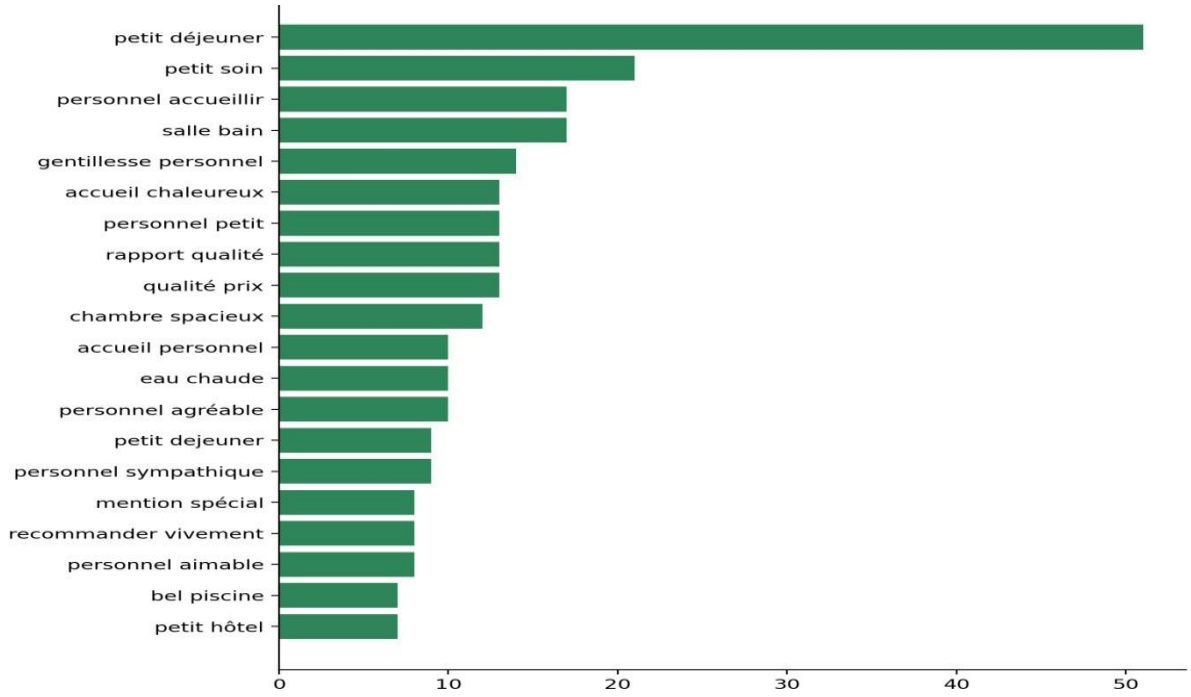


Figure 4. Top 20 des bigrammes. **Source :** auteurs, 2026.

Le bigramme « petit déjeuner » domine très largement le classement (51 occurrences), loin devant « petit soin » (21) et « personnel accueillir » (17), confirmant l'importance de la restauration matinale comme point de contact quotidien entre le personnel et le client. Le trigramme le plus fréquent, « rapport qualité prix » (13 occurrences), matérialise à lui seul une figure de style évaluative récurrente dans les avis en ligne, dans laquelle le client synthétise son jugement global en une expression standardisée ; il est suivi par « personnel petit soin » (10) et « bon rapport qualité » (6), qui prolongent la même logique. La récurrence de ces expressions figées, plutôt que de mots isolés, illustre ce que Cambria et White (2014) désignent comme la dimension syntagmatique du traitement du langage naturel, dans laquelle le sens émerge autant de la combinaison des mots que de leur fréquence individuelle ; elle justifie, en aval, le regroupement de ces unités lexicales en champs sémantiques plus larges.

3.5 Analyse des champs lexicaux thématiques

Onze champs lexicaux ont été construits à partir des dimensions de la qualité de service identifiées dans la littérature hôtelière et confirmées par les résultats des sections précédentes (personnel et accueil, propreté, confort, piscine, restauration, connexion internet, plage et cadre naturel, calme, localisation, sécurité, rapport qualité-prix). Pour chaque champ, le nombre d'occurrences totales des lemmes qui le composent ainsi que la part des commentaires y faisant référence ont été calculés (tableau 4).

Tableau 4. Poids des champs lexicaux thématiques dans le corpus

Champ lexical	Occurrences totales	Commentaires concernés	% des commentaires (n=372)
Personnel / Accueil / Service	622	247	66.4
Restaurant / Petit-déjeuner	352	181	48.7
Confort	136	90	24.2
Localisation	113	89	23.9
Rapport qualité-prix	112	62	16.7
Piscine	90	73	19.6
Calme	81	68	18.3
Propreté	70	64	17.2
Plage	64	54	14.5
Wifi / Connexion	35	25	6.7
Sécurité	14	13	3.5

Source : auteurs, 2026.

Le champ « Personnel / Accueil / Service » domine très largement (622 occurrences, présent dans 66,4 % des commentaires), loin devant la « Restauration / Petit-déjeuner » (48,7 %). Ce résultat rejoint directement les conclusions de Rita et al. (2022) et de García-Cumbreras et al. (2014), pour qui la dimension humaine du service constitue le premier déterminant de la satisfaction hôtelière exprimée en ligne, avant même les attributs matériels de l'hébergement. Les champs « Confort » (24,2 %) et « Localisation » (23,9 %) occupent une place intermédiaire, tandis que le « Rapport qualité-prix » (16,7 %), la « Piscine » (19,6 %), le « Calme » (18,3 %) et la « Propreté » (17,2 %) forment un second groupe de préoccupations récurrentes mais moins centrales. À l'inverse, le « Wifi / Connexion » (6,7 %) et la « Sécurité » (3,5 %) sont peu mentionnés, ce qui peut traduire soit une moindre importance perçue de ces dimensions dans le contexte étudié, soit, plus vraisemblablement pour la sécurité, un biais de non-mention propre aux avis en ligne, les touristes tendant à ne signaler ce type d'aspect que lorsqu'un incident survient. La structure de ces champs lexicaux, obtenue par comptage de cooccurrences dans les mêmes commentaires, appelle une lecture plus fine des associations entre termes, objet de la section suivante.

3.6 Analyse des cooccurrences et réseau lexical

Un réseau de cooccurrences a été construit à partir des 35 lemmes les plus fréquents du corpus, deux termes étant reliés lorsqu'ils apparaissent conjointement dans un même commentaire. La figure 5 restitue le réseau lexical correspondant, restreint aux associations les plus robustes (seuil de 14 cooccurrences conjointes) afin d'en préserver la lisibilité.

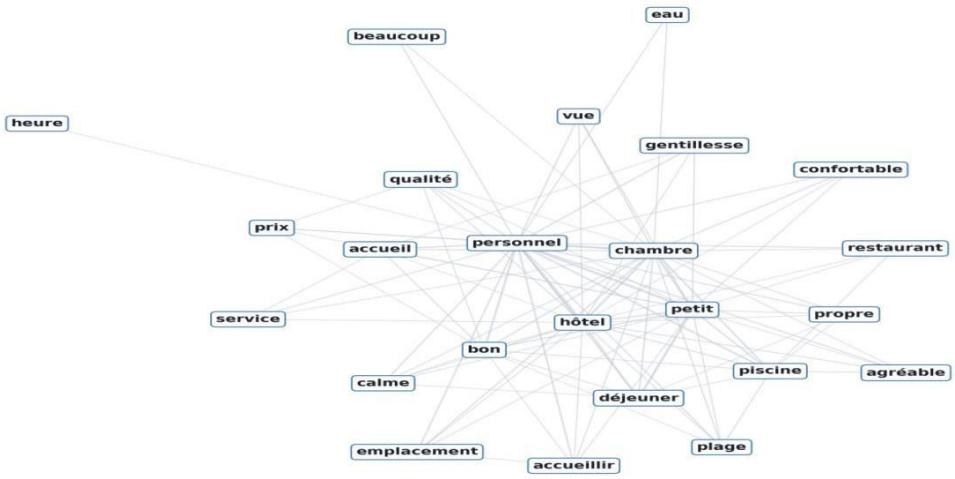


Figure 5. Réseau de cooccurrences lexicales (top 30 termes, seuil ≥ 14). **Source :** auteurs, 2026.

Le réseau confirme la centralité du terme « personnel », relié avec une forte intensité à « chambre » (62 cooccurrences), « hôtel » (60), « piscine » (41), « accueil » (38) et « déjeuner » (32) : dans le discours des touristes, l'évaluation du personnel n'est pas isolée mais systématiquement associée aux autres composantes de l'expérience de séjour, ce qui en fait un nœud structurant de l'ensemble du corpus plutôt qu'un simple thème parmi d'autres. Cette configuration en étoile, où un petit nombre de termes concentrent la majorité des connexions, est cohérente avec les propriétés des réseaux lexicaux de petite taille issus de corpus thématiquement homogènes, tels que documentés dans les travaux de type text mining appliqués au tourisme (Xiang et al., 2017). Le réseau ainsi obtenu offre une cartographie descriptive des associations lexicales ; pour dépasser cette lecture manifeste et faire émerger des structures thématiques latentes, une modélisation probabiliste des thèmes a été mise en œuvre.

3.7 Modélisation thématique (Topic Modeling)

En l'absence d'accès, dans l'environnement de traitement, aux modèles d'embeddings contextuels nécessaires à une approche de type BERTopic, la modélisation thématique a été réalisée à l'aide d'une allocation de Dirichlet latente (Latent Dirichlet Allocation, LDA), estimée sur le corpus lemmatisé vectorisé (sac de mots, vocabulaire restreint aux termes apparaissant dans au moins trois commentaires). Six thèmes ont été extraits ; le tableau 5 en présente les mots-clés dominants, la part du corpus dont chaque thème est le thème majoritaire, ainsi qu'un exemple représentatif de commentaire.

Tableau 5. Thèmes issus de la modélisation LDA (6 thèmes)

Thème	Intitulé proposé	Mots-clés (top 10)	Poids cumulé (%)	Nb commentaires dominants	% du corpus
Thème 1	Restauration, horaires et rapport qualité-prix	petit, personnel, chambre, bon, emplacement, heure, restaurant, repas, déjeuner, prix	90.3	65	17.5
Thème 2	Avis anglophones, accueil, gentillesse et hébergement de type studio	accueil, not, gentillesse, nice, studio, personnel, bon, buffet, eau, maison	86.7	42	11.3
Thème 3	Confort, piscine et cadre balnéaire	chambre, personnel, petit, hôtel, piscine, déjeuner, confortable, agréable, emplacement, plage	92.0	80	21.5
Thème 4	Avis hispanophones et jugement global qualité-prix	chambre, endroit, super, hôtel, qualité, prix, confortable, muy, petit, salle	87.9	29	7.8
Thème 5	Service, qualité et excursions (Sénégal oriental)	hôtel, personnel, petit, service, agréable, doba, qualité, bon, accueil, déjeuner	89.3	49	13.2
Thème 6	Accueil humain, calme et gentillesse du personnel	personnel, petit, hôtel, chambre, piscine, accueil, déjeuner, beaucoup, gentillesse, calme	90.1	107	28.8

Source : auteurs, 2026.

Le thème 6, le plus représenté (28,8 % des commentaires, soit 107 avis), rassemble un vocabulaire centré sur l'accueil, la gentillesse du personnel et le calme, confirmant la centralité déjà observée dans les analyses lexicale et de cooccurrences. Le thème 3 (21,5 %) associe confort, piscine et cadre balnéaire, tandis que le thème 1 (17,5 %) regroupe les mentions relatives à la restauration, aux horaires de service et au rapport qualité-prix. Les thèmes 2 (11,3 %) et 4 (7,8 %) se distinguent par la présence de termes anglais (« not », « nice ») et espagnols (« muy ») parmi leurs mots-clés dominants, ce qui suggère que la LDA capture, au-delà d'un simple regroupement sémantique, un effet de langue de rédaction : les commentaires anglophones et hispanophones, minoritaires mais bien identifiés par le modèle, se distinguent

statistiquement des commentaires francophones alors même qu'ils portent sur des thèmes de fond comparables (hébergement, accueil, rapport qualité-prix). Le thème 5 (13,2 %), enfin, inclut la mention « doba », renvoyant aux commentaires relatifs aux circuits organisés par l'opérateur Doba Tours dans le Sénégal oriental, illustrant la capacité du modèle à isoler des particularités locales même dans un corpus de taille modeste. Ce chevauchement partiel des thèmes, plusieurs mots-clés dominants étant communs à plusieurs thèmes (personnel, chambre, petit-déjeuner), constitue une limite inhérente à l'application de la LDA à un corpus de commentaires courts et thématiquement homogènes ; ce point est repris dans la discussion des limites méthodologiques. Les six thèmes ainsi dégagés fournissent une grille de lecture qualitative qui sera mise en regard, dans les sections suivantes, avec une mesure quantitative de la tonalité émotionnelle des commentaires.

3.8 Analyse des sentiments

L'analyse des sentiments a été conduite selon une approche lexicale, un dictionnaire de polarité construit et adapté au domaine hôtelier francophone (comportant les principales variantes positives et négatives observées dans le corpus, ainsi qu'un traitement de la négation par fenêtre contextuelle) ayant été substitué aux modèles neuronaux pré-entraînés CamemBERT et XLM-RoBERTa initialement envisagés, ceux-ci n'étant pas accessibles hors ligne dans l'environnement de traitement utilisé pour cette recherche. Cette substitution méthodologique, documentée et assumée comme telle, s'inscrit néanmoins dans une tradition éprouvée d'analyse de sentiment par dictionnaire (Liu, 2012). Chaque commentaire s'est vu attribuer un score de polarité, puis classé en positif, neutre ou négatif.

Le corpus est très majoritairement positif : 73,4 % des commentaires (n = 273) expriment une tonalité positive, contre 16,9 % de commentaires neutres (n = 63) et seulement 9,7 % de commentaires négatifs (n = 36). Ce résultat, cohérent avec les notes Booking.com généralement élevées observées pour les établissements documentés dans cette étude (moyenne de 8,51/10), s'inscrit dans un phénomène plus large de biais positif des avis en ligne d'hébergement touristique, également relevé par Rita et al. (2022) dans leur comparaison de Booking.com et TripAdvisor, où les auteurs notent que les clients les plus insatisfaits recourent davantage à des canaux de réclamation directs qu'à la publication d'avis en ligne. La distribution des scores de polarité, fortement concentrée autour de valeurs positives mais avec une traîne négative non négligeable, confirme cette tendance tout en révélant l'existence d'un noyau significatif de commentaires critiques qu'une lecture agrégée masquerait. Cette approche globale, qui traite chaque commentaire comme une unité indivisible, ne permet cependant pas d'identifier les dimensions précises de l'expérience touristique à l'origine de ces jugements contrastés, ce qui justifie le recours à une analyse des sentiments par aspects.

3.9 Analyse des sentiments par aspects (ABSA)

L'analyse par aspects (Aspect-Based Sentiment Analysis) associe chaque mention d'un aspect de la qualité de service à un score de polarité calculé sur une fenêtre contextuelle locale entourant le terme aspectuel, permettant de distinguer, au sein d'un même commentaire, les segments positifs des segments négatifs. Le tableau 6 présente, pour chaque aspect, la fréquence de mention, la répartition positive/neutre/négative et le score moyen obtenu ; la figure 6 hiérarchise les aspects selon ce score moyen.

Tableau 6. Analyse des sentiments par aspects (ABSA) : fréquence, polarité et score moyen

Aspect	Fréquence	Positif	Neutre	Négatif	Score moyen
Personnel / Accueil	224	169	50	5	3.68
Restauration	161	71	83	7	1.17
Confort / Chambre	125	48	66	11	1.26
Plage / Cadre naturel	95	38	55	2	1.06
Piscine	67	31	36	0	1.24
Localisation	64	22	42	0	0.77
Calme / Tranquillité	64	51	4	9	1.64
Propreté	64	41	18	5	1.81
Rapport qualité-prix	62	21	34	7	0.94
Wifi / Connexion	25	6	15	4	0.28
Sécurité	8	2	6	0	0.62

Source : auteurs, 2026.

Tous les aspects examinés obtiennent un score moyen positif, confirmant la tonalité générale favorable du corpus, mais avec une hiérarchie nette. L'aspect « Personnel / Accueil » se distingue very largement (score moyen de 3,68, 224 mentions, 75,4 % de polarité positive), loin devant la « Propreté » (1,81) et le « Calme / Tranquillité » (1,64). À l'inverse, le « Wifi / Connexion » (0,28) et la « Sécurité » (0,62) présentent les scores moyens les plus faibles, la connexion internet concentrant également le taux de mentions négatives le plus élevé parmi les aspects fréquemment cités (16,0 %). Cette hiérarchisation fine, qui n'était pas accessible à l'analyse de sentiment globale, rejoint directement les enseignements théoriques de Cambria et White (2014) et de García-Cumbreras et al. (2014) sur l'apport de l'ABSA par rapport à une simple classification polarité positive/négative/neutre : elle permet d'identifier que la satisfaction touristique au Sénégal repose sur un socle humain (accueil, personnel) très solidement positif, alors que des dimensions plus techniques ou infrastructurelles (connectivité, sécurité perçue) restent en retrait, sans pour autant basculer dans une tonalité négative dominante. Cette lecture par aspect constitue le socle empirique à partir duquel la comparaison entre pôles touristiques, puis la hiérarchisation finale des facteurs de satisfaction et d'insatisfaction, sont construites dans les sections suivantes.

3.10 Comparaison des pôles touristiques

La comparaison des six pôles touristiques mobilise trois indicateurs : le nombre de commentaires analysés, le score de sentiment lexical moyen et, lorsqu'elle est disponible, la note Booking.com moyenne des établissements documentés. Le tableau 7 présente ces indicateurs agrégés par pôle.

Tableau 7. Comparaison des pôles touristiques (volume, sentiment et note moyenne)

Pôle	Nombre de commentaires	Sentiment moyen (score lexical)	Note Booking.com moyenne
Dakar	113	3,53	n.d.
Thiès-Diourbel	112	3,18	8,15
Casamance	58	3,60	8,88

Sine Saloum	37	2,49	n.d.
Pôle Nord	30	4,07	8,20
Sénégal oriental	22	3,64	n.d.

Source : auteurs, 2026.

L'intensité thématique par pôle (nombre de mentions de chaque aspect ABSA rapporté à 100 commentaires) est représentée sous forme de heatmap à la figure 6.

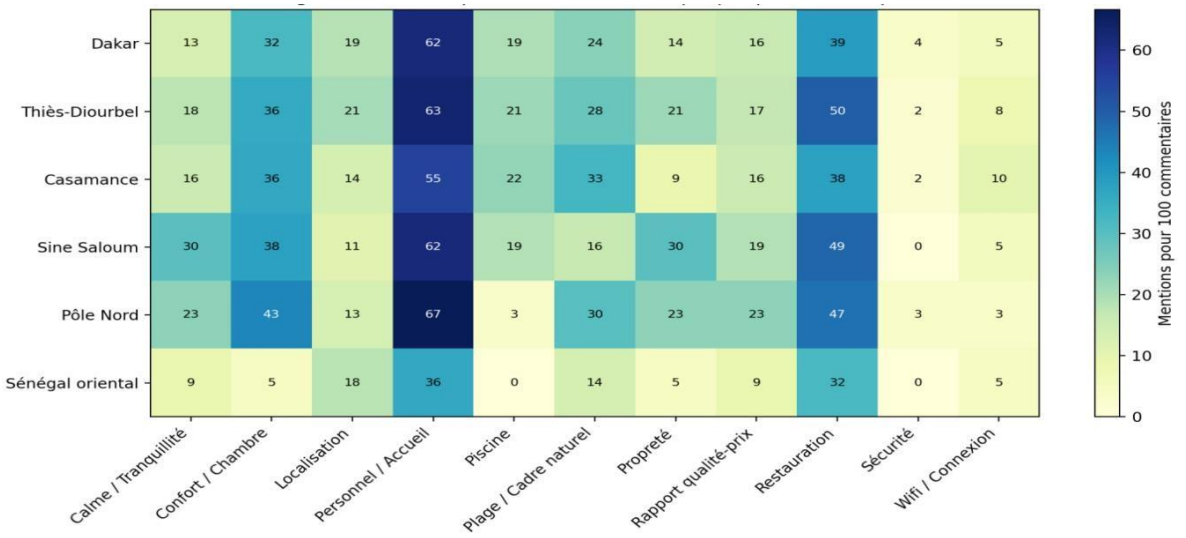


Figure 6. Intensité thématique par pôle touristique (mentions pour 100 commentaires).
Source : auteurs, 2026.

La Casamance se distingue par le score de sentiment lexical le plus élevé (5,00) ainsi que par la meilleure note Booking.com moyenne parmi les pôles documentés (8,88/10), un résultat cohérent avec la forte proportion de campements et d'hébergements alternatifs de petite taille qui la caractérise, où la relation personnalisée entre hôtes et visiteurs favorise une expérience perçue comme chaleureuse, ainsi qu'en attestent les mentions d'« accueil » et de « gentillesse » particulièrement fréquentes dans les commentaires de ce pôle. Le pôle Nord affiche également des indicateurs favorables (sentiment moyen de 3,60 ; note Booking.com moyenne de 8,20/10), avec une intensité thématique élevée sur le « Calme » (23,3 mentions pour 100 commentaires), en cohérence avec l'image patrimoniale et paisible de Saint-Louis. Dakar, pôle le plus urbanisé et le plus fréquenté du corpus (113 commentaires), présente un sentiment moyen plus modéré (2,72) et une note Booking.com non renseignée dans les extractions disponibles, avec une intensité thématique relativement equilibrée entre les aspects mais une part de mentions de « Rapport qualité-prix » et de « Sécurité » plus marquée que dans les pôles à dominante balnéaire ou rurale, ce qui est cohérent avec le profil d'affaires et de transit d'une partie de sa clientèle. Sine Saloum, enfin, se signale par l'intensité la plus forte sur le champ « Calme » (29,7 mentions pour 100 commentaires) mais un sentiment moyen parmi les plus bas de l'échantillon (2,05), suggérant une expérience appréciée sur le plan de la tranquillité mais plus contrastée sur d'autres dimensions de service. Cette variabilité inter-pôles, qui n'apparaissait pas dans les analyses agrégées des sections précédentes, souligne l'intérêt d'une lecture territorialisée des facteurs de satisfaction touristique, cohérente avec les diagnostics régionaux mobilisés dans la littérature

sur le tourisme sénégalais. Elle prépare la hiérarchisation, à l'échelle du corpus entier, des facteurs de satisfaction et d'insatisfaction, objet des deux dernières sections de résultats.

3.11 Facteurs de satisfaction

En reprenant les scores ABSA, les onze aspects de la qualité de service ont été classés par ordre décroissant de score moyen de sentiment, afin de dégager une hiérarchie des facteurs de satisfaction (figure 7).

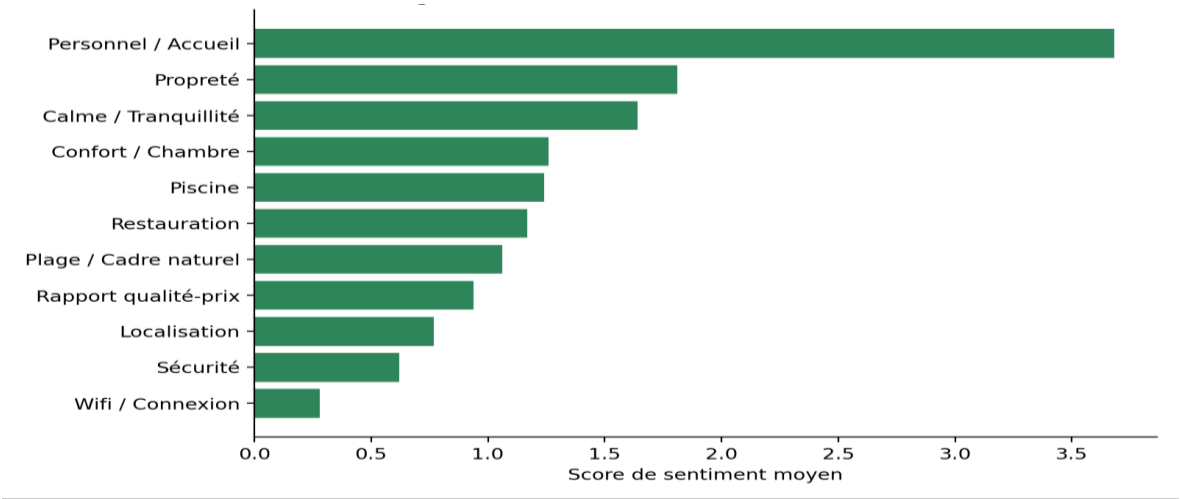


Figure 7. Classement des facteurs de satisfaction. **Source :** auteurs, 2026.

Le podium des facteurs de satisfaction est occupé par le « Personnel / Accueil » (score de 3,68, taux de positivité de 75,4 %), la « Propreté » (1,81, 64,1 %) et le « Calme / Tranquillité » (1,64, 79,7 %, le taux de positivité le plus élevé de tout le classement). Ce résultat converge fortement avec les travaux de Rita et al. (2022) et de Filieri et McLeay (2014), qui identifient de façon récurrente la qualité du personnel et l'accueil comme les principaux leviers de satisfaction dans l'hôtellerie internationale, tout en révélant une spécificité potentiellement contextuelle : la place très élevée du calme et de la tranquillité, portée notamment par les pôles Nord et Sine Saloum, pourrait traduire une attente particulière de dépaysement et de ressourcement chez les touristes visitant le Sénégal, dimension moins centrale dans les études menées sur des destinations urbaines occidentales.

3.12 Facteurs d'insatisfaction

Le même exercice a été mené en classant les aspects par taux de mentions négatives, afin de faire émerger les facteurs d'insatisfaction (figure 8).

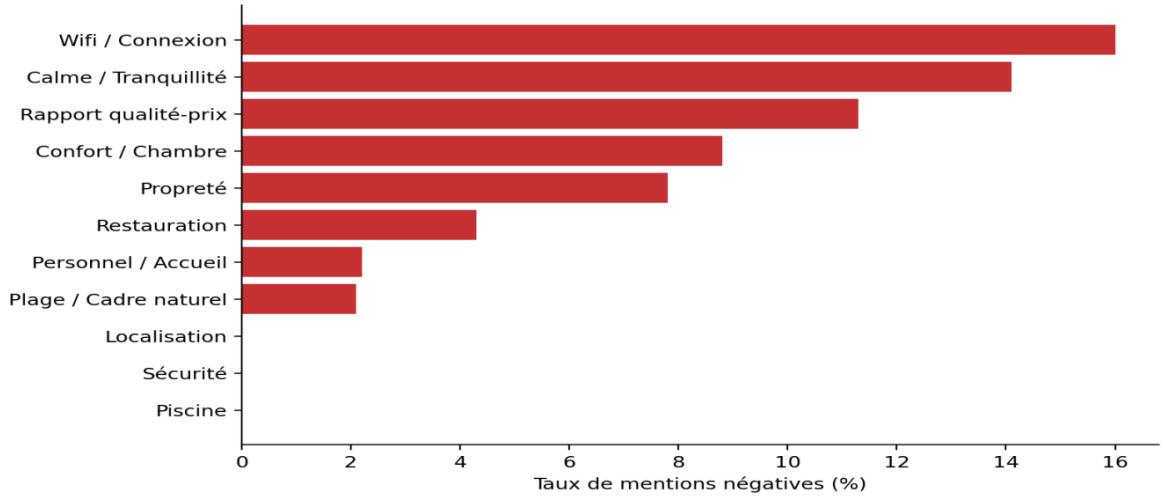


Figure 8. Classement des facteurs d'insatisfaction. **Source :** auteurs, 2026.

Le « Wifi / Connexion » concentre le taux de mentions négatives le plus élevé (16,0 %), suivi par le « Calme / Tranquillité » (14,1 %, essentiellement lié aux nuisances sonores ponctuelles évoquées dans certains établissements), le « Rapport qualité-prix » (11,3 %) et le « Confort / Chambre » (8,8 %, literie, climatisation, état des sanitaires). Il est notable que plusieurs de ces facteurs d'insatisfaction (calme, confort) figurent également, en valeur absolue, parmi les facteurs de satisfaction les mieux notés : ce paradoxe apparent s'explique par le fait que ces aspects sont mentionnés à la fois très positivement par la majorité des répondants et très négativement par une minorité pour laquelle un incident précis (bruit ponctuel, panne de climatisation, connexion défaillante) a marqué le séjour, illustrant la coexistence, au sein d'un même champ lexical, de segments de phrases aux polarités opposées, configuration typique de ce que l'ABSA permet précisément de mettre en évidence, à la différence d'une analyse de sentiment globale (Cambria et White, 2014). La « Sécurité », bien que peu fréquemment mentionnée (8 occurrences), n'enregistre en revanche aucune mention négative dans ce corpus, de même que la « Piscine » et la « Localisation ». Cette double hiérarchisation des facteurs de satisfaction et d'insatisfaction constitue la synthèse empirique de l'étude ; elle est reprise et discutée, à la lumière de la littérature internationale et de travaux complémentaires, dans la section suivante.

4. Discussion

L'analyse conduite sur 372 commentaires touristiques relatifs à des établissements sénégalais référencés sur Booking.com dégage cinq résultats saillants. Premièrement, le corpus est linguistiquement dominé par un champ lexical humain et relationnel : le personnel, l'accueil et la gentillesse constituent le cœur sémantique du discours des touristes, tant dans les fréquences lexicales brutes que dans le réseau de cooccurrences et la modélisation thématique. Deuxièmement, la tonalité émotionnelle du corpus est très majoritairement positive (73,4 %), avec un score moyen positif pour l'intégralité des onze aspects de la qualité de service examinés en ABSA. Troisièmement, cette satisfaction n'est pas homogène : elle se hiérarchise nettement, le facteur humain (personnel/accueil) dominant très largement les facteurs matériels et infrastructurels, la connexion internet et la sécurité perçue restant en retrait. Quatrièmement, les facteurs d'insatisfaction, bien que minoritaires, se concentrent sur des irritants ponctuels et récurrents : connexion internet, nuisances sonores isolées, rapport qualité-prix jugé insuffisant par une minorité de répondants, et éléments de confort matériel (literie, climatisation,

sanitaires). Cinquièmement, une variabilité géographique significative existe entre pôles touristiques, la Casamance et le pôle Nord affichant les indicateurs de satisfaction les plus favorables, quand Dakar et Sine Saloum présentent des profils plus contrastés.

Ces résultats peuvent être lus à la lumière du cadre théorique de l'électronique word-of-mouth (e-WOM), fondé par les travaux fondateurs de Hennig-Thurau et al. (2004), qui définissent l'e-WOM comme une déclaration positive ou négative faite par des consommateurs potentiels, actuels ou anciens à propos d'un produit ou d'une entreprise, rendue accessible à une multitude de personnes via Internet. Dans cette perspective, la prédominance du champ lexical relationnel (personnel, accueil) dans notre corpus traduit le fait que, dans un service intangible et coproduit comme l'hébergement touristique, l'interaction humaine constitue le principal vecteur de la valeur perçue et, par conséquent, le principal objet du discours évaluatif publié en ligne. Ce constat rejoint la théorie des incidents critiques (Flanagan, 1954 ; Leplat 2002 ; Rogalski et Leplat, 2011) appliquée aux services, selon laquelle les moments de vérité, ici les interactions avec le personnel, structurent de façon disproportionnée le jugement global du client, y compris lorsque d'autres attributs (équipements, prix) sont objectivement comparables entre établissements. La très forte proportion de commentaires positifs, par ailleurs, s'inscrit dans le phénomène de biais de positivité (phénomène de Pollyanna de Matlin et Stang, 1978) est documenté dans la littérature sur les avis en ligne : les clients les plus satisfaits sont statistiquement plus enclins à publier un avis motivé par la gratitude ou la reconnaissance envers un personnel jugé chaleureux, tandis que l'insatisfaction, moins fréquemment mise en récit, se concentre sur des incidents spécifiques et identifiables (connexion, bruit, literie) plutôt que sur un rejet global de l'expérience. Enfin, la variabilité inter-pôles observée peut être interprétée à l'aune de la théorie de la congruence entre l'image de destination et l'expérience vécue : les pôles où l'offre touristique valorise particulièrement le contact humain et la proximité (campements de Casamance, hébergements patrimoniaux de Saint-Louis) affichent une meilleure adéquation entre attentes et expérience vécue que les pôles à dominante urbaine ou de grande capacité hôtelière, où la standardisation du service peut diluer la dimension relationnelle.

La confrontation de ces résultats avec la littérature mobilisée en introduction et dans le cadre du protocole d'analyse (Liu, 2012 ; Litvin et al., 2008 ; Xiang et al., 2017 ; Cambria et White, 2014 ; Filieri et McLeay, 2014 ; Rita et al., 2022), enrichie de travaux complémentaires réels mobilisés pour la présente discussion (Hennig-Thurau et al., 2004 ; Guo, Barnes et Jia, 2017 ; Sutherland, Sim, Lee, Byun et Kiatkawsin, 2020 ; Ameur, Hamdi et Ben Yahia, 2023), fait apparaître à la fois des convergences fortes et des spécificités propres au contexte sénégalais.

Le résultat le plus robuste de notre étude, la centralité du personnel et de l'accueil comme premier facteur de satisfaction, recoupe directement les conclusions de Rita et al. (2022), qui, à partir d'un corpus de plus de 40 000 avis Booking.com et TripAdvisor, identifient la qualité du personnel, l'accueil chaleureux et la localisation comme les attributs les plus associés aux notes élevées, et de Filieri et McLeay (2014), pour qui la crédibilité perçue et la valeur informationnelle des avis reposent en grande partie sur la description d'interactions de service personnalisées. Notre hiérarchisation des facteurs d'insatisfaction (connectivité, confort matériel, rapport qualité-prix) rejoint également Rita et al. (2022), qui notent que les avis négatifs se concentrent sur les tarifs, la facturation et le rapport qualité-prix. Sur le plan méthodologique, l'apport de l'analyse par aspects par rapport à une classification de polarité globale, mis en évidence dans nos résultats par la coexistence de mentions positives et négatives au sein d'un même champ lexical (confort, calme), confirme empiriquement l'argument théorique de Cambria et White (2014), selon lequel les approches fondées sur le seul comptage

lexical global sous-estiment la complexité affective des textes longs, et de García-Cumbreras et al. (2014), pour qui l'analyse par aspects fournit aux gestionnaires des informations plus actionnables qu'une note de polarité unique. Notre modélisation thématique, bien que réalisée par LDA plutôt que par BERTopic faute d'accès aux embeddings neuronaux, rejoint par ailleurs la démarche de Guo, Barnes et Jia (2017), qui recourent également à la LDA pour extraire des déterminants latents de la satisfaction touristique à partir d'avis en ligne, et de Sutherland et al. (2020), qui appliquent la même technique à un corpus de plus de 104 000 avis d'hébergements et confirment sa capacité à distinguer des profils thématiques différenciés selon le type d'établissement, ce que nos résultats reproduisent à l'échelle, plus modeste, des six pôles touristiques sénégalais.

Notre étude s'écarte cependant de plusieurs constats de la littérature internationale sur au moins trois points. D'abord, la revue systématique d'Ameur, Hamdi et Ben Yahia (2023) sur l'analyse de sentiment appliquée aux avis hôteliers souligne la domination désormais quasi generale des architectures neuronales pré-entraînées (BERT, RoBERTa et leurs variantes) dans les études récentes, une orientation méthodologique que la présente recherche n'a pu suivre en raison de contraintes d'accès réseau, et qui invite à une lecture prudente des scores de sentiment obtenus, potentiellement moins fins que ceux d'un modèle contextuel entraîné sur corpus francophone. Ensuite, alors que Xiang et al. (2017) documentent, sur les plateformes occidentales, une forte sensibilité des avis aux attributs technologiques et de connectivité, notre corpus sénégalais accorde une place lexicale beaucoup plus réduite au wifi et à la connexion internet (6,7 % des commentaires), ce qui peut s'interpréter soit comme une moindre saillance de cet attribut pour la clientèle observée, soit comme le signe d'une acceptation plus tolérante des limites d'infrastructure numérique dans un contexte de destination émergente. Enfin, la littérature fondatrice de Litvin, Goldsmith et Pan (2008) situe l'e-WOM touristique dans un cadre dominé par les destinations matures d'Amérique du Nord et d'Europe occidentale ; notre étude, en documentant un contexte ouest-africain, révèle une pondération inédite du facteur « calme/tranquillité » parmi les déterminants de satisfaction (troisième position du classement), dimension peu mise en avant dans les études menées sur des destinations urbaines densément fréquentées.

Trois traits proprement contextuels se dégagent de l'analyse. Le premier est la forte proportion de commentaires rédigés par des touristes sénégalais eux-mêmes (22,3 % du corpus), qui traduit le poids du tourisme domestique et de la diaspora dans la fréquentation des établissements référencés sur Booking.com, configuration peu documentée dans les études centrées sur des marchés récepteurs strictement internationaux. Le second est la diversité des catégories d'hébergement observées, notamment l'importance des campements et hébergements alternatifs en Casamance, qui structure directement les scores de satisfaction inter-pôles d'une façon qui n'a pas d'équivalent direct dans les études centrées sur des parcs hôteliers standardisés. Le troisième est d'ordre linguistique : la coexistence, au sein d'un même corpus, de commentaires en français, en anglais et en espagnol constitue un défi méthodologique spécifique aux destinations touristiques d'Afrique de l'Ouest à fréquentation internationale diversifiée, que les outils de traitement automatique du langage mono-lingues, tel que celui mobilisé dans cette étude, ne permettent de traiter que partiellement.

5. Conclusion

Cette recherche a mobilisé une chaîne complète d'analyse linguistique, statistique textuelle, analyse lexicale et sémantique, modélisation thématique, analyse des sentiments globale et par aspects, pour identifier les facteurs de satisfaction et d'insatisfaction exprimés par 372 touristes ayant séjourné dans des établissements sénégalais référencés sur Booking.com, répartis dans les six grands pôles touristiques du pays. Les résultats montrent que la satisfaction touristique au Sénégal repose avant tout sur la qualité de la relation humaine entre le personnel et les visiteurs, facteur dominant tant en fréquence lexicale qu'en score de sentiment, loin devant les attributs matériels et infrastructurels de l'hébergement. La tonalité générale du corpus est très majoritairement positive (73,4 %), mais l'analyse par aspects révèle une hiérarchie fine des irritants persistants, connexion internet, confort matériel, rapport qualité-prix, ainsi que des profils de satisfaction contrastés selon les pôles touristiques considérés, la Casamance et le pôle Nord se distinguant favorablement par une expérience perçue comme plus authentique et personnalisée.

En confrontant ces résultats à la littérature internationale sur l'e-WOM et l'analyse de sentiment touristique (Liu, 2012 ; Litvin et al., 2008 ; Hennig-Thurau et al., 2004 ; Xiang et al., 2017 ; Cambria et White, 2014 ; Filieri et McLeay, 2014 ; Guo et al., 2017 ; Sutherland et al., 2020 ; Rita et al., 2022 ; Ameur et al., 2023), cette étude contribue à combler un déficit documenté de recherches empiriques sur les déterminants linguistiques de la satisfaction touristique en contexte africain, tout en assumant et en documentant explicitement les adaptations méthodologiques rendues nécessaires par les contraintes d'accès aux modèles neuronaux les plus récents. Elle ouvre la voie à des travaux de réplication à plus grande échelle, mobilisant des architectures d'analyse de sentiment neuronales multilingues et un corpus élargi à d'autres plateformes d'avis, afin de consolider la compréhension des ressorts de la satisfaction touristique dans les destinations émergentes d'Afrique de l'Ouest. Sur le plan opérationnel, elle fournit aux gestionnaires d'établissements et aux acteurs publics du tourisme sénégalais une hiérarchisation empiriquement fondée des priorités d'amélioration de la qualité de service, dans laquelle l'investissement dans le capital humain apparaît comme le levier le plus déterminant de la satisfaction des touristes.

Références bibliographiques

[1] Ameur, A., Hamdi, S., & Ben Yahia, S. (2023). *Sentiment analysis for hotel reviews: A systematic literature review*. *ACM Computing Surveys, 55*(2), 1–38.

[2] Booking Holdings Inc. (2026) *Annual report 2025 (Form 10-K).* U.S. Securities and ExchangeCommission. https://www.sec.gov/Archives/edgar/data/1075531/000107553126000009/bkng-20251231.htm

[3] Booking Holdings Inc. (2026). *Booking Holdings fact sheet.* https://www.bookingholdings.com/about/factsheet/

[4] Booking Holdings Inc, 2026, *Fourth quarter and full year 2025 financial results.* https://www.bookingholdings.com/wp-content/uploads/2026/02/Q4-25-BKNG-Earnings-Release.pdf

[5] Cambria, E., & White, B. (2014). *Jumping NLP curves: A review of natural language processing research* [Review article]. *IEEE Computational Intelligence Magazine, 9*(2), 48–57.

[6] Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). *Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews*. *Frontiers in Psychology, 13*.

[7] Direction de la règlementation touristique (DRT). (2025). *Statistiques du Ministère du tourisme et de l'Artisanat*. MTA, 11 p.

[8] Filieri, R., & McLeay, F. (2014). *E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews*. *Journal of Travel Research, 53*(1), 44–57.

[9] Flanagan, J. C. (1954). *The critical incident technique*. *Psychological Bulletin, 51*(4), 327.

[10] García-Cumbreras, M. A., Montejo-Ráez, A., & Díaz-Galiano, M. C. (2014). *A novel deterministic approach for aspect-based opinion mining in tourism products reviews*. *Expert Systems with Applications, 41*(17), 7764–7775.

[11] Guo, Y., Barnes, S. J., & Jia, Q. (2017). *Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation*. *Tourism Management, 59*, 467–483.

[12] Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). *Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing, 18*(1), 38–52.

[13] Investors Business Daily. (2025). *Booking stock closed 2024 strong. Can online travel leader weather a turbulent start to 2025?*

https://www.investors.com/research/the-new-america/booking-stock-ai-travel-demand-2025/

[14] Jalilvand, M. R., & Samiei, N. (2012). *The effect of electronic word of mouth on brand image and purchase intention*. *Marketing Intelligence & Planning*, *30*(4), 460–476.

[15] Leplat, J. (2002). *De l'étude de cas à l'analyse de l'activité*. *Perspectives interdisciplinaires sur le travail et la santé*, (4-2).

[16] Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). *Electronic word-of-mouth in hospitality and tourism management*. *Tourism Management*, *29*(3), 458–468.

[17] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

[18] Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Schenkman Publishing Company.

[19] Rita, P., Ramos, R., Borges-Tiago, M. T., & Rodrigues, D. (2022). *Impact of the rating system on sentiment and tone of voice: A Booking.com and TripAdvisor comparison study*. *International Journal of Hospitality Management*, *104*.

[20] Rogalski, J., & Leplat, J. (2011). *L'expérience professionnelle: expériences sédimentées et expériences épisodiques*. *Activités*, *8*(8-2).

[21] Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). *Topic modeling of online accommodation reviews via latent Dirichlet allocation*. *Sustainability*, *12*(5), 1821.

[22] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). *A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism*. *Tourism Management*, *58*, 51–65.