ELAD-SILDA
ISSN : 2609-6609
Éditeur : Université Jean Moulin Lyon 3

12 | 2025
English for Specific Purposes at the Interfaces

Être un genre ou ne pas être : le cas des sites web d'agences de voyages

To be or not to be a genre: The case of travel agency websites

Trang Tran Hanh Pham, Rémi Cardon, Rodrigo Wilkens, Van Công Trân et Thomas François

https://publications-prairial.fr/elad-silda/index.php?id=1791

DOI : 10.35562/elad-silda.1791

Référence électronique
Trang Tran Hanh Pham, Rémi Cardon, Rodrigo Wilkens, Van Công Trân et Thomas François, « Être un genre ou ne pas être : le cas des sites web d'agences de voyages », *ELAD-SILDA* [En ligne], 12 | 2025, mis en ligne le 18 décembre 2025, consulté le 30 septembre 2026. URL : https://publications-prairial.fr/elad-silda/index.php?id=1791

Droits d'auteur
CC BY 4.0 FR

Prairial
Pôle éditorial
Lyon Saint-Étienne

Être un genre ou ne pas être : le cas des sites web d'agences de voyages

To be or not to be a genre: The case of travel agency websites

Trang Tran Hanh Pham, Rémi Cardon, Rodrigo Wilkens, Van Công Trân et Thomas François

PLAN

Introduction
1. Analyse du genre
   1.1. L'analyse du genre du discours numérique
   1.2. L'analyse du genre de discours numérique touristique
2. Méthodologie
   2.1. Présentation des corpus
   2.2. Les caractéristiques du genre pour les analyses
   2.3. Identification des caractéristiques de genre
3. Résultats
   3.1. Traits spécifiques des discours touristiques numériques
   3.2. Traits spécifiques du discours touristique numérique des agences de voyages
      3.2.1. Niveau Genre, sans prise en compte des tours (avec frWac, et trois autres genres)
      3.2.2. Niveau Genre, avec des tours (avec trois autres genres)
4. Discussion
Conclusion

TEXTE

Introduction

1 Avec l'accroissement du nombre de voyages internationaux et de l'utilisation d'internet pour l'organisation des voyages, la communication par le biais de sites web à visée touristique – souvent mise en œuvre par des agences de voyages – est devenue cruciale. Les caractéristiques de ces sites web d'agences touristiques sont non seulement étroitement liées aux stratégies commerciales de ces agences, mais influencent aussi les choix des touristes et leur satisfaction globale (Buhalis et Law, 2008 ; Constantinides *et al.*,

2010 ; Gémar, 2014). Par conséquent, les agences désireuses de développer le commerce électronique doivent surmonter les barrières culturelles et linguistiques afin de mettre en œuvre des stratégies de communication adaptées sur leurs sites web (Violino, 2001). Selon Irimiea (2018), pour ce faire, il ne suffit pas d'examiner les résultats de recherche en marketing, mais il faut aussi prendre en considération les contributions d'autres domaines de recherche, parmi lesquels la sociolinguistique et la linguistique.

2 Dans les études du discours touristique sur les sites web, seule une petite partie d'entre elles font appel à la notion de genre de discours (Malenkina et Ivanov, 2018). L'analyse de genre se définit comme un cadre pour l'étude du discours spécialisé (Adam, 2001 ; Bhatia, 1993, 2004 ; Maingueneau, 2007 ; Swales, 1990 ; Richer, 2011). Le genre est « une classe d'événements communicatifs : le critère principal qui fait d'une collection d'événements communicatifs un genre est un ensemble partagé de buts communicatifs » (Swales, 1990 : 45). L'approche par le genre permet tout d'abord de systématiser la description des discours spécialisés en mettant en évidence leurs caractéristiques linguistiques. En ce qui concerne plus particulièrement le domaine du tourisme numérique, une compréhension du genre et de ses éléments clés permet aux agences de voyages de 1) rédiger des textes efficaces grâce à une mise en scène de la communication avec des ressources verbales et numériques variées et 2) assurer les trois fonctions de communication typiques des textes touristiques promotionnels : attirer l'attention, informer et persuader (Bhatia, 1993, 2004).

3 En réalité, si les chercheurs ont abordé divers genres de discours touristique en ligne, les sites web touristiques des agences de voyages ne sont que peu analysés. Les analyses portent souvent sur des sites web des offices de tourisme (Cheong, 2013), des destinations touristiques spécifiques (Huang, 2015 ; Hui *et al.*, 2020 ; Nguyen et Modehiran, 2023), des guides touristiques en ligne (Alali *et al.*, 2019), ou des brochures en ligne (Öztürk et Şafak, 2014). En outre, les caractéristiques linguistiques analysées dans des études précédentes sont généralement limitées (tantôt ce sont les étiquettes morphosyntaxiques, tantôt des dimensions syntaxiques ou lexicales). Elles ne représentent donc qu'une partie des caractéristiques pertinentes pour l'analyse du discours et, de plus, les caractéristiques

numériques typiques des genres en ligne semblent être négligées. Seuls deux articles (Askehave et Nielsen, 2005 ; Santini *et al.*, 2010) semblent analyser les hyperliens, bien que cette caractéristique joue un rôle significatif dans l'analyse du discours numérique, apportant une dimension supplémentaire à la compréhension de la structure et du contenu des textes en ligne.

4 L'objectif de la recherche est dès lors double. Tout d'abord, montrer que les sites web des agences de voyages constituent un genre de discours en soi, dont les caractéristiques linguistiques diffèrent d'autres discours touristiques numériques. Une double étude quantitative et comparative est réalisée au moyen d'un jeu de caractéristiques pertinentes pour définir un genre de discours. Elle vise à contraster les sites web des agences avec 1) un corpus numérique général, frWac (Baroni *et al.*, 2009), pour faire ressortir les traits typiques des discours touristiques, et avec 2) un corpus de trois discours touristiques numériques proches, afin d'affiner la détection des caractéristiques propres au discours web des agences de voyages.

5 Dans un deuxième temps, pour mettre en évidence l'articulation entre les caractéristiques et les fonctions communicatives visées par les sites web des agences de voyages, l'analyse prend en compte la structure rhétorique du discours, par la notion de « tours » (*move* en anglais) de Swales (1990). Une analyse comparative – cette fois quantitative et qualitative – est menée sur quatre genres touristiques numériques. L'objectif est de faire ressortir les expressions linguistiques et les pratiques numériques propres à certains « tours ».

6 Les contributions de cette étude sont multiples. La première contribution consiste à éclairer la problématique de la catégorisation d'un genre, en l'occurrence les sites web d'agences de voyages, à l'aide d'une approche empirique et à large spectre (au niveau des caractéristiques textuelles considérées). En effet, cette étude se base sur un jeu étendu de caractéristiques linguistiques, grâce à des outils de traitement automatique du langage (TAL), qui permettent de dépasser les « classiques » étiquettes morphosyntaxiques. Une seconde contribution concerne la prise en compte de caractéristiques typiques des pratiques numériques sur le web, au travers d'une analyse des hyperliens. Troisièmement, l'analyse

dépasse le niveau textuel et exploite également la notion de « tours » pour mieux caractériser l'association entre les fonctions communicatives et la structure rhétorique des discours étudiés. Cette tentative d'appliquer une approche fréquemment utilisée pour examiner le langage en anglais, à l'analyse du genre de discours promotionnel en français, représente une ultime contribution.

7 Le plan de l'article est organisé comme suit : les travaux existants pour l'analyse du genre des discours touristiques sont discutés dans la section 1. La méthodologie de l'étude est présentée en détail dans la section 2. Les résultats sont détaillés et commentés dans la section 3. La section 4 propose une discussion des résultats et des pistes pour de futurs travaux possibles.

1. Analyse du genre

1.1. L'analyse du genre du discours numérique

8 Dans le domaine de l'anglais sur objectifs spécifiques (AOS)[1], les genres sont considérés comme des outils de communication situés dans un contexte social. Grâce à ce contexte, on peut 1) examiner les objectifs de la communauté de discours et 2) étudier comment les caractéristiques du genre aident les membres de la communauté de discours à atteindre leurs objectifs de communication.

9 Swales (1990) et Bhatia (1993) – deux auteurs s'inscrivant dans le domaine de l'AOS – ont été les premiers à axer l'analyse de genre à l'aide des concepts de « tours » et « pas »[2] (*move* et *step* en anglais). Les tours sont des « unités discursives ou rhétoriques remplissant des fonctions communicatives cohérentes dans les textes », dont les réalisations linguistiques peuvent varier considérablement en termes de longueur et autres caractéristiques (Swales, 2004 : 228-229). Par exemple, pour l'analyse de l'introduction des articles scientifiques, Swales (1990) définit trois types de tours, tels que 1) « établir un territoire », 2) « établir une niche » et 3) « occuper la niche ». Chaque tour remplit une fonction communicative particulière et contribue à la structuration globale du texte. L'un des résultats les plus intéressants de ce type d'analyse est l'exploration du « répertoire

commun » des stratégies rhétoriques, c'est-à-dire les différentes possibilités de réalisation langagière d'un même tour, et ainsi d'un même objectif communicatif. En particulier, on peut déterminer si certaines expressions sont préférées et donc plus spécifiques à un genre qu'à d'autres. L'analyse des tours est souvent réalisée selon une approche ascendante (recherche de caractéristiques lexicales, grammaticales ou syntaxiques pour caractériser le tour) ou une approche descendante (lecture attentive du texte pour repérer les ruptures de sujet ou les changements de contenu) (Moreno et Swales, 2018). Cependant, cette théorie de structure des tours n'est guère répandue dans le courant francophone d'analyse du genre. Seul Richer (2011) dans le domaine du français sur objectifs spécifiques et quelques auteurs francophones étudiant les discours anglophones (Resche, 2010 ; De Cock et Hanegreefs, 2013 ; Gaillard, 2015 ; etc.) semblent reprendre cette théorie.

10 Le concept de genre numérique est perçu de diverses manières parmi les analystes de genre. Tout d'abord, la dénomination du concept n'est pas stable (Pleau, 2017). Caballero (2008) parle de « cybergenres », Bonhomme (2015) de « genres du web » ou de « genres informatiques », pour désigner un nouveau type de genre qui se distingue du genre traditionnel non numérique, ou des « genres pré-numériques ». Toujours selon Bonhomme (2015), le fait d'avoir le nouveau genre numérique est à l'origine d'un genre traditionnel adapté en format numérique par « l'intégration du maillage hypertextuel, la dynamisation des entrées ou des documents sonores et iconographiques » (Bonhomme, 2015 : 39). D'ailleurs, parmi les chercheurs anglophones, les termes souvent utilisés sont « *digital genre* » ou « *virtual discourse genre* » (Suciu, 2019).

11 Quoi qu'il en soit, plusieurs chercheurs en linguistique appliquée ont proposé des raisons convaincantes pour orienter notre regard vers les genres numériques, dont la plus fréquemment mentionnée est peut-être leur « proéminence dans la communication contemporaine [3] » (Hafner et Pun, 2020 : 4). Il y a plus d'une décennie, Giltrow et Stein (2009) ont observé qu'internet nous a donné un nouveau cadre de communication qui reconfigure les conditions rencontrées par les caractéristiques pragmatiques du langage. Le résultat de cette reconfiguration a été une prolifération si régulière et rapide des genres – ou des formes de communication qui

sont candidates à être un genre – que nous sommes mis au défi de repenser notre compréhension théorique du genre. Plus récemment, Hafner a noté que, malgré le potentiel des médias numériques pour favoriser l'innovation et le changement de genre dans la communication savante, la linguistique appliquée vient seulement de commencer à étudier « l'innovation dans les genres de recherche médiatisés numériquement [4] » (Hafner, 2018 : 15).

12 Parlant plus spécifiquement de « l'explosion des genres numériques » dans la communication scientifique, Luzón et Pérez-Llantada soulignent la nécessité de prendre en compte le contexte sociopolitique large, dans lequel il y a maintenant une « poussée vers la démocratisation de la science et le partage de la recherche [5] » (Luzón et Pérez-Llantanda, 2019 : 2-3), un contexte dans lequel l'intérêt des chercheurs pour la compréhension des genres utilisés pour communiquer de nouvelles connaissances en ligne s'est considérablement accru. En conséquence, Hafner et Pun ont proposé une définition des genres numériques comme suit :

> Lorsque nous parlons de genres et de pratiques numériques, nous nous référons aux événements communicatifs récurrents, orientés vers un but (Swales, 1990) médiatisés par des outils numériques de cette manière, ainsi qu'à la constellation d'activités qui entourent ces événements. (Hafner et Pun, 2020 : 4) [6]

13 La première évolution majeure identifiée dans les études de genre est l'internalisation de la numérisation. Selon Giltrow et Stein, les technologies numériques créent « un nouveau cadre de communication qui reconfigure les conditions auxquelles les caractéristiques pragmatiques du langage répondent [7] » (Giltrow et Stein, 2009 : 9). Cette nouveauté dans les contextes de communication peut, par exemple, se manifester à travers la diversité des utilisateurs rendue possible par internet ainsi que par l'interaction croissante entre l'utilisateur et l'auteur dans le processus de construction de certains types de textes. Ces nouveautés doivent être prises en compte par les chercheurs lorsqu'ils analysent le contexte dans lequel un genre est construit. Ces changements induits par la numérisation justifient des innovations théoriques dans les théories des genres pour s'adapter aux nouvelles situations. Par exemple, Heyd (2008) présente un cadre à quatre paramètres pour

l'analyse des genres numériques. Tout d'abord, (1) le paramètre vertical demande aux analystes de considérer le niveau de généralité dans la catégorisation des genres, et de distinguer des notions telles que les supergenres, les sous-genres et les types de textes. Les genres peuvent également être étudiés (2) horizontalement, de manière que les chercheurs puissent observer comment un genre est lié à d'autres genres. Par exemple, une publicité contextuelle sur l'internet peut mener à un site web touristique, qui à son tour peut mener à une lettre de confirmation. Les notions pertinentes à cet égard comprennent le répertoire de genres (Orlikowski et Yates, 1994), qui se concentre sur les genres utilisés pour réaliser des activités dans des contextes organisationnels, ou l'écologie des genres (Heyd, 2008), qui se réfère à une constellation générique qui se forme naturellement. Le troisième paramètre est (3) la perspective ontologique qui aide les chercheurs à identifier un genre dans une perspective descendante ou ascendante. Une approche descendante de l'identification des genres suppose que la classification des genres dépend d'une liste préfabriquée de caractéristiques saillantes, tandis qu'une approche ascendante (également connue sous le nom de taxonomie populaire) s'appuie sur la perception de la communauté discursive concernée pour identifier les genres. Le dernier paramètre (4) guide les chercheurs pour qu'ils considèrent les genres dans une perspective dynamique et diachronique afin de comprendre comment les facteurs sociotechniques, tels que le développement des plates-formes numériques, influencent le processus d'évolution des genres. Ce faisant, le genre n'est plus considéré comme une entité isolée et statique, mais plutôt comme un phénomène social interconnecté, dynamique et résilient. Plus précisément, un genre n'existe pas dans le vide, mais plutôt dans une relation verticale ou horizontale avec d'autres genres. Il réagit également aux progrès technologiques constants et en bénéficie. Ces caractéristiques interconnectées et dynamiques permettent aux genres d'atteindre différents objectifs de communication dans différentes circonstances sociales.

14 Santini *et al.*, (2010) ont développé un cadre analytique à trois niveaux pour analyser les documents web. Compte tenu de la structure multiniveaux permise par les liens hypertextes sur les sites web, ce cadre demande aux analystes d'étudier les genres web aux niveaux

micro, méso et macro. L'étude au niveau micro se concentre sur les unités au niveau de la page et leurs constituants. Bien que ce cadre ne précise pas les ressources à étudier à ce niveau, il est raisonnable de supposer qu'il inclut les caractéristiques sémiotiques d'une page web spécifique, telles que l'utilisation du texte, de l'image et de la mise en page. Au niveau méso, l'attention se porte sur la structure interne du site web, c'est-à-dire le plan du site. Enfin, au niveau macro, les analystes sont chargés d'étudier le référencement intertextuel impliqué dans un site web.

15 Une autre tentative de prendre en compte la numérisation dans l'analyse des genres a été faite par Askehave et Nielsen (2005), qui ont construit un modèle bidimensionnel pour décrire les genres numériques en tant que texte et, plus important encore, en tant que support. Dans ce modèle, les utilisateurs de genres numériques sont dotés de deux identités en fonction de la manière dont ils utilisent le genre cible. Si l'utilisateur ne s'intéresse qu'au texte, il assume le rôle conventionnel d'un lecteur qui « zoome sur le texte » (Askehave et Nielsen, 2005) ; en revanche, si l'utilisateur se sert du genre numérique comme d'un portail vers d'autres sources d'information, son identité passe de celle d'un lecteur à celle d'un navigateur, et le texte devient un médium (un support). Ce modèle est applicable à la description des caractéristiques de certains genres numériques dominés par les hyperliens, comme la page d'accueil. Il combine le modèle de genre de Swales (1990) avec le concept de mode de navigation de Finnemann (1999).

16 À l'heure actuelle, avec le développement de la technologie, d'autres éléments alimentent le discours, par exemple des éléments numériques (hyperliens, boutons cliquables, barre de navigation, outil de recherche, etc.) ou des éléments graphiques (photos/images, vidéos, symboles, émoticônes). L'émergence de ces éléments discursifs fait naître l'analyse multimodale du discours, une analyse combinant diverses sémiotiques de communication. Kress et van Leeuwen (2021) affirment que la multimodalité est l'une des caractéristiques de la société moderne. Les auteurs développent la grammaire visuelle, qui considère que la langue est une forme sémiotique sociale. Le modèle se compose de trois dimensions pour l'analyse des images : signification représentationnelle, signification interactionnelle et signification compositionnelle. Tout d'abord, la

signification représentationnelle fait référence à la reproduction fidèle du monde objectif des personnes, des activités humaines, des lieux et même du monde intérieur des personnes. Elle se compose de deux processus : le processus narratif et le processus conceptuel. Le processus narratif comprend le processus d'action, le processus de réaction, le processus verbal et mental, tandis que le processus conceptuel comprend le processus relationnel et le processus existentiel. Ensuite, la signification interactionnelle est la description de la relation entre le marqueur d'image, la représentation de l'image et les lecteurs. Les éléments inclus dans le sens interactif sont le contact, l'attitude, la distance et la modalité. Dans la grammaire visuelle, la modalité est classée en trois catégories : la modalité élevée, la modalité moyenne et la modalité faible, qui peut être démontrée par la couleur, l'éclairage, etc. Enfin, la signification compositionnelle désigne l'ensemble de la présentation du discours multimodal. Les éléments qui composent le sens compositionnel sont la valeur informative, le cadrage et la saillance. La valeur informative correspond à l'emplacement d'informations particulières, telles que des informations nouvelles ou anciennes. Le cadrage signifie qu'il existe ou non des dispositifs permettant de relier les éléments du discours. La saillance fait référence aux différents degrés d'importance des éléments pour les lecteurs, tels que l'arrière-plan et le premier plan. Kress et van Leeuwen considèrent que le modèle des trois dimensions fonctionne bien non seulement pour le langage, mais aussi pour « la réflexion sur tous les modes de représentation, donc les images » (Kress et van Leeuwen, 2021 : 20).

1.2. L'analyse du genre de discours numérique touristique

17 La structure des tours et la caractérisation du genre sont appliquées à différents types de genres écrits. Divers chercheurs ont appliqué la théorie de la structure des tours afin d'analyser le discours touristique numérique, démontrant ainsi son utilité dans leurs travaux respectifs. Les objectifs communs des études existantes consistent à 1) identifier les tours et éventuellement les pas – des subdivisions plus fines au sein des tours (Swales, 1990) –, 2) définir les caractéristiques linguistiques (ou visuelles) associées au tour.

Cependant, il n'existe pas d'étude portant spécifiquement sur les sites web des agences de voyages, et, de façon générale, seul un nombre limité de caractéristiques linguistiques a été analysé. Le tableau ci-dessous présente quelques exemples d'études représentatives, avec le type de corpus et les caractéristiques analysées :

Tableau 1 : Des études sur l'analyse du genre de discours touristique en ligne

Auteur(s)	Corpus	Caractéristiques analysées
Henry et Roseberry (1996)	44 exemples typiques de « Brèves informations touristiques » en ligne	• Les tours • Caractéristiques linguistiques : longueur, termes d'adresse, modalité, expressions idiomatiques, phrases lexicales et expressions lexicales communes.
Ip (2008)	Brochure de voyage disponible au centre d'accueil de l'office du tourisme de Hong Kong.	• Caractéristiques linguistiques : le choix des mots, les figures de style, les structures grammaticales • Les caractéristiques visuelles
Cheong (2013)	16 pages d'accueil des offices de tourisme malaisiens 49 publireportages	• Les tours • Les contenus discursifs apparus dans chaque tour (ex. : slogan, logo, informations diverses)
Öztürk et Şafak (2014)	Brochure Kemer en ligne	• Les tours • Éléments linguistiques : utilisation d'autoréférences (référence personnelle, impersonnelle et interactive) ; d'adjectifs, de verbes et de temps • Les caractéristiques visuelles (images)
Huang (2015)	30 textes de « Brèves informations touristiques » des sites web des destinations touristiques	• Les tours • Les pas

Yaemwannang (2017)	24 pages d'accueil d'hôtels, 24 pages d'accueil de restaurants et 24 pages d'accueil de destinations touristiques de 8 pays différents	• Les tours • Caractéristiques linguistiques : phrases nominales, adjectifs, superlatifs et phrases ciblant l'ego • Caractéristiques visuelles
Alali *et al.* (2019)	Guides touristiques en ligne de trois pays asiatiques.	• Les tours • Les degrés de tours (tour périphérique/central)
Hui *et al.* (2020)	63 textes de promotion d'attractions touristiques	• Les tours • Les caractéristiques linguistiques : des adjectifs
Nguyen et Modehiran (2023)	24 brochures de 24 attractions publiées sur les sites web officiels vietnamiens	• Les tours et les stratégies de communication pour réaliser les tours

18 Ces analyses de genre tendent à négliger les caractéristiques numériques typiques des sites web. En effet, seuls deux articles semblent aborder le rôle des hyperliens – une caractéristique numérique importante pour le genre web. Il s'agit des études de Askehave et Nielsen (2005) et de Santini *et al.* (2010), abordées dans la section précédente.

2. Méthodologie

19 En vue de déterminer la spécificité du genre des sites web des agences de voyages francophones et de caractériser ce genre sur les plans linguistique et numérique, les corpus utilisés pour les analyses comparatives sont présentés dans la section 2.1. Ensuite, les caractéristiques linguistiques et numériques sont détaillées dans la section 2.2. Les sections 2.3 et 2.4 exposent les analyses ainsi que les tests statistiques utilisés pour l'étude.

2.1. Présentation des corpus

20 Le principal corpus de l'étude comprend 80 sites web touristiques d'agences de voyages francophones. Afin d'établir ce corpus, nous

avons limité l'espace géographique étudié aux pays dont le français est l'une des langues officielles et qui, selon l'Organisation mondiale du tourisme (2010, cité par Viallon et Henneke-Lange, 2011), ont en commun d'avoir des PIB élevés, une forte activité – et donc une forte professionnalisation –, aussi bien dans le tourisme émetteur que récepteur. Ces pays sont la France, le Canada, la Belgique et la Suisse. Pour réaliser l'échantillonnage d'agences de ces quatre pays, nous nous sommes concentrés sur leurs capitales ou métropoles (Paris, Montréal, Bruxelles, Genève). Ces villes, selon Viallon et Henneke-Lange (2011), sont non seulement celles qui ont le plus de concurrence au niveau mondial, mais aussi les premières à avoir ainsi la possibilité d'améliorer leur offre existante. Ensuite, les agences de ces métropoles possédant un site web francophone ont été sélectionnées. Afin d'établir la liste des sites web d'agences de voyages, nous avons recouru aux listes officielles émises entre 2021 et 2022 par les institutions nationales de chaque pays : le Registre national des opérateurs de voyages et de séjours de France, la liste des agences de voyages du Gouvernement du Québec, la liste complète des agences certifiées au label Certa de Belgique et la liste des agences de la Fédération suisse du voyage. Sur la base de ces listes, nous avons effectué un échantillonnage aléatoire et vérifié si les sites web retenus étaient bien fonctionnels. Seules les pages d'accueil de ces sites ont été prises en considération, pour un total de 80 pages d'accueil, dont 21 de Paris, 11 de Montréal, 26 de Bruxelles et 22 de Genève. En effet, la page d'accueil peut être définie comme un genre distinctif (Askehave et Nielsen, 2005), elle est la première page d'information que voit un internaute, lorsque ce dernier accède à un site web. La page d'accueil représente un élément très important dans la création de la première impression chez le lecteur ; elle est soigneusement conçue pour non seulement stimuler l'attention du visiteur et l'encourager à explorer davantage le site web, mais aussi pour établir un contact avec le visiteur de la page.

21 Par ailleurs, nous avons collecté trois autres corpus représentant trois genres de discours proches de celui des agences de voyages : 30 pages d'accueil de blogs de voyage, 30 pages d'accueil de sites d'offices de tourisme et 30 pages d'accueil des sites de destinations touristiques (par ex. : musées, châteaux, parcs naturels, etc.). La sélection de ces pages est aléatoire ; il s'agit de pages d'accueil en

français, créées par des émetteurs francophones des quatre pays (France, Canada, Belgique et Suisse). Il est nécessaire de souligner à la fois la similitude et la spécificité du corpus des agences de voyages avec ces trois corpus. Tout d'abord, en termes de contenu, les thèmes abordés dans ces corpus sont semblables (présentation et promotion de la/des destination(s), événements touristiques, etc.). Cependant, le corpus des agences de voyages relève du type de discours promotionnel/commercial, l'émetteur relevant du secteur commercial, dont l'objectif est de vendre des produits/services touristiques. Le corpus des offices de tourisme, quant à lui, relève davantage du discours institutionnel, fournissant des informations sur un lieu géographique spécifique (par exemple, la Wallonie en Belgique ou Montréal au Canada), dans le but de rendre le lieu attrayant. Le corpus de destinations spécifiques aborde des lieux touristiques uniques, afin d'informer sur l'endroit et aussi d'inciter des visiteurs à venir. Finalement, le corpus des blogs de voyage est un discours personnel et combine deux objectifs des deux autres genres : fournir des informations et/ou faire de la publicité pour des services.

22 Le dernier corpus utilisé est frWac (Baroni *et al.*, 2009) – un corpus de 1,6 milliard de mots construit à partir du web en limitant l'exploration au domaine .fr. Le corpus frWac aborde des thèmes très variés, que ce soit la politique, l'économie ou la vie en général (des blogs personnels, forums divers, etc.). Sur la base du corpus original de 8000 sites web, nous avons filtré (1) les pages d'accueil qui (2) sont toujours en ligne, ce qui revient à un total de 152 pages d'accueil. Le fait de ne conserver que des pages d'accueil assure une cohérence dans la comparaison avec notre corpus de tourisme. FrWac sert de corpus de comparaison avec les corpus touristiques. Le tableau suivant résume nos différents corpus. Une abréviation a été associée à chaque corpus pour faciliter leur référencement futur.

Tableau 2 : Présentation des corpus de l'étude

Corpus	Abréviation	Nombre de pages d'accueil	Nombre de mots totaux
Français général numérique	frWac	152	64332
Blogs de voyage	BLOG	30	27862

Offices de tourisme	OFF	30	23271
Destinations touristiques spécifiques	DEST	30	14336
Agences de voyages	AGENCE	80	46041

2.2. Les caractéristiques du genre pour les analyses

23 Tout d'abord, nous établissons une série de caractéristiques linguistiques et numériques pertinentes pour caractériser les discours touristiques numériques promotionnels, en nous basant sur les résultats des études précédentes ainsi que sur des théories liées au langage du tourisme. Nous utilisons FABRA (Wilkens *et al.*, 2022), une API web qui permet de calculer automatiquement 509 caractéristiques linguistiques, parmi lesquelles nous avons repris les suivantes :

24 La structure des tours. En identifiant les tours pour chaque page d'accueil, on trouve les objectifs communicatifs liés au genre. Cette caractéristique est la seule pour laquelle nous n'avons pas recours à FABRA. Nous avons réalisé, lors d'une étude antérieure (Cardon *et al.*, 2024), une campagne d'annotation pour l'identification des tours dans les quatre corpus BLOG, OFF, DEST, AGENCE du domaine touristique. Pour ce faire, deux annotateurs (1) ont tout d'abord identifié chaque bloc de texte de la page d'accueil, en suivant l'ordre de lecture (de gauche à droite, de haut en bas) et ensuite, (2) ils ont associé un tour à chaque bloc de texte, en se basant sur le modèle rhétorique des tours proposé par Askehave et Nielsen (2005). Cependant, nous avons modifié et adapté les définitions de certains tours, en accord avec notre objectif d'annotation. Un guide d'annotation a été développé sur cette base, qui, en plus de la définition, inclut des informations complémentaires pour chaque tour telles que des exemples et un bref organigramme. Le taux d'accord entre les annotateurs, mesuré à l'aide du κ de Cohen (Cohen, 1960) est de 0,73, ce qui suggère que le guide d'annotation produit une annotation fiable. Voici une brève description de chaque tour :

- Tour 1 : Attirer l'attention. Attirer l'attention du lecteur lorsqu'il entre sur la page d'accueil.

- Tour 2 : Saluer. Souhaiter la bienvenue aux internautes.
- Tour 3 : Identifier l'expéditeur. Identifier le propriétaire du site web et est souvent réalisé à l'aide d'un logo.
- Tour 4 : Indiquer la structure du contenu. Fournir un aperçu clair de la structure du contenu, souvent appelé le menu principal.
- Tour 5 : Détailler le contenu (sélectionné). Détailler les sujets énumérés dans le menu principal. Outre les informations détaillées, ce tour fonctionne également comme un outil de présentation de nouvelles et de création d'une image publique.
- Tour 6 : Asseoir la crédibilité. Établir une image de confiance du propriétaire du site web.
- Tour 7 : Établir le contact. Donner au visiteur les moyens de contacter l'expéditeur.
- Tour 8 : Établir une communauté (de discours). Permettre aux utilisateurs fidèles ou fréquents d'établir des communautés au sein du site web (ex. par le biais de réseaux sociaux externes ou de compte d'utilisateur sur le site).
- Tour 9 : Promouvoir une organisation externe. Faire la promotion d'une autre entreprise, produit, etc. Cela prend généralement la forme d'une bannière publicitaire.

25 La longueur du texte (4 variables). La longueur des phrases est une variable traditionnelle de la lisibilité, ce qui constitue une caractéristique désirable des textes promotionnels. Sa justification est essentiellement empirique (François, 2011). Concernant la lisibilité sur le web, la longueur des phrases a beaucoup été étudiée, car elle constitue une ligne directrice simple pour les rédacteurs (Kadayat et Eika, 2020). Pour des textes promotionnels, les phrases courtes et simples augmentent le degré de lisibilité et sont préférées (Teodorescu, 2015).

26 La diversité lexicale (118 variables). La diversité lexicale évalue dans quelle mesure le vocabulaire d'un texte est varié. Dans un texte à grande diversité lexicale, l'auteur utilise de nombreux mots différents, avec peu de répétitions (Johansson, 2008). Ces variables

prennent en compte soit tous les mots du texte, soit uniquement certaines catégories du discours (ex. adjectifs, noms, verbes, etc.).

27 La fréquence lexicale (28 variables). La fréquence lexicale est un puissant prédicteur de la complexité lexicale et de la lisibilité d'un texte (Rayner et Duffy, 1986). Cette caractéristique lexicale est calculée sur la base des fréquences de la ressource Lexique3 (New *et al.*, 2007). La fréquence lexicale est calculée pour chacune des catégories suivantes : tous les mots, les mots de contenu, les mots fonctionnels, les noms (à l'exception des noms propres), les verbes et les adjectifs.

28 La sophistication lexicale (18 variables). La sophistication lexicale renvoie à l'utilisation de mots difficiles ou rares dans un texte (Bonvin et Lambelet, 2019). Dans FABRA, les 9000 mots les plus fréquents dans Lexique3 sont répartis en 9 « tranches » égales de 1000 mots. Ainsi, la 1re tranche englobe les 1000 mots les plus fréquents, la 2e contient les 1000 mots situés entre le rang 1001 et 2000, etc. Un texte comportant plus de mots dans des bandes de fréquence plus avancées peut être considéré comme ayant un lexique plus sophistiqué.

29 Autres propriétés lexicales (34 variables). Les autres propriétés lexicales que nous essayons de mesurer sont la familiarité lexicale, l'imageabilité de chaque mot, le caractère concret/abstrait et le degré de polysémie lexicale. Ces variables correspondent à des normes lexicales typiquement utilisées en psycholinguistique. La familiarité est une mesure de la capacité de reconnaissance d'un mot par une personne. L'imageabilité fait référence à la facilité avec laquelle on peut former une image mentale ou une représentation mentale du concept auquel ce mot se réfère. De même, le caractère concret d'un mot indique dans quelle mesure il se réfère à une entité réelle, tangible, décrivant des objets et des comportements d'une manière qui semble plus spécifique, familière et perceptible pour l'individu ou l'esprit (Brysbaert *et al.*, 2014).

30 Les marqueurs de développement du langage (27 variables). Les marqueurs de développement du langage peuvent (1) indiquer la structure syntaxique, et aussi (2) évaluer si la structure syntaxique est complexe. Les mesures comme l'index Yngve (voir François, 2011) ou la profondeur moyenne des constituants sont souvent utilisées pour

évaluer la complexité de la structure des phrases en lien avec le nombre de constituants syntaxiques.

31 Les caractéristiques des phrases (5 variables). Nous calculons leur longueur, la proportion de phrases simples, complexes et composées dans les textes.

32 Les caractéristiques morphologiques (84 variables). Les variables liées aux caractéristiques morphologiques des mots du texte sont également révélatrices du genre de texte. Nous avons ainsi identifié divers types de propriétés, par exemple la proportion de mots d'un genre ou d'un nombre donné, la proportion de définis, etc. Plus particulièrement, les propriétés verbales (temps, mode, et voix) ont été prises en compte, car étudiées par plusieurs chercheurs dans le domaine du genre touristique promotionnel (section 1).

33 16 variables calculant la proportion de chacune des catégories de discours définies dans le jeu d'étiquettes universel du projet UD (de Marneffe, 2021), sont également identifiées. Cependant, nous accordons une attention particulière aux cas des adjectifs et des pronoms personnels, qui sont décrits dans les sections suivantes.

34 Les adjectifs. Les adjectifs sont particulièrement fréquents dans le registre écrit. Les adjectifs sont essentiels pour « présenter la beauté, l'attrait et le caractère unique des destinations ou des attractions » dans le tourisme (Manca, 2016). L'usage des adjectifs renforce également la persuasion dans les discours touristiques (Pierini, 2009) et les adjectifs évaluatifs, en particulier, peuvent exprimer le point de vue de l'auteur ou du locuteur (Edo-Marzá, 2017) et provoquer chez le lecteur une sorte de sentiment esthétique (Muñoz, 2019). Afin d'aller au-delà d'un simple décompte des adjectifs, nous avons décidé de distinguer (1) les adjectifs objectifs, (2) les adjectifs évaluatifs et (3) les adjectifs émotionnels. Une autre campagne d'annotation a été réalisée dans le cadre de cette étude ; pour rédiger le guide d'annotation, nous nous sommes basés sur la classification psychopragmatique des adjectifs de Kerbrat-Orecchioni (1980). Le κ de Cohen (Cohen, 1960) obtenu pour cette campagne est de 0.87, ce qui représente un excellent accord.

35 Les pronoms personnels. Selon Callaway et Lester (2001), sans pronoms, les textes deviennent rapidement confus, car les lecteurs

commencent à prêter plus d'attention au style d'écriture qu'au contenu. Dans les textes promotionnels, l'utilisation des pronoms peut néanmoins être l'une des caractéristiques les plus distinctives. Selon Cook (2001), la caractéristique la plus frappante et la plus fréquente qui distingue les textes promotionnels des autres types de discours est l'utilisation omniprésente des pronoms personnels de la 2e personne (*vous* et *tu*).

36 Les mesures de cohérence du texte (8 variables). Ce sont des mesures pour calculer la cohérence textuelle. Ces variables permettent de mesurer la cohérence au niveau des champs lexicaux et/ou du lexique entre une phrase et toutes les autres phrases, mais aussi entre chaque phrase et la phrase suivante, en se basant sur la forme fléchie des mots et leur lemme.

37 Les expressions référentielles (6 variables). Ces mesures sont la proportion de pronoms entre tous les noms et tous les mots, le ratio des possessifs, le ratio des pronoms personnels et la proportion des pronoms personnels dans le texte.

38 Les hyperliens. Il s'agit d'une caractéristique numérique du genre numérique. Les hyperliens sont stratégiques pour la communication touristique numérique, dans la mesure où ils reflètent un choix communicatif des expéditeurs (Maci, 2007). Dimitrova *et al.* (2005) notent que les liens peuvent rendre le texte plus compréhensible. Nous décrivons ici la méthode de récupération des hyperliens. Nous récupérons d'abord la source HTML de chaque page d'accueil du corpus. Pour chacune de ces pages, nous repérons et extrayons toutes les balises <a>. Pour chacune de ces balises, nous effectuons une recherche de son contenu textuel s'il existe (le texte entre la balise ouvrante et la balise fermante) dans chacun des tours identifiés pour le site en question. Cela nous permet d'établir la correspondance entre les tours et les liens automatiquement. Nous traitons manuellement les cas où le texte est présent dans plusieurs tours, ce qui ne permet pas d'établir la correspondance automatiquement.

2.3. Identification des caractéristiques de genre

39 La méthode utilisée pour identifier les caractéristiques des sites web des agences de voyages francophones s'inspire des techniques statistiques classiquement utilisées pour extraire les termes de spécialité dans un corpus. Les caractéristiques sont calculées à l'aide de FABRA. Pour faire ressortir les caractéristiques typiques des genres considérés, chaque caractéristique est comparée en fonction de sa distribution dans le corpus d'intérêt (ex. AGENCE) et de sa distribution dans un corpus représentatif des discours numériques (ex. frWac). L'utilisation d'un corpus de contraste est une approche courante dans des travaux tels que l'identification de mots-clés (Rehm *et al.*, 2008), et au niveau des caractéristiques, elle a été utilisée dans des travaux comme Wong et Dras (2009) et Hai *et al.* (2014). Les distributions sont comparées à l'aide d'un test U de Mann-Whitney, qui permet de déterminer si le suremploi ou le sous-emploi d'une caractéristique dans le corpus d'intérêt est significatif. Une fois une différence significative identifiée, nous utilisons la médiane des distributions comme outil diagnostic pour identifier s'il s'agit d'un suremploi ou d'un sous-emploi. Si la médiane est plus élevée dans le corpus d'intérêt, cette caractéristique linguistique est suremployée dans le genre étudié ; à l'inverse, la caractéristique y est sous-employée. Pour certaines distributions atypiques où la médiane ne permet pas de clarifier suffisamment la nature de l'emploi, nous avons alors recours à la moyenne. Au vu du nombre élevé de caractéristiques considérées et donc de comparaisons statistiques réalisées, nous avons opté pour une p-valeur plus élevée (0,01) afin de limiter les erreurs de type I (identifier à tort une caractéristique comme typique du genre). Le tableau suivant résume les trois comparaisons que nous avons réalisées à partir de la méthodologie décrite ci-dessus :

Tableau 3 : Présentation des analyses

Analyse	Comparaison (corpus)	Objectif
1re	(AGENCE, BLOG, OFF, DEST) vs frWac	Faire ressortir les caractéristiques des discours touristiques

2e	AGENCE vs frWac	Faire ressortir les caractéristiques des sites web des agences de voyages
3e	AGENCE vs (BLOG, OFF, DEST)	Identifier les spécificités de ces sites web par rapport aux autres discours touristiques

40 Dans la première analyse, nous comparons les quatre genres touristiques avec le corpus frWac, à l'aide des caractéristiques linguistiques décrites à la section 2.2 (sauf les hyperliens et la structure des tours). Le but est de confirmer les spécificités des discours touristiques numériques promotionnels, mais aussi d'enrichir cette description à l'aide de nos nouvelles variables.

41 Ensuite, nous cherchons à caractériser plus finement le discours des agences (corpus AGENCE) en le comparant à une diversité de discours numériques en français (frWac), toujours à l'aide des variables de la section 2.2 (sauf les hyperliens et la structure des tours). Cette analyse vise à vérifier si le mélange des quatre discours touristiques numériques n'a pas occulté certaines caractéristiques typiques des sites web des agences.

42 Enfin, notre troisième analyse consiste à caractériser le genre des agences de voyages (corpus AGENCE) par rapport aux trois autres genres du même domaine (corpus BLOG, OFF et DEST). Nous comparons plusieurs niveaux : corpus AGENCE avec l'ensemble des corpus BLOG, OFF et DEST, et corpus AGENCE avec chacun des corpus BLOG, OFF et DEST. Ici, le jeu de caractéristiques analysées est plus riche : toutes les caractéristiques linguistiques et numériques présentées à la section 2.2.

43 Pour l'analyse, les caractéristiques ont été identifiées et comparées, dans un premier temps, au niveau de l'ensemble des textes (sur chaque page d'accueil). Cette comparaison se distingue de la comparaison qui prend en compte les tours. Pour chaque page d'accueil, nous calculons les variables sur chaque fragment de texte associé à un tour. Pour les analyses 1 et 2, la comparaison se fait au niveau Genre ; l'analyse 3 implique une comparaison à la fois au niveau Genre avec et sans la présence des tours. Le fait de mener une comparaison au niveau du genre (sans tour) aide à déterminer les phénomènes linguistiques spécifiques du genre. Cependant, ces phénomènes doivent être mis en relation avec les tours, afin de clarifier leur contribution à chaque objectif communicatif.

3. Résultats

44 Dans cette section, nous présentons les résultats de nos expériences décrites à la section 2. À la section 3.1, nous rapportons nos observations sur les caractéristiques qui distinguent le discours touristique numérique des discours numériques en général (l'analyse 1), avant de présenter celles qui distinguent le discours des sites d'agences de voyages des autres genres de discours touristique numérique (analyses 2 et 3) à la section 3.2.

3.1. Traits spécifiques des discours touristiques numériques

45 Les organisations touristiques doivent vendre le même produit sur différents marchés et à différents clients, et donc fournir différents types d'informations (Osti et Pechlaner, 2001). En effet, nos résultats montrent que le texte touristique a un vocabulaire plus riche que les textes numériques en général. Par exemple, une des variables statistiques souvent utilisées pour illustrer la diversité lexicale est le type-token ratio au carré (Squared TTR), soit le nombre de vocables divisé par le nombre d'occurrences (*token* en anglais) au carré. La médiane est ≈ 873 dans le corpus touristique, contre ≈ 564 dans le corpus frWac. La sophistication se traduit également par un plus grand nombre de mots polysyllabiques des textes touristiques.

46 En sus de cette diversité lexicale, les textes touristiques numériques semblent moins accessibles, lexicalement parlant, à un public large que celui des textes numériques de français général. Les résultats de fréquence et sophistication lexicale montrent que les formes du lemme des noms, noms propres, verbes, adjectifs et adverbes tendent à être moins fréquentes dans le corpus touristique (médiane ≈ 296 occurrences par million) que dans celui frWac (médiane ≈ 410).

47 Sur le plan syntaxique, les textes touristiques se distinguent par une syntaxe plus complexe que celle des textes numériques issus de frWac, ce qui peut en réduire l'accessibilité. En effet, l'indice Yngve – qui mesure la profondeur moyenne des syntagmes dans un texte – met en évidence une syntaxe plus élaborée dans les textes

touristiques (médiane ≈ 3,28) par rapport aux textes numériques (médiane ≈ 2,77).

48 Dans la communication touristique numérique, nous remarquons une relation plus dynamique entre les interlocuteurs (les rédacteurs et les internautes). En effet, les pronoms personnels et possessifs sont beaucoup plus présents dans le corpus touristique que dans le corpus frWac. On peut par exemple observer un emploi plus marqué du pronom personnel à la première personne du pluriel (*nous*) dans le corpus touristique (médiane ≈ 0,33) que dans frWac (médiane ≈ 0,09). Cette présence montre une volonté d'identifier l'émetteur dans la communication touristique (Askehave et Nielsen, 2005). Les textes touristiques visent à inclure explicitement le public visé dans la communication. Ainsi, le pronom personnel à la deuxième personne au pluriel (*vous*) est aussi davantage employé dans le corpus touristique. La stratégie adoptée repose sur la technique de ciblage de l'ego (*ego-targeting*), visant à offrir aux touristes une expérience personnalisée et exclusive (Williamson, 1983 ; Boyer et Viallon, 1994). L'utilisation de ces pronoms personnels rend le message chaleureux, réduit le fossé entre l'émetteur et le destinataire, jouant ainsi sur un marketing de proximité (Gahlot Sarkar et Sarkar, 2016). Cela fait en sorte que le public perçoit le message comme étant conçu pour l'aider (Teodorescu, 2015). Voici quelques exemples de l'usage davantage des pronoms personnels :

(1) Voyages exceptionnels : Nous vous faisons découvrir le monde comme vous ne l'avez jamais vu.

(2) Nous nous occupons de tout : vol, hôtel & expérience et vous aidons en cas de besoin.

(3) Laissez-nous travailler pour vous. Nous sommes ravis de vous conseiller.

(4) Les voyages d'affaires conçus pour vous, avec vous partout où vous devez être.

49 En complément de l'usage des pronoms personnels, l'élément qui signale le plus la relation dynamique entre les interlocuteurs est le souci d'une communication interactive sur les sites web touristiques. Nous trouvons qu'il s'agit d'une évolution du discours touristique, souvent considéré comme une forme d'un monologue (Dann, 1996). Cette interaction dynamique, grâce au web actuel, se traduit par une série de propositions communicatives de la part des propriétaires de sites web aux internautes : contacter l'expéditeur, s'abonner au site

web, se connecter en tant qu'utilisateur du site, commenter et aimer des articles présentés sur la page, etc. Nous avons trouvé plusieurs formules impératives ou formules de phrases infinitives (médiane ≈ 1,72, contre celle à ≈ 0,71 de frWac), qui sont plus fréquentes pour le genre du discours touristique promotionnel. Voici quelques exemples :

(1) Nous contacter !
(2) Réservez votre voyage dès aujourd'hui !
(3) Abonnez-vous à notre newsletter !
(4) Discutons, rêvez, construisons… découvrez !
(5) Partagez cet article !

50 À partir des constats sur les caractéristiques linguistiques citées ci-dessus, nous pouvons affirmer que le genre de discours touristique numérique possède bien des particularités linguistiques par rapport à l'ensemble des discours numériques français à plusieurs niveaux. Nous allons, dans la section suivante, caractériser le genre de discours des agences de voyages.

3.2. Traits spécifiques du discours touristique numérique des agences de voyages

3.2.1. Niveau Genre, sans prise en compte des tours (avec frWac, et trois autres genres)

51 Il est intéressant de noter que la comparaison du corpus AGENCE avec le corpus frWac (analyse 2) fait ressortir les mêmes variables qu'à la section 3.1. Concernant la comparaison entre le genre des sites web des agences (AGENCE) avec l'ensemble des trois autres genres (BLOG, OFF et DEST) ainsi que les sites web (AGENCE) avec chaque genre séparément (BLOG), (OFF) et (DEST), les résultats sont les suivants :

52 Tout d'abord, les sites web des agences de voyages utilisent un vocabulaire riche, diversifié et sophistiqué ; leurs textes semblent moins accessibles, en raison d'un lexique moins fréquent et une syntaxe plus complexe. En effet, le corpus AGENCE a des mots plus

longs, contenant plus d'affixes, préfixes et suffixes que des mots dans les trois autres genres. La complexité syntaxique se traduit par un usage de plus de phrases composées (moyenne ≈ 6,67 contre ≈ 0,06).

53 Les discours des agences de voyages soulignent le temps présent et projettent les lecteurs dans l'avenir. En effet, le corpus AGENCE utilise plus de temps au présent de l'indicatif et au futur. Ce sont des formes verbales qui orientent la lecture des internautes et qui influencent leur perception du voyage (Dann, 1996). Les internautes sont amenés à se projeter dans le temps présent, via des présentations de produits touristiques ou de l'agence et dans le temps futur, avec des propositions précises de voyages ou d'actions. Les exemples ci-dessous, tirés du corpus AGENCE, illustrent bien cette tendance. Le futur proche peut également apporter une nuance dans le temps court, selon laquelle l'agence veut se montrer disponible dans l'interaction avec l'internaute.

(1) Si vous souhaitez savoir comment (*nom de l'agence*) peut vous aider et recevoir des communications sur nos produits et services, veuillez remplir ce formulaire. Nous vous contacterons pour prendre un rendez-vous et échanger ensemble.

(2) EXCLUSIF. De la préparation à la réalisation de votre voyage, vous serez encadrés par une équipe professionnelle forte d'une longue expérience.

(3) Depuis plus de 40 ans, cette entreprise indépendante propose un service sur mesure et personnalisé. Ses conseillers expérimentés et proactifs seront à votre écoute et vous feront bénéficier de leur savoir-faire, de leur connaissance des destinations…

54 Les agences de voyages ont tendance à ajouter une touche émotionnelle et esthétique, en utilisant plus d'adjectifs évaluatifs que les autres genres. Il s'agit d'une stratégie de communication qui vise à convaincre de la crédibilité de l'agence. Cette stratégie implique également le lecteur dans la communication et augmente la valeur perçue d'une destination touristique (Mocini, 2013). Voici quelques exemples de l'usage des adjectifs évaluatifs :

(1) Un service complet, des conseils personnalisés et compétents en temps de crise.

(2) Des paysages très variés, les destinations extraordinaires du monde.

(3) De magnifiques villes et villages vous attendent pour vous faire découvrir des régions inestimables.

55 Une autre remarque importante concernant les sites des agences de voyages est la volonté d'engager le public, d'encourager l'action et de

rendre le texte plus accessible. Cette stratégie de communication des agences de voyages est illustrée par une fréquence importante des hyperliens dans le corpus AGENCE. En effet, la moyenne d'hyperliens par site web du genre AGENCE est à 8,2, tandis que celle des blogs est à 4,5, des offices de tourisme à 2,5 et des destinations touristiques à 1,6. Le fait de présenter un hyperlien dans le texte montre un effort des agences de voyages dans l'implication communicative des internautes. Le lecteur, en cliquant sur l'hyperlien, devient le locuteur, qui décide de son schéma de communication. Il s'agit d'une des fonctions communicatives des hyperliens : donner de la crédibilité à l'auteur du lien en offrant aux lecteurs un moyen transparent de diriger eux-mêmes la communication (Coddington, 2012). Dans le texte touristique des agences de voyages, cette fonction se manifeste notamment à travers les menus qui indiquent le chemin de navigations, ainsi que les boutons d'appel à l'action (en anglais : *call to action*) : « réserver un voyage », « télécharger la brochure », « lire la suite », « s'inscrire à la newsletter », etc.

3.2.2. Niveau Genre, avec des tours (avec trois autres genres)

56 Concernant les fonctions communicatives de l'ensemble des quatre catégories (AGENCE, BLOG, OFF et DEST), nous remarquons que les émetteurs de discours touristiques numériques font attention à 1) l'identification et à la construction de l'image de soi (ethos) et 2) à la communication claire des informations. Ces stratégies se traduisent par un emploi récurrent des tours 3 (identifier l'expéditeur), 4 (indiquer la structure du contenu) et 5 (détailler le contenu), selon notre résultat statistique. En effet, quand l'internaute va sur une page web, pour ne pas se perdre, il doit connaître le propriétaire du site, ainsi que la structure globale du site et recevoir des informations détaillées sur le contenu. Il existe cependant une grande différence dans l'usage du tour 5 dans le genre des blogs par rapport aux autres genres : les auteurs des blogs présentent plus d'informations détaillées sur leur page d'accueil. Afin d'indiquer la structure du contenu des pages d'accueil, les sites des agences utilisent des menus plus riches en noms (médiane ≈ 0,92, plus que celle ≈ 0,20 des autres catégories). Nous remarquons également le même résultat concernant l'usage des noms et de noms propres des agences de

voyages, afin de détailler des informations. Le fait d'utiliser plus de noms et de noms propres montre une dimension plus informative et explicative dans le discours des agences de voyages.

57 Les agences de voyages ont plus tendance à attirer l'attention des lecteurs à leur arrivée sur la page. Cela se traduit par un usage plus fréquent du tour 1 (attirer l'attention). En effet, beaucoup de phrases sous forme de slogans apparaissent à la première instance de l'ouverture du site, et les caractéristiques linguistiques utilisées pour former ce tour sont souvent des phrases nominales ou des formules de phrase impérative. Voici quelques exemples :

(1) L'art de voyager.
(2) Soyez prêts à voyager.
(3) Voyager léger.
(4) Ouvrez les portes du monde – Circuit francophone
(5) Découvrez l'Iran – Aux milles et un visages

58 Les agences de voyages n'ont pas tendance à souhaiter la bienvenue aux internautes. En effet, seuls sept sites web d'agences de voyages de notre corpus suivent cette pratique, et tous les sept utilisent le mot « bienvenue ». Cet acte de parole se retrouve le plus dans des blogs de voyage, par exemple :

(1) Bienvenue sur (*nom du blog*).
(2) Je m'appelle (*nom de la personne*) et je suis très heureuse de t'accueillir sur mon blog de voyage !

59 De même, les agences ne cherchent pas particulièrement à créer des communautés de discours, ce qui se traduit par une présence moyenne moins élevée (0,95) du tour 8 (établir une communauté de discours). À l'inverse, les agences semblent très soucieuses d'asseoir leur crédibilité. Cette pratique se traduit par l'usage très fréquent du tour 6 dans le corpus AGENCE. La présence moyenne de ce tour dans le corpus AGENCE est à 1,23, alors qu'elle se situe entre 0,16 et 0,33 dans les trois autres corpus. Les agences de voyages utilisent un discours à la fois informatif et persuasif, impliquant activement les internautes. Cela se traduit par l'usage de noms (médiane ≈ 0,93 contre 0,1), d'adjectifs émotionnels (médiane ≈ 0,5 contre 0,02),

d'adjectifs évaluatifs (médiane ≈ 0,8 contre 0,25) et une grande proportion de possessifs dans le texte (médiane ≈ 0,3 contre 0,01). En voici un exemple :

(1) Pourquoi (*nom de l'agence*) ? Des conseils personnalisés et compétents en temps de crise. Un service complet avant, pendant et après votre voyage. Les meilleures offres de vacances à des prix raisonnables. Une permanence téléphonique 24 h/24 et 7 j/7.

60 Ce sont les agences de voyages qui essaient le plus d'établir le contact avec les internautes. Cela s'observe par une proportion plus élevée du tour 7 (établir le contact) dans le genre des agences, tandis que le genre des blogs ne comporte pratiquement pas ce tour. Souvent, le tour 7 est formé par des formules impératives ou infinitives (« Contactez-nous ! », « Parlons-nous ! » ou « Nous contacter »), sous la forme d'un formulaire de contact que le lecteur doit remplir ou encore sous la forme de texte reprenant les coordonnées de l'agence.

61 Finalement, la caractéristique numérique au sein du genre AGENCE se présente de la même manière que celle dans tout le corpus tourisme, c'est-à-dire une concentration des hyperliens pour indiquer la structure de la page (tour 4 – 10,32 en moyenne) et détailler des informations (tour 5 – 4,4 en moyenne). Les hyperliens sont également plus fréquents en vue d'établir un contact entre le visiteur du site web avec les agences. On y trouve des boutons *call-to-action*.

4. Discussion

62 Les résultats de notre recherche montrent que les sites web des agences de voyages constituent un genre de discours en soi, dont les caractéristiques linguistiques et numériques diffèrent bien d'autres discours touristiques numériques.

63 Tout d'abord, les agences de voyages se démarquent par l'usage d'un langage riche, diversifié et sophistiqué, qui semble moins accessible au grand public. Les caractéristiques tant lexicales que syntaxiques du corpus des textes touristiques en général et celui des agences de voyages en particulier, s'avèrent moins lisibles et pourraient nuire au traitement de l'information présentée dans un texte promotionnel (Lowrey, 1998). Toutefois, il est à noter que les textes promotionnels

sont conçus pour être compréhensibles par le plus grand public possible (Lowrey, 1998) ; il est donc nécessaire de regarder d'autres facteurs qui participent à l'accessibilité d'un texte. Cela est d'autant plus important que, selon Chamblee *et al.* (1993), les textes promotionnels complexes sur le plan lexical ont un effet positif sur la lecture des lecteurs, car ces derniers sont amenés à réfléchir davantage aux contenus et à traiter le message de manière plus profonde. En d'autres termes, cette complexité lexicale stimule la rétention et la mémorisation des informations, renforçant ainsi l'impact du message sur le lecteur.

64 Ensuite, les agences de voyages accordent une grande importance à leur image. L'emploi fréquent d'une touche émotionnelle et esthétique – avec l'utilisation abondante d'adjectifs émotionnels et évaluatifs, de noms et de possessifs – représente une stratégie de communication visant à convaincre les lecteurs de la fiabilité et du professionnalisme de l'agence. La dimension temporelle dans le discours des agences de voyages participe également à l'exercice de séduction des internautes, en leur montrant un temps présent, mais aussi une vision tournée vers l'avenir, grâce à l'usage des temps du présent et du futur proche.

65 Les agences de voyages cherchent également à impliquer activement les internautes dans la communication. Les éléments notables sont l'inclusion d'hyperliens – qui aident à indiquer la structure web et à détailler les informations – et des embrayeurs de l'énonciation. Bien que les agences de voyages cherchent à capter l'attention des lecteurs dès leur arrivée, elles ne souhaitent pas nécessairement la bienvenue aux internautes. De plus, il semble qu'elles ne cherchent pas spécifiquement à créer des communautés de discours.

66 Une limite de notre étude est que nous avons considéré les 156 pages d'accueil de frWac comme un ensemble représentatif du discours numérique en général. Il est possible que cet échantillon ne soit pas suffisamment représentatif, ce qui pourrait altérer certaines de nos conclusions. Cependant, le fait de nous concentrer sur les pages d'accueil avait pour but d'assurer une cohérence au niveau des types de documents comparés.

67 Notre travail bénéficierait idéalement d'une analyse multidimensionnelle. En effet, face à l'émergence d'un nouveau type

de discours numérique complexe – en l'occurrence le discours touristique dans ce cas – nous devons (1) réexaminer ou revisiter les modèles de genre traditionnels, et (2) adopter une approche multidimensionnelle. Dans cette perspective, nous portons notre attention non seulement sur le texte lui-même, mais également sur d'autres éléments associés (tels que les éléments graphiques visuels, ainsi que d'autres éléments non textuels, discursifs et commerciaux). Nous prévoyons d'effectuer cette analyse lors d'une étape ultérieure de notre travail.

Conclusion

68 Dans cette étude, nous avons caractérisé le genre de discours touristique numérique des agences de voyages. Pour ce faire, nous avons comparé ce genre à d'autres genres de discours, grâce à des approches mixtes quantitatives, qualitatives et comparatives, afin de faire ressortir les caractéristiques linguistiques et numériques typiques de ce genre. Les résultats montrent que les sites web des agences de voyages constituent un genre de discours spécifique, non seulement au niveau des fonctions communicatives du discours, mais aussi au niveau des stratégies linguistiques utilisées. Ces résultats contribuent à l'établissement d'un modèle de genre pour les discours publiés sur les sites web touristiques francophones, en tenant compte des stratégies linguistiques, communicationnelles et numériques, ainsi que de leur efficacité. Ce modèle pourrait servir de référence pour les professionnels du tourisme souhaitant améliorer leur communication sur les pages d'accueil. Enfin, il serait intéressant d'analyser en amont la mise en scène culturelle dans les discours touristiques, afin de comprendre comment le genre est conçu en fonction des différentes cultures et communautés de discours. En effet, Jin, Park et Kim (2008) ont montré que les différences culturelles peuvent expliquer certaines stratégies appliquées par les entreprises touristiques à l'échelle mondiale, mais également pourquoi leurs produits touristiques rencontrent le succès dans certaines communautés et pas dans d'autres.

BIBLIOGRAPHIE

Adam, Jean-Michel. 2001. En finir avec les types de textes. Dans Michel Ballabriga (dir.), *Analyse des discours. Types et genres : Communication et interprétation*, 42-59. Toulouse : Éditions universitaires du Sud.

Alali, Bashar Abdulkareem, Afiza Mohamad Ali & Afida Mohamad Ali. 2019. Genre-based analysis of travel guides: A study on Malaysia, Thailand and the Philippines. *LSP International Journal* 6(2). 10.11113/lspi.v6n2.86.

Askehave, Inger & Anne Ellerup Nielsen. 2005. Digital genres: A challenge to traditional genre theory. *Information Technology & People* 18(2). 120-141. 10.1108/09593840510601504.

Banks, David. 2001. Analyses des discours spécialisés : le domaine anglais. *Revue Française de Linguistique Appliquée*, 6(2). 7-16. 10.3917/rfla.062.0007.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209-226.

Bhatia, Vijay K. 1993. *Analysing genre: Language use in professional setting*. Londres & New York : Longman.

Bhatia, Vijay K. 2004. *Worlds of written discourse: A genre-based view*. Londres : Continuum. 10.5040/9781474212038.

Bonhomme, Marc. 2015. La problématique des genres de discours dans la communication sur Internet. *Travaux Neuchâtelois de Linguistique* 63. 31-47.

Bonvin, Audrey & Amelia Lambelet. 2019. Exploration empirique de la richesse lexicale : la perception humaine. *Linguistik Online* 100(7). 65-94. 10.13092/lo.100.6018.

Boyer, Marc & Philippe Viallon. 1994. *La communication touristique*. Paris : Presses universitaires de France.

Brysbaert, Marc, Amy B. Warriner & Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3). 904-11. 10.3758/s13428-013-0403-5.

Buhalis, Dimitrios & Rob Law. 2008. Progress in information technology and tourism management: 20 years on and 10 years after the Internet – The state of eTourism research. *Tourism management* 29(4). 609-623. 10.1016/j.tourman.2008.01.005.

Caballero, Rosalio. 2008. Theorizing about genre and cybergenre. *Computer Resources for Language Learning* 2. 14-27.

Callaway, Charles B. & James C. Lester. 2001. Evaluating the effects of natural language generation techniques on reader satisfaction. *Annual Meeting of the*

Cognitive Science Society 23. 164-169.

Cardon, Rémi, Trang Tran Hanh Pham, Julien Zakhia Doueihi & Thomas François. 2024. Contribution of move structure to automatic genre identification: An annotated corpus of French tourism websites. *International Conference on Computational Linguistics Language Resources and Evaluation* (LREC-COLING), 3916-3926.

Chamblee, Robert, Robert Gilmore, Gloria Thomas & Gary Soldow. 1993. When copy complexity can help ad readership. *Journal of Advertising Research* 33(3). 23-29. 10.1080/00218499.1993.12466890.

Cheong, Cecilia Yin Mei. 2013. A *multi-dimensional genre analysis of tourism homepages and web-mediated advertorials*. Kuala Lumpur, MYS: Universiti Malaya, dissertation.

Coddington, Mark. 2012. Building frames link by link: The linking practices of blogs and news sites. *International Journal of Communication* 6. 2007-2026.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1). 37-46. 10.1177/001316446002000104.

Constantinides, Efthymios, Carlota Lorenzo-Romero & Miguel A. Gómez. 2010. Effects of web experience on consumer choice: A multicultural approach. *Internet Research* 20(2). 188-209.

Cook, Guy. 2001. *The discourse of advertising*, 2nd éd. Londres & New York : Routledge.

Dann, Graham M. S. 1996. The language of tourism: A sociolinguistic perspective. Wallingford: CAB International.

De Cock, Barbara & Hilde Hanegreefs. 2013. On the representation of observations in business discourse vs. academic discourse. *Global Advances in Business communication*. http://hdl.handle.net/2078.1/134668 (22 octobre 2025).

Dimitrova, Daniela V., Lynda Lee Kaid, Andrew Paul Williams & Kaye D. Trammell. 2005. War on the Web: The immediate news framing of Gulf War II. *Harvard International Journal of Press/Politics* 10(1). 22-44. 10.1177/1081180X05275595.

Edo Marzá, Nuria. 2017. A comprehensive corpus-based study of the use of evaluative adjectives in promotional hotel websites. *Odisea* 12. 97-123. 10.25115/odisea.v0i12.222.

Finnemann, Niels O. 1999. Hypertext and the representational capacities of the Binary Alphabet. *Arbejdspapirer* 77(99). Aarhus: Centre for Cultural Research.

François, Thomas. 2011. *Les apports du traitement automatique des langues à la lisibilité du français langue étrangère*. Louvain, CHE : Université catholique de Louvain, thèse.

Gaillard, Mathilde. 2015. Entre parole d'expert et visée programmatique : les policy briefs, genre discursif propre aux think tanks américains. *ASp* 67. 55-80.

10.4000/asp.4647.

Gahlot Sarkar, Juhi G. & Abhigyan Sarkar. 2016. Up, close and intimate: Qualitative inquiry into brand proximity amongst young adult consumers in emerging market. *Young Consumers: Insight and Ideas for Responsible Marketers* 17(3). 256-273. 10.1108/YC-04-2016-00593.

Gémar, Germán. 2014. Influence of cultural distance on the internationalization of Spanish hotel companies. *Tourism & Management Studies* 10 (1). 31-36.

Giltrow, Janet & Dieter Stein. 2009. Genre in the Internet: Innovation, evolution, and genre theory. Dans Janet Giltrow & Dieter Stein (éd.), *Genres in the Internet: Issues in the theory of genre*, 1-26. Amsterdam & Philadelphie : John Benjamins. 10.1075/pbns.188.01gil.

Hafner, Christopher A. 2018. Digital discourses research and methods. Dans Aek Phakiti, Peter De Costa, Luc Plonsky & Sue Starfield (éd.), *The Palgrave handbook of applied linguistic research methodology*, 375-390). Londres : Palgrave Macmillan. 10.1057/978-1-137-59900-1_18.

Hafner, Christopher. A. & Jack Pun. 2020. Editorial: Introduction to this special issue: English for academic and professional purposes in the digital era. *RELC Journal* 51(1). 3-13. https://doi.org/10.1177/0033688220917472.

Hai, Zen, Kuiyu Chang, Jung-Jae Kim & Christopher C. Yang. 2014. Identifying features in opinion mining via intrinsic and extrinsic domain relevance. *IEEE Transactions on Knowledge and Data Engineering* 26(3). 623-634. 10.1109/TKDE.2013.26.

Henry, Alex & Robert L. Roseberry. 1996. A corpus-based investigation of the language and linguistic patterns of one genre and the implications for language teaching. *Research in the Teaching of English* 30(4). 472-489.

Heyd, Theresa. 2008. *Email hoaxes: Form, function, genre ecology*. Amsterdam & Philadelphie : John Benjamins.

Heyd, Theresa. 2016. Digital genres and processes of remediation. Dans Alexandra Georgakopoulou & Tereza Spilioti (éd.), *The Routledge handbook of language and digital communication*, 87-102. Londres & New York: Routledge.

Huang, Shuang. 2015. A genre-based analysis of brief tourist information texts. Dans *Proceeding of Joint International Social Science, Education, Language, Management and Business Conference*, 191-202. Atlantis Press. 10.2991/jisem-15.2015.40.

Hui, Wang, Natthaphon Santhi & Sornchai Mungthaisong. 2020. A genre analysis of online English tourist attraction promotional texts in Lijiang Yunnan Province. Dans *วารสาร สห วิทยาการ สังคมศาสตร์ และ การ สื่อสาร* 3(1). 83-106.

Ip, Janice Y. L. 2008. Analyzing tourism discourse: A case study of a Hong Kong travel brochure. *LCOM papers* 1. 1-19.

Irimiea, Silvia Blanca. 2018. The relationship between the language of tourism, tourism and sociology. *European Journal of Social Science Education and Research* 5(2). 128-141.

Jayes, Lewis T., Fitzsimmons Gemma, Weal Mark J., Kaakinen Johanna K. & Drieghe Denis. 2022. The impact of hyperlinks, skim reading and perceived importance when reading on the Web. *PLoS One* 17(2). 10.1371/journal.pone.0263669.

Jin, Byoungho, Jin-Yong Park & Jiyoung Kim. 2008. Cross-cultural examination of the relationships among firm reputation, e-satisfaction, e-trust, and e-loyalty. *International Marketing Review* 25(3). 324-337. 10.1108/02651330810877243.

Johansson, Victoria. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics* 53. 61-79.

Kadayat, Bam Bahadur & Eika Evelyn. 2020. Impact of sentence length on the readability of web for screen reader users. Dans Margherita Antona & Constantine Stephanidis, *Universal Access in Human-Computer Interaction. Design Approaches and Supporting Technologies*, 261-271. Cham : Springer. 10.1007/978-3-030-49282-3_18.

Kerbrat-Orecchioni, Catherine. 1980. *L'énonciation de la subjectivité dans le langage*. Colin.

Kress, Gunther. 2003. *Literacy in the new media age*. Londres : Routledge.

Kress, Gunther & Theo van Leeuwen Theo. 2021. *Reading images: The grammar of visual design*, 3e éd. Routledge.

Lowrey, Tina M. 1998. The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology* 7(2). 187-206. 10.1207/s15327663jcp0702_04.

Luzón, Maria José & Carmen Pérez-Llantada. 2019. *Science communication on the Internet* (Pragmatics and Beyond 308). Amsterdam & Philadelphie : John Benjamins Publishing Company.

Maci, Stefania-Maria. 2007. Virtual touring: The Web-Language of tourism. *Linguistica e Filologia* 25. 41-65. 10.6092/10446_102.

Maingueneau, Dominique. 2007. Genres de discours et modes de généricité. *Le Français Aujourd'hui* 159(4). 29-35. 10.3917/lfa.159.0029.

Malenkina, Nadezhda & Ivanov Stanislav. 2018. A linguistic analysis of the official tourism websites of the Seventeen Spanish autonomous communities. *Journal of Destination Marketing & Management* 9. 204-233. 10.1016/j.jdmm.2018.01.007.

Manca, Elena. 2016. *Persuasion in tourism discourse: Methodologies and models*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Marneffe, Marie-Catherine de, Christopher D. Manning Christopher, Joakim Nivre Joakim & Daniel Zeman. 2021. Universal dependencies. *Computational Linguistics* 47(2). 255-308. 10.1162/coli_a_00402.

Mocini, Renzo. 2013. The promotional functionality of evaluative language in tourism discourse. *Lingue e Linguaggi* 9. 157-171.

Moreno, Ana I. & John M. Swales. 2018. Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes* 50. 40-63. 10.1016/j.esp.2017.11.006.

Durán Muñoz, Isabel. 2011. Tourist translations as a mediation tool: Misunderstandings and difficulties. *Cadernos de Tradução* 1(27). 29-49. 10.5007/2175-7968.2011v1n27p29.

Durán Muñoz, Isabel. 2019. Adjectives and their keyness: A corpus-based analysis of tourism discourse in English. *Corpora* 14(3). 351-378. 10.3366/cor.2019.0178.

New, Boris, Marc Brysbaert, Jean Veronis & Christophe Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28(4). 661-677. 10.1017/S014271640707035X.

Nguyen, Thi Dung & Prannapha Modehiran. 2023. A genre-based analysis of Vietnam tourist attraction brochures with pedagogical purposes. *ABAC ODI Journal Vision. Action. Outcome* 11(1). 43-55. 10.14456/abacodijournal.2023.29.

Orlikowski, Wanda J. & JoAnne Yates. 1994. Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly* 39(4). 541-574. 10.2307/2393771.

Osti Bozen, Linda & Harald Pechlaner. 2001. Communication issues in NTO distribution strategies. Dans Dimitrios Buhalis & Eric Laws (éd.), *Tourism distribution channels: Practices, issues and transformations*, 231-242. Londres & New York : Continuum.

Öztürk, Bilge & Zafer Şafak. 2014. Genre analysis of a Turkish tourism brochure. Dans Azamar Akbarov (éd.), *Linguistics, Culture and Identity in Foreign Language Education*, 351-364.

Pierini, Patrizia. 2009. Adjectives in tourism English on the web: A corpus-based study. *CÍRCULO de Lingüística Aplicada a la Comunicación* 40. 93-116. https://revistas.ucm.es/index.php/CLAC/article/view/41886 (6 octobre 2025).

Pleau, Joannie. 2017. Le texte à l'ère du numérique : analyse du concept de genre numérique. *Canadian Journal for New Scholars in Education* 8(1). 144-149. https://journalhosting.ucalgary.ca/index.php/cjnse/article/view/30798 (6 octobre 2025).

Rayner, Keith & Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14(3). 191-201. 10.3758/BF03197692.

Rehm, Georg, Marina Santini, Alexander Mehler, Pavel Braslavski, Rüdiger Gleim, Andrea Stubbe, Svetlana Symonenko, Mirko Tavosanis & Vedrana Vidulin. 2008. Towards a reference corpus of Web genres for the evaluation of genre identification systems. *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (LREC'08). 351-358. http://www.lrec-conf.org/proceedings/lrec2008/pdf/94_paper.pdf. (24 octobre 2025).

Resche, Catherine. 2010. Analyse d'un genre de discours spécialisé : les lettres annuelles de Warren Buffett aux actionnaires de Berkshire Hathaway et leur spécificité. *e-Rea* 8(1). https://doi.org/10.4000/erea.1365.

Richer, Jean-Jacques. 2011. Les genres de discours : une autre approche possible de la sélection de contenus grammaticaux pour l'enseignement/apprentissage du F.L.E. ?. *Linx* 64-65.15-26. 10.4000/linx.1396.

Santini, Marina, Alexander Mehler & Serge Sharoff. 2010. Riding the rough waves of genre on the web: Concepts and research questions. Dans Alexander Mehler, Serge Sharoff & Marina Santini (éd.), *Genre on the web: Computational models and empirical studies*, 3-30. Dordrecht, Heildeberg & Londres : Springer. 10.1007/978-90-481-9178-9_1.

Swales, John M. 1990. *Genre analysis: English in academic and research settings*. Cambridge & New York: Cambridge University Press.

Swales, John M. 2004. Research genres: Explorations and applications. Cambridge & New York: Cambridge University Press.

Suciu, Lavinia. 2019. Introductory chapter: Discourse and discourse analysis: A retrospective approach. Dans Lavinia Suciu (éd.), *Advances in discourse analysis*. IntechOpen. 10.5772/intechopen.82823.

Teodorescu, Adriana. 2015. Linguistic patterns in advertising messages. *Knowledge Horizons. Economics* 7(3). 115-118.

Viallon, Philippe & Sandrine Henneke-Lange. 2011. L'approche interculturelle comme critère de l'excellence des sites web des métropoles touristiques. *Téoros* 30(1). 64-72. http://journals.openedition.org/teoros/1259 (6 octobre 2025).

Violino, Bob. 2001. E-business lurches abroad. *Internet Week* 19.

Wilkens, Rodrigo, David Alfter, Xiao Wang, Alice Pintard, Anaïs Tack, Kevin P. Yancey & Thomas François. 2022. Fabra: French aggregator-based readability assessment toolkit. *Language Resources and Evaluation Conference*, 1217-1233.

Williamson, Judith. 1983. *Decoding advertisements: Ideology and meaning in advertising*. Londres : Marion Boyars.

Wong, Sze-Meng Jojo & Mark Dras. 2009. Contrastive analysis and native language identification. *Proceedings of the Australasian Language Technology Association Workshop* 2009. 53-61. https://aclanthology.org/U09-1008.pdf (24 octobre 2025).

Yaemwannang, Chalita. 2017. Online promotional genre set in tourism business: multimodal discourse analysis of international hotel, restaurant, and tourist destination homepages. Nakhon Ratchasima, THA : School of Foreign Languages Institute of Social Technology Suranaree University of Technology, dissertation.

NOTES

1 En anglais : English for Specific Purpose (ESP).

2 Les termes « tours » et « pas » sont abordés dans Banks (2001).

3 « prominence in contemporary communication ». Nous traduisons, sauf mention du contraire.

4 « innovation in digitally mediated research genres ».

5 « a push towards the democratization of science and the sharing of research ».

6 « When we talk about digital genres and practices we are referring to the recurrent, goal-directed, communicative events (Swales, 1990) mediated by digital tools in this way, along with the constellation of activities that surround such events. »

7 « a new communication setting which reconfigures the conditions to which pragmatic features of language respond ».

RÉSUMÉS

Français
L'étude présentée vise à analyser les pages d'accueil des sites web d'agences de voyages francophones afin de déterminer s'il s'agit d'un genre de discours distinct et d'en caractériser les spécificités. Une série d'expériences a été conduite à cette fin. Un corpus de français numérique bien connu, frWac, ainsi qu'un corpus de sites web touristiques collecté pour cette étude ont été utilisés. Les analyses effectuées consistent, d'une part, en une comparaison des discours touristiques numériques à un échantillon de discours numériques variés, et, d'autre part, en une comparaison des discours des sites d'agences de voyages avec d'autres discours numériques touristiques, tels que les blogs, les sites d'offices de tourisme et les sites de points d'intérêt touristiques. Pour ces comparaisons, des outils de traitement automatique des langues ont été mobilisés pour extraire une série de caractéristiques linguistiques, complétée par une analyse rhétorique basée sur la structure des tours (moves en anglais). Il est démontré que le discours des agences de voyages

constitue un genre distinct, dont les caractéristiques spécifiques sont décrites.

English
The aim of this study was to analyze the homepages of French-language travel agency websites in order to determine whether they constitute a distinct discourse genre and to characterize their specific features. To this end, a series of experiments were carried out. A well-known French digital corpus, frWac, and a corpus of tourism websites collected for this study were used. The analyses carried out consist, on the one hand of a comparison of digital tourism discourses with a sample of different digital discourses and, on the other hand, of a comparison of the discourses of travel agency websites with other digital tourism discourses such as blogs, tourist office websites and tourist attraction websites. For these comparisons, automatic language processing tools were used to extract several linguistic features, complemented by a rhetorical analysis based on the moves structure. It is shown that travel agency discourse constitutes a distinct genre, the specific characteristics of which are described.

INDEX

Mots-clés
genre de discours, discours numérique touristique, structure des tours, traitement automatique du langage

Keywords
discourse genre, tourist digital discourse, move structure, natural language processing

AUTEURS

Trang Tran Hanh Pham
Université catholique de Louvain, Institut Langage et Communication (IL&C), Centre de traitement automatique du langage (CENTAL).
tran.pham@uclouvain.be
IDREF : https://www.idref.fr/292522207
ORCID : http://orcid.org/0009-0009-7366-3554

Rémi Cardon
Université de Gand, Département de traduction, d'interprétation et de communication.
remi.cardon@ugent.be
IDREF : https://www.idref.fr/255501498
ORCID : http://orcid.org/0000-0002-7858-3504

Rodrigo Wilkens
Université d'Exeter, Institut pour la science des données et l'intelligence artificielle.
r.wilkens@exeter.ac.uk
IDREF : https://www.idref.fr/292522657
ORCID : http://orcid.org/0000-0003-4366-1215

Van Công Trân
Université de Hanoï, Département de français.
congtv@hanu.edu.vn

Thomas François
Université catholique de Louvain, Institut Langage et Communication (IL&C), Centre de traitement automatique du langage (CENTAL).
thomas.francois@uclouvain.be
IDREF : https://www.idref.fr/255508506