

# Structuration argumentative des débats parlementaires par IA : vers une modélisation interprétable des dynamiques discursives

Adam Faci 

Huma-Num IR\* (UAR 3598), CNRS, Paris, France

## Abstract

This paper proposes an interpretable computational framework for structuring parliamentary debates through argumentative annotation, semantic clustering, and relational graph modeling. Drawing on open data from the French National Assembly, we first review existing work in argument mining, computational political analysis, stylometry, and document structuring in the context of digital humanities. We then describe a processing pipeline based on large language models for sentence-level argumentative categorization, complemented by sentiment analysis and semantic embeddings. The resulting structured representation enables the construction of an argument-centered knowledge graph, onto which parliamentary actors are projected according to their discursive practices. The analysis highlights both the internal argumentative coherence of individual interventions and the relative rarity of explicit argumentative relations between interventions, reflecting the institutional nature of parliamentary discourse. Beyond these empirical findings, the paper demonstrates how such a structuring approach enhances the interpretability, explorability, and downstream usability of political corpora, particularly in the context of Retrieval-Augmented Generation systems. More broadly, it argues that document structuring should be considered not merely as a technical preprocessing step, but as an interpretative infrastructure for computational hermeneutics in the humanities and social sciences.

**Mots-clés:** grands modèles de langage, débats parlementaire, détection d’arguments

**Keywords:** large language models, parliamentary debates, argument mining

## 1 Introduction

Les débats parlementaires constituent une source centrale pour l’analyse politique, historique et sociologique. Ils permettent d’observer la construction des positions, les stratégies rhétoriques, les rapports de force institutionnels et l’évolution des thématiques dans le temps. Les sciences humaines et sociales (SHS) mobilisent ces corpus pour analyser les formes de délibération démocratique et de conflictualité politique [7 ; 10].

Avec l’essor des humanités numériques, ces corpus ont fait l’objet d’analyses computationnelles variées : lexicométrie, *topic modeling*, analyse de réseaux, détection de polarité ou de position idéologique [5 ; 9]. Toutefois, ces approches reposent encore majoritairement sur des unités textuelles peu structurées et peinent à rendre compte de la logique argumentative interne des débats.

Or, la structuration du discours constitue une condition fondamentale de l’interprétation des textes. Comme le souligne [18], le sens ne réside pas uniquement dans les mots, mais dans l’organisation discursive, argumentative et relationnelle des énoncés. Dans le cas des débats parle-

---

Adam Faci. “Structuration argumentative des débats parlementaires par IA : vers une modélisation interprétable des dynamiques discursives.” *Actes de la Conférence Humanistica*, éd. par Serena Crespi, Simon Gabay, Martin Grandjean, Ariane Pinche, Marie Puren et Léa Saint-Raymond. Vol. 4. Anthology of Computers et the Humanities. 2026, 197–204. <https://doi.org/10.63744/eSaQ67MAMJSL>.

© 2026 par les auteurs. Sous licence Creative Commons Attribution 4.0 International (CC BY 4.0).

mentaires, cette structuration est essentielle pour comprendre non seulement ce qui est dit, mais comment et pourquoi cela est dit.

Nous faisons l'hypothèse que la structuration automatique des débats parlementaires en unités argumentatives, thématiques et relationnelles constitue une étape cruciale pour transformer ces corpus en objets pleinement exploitables par des méthodes computationnelles interprétables. Cette structuration ne vise pas uniquement une amélioration technique du traitement automatique, mais une reconfiguration épistémologique de l'objet *débat parlementaire* comme espace de confrontation de formes argumentatives et de modes d'expression.

Nous proposons ainsi une chaîne de traitement combinant l'usage de grands modèles de langage, des méthodes de *clusterisation* sémantique et une modélisation par graphe afin de représenter les débats non plus seulement comme une succession d'interventions d'acteurs, mais comme une dynamique argumentative structurée.

Une telle approche permet d'envisager une autre manière de raconter les débats parlementaires : non seulement comme l'expression de positions politiques portées par des représentants du peuple, mais également comme une confrontation de styles discursifs, de stratégies argumentatives et de registres rhétoriques propres à des individus ou à des familles politiques. Cette perspective s'inscrit pleinement dans les objectifs des humanités numériques, qui visent à produire des instruments d'analyse permettant d'articuler puissance computationnelle et interprétabilité herméneutique.

## 2 État de l'art

### 2.1 *Argument mining* et structuration argumentative

L'*argument mining* vise à identifier automatiquement les structures argumentatives dans les textes, notamment les prémisses, conclusions et relations entre arguments. Les travaux fondateurs de [17] et [16] ont posé les bases méthodologiques de ce champ, principalement à partir de modèles supervisés.

Cependant, l'application de ces approches aux débats parlementaires reste limitée en raison de la complexité discursive de ces corpus, marqués par l'ironie, l'implicite, les références contextuelles et la polyphonie politique. Les performances des modèles classiques de type BERT restent souvent insuffisantes pour capter les relations argumentatives complexes.

Des travaux récents montrent que les grands modèles de langage *fine-tunés* offrent des performances qualitatives supérieures pour l'*argument mining* relationnel [3; 4; 8]. Ces modèles permettent notamment de détecter des relations d'opposition, de soutien ou de complément avec une finesse accrue, tout en réduisant les besoins en annotation manuelle.

### 2.2 Analyse politique computationnelle

Dans le domaine de l'analyse politique computationnelle, de nombreux travaux se sont concentrés sur la détection de position (*stance detection*), la polarisation et la classification idéologique [5; 9]. D'autres approches ont exploré les réseaux d'acteurs à partir de cooccurrences ou de citations.

Des travaux récents ont mobilisé des méthodes computationnelles pour cartographier l'espace parlementaire français. Boelaert et Ollion [2] par exemple, utilisent des cartes auto-organisatrices afin de représenter la structuration du champ parlementaire et les proximités discursives entre acteurs politiques.

À l'échelle européenne, l'initiative ParlaMint<sup>1</sup> a contribué à structurer de larges corpus parlementaires multilingues afin de faciliter leur analyse computationnelle et comparative.

Cependant, ces travaux modélisent rarement la structure argumentative du discours. Ils analysent principalement les acteurs ou les thèmes, mais peu les logiques argumentatives qui relient

---

1. <https://www.clarin.eu/parlamint>.

les énoncés entre eux.

### 2.3 Stylométrie et auctorialité

La stylométrie a montré depuis longtemps que les caractéristiques linguistiques permettent de caractériser des styles d'écriture spécifiques [13; 14; 20]. Les travaux récents en attribution d'auteur ont confirmé l'efficacité de ces approches [1; 19], notamment par l'utilisation de grands modèles de langage [12].

Dans le cadre parlementaire, l'auctorialité n'est pas un problème d'identification, puisque l'auteur est connu. En revanche, elle devient un outil pour caractériser les manières de s'exprimer et d'argumenter des différents acteurs et groupes politiques. L'analyse argumentative apporte ainsi une dimension complémentaire à la stylométrie en permettant d'intégrer non seulement la forme linguistique, mais aussi la structure discursive.

### 2.4 Graphes de connaissances et humanités numériques

Les graphes de connaissances sont largement mobilisés en humanités numériques pour modéliser des relations complexes entre entités, concepts et textes [11]. Appliqués aux débats parlementaires, ils permettent de dépasser la linéarité textuelle pour représenter des dynamiques relationnelles multidimensionnelles.

## 3 Corpus

Le corpus utilisé provient de la plateforme open data de l'Assemblée nationale française<sup>2</sup>. Ces données sont diffusées sous licence ouverte. Le corpus comprend les transcriptions complètes des séances parlementaires, incluant les identifiants des locuteurs, les dates et les contextes institutionnels.

Les interventions ont été segmentées en phrases, constituant l'unité d'analyse principale. Les analyses traitent par la suite chaque intervention soit phrase par phrase, soit par une prise en compte du contexte de la phrase. Ce choix permet une granularité suffisante pour l'analyse argumentative tout en conservant la cohérence discursive.

## 4 Méthodologie

La typologie retenue s'inspire des schémas classiques de l'*argument mining* distinguant *claim*, *premise* et unités non argumentatives [16; 17]. Afin de simplifier l'annotation automatique et de mieux correspondre aux formes discursives observées dans les débats parlementaires, ces catégories ont été adaptées en quatre classes : affirmation, argument, opposition et contexte.

Le Tableau 1 donne la définition de chaque catégorie ainsi qu'un exemple. Ainsi, toute l'annotation est centrée autour de la catégorie affirmation, les autres étant définies par rapport à cette dernière. La chaîne de traitements qui suit est résumée dans la Figure 1.

Les débats sont d'abord segmentés en interventions correspondant à une prise de parole continue d'un locuteur sur un point de l'ordre du jour. Chaque segment est ensuite soumis à un modèle

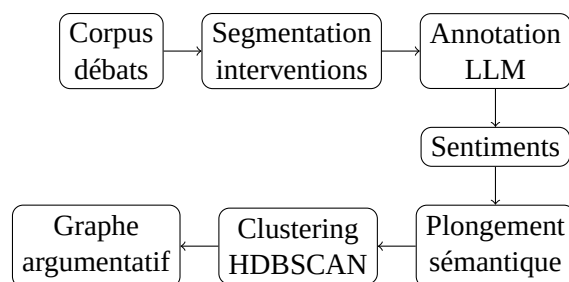


FIGURE 1 – Chaîne de traitements de structuration argumentative.

2. <https://data.assemblee-nationale.fr/travaux-parlementaires/debats>.

Catégorie	Définition	Exemple
Contexte	Une précision ou une information descriptive sans défendre ni contester une affirmation.	La parole est à M. Marcellin Nadeau.
Affirmation	Un fait, une position ou une proposition.	Il est effectivement essentiel de redresser nos comptes publics : je crois que tout le monde s'accorde là-dessus.
Argument	Une raison qui soutient une affirmation, présente ou implicite.	Le risque pour l'emploi, comme en 2012, n'est pas négligeable.
Opposition	Une raison qui conteste une affirmation, présente ou implicite.	C'est à cause de vous!

**TABLEAU 1** – Typologie argumentative utilisée pour l'annotation des phrases.

Llama 3<sup>3</sup> via un prompt contrôlé. Le modèle segmente le texte en phrases et attribue à chacune une étiquette argumentative parmi les quatre catégories définies précédemment. Ce traitement par segment, plutôt que phrase isolée, permet au modèle de prendre en compte le contexte discursif immédiat. Une vérification manuelle rapide sur un échantillon montre une qualité satisfaisante de la classification avec cette stratégie.

Cette approche repose sur une stratégie de *few-shot prompting*, permettant d'obtenir une annotation argumentative de qualité acceptable sans disposer d'un corpus préalablement annoté. Conformément aux résultats de [8], l'usage d'un modèle génératif se révèle qualitativement plus performant qu'un modèle discriminatif de type BERT [6] pour la détection des relations argumentatives, tout en évitant une phase d'annotation supervisée.

Une analyse de sentiment est ensuite appliquée à l'aide d'un modèle RoBERTa<sup>4</sup> afin d'identifier les zones d'intensité émotionnelle. Les phrases sont chacune représentées par des plongement sémantiques.<sup>5</sup>

Une *clusterisation* non supervisée est réalisée sur l'ensemble du corpus à l'aide de HDBSCAN, puis visualisée par réduction de dimension UMAP, afin d'identifier des regroupements thématiques au sein des catégories argumentatives. Chaque cluster est décrit automatiquement en fonction des termes utilisés et des catégories argumentatives reconnues.

Une *clusterisation* non supervisée supplémentaire est appliquée afin de détecter des sous-groupes thématiques au sein de chaque catégorie argumentative. Cette étape révèle une typologie émergente des modes d'argumentation qui sera étudiée prochainement et comparée à l'état de l'art.

Enfin, un graphe de relations argumentatives est construit. Les nœuds représentent des unités argumentatives, i.e. des phrases préalablement classées, et les relations sont établies selon la proximité thématique, la relation argumentative et la polarité émotionnelle. La proximité thématique est établie par l'appartenance à un même cluster thématique. La relation argumentative est établie si une affirmation est suivie par une opposition ou un argument, ou si une opposition est suivie par un argument. La plausibilité est vérifiée par un modèle Llama 3 de chat via *few-shot prompting* avec exemples positifs et négatifs. La polarité émotionnelle est établie si deux phrases sont reliées et une des deux au moins est intense au regard de l'analyse de sentiments. Les acteurs parlementaires sont projetés sur ce graphe sous forme d'hyperarcs et de centroïdes en fonction des

3. <https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct>.

4. <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>.

5. *Embeddings* générés avec *paraphrase-multilingual-MiniLM-L12-v2*

(<https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2>).

arguments mobilisés.

## 5 Résultats et discussion

La première analyse quantitative des annotations produites par le modèle montre une distribution fortement déséquilibrée des catégories argumentatives. Le Tableau 2 donne les proportions d'apparition pour chaque catégorie. Les oppositions et les arguments apparaissent ainsi nettement sous-représentés. Cette distribution suggère que le discours parlementaire est majoritairement structuré autour de la contextualisation et de l'énonciation de positions, plutôt que par une confrontation explicite d'arguments ou d'oppositions formalisées.

Catégorie	Proportion
Contexte	43%
Affirmation	38%
Opposition	10%
Argument	9%

**TABLEAU 2** – Distribution des catégories argumentatives dans le corpus.

La faible proportion d'oppositions semble en partie liée à une proximité sémantique et discursive avec la catégorie des affirmations, rendant parfois la frontière entre ces deux classes difficile à tracer. Cette observation souligne à la fois une limite de la typologie retenue et une piste d'amélioration méthodologique, notamment par l'introduction de sous-catégories ou de relations argumentatives plus fines.

Une évaluation manuelle réalisée sur un échantillon de phrases annotées indique un taux de validité d'environ 80%. Ce résultat confirme la qualité globale de l'annotation automatique, tout en appelant à la prudence. Ce taux doit en effet être interprété à la lumière de la simplicité relative de la typologie utilisée et de la méthode de *few-shot prompting* relativement peu robuste. Une catégorisation plus fine serait plus long à paramétrer, mais offrirait en contrepartie une richesse interprétative accrue.

L'analyse de la progression temporelle des débats ne permet pas de faire émerger de tendances claires entre les interventions successives. L'ordre des prises de parole semble avoir un impact limité sur la structure argumentative des interventions suivantes. Les discours apparaissent ainsi largement indépendants les uns des autres du point de vue de leur construction argumentative.

Cette observation est toutefois nuancée par une analyse qualitative de débats d'origine. Celle-ci montre que les réactions aux interventions précédentes sont souvent intégrées de manière subtile dans le discours suivant, sous forme d'allusions, de reformulations ou de références implicites. Par ailleurs, certaines réactions prennent la forme d'interjections ou de prises de parole brèves, lorsque la personne n'a pas la parole, qui ne sont pas toujours systématiquement retranscrites dans le jeu de données utilisé. Ces éléments expliquent en partie la faible visibilité computationnelle des relations inter-interventions.

L'analyse révèle en revanche une forte structuration argumentative interne à chaque intervention. Les arguments, affirmations et éléments de contexte s'organisent principalement à l'échelle du discours individuel. Les relations entre interventions, et plus particulièrement entre arguments portés par des locuteurs différents, apparaissent plus rares mais également plus significatives lorsqu'elles sont détectées. Ces relations constituent ainsi des points d'ancrage privilégiés pour l'analyse des dynamiques discursives du débat.

Ces résultats peuvent s'expliquer par la nature même du discours parlementaire. Les interventions sont en grande partie préparées en amont, avec parfois des réponses anticipées, mais surtout avec un souci marqué de cohérence interne. Le discours est ainsi conçu comme un objet autonome, davantage que comme une réplique immédiate à une intervention précédente. Cette caractéristique institutionnelle limite mécaniquement la visibilité des chaînes argumentatives inter-discursives.

Ces observations invitent à relativiser l'idée d'un débat parlementaire comme interaction argumentative continue, et à le considérer plutôt comme une juxtaposition de prises de position discursivement cohérentes, reliées par des points de contact ponctuels mais structurants.

Enfin, ces résultats ouvrent des perspectives méthodologiques importantes. Il serait particulièrement pertinent de tester cette approche dans des contextes discursifs moins structurés, tels que les débats en ligne, les forums politiques ou les échanges sur les réseaux sociaux. Dans ces espaces, les réactions immédiates, les confrontations explicites et les chaînes argumentatives inter-interventions sont plus fréquentes, ce qui permettrait d'évaluer plus finement la capacité de la méthode à détecter et modéliser les relations argumentatives dynamiques.

Ainsi, les résultats obtenus ne constituent pas seulement une analyse du débat parlementaire, mais également un révélateur des propriétés discursives propres à ce type d'institution, tout en soulignant l'intérêt d'étendre la méthodologie à d'autres formes d'expression politique.

Les clusters montrent une cohérence thématique significative, confirmant la pertinence de la segmentation argumentative. Certains thèmes sont principalement associés à des oppositions, tandis que d'autres sont dominés par des arguments justificatifs ou contextuels.

La projection des acteurs révèle des proximités discursives parfois indépendantes des appartenances partisans, suggérant l'existence de styles argumentatifs transversaux. Inversement, des divergences apparaissent au sein de mêmes groupes politiques.

Cette modélisation permet ainsi de lire les débats comme une confrontation de modes d'expression et de stratégies argumentatives, et non uniquement comme une confrontation de positions politiques. Elle enrichit la compréhension des dynamiques parlementaires en révélant des dimensions discursives souvent invisibles dans les analyses traditionnelles.

## 6 Perspectives

La détection automatique des arguments et de leurs relations demeure, dans le contexte actuel, un champ scientifique riche et encore largement ouvert. Malgré les avancées récentes, la modélisation fine des structures argumentatives, en particulier dans des corpus complexes et fortement contextualisés comme les débats parlementaires, soulève encore de nombreuses questions théoriques et méthodologiques. Les résultats obtenus dans ce travail, tout comme ceux rapportés dans la littérature récente [3 ; 8], montrent que la qualité des annotations et des relations détectées reste perfectible, mais suffisamment robuste pour ouvrir de nouvelles pistes d'exploration.

Dans ce contexte, les grands modèles de langage apparaissent comme une méthode particulièrement efficace pour l'*argument mining*, en raison de leur capacité à produire des annotations qualitatives sans nécessiter de corpus annotés, ou seulement à partir de quelques exemples. Cette propriété constitue un atout majeur pour les humanités numériques, où les ressources annotées sont souvent rares. Néanmoins, la robustesse, la stabilité et la reproductibilité de ces modèles doivent encore être évaluées de manière systématique, notamment face aux variations de corpus, de domaines et de formulations discursives.

Au-delà de la seule détection argumentative, l'enjeu principal réside dans l'articulation entre ces avancées méthodologiques et les usages concrets des documents structurés. Les systèmes de *Retrieval-Augmented Generation* [15], largement mobilisés aujourd'hui, ne sont pas encore considérés comme des outils fiables d'analyse scientifique, mais plutôt comme des interfaces exploratoires permettant d'interagir en langage naturel avec des corpus documentaires. Dans ce cadre, la structuration argumentative et relationnelle des documents apparaît comme un levier essentiel pour enrichir les possibilités exploratoires offertes par ces systèmes.

Une telle structuration permet en effet de multiplier les niveaux de lecture : par thématique, par type d'argument, par relation argumentative, par tonalité émotionnelle ou encore par style discursif. Elle autorise des parcours guidés dans les corpus, favorisant une exploration raisonnée plutôt qu'une simple récupération textuelle. De plus, une recherche documentaire tenant compte de cette structuration améliore en aval la qualité des générations produites par les systèmes de RAG, en les rendant plus cohérentes, plus contextualisées et, surtout, plus explicables.

Ces perspectives invitent ainsi à concevoir la structuration des débats parlementaires non comme une fin en soi, mais comme une infrastructure interprétative au service de nouvelles formes d'exploration, de discussion et d'analyse des corpus politiques. Elles ouvrent la voie à une articulation renouvelée entre intelligence artificielle, structuration documentaire et herméneutique computationnelle, pleinement inscrite dans les objectifs des humanités numériques.

## Références

- [1] BJÖRKLUND, Johanna et ZECHNER, Niklas. « Syntactic Methods for Topic-Independent Authorship Attribution ». In : *Natural Language Engineering* 23, no. 5 (2017), p. 789-806. DOI : 10.1017/S1351324917000249.
- [2] BOELAERT, Julien et OLLION, Étienne. « Les sommets du Palais : Analyser l'espace parlementaire avec des cartes auto-organisatrices ». In : *Revue française de science politique* 70, no. 3 (2020), p. 373-398. DOI : 10.3917/rfsp.703.0373.
- [3] CABESSA, Jérémie, HERNAULT, Hugo et MUSHTAQ, Umer. « Argument Mining with Fine-Tuned Large Language Models ». In : *Proceedings of the 31st International Conference on Computational Linguistics*, sous la dir. d'Owen RAMBOW, Leo WANNER, Marianna APIDIANAKI, Hend AL-KHALIFA, Barbara DI EUGENIO et Steven SCHOCKAERT. Abu Dhabi, Émirats arabes unis : Association for Computational Linguistics, 2025, p. 6624-6635. URL : <https://aclanthology.org/2025.coling-main.442>.
- [4] CHEN, Guizhen, CHENG, Liying, LUU, Anh Tuan et BING, Lidong. « Exploring the Potential of Large Language Models in Computational Argumentation ». In : *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, sous la dir. de Lun-Wei KU, Andre MARTINS et Vivek SRIKUMAR. Bangkok, Thaïlande : Association for Computational Linguistics, 2024, p. 2309-2330. DOI : 10.18653/v1/2024.acl-long.126.
- [5] CONOVER, Michael, RATKIEWICZ, Jacob, FRANCISCO, Matthew, GONCALVES, Bruno, MENCZER, Filippo et FLAMMINI, Alessandro. « Political polarization on Twitter ». In : *Proceedings of the International AAAI Conference on Web and Social Media*. Barcelone, Espagne, 2011, p. 18-25. DOI : 10.1609/icwsm.v5i1.14126.
- [6] DEVLIN, Jacob, CHANG, Ming-Wei, LEE, Kenton et TOUTANOVA, Kristina. « BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding ». In : *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies (NAACL-HLT 2019)*, sous la dir. de Jill BURSTEIN, Christy DORAN et Tamar SOLORIO. Minneapolis, États-Unis : Association for Computational Linguistics, 2019, p. 4171-4186. DOI : 10.18653/v1/N19-1423.
- [7] FINLAYSON, Alan. « From Beliefs to Arguments : Interpretive Methodology and Rhetorical Political Analysis ». In : *The British Journal of Politics and International Relations* 9, no. 4 (2007), p. 545-563. DOI : 10.1111/j.1467-856X.2007.00269.x.
- [8] GORUR, Deniz, RAGO, Antonio et TONI, Francesca. « Can Large Language Models Perform Relation-based Argument Mining ? ». In : *Proceedings of the 31st International Conference on Computational Linguistics*, sous la dir. d'Owen RAMBOW, Leo WANNER, Marianna APIDIANAKI, Hend AL-KHALIFA, Barbara DI EUGENIO et Steven SCHOCKAERT. Abu Dhabi, Émirats arabes unis : Association for Computational Linguistics, 2025, p. 8518-8534. URL : <https://aclanthology.org/2025.coling-main.569>.
- [9] GRIMMER, Justin et STEWART, Brandon M. « Text as Data : The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts ». In : *Political Analysis* 21, no. 3 (2013), p. 267-97. DOI : 10.1093/pan/mps028.

- [10] HABERMAS, Jürgen. *De l'éthique de la discussion*. Paris : Éditions du Cerf, 1992.
- [11] HASLHOFER, Bernhard, ISAAC, Antoine et SIMON, Rainer. « Knowledge Graphs in the Libraries and Digital Humanities Domain ». 2018. DOI : 10.48550/arXiv.1803.03198.
- [12] HUANG, Baixiang, CHEN, Canyu et SHU, Kai. « Authorship Attribution in the Era of LLMs : Problems, Methodologies, and Challenges ». In : *SIGKDD Explorations Newsletter* 26, no. 2 (2025), p. 21-43. DOI : 10.1145/3715073.3715076.
- [13] JUOLA, Patrick. « Authorship attribution ». In : *Foundations and Trends in Information Retrieval* 1, no. 3 (2008), p. 233-334. DOI : 10.1561/1500000005.
- [14] KESTEMONT, Mike. « Function Words in Authorship Attribution : From Black Magic to Theory? » In : *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, sous la dir. d'Anna FELDMAN, Anna KAZANTSEVA et Stan SZPAKOWICZ. Göteborg, Suède : Association for Computational Linguistics, 2014, p. 59-66. DOI : 10.3115/v1/W14-0908.
- [15] LEWIS, Patrick et al. « Retrieval-augmented generation for knowledge-intensive NLP tasks ». In : *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Vancouver, Canada : Curran Associates Inc., 2020.
- [16] LIPPI, Marco et TORRONI, Paolo. « Argumentation Mining : State of the Art and Emerging Trends ». In : *ACM Transactions on Internet Technology* 16, no. 2 (2016). DOI : 10.1145/2850417.
- [17] PELDSZUS, Andreas et STEDE, Manfred. « Joint Prediction in MST-Style Discourse Parsing for Argumentation Mining ». In : *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, sous la dir. de Lluís MÀRQUEZ, Chris CALLISON-BURCH et Jian SU. Lisbonne, Portugal : Association for Computational Linguistics, sept. 2015, p. 938-948. DOI : 10.18653/v1/D15-1110.
- [18] RASTIER, François. *Arts et sciences du texte*. Formes sémiotiques. Paris : Presses Universitaires de France, 2001. DOI : 10.3917/puf.rast.2001.01.
- [19] SEGARRA, Santiago, EISEN, Mark et RIBEIRO, Alejandro. « Authorship Attribution Through Function Word Adjacency Networks ». In : *IEEE Transactions on Signal Processing* 63, no. 20 (2015), p. 5464-5478. DOI : 10.1109/TSP.2015.2451111.
- [20] STAMATATOS, Efstathios. « A Survey of Modern Authorship Attribution Methods ». In : *Journal of the American Society for Information Science and Technology* 60, no. 3 (2009), p. 538-556. DOI : 10.1002/asi.21001.