



Legislating to Control Online Hate Speech: A Corpus-Assisted Semantic Analysis of French Parliamentary Debates

Nadia Makouar^{1,2} · Lauren Devine¹ · Stephen Parker¹

Accepted: 12 May 2023
© The Author(s) 2023

Abstract

This corpus analysis of linguistic and semantic features in French parliamentary debates concerning online hate speech regulation, highlights tensions between state powers and private rights. Two key themes are identified: first, the *problem of definition*: how such online content is defined in the debates, and second, the *problem of regulation*: how the debates negotiate the supra-jurisdictional and individual jurisdiction issues involved, in regulating both the global online content and the responsibilities of the owners of the platforms who manage the content. For this analysis, the authors created a corpus of French National Assembly Examination Committee discussions and public sessions between July 2019 to January 2020 discussing the *Loi Avia* (Avia Law). The corpus was analysed using quali-quantitative methods. Overall, the study combines a top-down (global) and a bottom-up (local) analysis of the data, applying corpus semantic approaches (Lecolle et al. in *Langages* 210(2):35–54, 2018; Rastier in *The Routledge handbook of semantics*, Routledge, pp 507–522, 2015) to contextualize the analysis of the meta-discourse (Hyland in *Meta-discourse: exploring interaction in writing*, 1st ed. Bloomsbury Academic, Bloomsbury Collections, 2018. <https://doi.org/10.5040/9781350063617>). The analysis highlights the linguistic features revealing the power dynamics inherent within the discourse and counter-discourse of the dominant and opposition political parties respectively.

Keywords Online hate speech (cyberhate) · Parliamentary debate · Regulation · Corpus linguistics · Semantics

French to English translations in this Article are by Makouar.

✉ Lauren Devine
l.devine@aston.ac.uk

¹ Centre for Language and Law, Aston Institute for Forensic Linguistics, Aston University, Birmingham, UK

² Université de Montpellier 3 Paul Valéry, Montpellier, France

1 Introduction

This article presents a corpus-assisted, semantic analysis of discourse in a sample of French parliamentary debates. The debates concern regulation of two relatively recent phenomena; online hate speech (*haine en ligne*) and the related question of online fake news (*fausses nouvelles*). The focus of this article is online hate speech, further analysis will examine the phenomenon of online fake news.

In addition to the social need to control online hate speech, governments are also interested in its control where hate speech occurs in conjunction with fake news. Together these phenomena threaten the integrity of elections as they aim to influence voters. As such, the phenomena represent both a societal and a political issue. Concern for elections to be protected from interference is reflected in European Court of Human Rights jurisprudence which holds that “elections must be protected as they are a characteristic principle of democracy” [4, 5]

Legislating to control the phenomena is problematic. The production and dissemination of online hate speech and fake news can arguably be protected by the principles of free speech contained in Article 10 of the *Convention for the Protection of Human Rights and Fundamental Freedoms* [6, p.12]. Consequently, discussion in parliamentary debates undertaken by European states about the legality of restriction must be balanced against this principle. These inter-connecting legal, social, and linguistic strands complicate the discussions, both in respect of defining the parameters of any proposed legislation (establishing what falls within a legislated definition, and what falls outside it), and the need to balance the right to protect democratic elections with the right of individual citizens for their freedom of expressions to be protected. This tension creates a dichotomy between two foundational legal principles, creating a significant complexity within the debates; that of establishing an inclusive (what is controlled) and exclusive (what is free speech, and thus uncontrolled) definition. The European Convention on Human Rights does not recognise the term *hate speech*, but the European Court of Human Rights recognizes it as encompassing “all forms of expression which spread, incite, promote or justify hatred based on intolerance.” [7, 8] *problem of definition*, seen within the debates as illustrating a *grey zone* (*zone gris*).

The focus of this investigation concerns regulation of *online* content as a specific issue, rather than hate speech as a wider phenomenon. In discussions concerning such regulation, complex issues arise in relation to establishing and enforcing supra-jurisdictional control of the online content. This is a difficult legal area, noted to be an unsettled matter as it concerns multiple jurisdictions, causing a “human rights nightmare” [9, p. 375]. The online feature of these phenomena can thus be singled out as creating a second significant complexity; the need for inter-jurisdictional co-operation and common aim in creating a regulatory framework. This second complexity can be expressed as a *problem of regulation*.

The objective of our analysis is to offer new insights into how both the *problem of definition* and the *problem of regulation* are discussed during the debates by identifying linguistic features. Using corpus-assisted methods, we identify the main themes within the debates, examine the discursive practices of the

individual political enunciators in the context of their political affiliations, and contextualise the discourse within the meta-discursive activity of the overall debates. These methods identify individual political party affiliations, and collective parliamentary contributions. In addition to revealing the linguistic features of the debates, this dual approach illuminates features of the entextualisation process of parliamentarians in converting social issues into legislation via political debate. Entextualisation refers to “the process by which texts are produced by extracting discourse from its original context [and] is a fundamental process of power and authority” [10, p.486].

France has been chosen for this analysis because in May 2020 it made a strenuous but ultimately unsuccessful attempt to legislate by enacting the *Loi Avia* and *Loi No. 2018–1202* [11]. In so doing, it claimed to have resolved both the *problem of definition* and the *problem of regulation* despite substantive criticism from the French opposition parties, the European Commission, the Czech Republic, digital and anti-racist organisations. As soon as it was adopted, the French Senate brought the law before the French Constitutional Council, that declared the main provisions of the *Loi Avia* unconstitutional on 18th June 2020. The Council considered that the *Loi Avia* placed undue restrictions on “the exercise of freedom of expression and communication which is not necessary, appropriate and proportionate” [12, paras. 8 and 19].

The drive behind the *Avia Bill* was the criticism of existing laws, and global acknowledgement that online hate speech is a serious problem. Arsène and Mabi [13] raise questions about the applicability of existing laws, concerns shared by the Council of State [14], the National Consultative Commission on Human Rights [15], and the European Commission [16]. This criticism is not restricted to France: Bakalis for example suggests that the UK’s “current legislation on cyberhate is inadequate. It does not reflect the real nature of the harm caused...” [17, p. 109]. We therefore recognize and frame the problem in the context of an unresolved global issue with no model solution and consider it likely that elements of the debates we analyse here will mirror debates in other jurisdictions. Ultimately, such debates concern the vulnerability of States to the destabilization of democracy through insidious infiltration of individual voters’ online news exposure. Such infiltration is designed to operate at an a-rational, subliminal level, influencing voters’ ability to make informed and rational choices, specifically targeting Western, democratic nations.

Debates in the UK Parliament, for example, exemplify the complexity involved in finding a solution. The UK’s government conducted an eighteen-month parliamentary investigation into “disinformation and ‘fake news’” which resulted in two reports [18, 19]. Although the UK’s Hansard debates are outside the scope of this specific study, the approach and methods used for this analysis may be useful in grounding future research into debates of other jurisdictions where there is sufficient published parliamentary material on the topics for corpus-assisted analysis.

In addition to its strenuous attempts to legislation, France was also selected for this study as it is a member of the European Union. Thus, French internal and external sovereignty aligns with the notion of an *international* rule of law. This differs from the national context:

In the context of national law, the rule of law means that citizens are not there to serve the state, but the state is there to serve its citizens. In the context of international law, the rule of law means that states serve as the trustees of their citizens, bound to the rule of international law not for the sake of their own sovereignty, but for the sake of the people whose well-being they are entrusted with. Because this fiduciary position of states depends on the international legal order, to some extent they are also officials of the international legal order. Ultimately, this may entail a responsibility of states for subjects of other states [20, p. 94].

Inherent within the parliamentary debates, therefore, is an acknowledgement of the complexities of the relationship between national constitutions, the binding force of international treaties, the parameters of consent and custom in the force of international legislation, and the effect of fundamental principles acknowledged by all states. As France is a member of the European Union, the debates must acknowledge the supranational EU jurisdiction, including the distribution of sovereignty between the member states and the Union including the legislative instruments, regulations, and directives [20, Ch. 4]. As such, there must be a central, international alignment of the language and meaning of it in any regulatory framework.

2 Context and Conceptual Framework

2.1 The Politico-Legal Context of the Debates

Online hate speech and fake news is read by citizens within specific physical jurisdictions (in our example, France), accessed via online social media platforms which can be hosted anywhere in the world. Thus, controlling such speech and news is the problem of both individual States [17], and the digital platforms hosting the hate speech [5, 21]. This fast-moving technico-legal phenomenon, coupled with the complexity of the supranational laws required to address it, has led Parliaments worldwide to debate whether their territorial laws are sufficient [22] and whether these laws sufficiently balance the right to free speech [23, 24]. As this area is under-researched and under-regulated, the ability of online platforms to disseminate politically de-stabilising content with intent to manipulate elections is serious. It has become a global issue of concern, transcending to a supranational question of how such online “soft invasions” of government can, and should be, regulated and prevented [25].

Despite the commonalities across jurisdictions, there will be differences in the nature of the debates, attributable to the individual or supranational jurisdictional structures in place. France is part of the European Union and thus its laws are not entirely separate from those of other EU member states. French law raises important structural issues for the European Commission and its institutional and associative actors in respect of the regulation of online hate speech and fake news [26, 27]. This includes the complexity of how, or whether, to transfer arbitration and moderation of the content to the online platforms, with the associated risks of de-judicialization,

the risk of unnecessary censorship of free speech, and the consequential risk of over-regulation of the freedom of expression; effectively outsourcing one of the State's primary responsibilities: to defend its democratic security.

The choice of French debates as the focus for this analysis was to acknowledge the attempts that the French Parliament have made to legislate against the problem. The *Avia Bill* against online hate speech was adopted in May 2020 [28]. The *Loi Avia* and *Loi No. 2018–1202* [11] were introduced in the context of discussion around the issue of the responsibility of online platforms in several European countries. *Inter alia*, it required online platforms to remove terrorist or child pornographic content within one hour and remove "manifestly illegal" hateful content within 24 h. However, the legislation raised several problems: firstly, the speed of action required may be unfeasibly swift, for example, the requirement for referral to the judge within one hour to qualify the facts. Secondly, as we note in the introduction, the French Constitutional Council censured the law on the grounds that it constitutes a disproportionate attack on freedom of expression. This reaction exemplifies the complex tensions between freedom and restriction of speech regardless of its online hosting, and suggests the French parliamentary debates provide an ideal source of material for analysis if only to illuminate the shortcomings in the debates. Such analysis can be used to inform future debates, highlighting where there was insufficient consideration of the issues which led to the censure, and where there was over-emphasis which led to the French Constitutional Council's decision that the *Loi Avia* did not strike the right balance between censorship and freedom of speech. Consequently, following its censure, only a few provisions of the *Loi Avia* remain. This legislative episode provides an example of the complexity, not least as the parliamentary and state actors have potentially competing interests including commercial, legal, administrative, and political.

The *Loi Avia* failed to strike an appropriate balance between freedom and restriction and did not address the problem of controlling marginal cases. There are well documented concerns about the precise boundaries of the hate speech regulation, whether online or in other forums [30]. Some instances of hate speech may be relatively uncontroversial, others may fall into a *grey zone*, exemplifying the question of the appropriate balance between state powers and private rights [29]. This is both a philosophical and practical question, creating tension between fundamental powers and freedoms. The lack of firm definition makes it difficult to establish where, exactly, the boundaries should be set, exemplifying the tensions between online content regulation (state power) and the right to free speech (private rights). That which is "manifestly illegal" is uncontroversial, and that which is not so easily categorized may interfere with the boundaries of free speech.

The extent of the state's control is exhibited in the language of the law. State power contained in written regulations and legislation, the interpretation of which becomes the responsibility of law enforcement agencies and ultimately the courts. The question of what falls 'inside' and 'outside' the locus of control forms a liminal space, or *grey zone*. Such space is often left deliberately vague in the language of the regulatory provisions so as not to curtail judicial flexibility in how provisions are interpreted. Thus, over time, broadening and contracting of the boundaries of legislation is possible and the language used in the regulatory and legislative provisions

is revealed to be both certain and uncertain. This duality can be used to either strengthen or weaken rights of both state and citizens depending up the interpretation of matters within the liminal space of the *grey zone*. The European Committee of Ministers exemplify this point, noting that member states should ensure that:

A range of properly calibrated measures is in place to effectively prevent and combat hate speech. Such a comprehensive approach should be fully aligned with the European Convention on Human Rights and the relevant case law of the European Court of Human Rights (the Court) and should differentiate between:

- a. hate speech that is prohibited under criminal law; and
- b. hate speech that does not attain the level of severity required for criminal liability, but is nevertheless subject to civil or administrative law; and
- c. offensive or harmful types of expression which are not sufficiently severe to be legitimately restricted under the European Convention on Human Rights, but nevertheless call for alternative responses, as set out below, such as: counter-speech and other countermeasures; measures fostering intercultural dialogue and understanding, including via the media and social media; and relevant educational, information-sharing and awareness-raising activities. [30, p. 4]

Interpreting these criteria is complex: online discourse falling within each criterion can only be established following reference to multiple legislative materials and identifying legal precedents. Even when these tasks are completed, the *grey zone* exists, particularly in relation to the types of expression described in category c.

2.1.1 Parliamentary Discourse: The Liminal Space Between a Societal Issue and the Law

Parliamentary debates on any topic occupy a unique position in socio-politico-legal phenomena representing a liminal, discursive space between a social issue and its trajectory into legislative control. In addition to the issue itself, the debates are also inevitably mediated by political allegiances and the associated, inherent power dynamics between the party in power, and the opposition parties. Galembert [31] describes parliamentary discourse as a unique composite of parliamentary and legislative discourses, where the discourses of legal issues and political sociology can be observed. We widen this definition to include discourse operating within the *intersections* between the society, politics, and the law, and as such create a currently under-observed and liminal space with unique characteristics of its own. As such, we posit the debates offer opportunities for deep analysis. They constitute a unique opportunity to explore language features, to observe how the discourse is controlled at an individual and collective level to identify the process of normative law making, and its reception and interpretation within the society which it serves. We observe the trajectory within the debates of social issue, translated into legislation shaped through politico-legal debate and discussion. The social issues thus form both the

foundation, and the final arbiter of acceptability of the subject matter during its process of reformulation from issue into regulatory framework.

Moor [32] observes an important distinction in relation to the actors involved, first by constructing the law, and second in interpreting the law. Judges and legislators constitute two distinct legal figures whose roles and functions differ. The judge is an arbiter of fact and is also responsible for interpreting the law, responding to social situations in individual cases. The role of the legislator is to construct the written texts, responding to the social need represented in the Parliamentary debates. However, their role is intrinsically linked by normative constructions in the interpretations, and thus the processes of constructing law and interpreting the law are conceptually and practically inter-related. Moor's point is fundamentally that one cannot exist effectively without the other, and the development of the law is dependent upon both actors.

Only the (general) norms adopted by the legislator and the (particular) norms adopted by the judge are legal — legislator and judge obviously being legal institutions, more precisely *legal figures*. [32, p. 77]

Indeed, this positioning of the separate roles of the judge and the legislator also highlights the idea of a discursive movement between social issues and their echo in firstly the parliamentary debate, and then a subsequent legislative framework. As Moor [33] notes:

There is a constant flow of information between the social system and the legal system, which passes through what we call legal figures. These - the legislator, the judge or the subject of law - are in a position to choose in the social context what they decide to take, to give it the status of a norm. [33, p. 783]

These insights point to the dialectic manifesting in the debates, where the truth of the opinions put forward by the parliamentarians is discussed. The debates take place in what can be described as the “parliamentary sphere” [34]. We consider this an example of liminality within which the intersectional, discursive practices mediating the translation of societal issues into law take place. The legislator thus has the role of transmitter of both information and claims across the two discourse genres: societal discourse, and legal discourse. Similarly, Karsenti posits that the legislator is a “discursive agent” who.

... conveys a type of discourse which precedes and conditions the fact that the people give themselves their law, that the general will, not wanting, but see, understand what it is tending towards, and aim for it willingly. [35, p. 319]

This foregrounds our preposition that the process of dialectic reasoning in the debates, followed by its transformation into legislation is characterized by a particular type of discourse within this very specific liminal space. This discourse not only represents the translation and integration of societal norms as it develops into legislation but also reveals the power-dynamics inherent within this process. This is a second *grey zone*, not specifically referred to in the debates, where the precise boundaries of socio-politico-legal discourse overlap, intersect and

ultimately position themselves through carefully mediated discourses of political power. Galembert [31] observes that this constitutes a dynamic of “legal transubstantiation.” Chevallier [36] sees the discourse as both legal and political, emphasising that legal and political discourses are linked, and communicate through this medium of parliamentary debates. The discursive development of societal issues into legal norms can be understood as a political product controlled by the power-dynamics observable within the debates.

3 Methods

Our methods include both qualitative and quantitative analysis of the debates, enabling identification and insights into both the individual level discursive, and the meta-discursive processes [3]. To conduct the analysis, we conducted a two-pronged quantitative exploration of the corpus. The quantitative analysis adopted a “global to local approach” [37]. The first prong of this analysis, the global approach, identified the major themes and characteristics within the debates. The second prong, the local approach, considered the stance of the individual enunciators, including their semantic, syntactic, and pragmatic features. To deepen the analysis, we conducted a qualitative, discursive semantic analysis. This semantic analysis acknowledges the multidimensionality and the dynamic nature of meaning, [1] highlighting the characteristics and semantic determinations in the debates.

3.1 Corpus Description and Software Used for the Analysis

A specialized corpus of French parliamentary texts was created for this study, consisting of seven extracts from the French National Assembly Examination Committee discussions, and the subsequent Public Sessions held during the first and second reading of the *Avia Bill*. These seven extracts were separated into seven sub-corpora. The debates took place between July 2019 to January 2020. The whole corpus contained 272,192 running words constructed from texts transcribed and made available on the National Assembly website.¹ We selected only debates that were held in the National Assembly during the first and the second reading. Figure 1 summarises the chronology of the legislative process:

The corpus contains seven sub-corpora, annotated differently according to the protocols required by the analytic tools. Two statistical software packages were used for the analyses: First, IRaMuTeQ [40] was used to produce the hierarchical clustering classifications (the Reinert method, explained). Second, Lexico 3 [41] was used to calculate the lexical specificities² of the sub-corpora from n-grams. For the IRaMuTeQ analysis, the texts were annotated by date, and for the Lexico analysis,

¹ Legislative dossier available at: [38].

² To obtain the specificities scores, the software conducts a hypergeometric test. This test determines whether a given word or n-grams occurs in a sub-corpus at a frequency more than would be expected by chance.

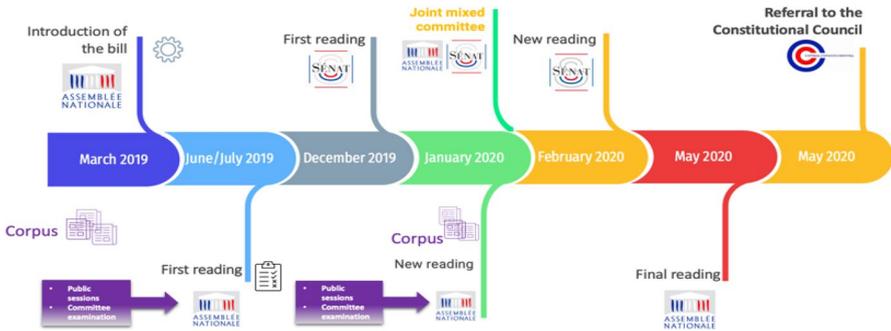


Fig. 1 French legislative process (Avia Bill)

the texts were annotated by date and political party. The latter annotation facilitated the identification of the linguistic features of the speech segments produced by the parliamentary majority, and those of the opposition parties.

All texts were extracted from the National Assembly website and include three texts discussed in the closed French National Assembly Examination Committee hearings (on 05/06/2019, 19/06/2019, and 14/01/2020) during which they discussed the *Avia Bill* before the debates in the public sessions:

As soon as they are submitted, the Bills are printed and distributed to all the MPs. Unless a special committee is created, the Bill is sent to one of the standing committees for evaluation; other interested committees may also examine the text. [39]

The sub-corpora also contained four debates in the public discussions (on 03/07/2019, 04/07/2019, 21/01/2020, and 22/01/2020).

A public discussion is held once the text has been placed on the agenda. This begins as a general discussion, with several participants: a member of the Government, the person who followed the Bill in the committee (the rapporteur) along with others consulted for information, as well as the MPs who, either in the name of their group or as individuals would like to indicate their point of view. The Assembly examines the articles one by one, along with any amendments that may be attached to each. [39]

We chose to keep the reactions of parliamentarians (e.g., “applause coming from all side”) as well as the elements that indicate the legislative process (e.g., “rejection of the amendment”).

3.2 The Global Approach: Identifying the Themes of the Debates

To identify the main themes of the debates we conducted a hierarchical cluster analysis, generating a dendrogram. These clusters highlight sets of identifiable meanings

that structure the representations in the corpus [43]. There are many different types of cluster analysis [44], but for corpus-assisted analysis a hierarchical agglomerative cluster analysis is typically used [45, p. 336ff], revealing the major topics in the whole corpus. The importance of identifying such themes as a discourse category is based on Van Dijk's consideration for semantic study of text. According to Van Dijk topics are:

[d]efined as semantic macrostructures and [...] represent what speakers find most important, they regulate overall coherence of discourse, how discourse is planned and globally controlled and understood, and what is best remembered by the recipients. [42, p. 234]

This method reveals themes that tend to be found in similar contexts [46]. Our corpus was lemmatized only for this clustering procedure step. Each text segment contained 40 occurrences (forms) which were clustered according to their vocabularies and distributed according to their reduced frequencies. This grouped the forms found in similar contexts [46] before they were divided into clusters, dichotomously at each step. The forms most likely to be grouped together in a cluster occur where there is a correspondingly high number appearing in more than one segment. Thus, where the number of forms common to more than one segment is high, the closer these two segments will be, and the likelihood of them being grouped together in a same cluster is higher.

3.3 The Local Approach: Collocates, n-Grams, Frequencies, and Semantic Features of the Debates

We then used the seven sub-corpora to determine the points of convergence and divergence between the enunciators within the debates. The analysis revealed the political affiliations of the individual enunciators, demonstrated in their attempts to determine what is, or is not, online hate speech, and how it should be controlled.

During this analysis, the semantic, syntactic, and pragmatic features of the discourse were revealed by identifying the collocates, n-grams, and frequencies within the debates. This second part of the two-pronged quantitative methods, combined with qualitative discursive semantic analysis, clarified the process of "meaning-making" as well as highlighting the dominant discourse and associated power dynamics of the political parties in the debates.

4 Findings

4.1 The Global Approach: Identifying Key Themes in the Debates

In this analysis, IRaMuTeQ split the whole corpus into four clusters, revealing the most salient themes within the debates. Two key themes emerged: the *problem of definition* and the *problem of regulation*. The *problem of regulation* received

Table 1 Summary of themes and sub-themes within the clusters

Cluster	Level of detail	Theme	Sub-theme
Cluster 1	Broad discussion	Problem of definition	Balance between online free speech and its control
Cluster 2	Detailed discussion	Problem of regulation	The specific French process of implementing new legislation
Cluster 3	Detailed discussion	Problem of regulation	The supra-jurisdictional issues of implementation
Cluster 4	Broad discussion	Problem of regulation	The power dynamics of legislation to restrict free speech

significantly more attention than the *problem of definition*. A brief categorisation of the clusters is provided below in Table 1.

The descending classifications shown in the dendrogram in Fig. 2 lists the most frequent and closest (based on Chi² scores) words within each cluster. On the left side of Fig. 2, cluster 3 (containing 26.4% of the words) contained discourse representing a detailed discussion of the problem of regulation. The sub-theme in this detailed discussion concerned the issues of *implementing* the legislation, rather than the specifics of the *content* of the legislation. This cluster contains a wide lexicon of discourse about the various actors involved in implementing legislation, including discussion of the online platforms, the judges, the operators, and the authorities. This cluster mentions the categorization of online content, such as *illicite* (illicit) and actions within the legislation, *signalement* (reporting), and *retrait* (retraction).

Appearing next to cluster 3, cluster 2 (containing 13.9% of the words) also concerns the problem of regulation. This cluster contains words relating to the detail of the *Avia Bill*, and its judicial and law enforcement implications. This cluster contains words such as *article* (article), *penal* (penal), *disposition* (disposal), *code* (code), and *infraction* (infringement).

At the right side of Fig. 2, cluster 1 (containing 32,5% of the words) is the only cluster concerning the problem of definition. It contains discussion of the online platforms and digital spaces hosting online content. Words such as *internet*, *réseau social* (social network), *Facebook*; *espace* (space) and phrases related to freedom and democracy are within this cluster: *haine* (hate), *liberté d'expression* (freedom of speech), *opinion* (opinion), *démocratie* (democracy).

Cluster 4 (containing 27.2% of the words) contains the more theoretical aspects of the debates, most closely linked to the problem of definition. It is more isolated than the other three clusters, evident in the factor analysis data visualisation in Fig. 3. It contains words relating to of the power dynamics of parliamentary regulation, and the more discursive elements of the debates: *amendement* (amendment), *avis* (opinion), *commission* (commission), *adopter* (adopt), *intervention* (intervention).

A factor analysis, illustrated in Fig. 3, shows the distribution and the relative proximities between the clusters. This analysis shows that clusters 1 and 2

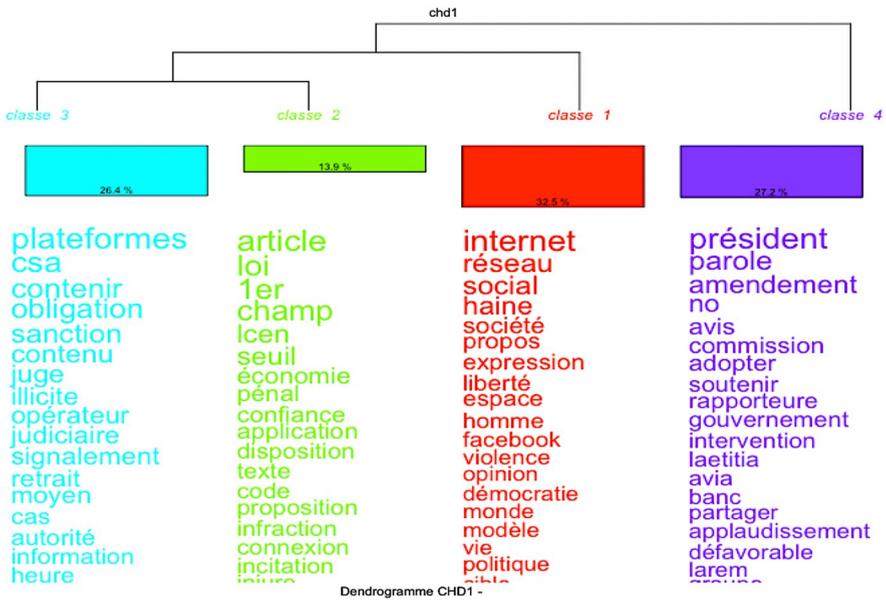


Fig. 2 Dendrogram showing the Hierarchical Clustering Analysis: Descending Classification

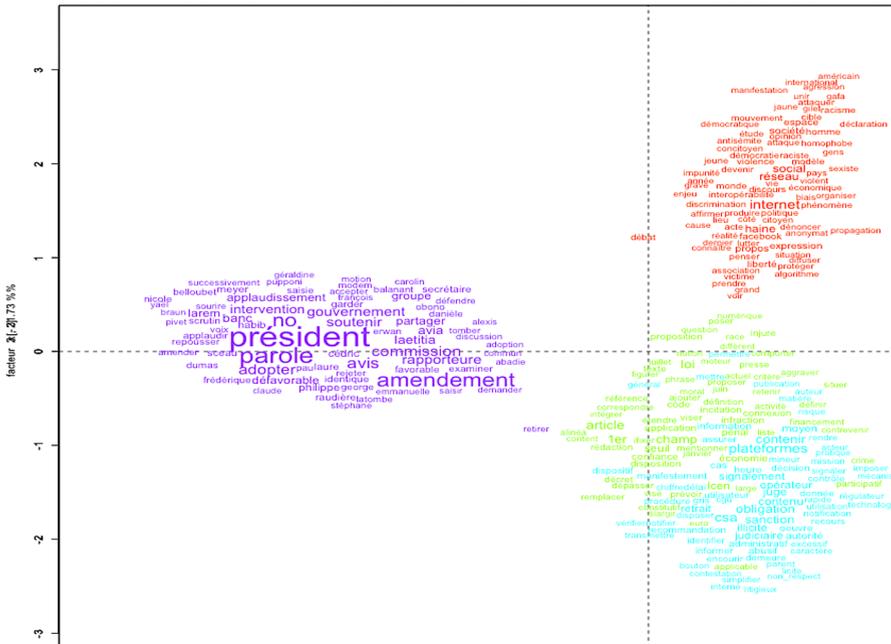


Fig. 3 Factor analysis of the clusters

Nombre d'occurrences:		175585	Nombre de formes:		14965		
Nombre d'hapax:		7502	Fréquence maximale:		9685		
	Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
✓	1	AE	5936	1660	1034	319	de
✓	2	DLF	242	172	145	14	de
✓	3	FI	3063	1294	942	145	de
✓	4	GDR	5399	1716	1108	299	de
✓	5	LFI	14707	3560	2217	857	de
✓	6	LR	15204	3578	2138	797	de
✓	7	LREM	81465	9025	4600	4653	de
✓	8	LT	6592	1927	1225	357	de
✓	9	MD	18374	3918	2305	934	de
✓	10	NI	590	300	206	53	de
✓	11	PC	786	387	290	50	de
✓	12	PS	6799	2038	1340	323	de
✓	13	RN	4125	1385	935	216	de
✓	14	SA	1495	667	474	79	de
✓	15	UDI	10808	2980	1903	589	de

Fig. 4 Frequency statistics within the corpora

largely overlap, revealing the inter-relatedness within the debates of the two key themes, the *problem of definition*, and the *problem of regulation*. This overlap shows that the combined, dominant focus of the debates was the specific methods of implementation of the legislation, coupled with a broad discussion of the *problem of definition*. Although overlapping, the dominant focus of the debates is thus revealed to be how to implement regulation, rather than how to define the content of it.

Cluster 3 is most closely related to cluster 2. These are the clusters containing the detailed debates about the *problem of regulation*. Cluster 2 is primarily concerned with the French single-jurisdiction procedural problems of implementation, whereas cluster 3 acknowledges the supra-jurisdictional complexity. Cluster 4 formed a large portion of the debates and is notably separated from the other clusters. The focus of cluster 4 reveals the power dynamics within the debates, and the concerns raised by the opposition parties which were ultimately upheld when the *Loi Avia* was censured. This positioning represents the general concern within the debates around the passing of restrictive legislation to restrict free speech, and the power dynamics between the political parties, which we expand upon below and in the discussion.

4.2 The Local Approach: the Collocates, n-Grams, Frequencies, and Semantic Features of the Debates

This analysis revealed the political affiliations within the debates. Figure 4 shows the frequencies within the entire corpus in respect of the speakers from each political party. This was calculated by conducting an n-gram analysis on each sub corpora and by identifying the collocates. The sub-corpora of the political majority, *La République En Marche* (LREM), contains about 50% of the words in the entire corpus. Compared to the other political parties, they dominate the

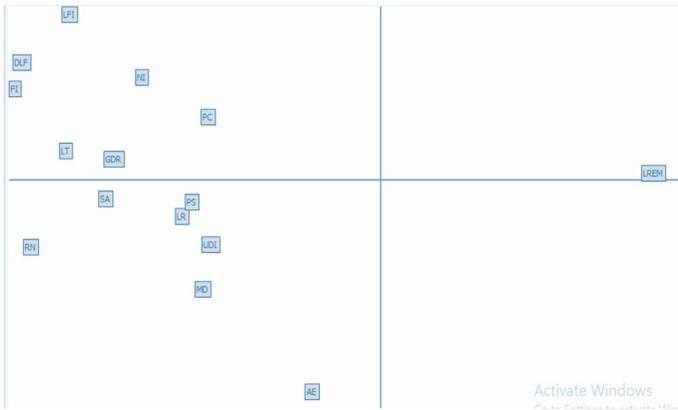


Fig. 5 Factorial analysis of political parties

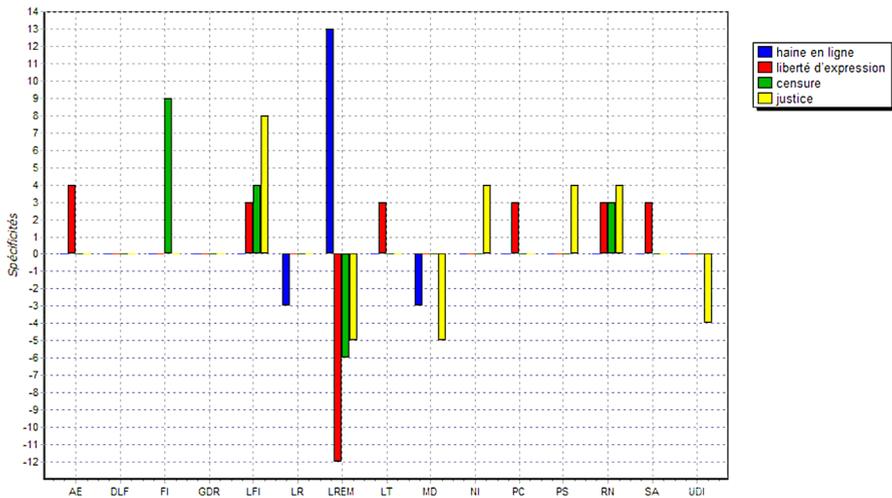


Fig. 6 Haïne en ligne, liberté d'expression, censure, justice: distribution by political party

debates. This is unsurprising as the political majority both lead and Chair most of the debates.

The factorial analysis in Fig. 5 shows the lexical closeness and distance between the sub-corpora, revealing that LREM has a different and distinctive lexical stock. The other parties share more lexical features. This illustrates the power dynamics within the debates, where the discourse of the dominant party has different lexical features to the debates of the opposition. We return to this point in the discussion.

Figure 6 shows that the phrase *haïne en ligne* (online hate speech) is overrepresented in the discourse of the parliamentary majority who propose the Bill (LREM) while the word *propos* (remarks) is overrepresented in the discourse of the opposition, as well as *liberté d'expression* (freedom of expression) and *censure* (censor).

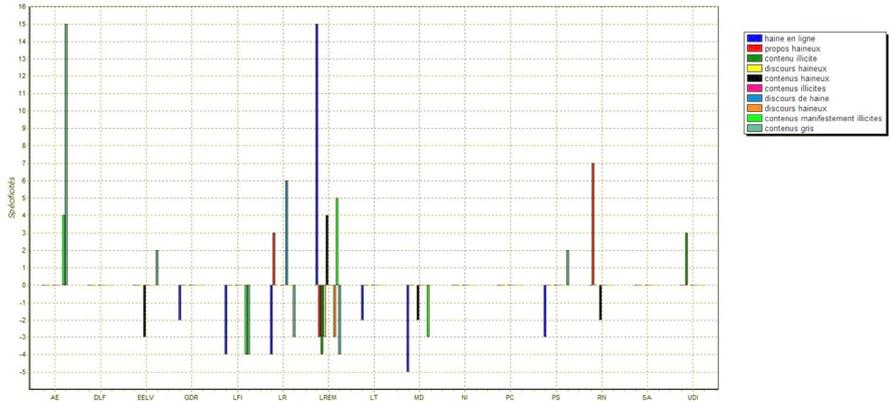


Fig. 7 *haine en ligne*: distribution of lexical variation per political party

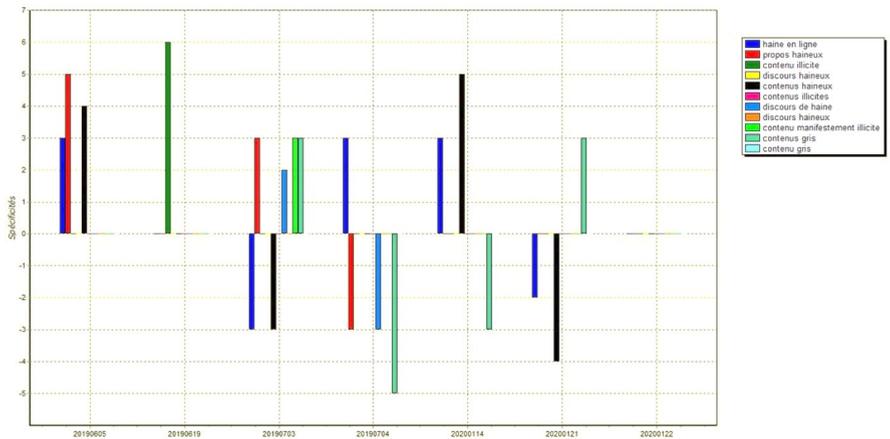


Fig. 8 *Propos haineux* and *contenus gris*: distribution, lexical variations, and dates

Figure 7 shows the lexical variation for each political party, using the most frequent collocates to *hate*. *Haine* (hate) most commonly co-occurs with *discours de haine* (hate speech) or *haine en ligne* (online hate). The most important immediate collocates of *haineux* (hateful) are *caractère* (character), *propos* (remarks), *propaganda* (propaganda), and *derives* (excesses). The n-gram analysis shows that *haine en ligne* appears mainly at the beginning and at the end of debates of the first reading. It then appears significantly on the second reading collocated with *contenus haineux* (hatful content) and appears less significantly afterwards. *Contenus gris* (grey content) is more significant at the start of the second reading, during the Public Discussion. *Haine* (hate) is much less prevalent in the last two debates (Fig. 8).

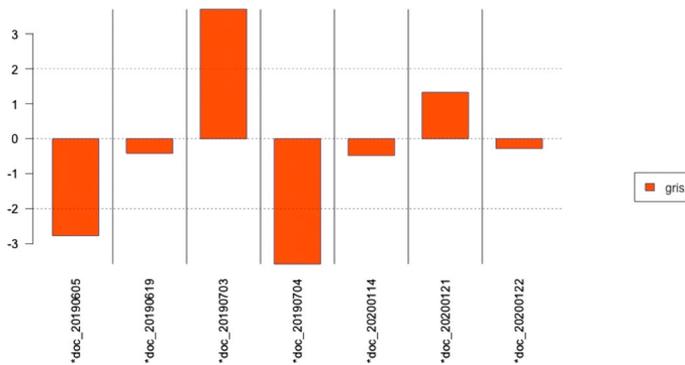


Fig. 9 Frequency of the inclusion of *gris* by date

Figure 9 shows that *grey content* dominates the debates on 3/07 and 21/01, mainly by several opposition parties to the LREM. This finding is consistent with Fig. 6 (above)

5 Discussion

5.1 The Meta-Discourse of Power Relations within the Debates

For the analysis in this paper, we assumed that the process of establishing the discourse norms and controls within the debates are contained within the meta-discourse, illuminating the “speakable” and the “unspeakable”. This meta-discourse activity is, in turn, revealed by the markers and process of identification in our specific analyses. Hyland explains that:

... acts of meaning-making are never neutral but always engaged in that they realize the interests, the positions, the perspectives and the values of those who enact them. [3, p. 14],

and that the meta-discourse is:

... the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community [3, p. 43].

Within the meta-discourse the stance of the enunciators is revealed. Stance markers are critical linguistic devices conveying personal attitudes, judgements, or assessments about the proposition of certain messages. According to Hyland, stance has three components: evidentiality, affect, and presence, identified as hedges, boosters, attitude markers, and self-mention [3, 3

Based on these tools, we suggest that attitude markers and boosters reveal the way in which the meta-discourse is developed, and the way in which semantic uncertainty, and its negotiations, manifest themselves in the interactions. The contrastive analysis between the different political parties identifies these semantic negotiations, grounded in the power relations within the discourse of the majority and of the opposition discourses.

As the debates represent a liminal space between empirical reality (what, precisely, is/is not hate speech) and its precise boundaries (the legislative parameters of the *grey zone*), the philosophical considerations of theorists who consider discourse in the context of power dynamics, such as Bourdieu [47] and Foucault [48] can be drawn on to highlight the nature of state control of both the discourses within institutional settings. This is revealed via the identification of the dominant political contributions during the debates, and the subject of these debates itself; the potential suppression of free speech. By adopting our broad description of parliamentary discourse to include social issues, and by developing Galembert's definition [31] to enable our focus on linguistic analysis of the *process* of translating socio-political issues into legislation, the power relations are revealed in the texts. Bourdieu describes the "things named" (in our analysis, creating a regulatory framework for online hate speech and fake news) as the product of symbolic power: "law is undoubtedly the form par excellence of the symbolic power of nomination and classification which creates the things named" [47, p.13]. This philosophical perspective guides our analysis, and grounds the study in the broader context of the power dynamics inherent with the translational law-making cycle, of converting social issues via debate into legislation which is, in turn, interpreted by judges when individual cases are brought before the courts.

These are the liminal spaces that can be observed in the debates operating at this translational intersection between society and government, exemplifying the tensions between state powers and private rights in law making and law interpreting/enforcing. The inherent discourse processes reveal how normative acts are constructed via the debates where the sayable and the unsayable, the acceptable and the unacceptable, together with their linguistic features can be observed. Thus, the study identifies the way in which the negotiations and the resistances around the discursive norms of the dominant political party are expressed. These norms are specifically expressed as defining the parameters of hate speech and fake news, in which contexts, and for what reasons the definitions have been decided upon. It is also important to understand how the different speeches synthesize the positions taken and to determine to what extent social issues as well as political interests are highlighted in the discussions.

Analysis of parliamentary debates inevitably interrogate discourses of power within state institutions, revealing power relations operating between the different actors in the debates, particularly the position and dominance of the political party in power, and the positions of the opposition parties. We set out to gain insights into the way in which these power relations are mediated within the discourses of the debates. Galembert considers this an inherent function of a "logic" that exists within the discourse, an unconscious process that almost 'mathematically' balances the competing voices of the enunciators:

Certainly, the passing of a law is a matter of balance of power and majority vote (and in this respect a matter of mathematics). However, it is up to the logic of parliamentary discourse to relativize this balance of power, and to proceed to the discursive erasure of the antagonisms it has aroused. [31]

LREM, the political party in power, dominates the debates and ultimately the “logic of parliamentary discourse” [49] failed to re-balance power during the debates themselves. In this example, the re-balancing occurred with the censure from the French Senate.

Before discourse can be controlled and suppressed, first it must be categorized into acceptable and unacceptable categories. This process, and who controls it, is evident within parliamentary debates. Foucault’s discourse analysis identifies the mechanisms that enable one to distinguish true and false statements [48]. Indeed, the Foucauldian perspective reveals the way in which the process for the elaboration of truthful facts is construed as an acceptable object within society. In our analysis, this was evident in the fact that although there were two key themes identified in the debates, the *problem of definition* and the *problem of regulation*, the regulatory issues were more prominent. It was also evident that although the supra-jurisdictional complexities were acknowledged, the dominant focus of the debates, driven by LREM (the French party in power), was that of the procedural aspects of implementing new regulations.

Although one of the clusters identified broader issues concerning the *problem of definition*, overall, there was far less focus on the fundamental legal questions surrounding the balance between the right to free speech and freedom of expression, and the need for individual and collective state control of it to prevent unwarranted interference in democratic electoral processes. The development of a regulatory framework to mediate and control discourse must first identify precisely *where* it is acceptable to intervene to control *what* can be said, by whom and where. This is insufficiently addressed in the sample in our corpus. A liberal, Foucauldian analysis legitimizes the concept of free speech by framing it as a necessary discourse, within which discursive space ideas and concepts can be developed. The counter argument is to make the distinction between free speech and debate about social issues, norms and values grounded in empirical reality as opposed to those which have no such grounding. The precise philosophical and legal boundaries should intersect, but the debates were dominated by the power-dynamics of regulatory control rather than the more philosophical issues.

Fundamentally, Foucault’s consideration of discourse explores the dimension of the socially and politically acceptable, the speakable, and the unspeakable, the thinkable and the unthinkable as well as that which is articulated and not articulated [49], p.31] In so describing, the notion of truth is revealed to be fluid and exploratory, controlled more by the power relations between the enunciators than by the empirical evidence upon which the utterances are based:

Truth is a thing of this world: it is produced only by virtue of multiple forms of constraint. And it induces regular effects of power. Each society has its regime of truth, its “general politics” of truth: that is, the types of discourse which it accepts and makes function as true; the mechanisms and instances which

enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true. [49, p. 131]

The precise space within the debates where the fluid and exploratory notion of truth is discussed is the liminal space to which we refer at the start of this article. Empirically grounded facts may guide interpretations of their meaning, but subjectivity is an element of interpretation, and this subjectivity is explored during the debates because the legislative boundaries are not set. This unique space is where the conceptualization of the limits of the *grey zone* takes place, grounded in the discourse of the dominant political party in power.

5.2 The Semantic Meanings within the Debates

Having established the existence of the meta-discourse and the associated power relations within the debates, we turn to semantic analysis of the texts from two positions: firstly, the multidimensional *character* of meaning, and secondly the dynamic *nature* of meaning. Our analysis considers the co-occurrences and specific semantic features of words in the meta-discourse, identifying the context of individual words by examining the immediate and nearest collocates. This thematic analysis shows that discussion of *hate* is associated most closely to the largest cluster (cluster 1, shown above) in terms of percentage and is directly related to social media and freedom of expression. The classifications show that *hateful* appears in contexts similar with those of cluster 2, relating to law and legal implication.

The following extracts focus on examples from the debates where *hate* or *hateful* is mentioned. For example, in extract 1, the speaker discusses *Article 1* of the *Avia Bill* and argues the precise use of words to delimit and define *haine*.

- (1) L'article 1er fixe un cadre; Tout est là. C'est l'occasion de rappeler combien il est important de veiller aux mots que nous employons. Ainsi, on ne lutte pas contre la haine. La **haine** est un **sentiment intime, personnel**. Je pense ce que je veux à l'égard de qui je veux. Cependant, je n'ai pas le droit de l'exprimer. De fait, c'est contre l'expression de la haine et non contre la haine elle-même que l'on peut lutter. C'est l'acte, et lui seul, que nous devons condamner. Prenons garde à ne pas laisser croire que le législateur s'imisce dans les consciences pour y réglementer d'autres passions, comme la colère ou l'amour. **Il était important de le rappeler.**

Article 1 sets a framework; Everything is here. This is an opportunity to remember how important it is to watch the words we use. Thus, we do not fight against hatred. **Hatred** is an **intimate, personal feeling**. I think what I want about whom I want. However, I have no right to express it. In fact, it is against the expression of hatred and not against hatred itself that one can fight. It is the act, and it alone, that we must condemn. Let us be careful not to suggest that the legislator interferes in consciences to regulate other passions, such as anger or love. **It was important** to remind that.

M. Hervé Saulignac. **** *doc_20190703 *SP_0307 *Lecture_1

In this example, the speaker (from the opposition) specifies and differentiates between hatred as a feeling and its verbal or written “expression”. To differentiate between them, he uses the term “expression”, emphasizing it is the act of expression that constitutes an offense, not the feeling that is being expressed. In doing so, he articulates a defining statement around the notion of the offence, opposing the use of the term *hate* to designate offense on the premise that this refers to a feeling (concept) rather than its expression (articulation). The expression of this opposition to the use of the term *hate* in an offense is a booster [3] where the speaker moves the attention of the object to be condemned from a feeling or an emotion to an expression. The speaker’s statements are assertive: “it is the act, and it alone, that we must condemn”; “it is important to remember”; “I think what I want regarding whom I want. However, I have no right to express it.”. He also mentions the role of the legislator “Let’s be careful not to suggest that the legislator interferes in consciences to regulate other passions, such as anger or love”. In this sense, the speaker brings the legislator’s attention to an identifiable and material object, namely an expression which is a stated and visible discourse.

Extracts 2 and 3, similarly, to extract 1, associate the need to define the parameters of *hate speech* with the question of the balance between the mechanisms of regulation and self-regulation. Extract 2 is also from the opposition:

- (B) Selon le CNNum, il faut tout d’abord, et **urgemment**, définir **très précisément** ce qu’est un **discours haineux**, tout comme il faut **impérativement prévoir** un juste équilibre entre le recours au mécanisme judiciaire, à la régulation et à l’autorégulation.

According to the CNNum, it is **urgently** necessary to define **very precisely** what **hate speech** is, just as it is **imperative to provide** for a fair balance between recourse to the judicial mechanism, to regulation and to self-regulation.

M. Michel Larive. **** *doc_20190605 *EC_0506 *Lecture_1

- (C) **La notion de haine est difficile à définir**; les juges eux-mêmes éprouvent des difficultés à qualifier l’incitation à la haine, dont la définition varie en fonction des jurisprudences. **De plus, un contenu peut être notifié comme illicite pris isolément alors que, replacé dans un contexte plus général, il prendra un tout autre sens, incontestablement légal et qui relève de la liberté d’expression.** La Cour de cassation a ainsi eu l’occasion de déclarer légal au regard de la notion de débat d’intérêt général un **contenu manifestement illicite**. Par ailleurs, l’article donne aux plateformes un pouvoir de police des mœurs, alors que ce n’est pas leur rôle mais bien celui du juge. La liberté d’expression ne peut être bridée par des opérateurs privés dont l’expérience prouve que leur appréciation des contenus illicites est à géométrie variable. Cet article me paraît donc déséquilibré: les risques d’atteinte disproportionnée à la liberté d’expression sont réels et il ne sera pas efficace pour lutter contre les contenus haineux en ligne.

The notion of hate is difficult to define; judges themselves find it difficult to qualify incitement to hatred, the definition of which varies according to case law.

In addition, content can be notified as illegal taken in isolation whereas, placed in a more general context, it will take on a completely different meaning, incontestably legal and which falls within the scope of freedom of expression. The Court of Cassation thus had the opportunity to declare legality with regard to the concept of general interest debate a clearly illicit content. In addition, the article gives platforms the power to police morals, when it is not their role but that of the judge. Freedom of expression cannot be restricted by private operators whose experience proves that their appreciation of illegal content is variable. This article therefore seems unbalanced to me: the risks of a disproportionate attack on freedom of expression are real and it will not be effective in combating hateful content online.

M. Yannick Favennec Becot. ** *doc_20190703 *SP_0307 *Lecture_1**

In extract 3, the speaker again expresses the difficulty of defining the notion of *hate*. He associates it with the “duty of the judge” and refers to “interpreting content in context”. Thus, the speaker highlights the need to contextualize the hateful discourse and tries to dissociate the task of interpretation away from being the responsibility of the online platforms. This distinction is made to distinguish between the power of online platform moderation which has the “power to police morals”, and the power of the legislature and the judiciary, who have the power to police adherence to unlawful expressions of hate. These opposition comments are made in the context of protecting the right to freedom of expression and preventing unwarranted interference with this right.

After the examination of the Bill in the Senate debates, it is examined in a second reading at the National Assembly. The following examples are taken from a Parliamentary Committee Review Session where the deputies discuss the amendments. Extract 4 concerns both the *problem of definition* and the *problem of regulation*:

(D) **Je défends** également les amendements CL13 et CL7 qui visent à s’interroger sur la rédaction de certains chapitres et du titre de la proposition de loi. Le terme de contenus haineux me gêne: la haine n’a pas de définition juridique. Encadrer l’utilisation d’internet en s’appuyant sur des **présupposés moraux et non objectifs** suppose que le législateur soit détenteur d’une vérité universelle et sans nuance. **Substituer aux mots « contenus haineux» les mots « ne répondant pas aux standards de la communauté» permettrait une adaptation des utilisateurs aux règles des plateformes sur lesquelles ils s’inscrivent.** Une telle rédaction permettrait de préserver la liberté des plateformes et de protéger l’utilisateur. Les standards de la communauté sont élaborés afin d’entraîner la suppression des atteintes à une caractéristique protégée juridiquement et définie de façon objective.

I also support amendments CL13 and CL7 which aim to question the wording of certain chapters and the title of the Bill. The term hate content **bothers me**: hate has no legal definition. Regulating the use of the Internet **based on moral and non-objective presuppositions presupposes that the legislator is the holder of a universal and unqualified truth.** Replacing the words “hate content” with the words “not meeting community standards” would allow users

to adapt to the rules of the platforms on which they register. Such wording would preserve the freedom of the platforms and protect the user. Community standards are developed to result in the removal of infringements of a legally protected and objectively defined characteristic.

Mme Marie-France Lorho. **** *doc_20200114 *EC_14012020 *Lecture_2

In extract 4, the utterance associates the attitude marker “bothers me” to the booster “hatred has no legal definition”. The lack of legal definition is linked to *moral* and *non-objective presuppositions*. The speaker proposes to “substitute” the terms in the Bill and to change the legislative discourse by transferring regulatory responsibility to the online platforms. She implicitly advances the argument that “truth” is based on a moral presupposition, has no universal meaning, and is better defined in the context of what is acceptable by individual, online “community standards”. In this extract, the role of the legislator is re-framed in the warning that they overstep their position if they become “holder of a universal truth without nuance”. The Deputy (Laetitia Avia who proposed the Bill) responds in extract 5 with the following:

(E) Avis défavorable. **Nous avons défini à l’article 1er le champ de ces contenus haineux.** Il ne s’agit précisément pas de se situer dans le cadre des standards de la communauté – cela impliquerait que la nudité, qui est refusée sur Facebook, devienne un motif de retrait de contenu –, mais bien de se cantonner à l’application de notre proposition de loi et à ce qui, dans notre pays, est considéré comme une infraction au regard de la loi de 1881.

Unfavorable opinion. **We have defined in Article 1 the scope of such hateful content.** It is precisely not a question of placing oneself within the framework of the standards of the community – this would imply that nudity, which is refused on Facebook, becomes a reason for removing content –, but of confining oneself to the application of our Bill and what, in our country, is considered an offense under the law of 1881.

Mme Laetitia Avia **** *doc_20200114 *EC_14012020 *Lecture_2

This interaction in extract 5 shows apprehension for online hate speech to be arbitrated by a process of moderation controlled by the online platforms, and takes a stance directly opposed to that of the opposition. Deputy Laetitia Avia specifies that it is precisely *not* a question of leaving the platforms responsible, and that such content must be referred to under the terms of the law of 1881.³ This position places the responsibility for adjudication firmly with the legislature and judiciary, arguing firstly that the parameters of online hate speech *are* sufficiently clearly defined, and secondly that it should be a matter for legislation. This does not, however, address the complex supra-jurisdictional questions, or the question of balancing competing rights which ultimately led to the censure of the *Loi Avia*.

³ The Law on the Freedom of the Press of 29 July 1881 (French: Loi sur la liberté de la presse du 29 juillet 1881) [50].

These extracts show the importance of the *problem of definition*, which our hierarchical clustering analysis (shown above) identifies as having been insufficiently considered within the debates. Defining what is, and is not, hateful content, occurs at all stages of contemplation, discussion, and formulation of legislation. If it is unclear *what* is being policed, the question of *who* polices and *how* it is policed become secondary considerations. The cluster analysis, however, reveals the dominance of the regulatory discussions (clusters 2,3, and 4) over discussions concerning definition (cluster 1), and was reinforced in our analysis of the meta-discursive practices and dominant discourse of the political party in power.

The *problem of definition* is undoubtedly a complex and difficult conceptual and practical issue that the debates failed to adequately resolve. The phrase used throughout the debates to denote the complexity is *grey zone* (zone gris), eloquently expressing both uncertainty and ambiguity. In our corpus, *gris* was most closely collocated with *contenus* (content) and *area* (zone). Its most frequent collocate is *obviously illicit*, which suggests some content is *not* obvious in its categorization as either illicit, or alternatively as legitimate. The use of *grey content* in the debates is intended to define the liminal space between that which is obviously illegal content, and that which could be defined as free speech, and thus protected from state (or online) control. Before resolving the problem of the *grey zone*, the debates re-focus towards the *problem of regulation*, including the power conferred in the French legislation to French judges, and the power of the online platforms to self-regulate (moderate) content, referring to both as “a situation of uncertain arbitration boundaries”.

The discursive antonym “obviously illicit” and particularly the adjective “obviously” amplifies the semantic uncertainty and the ambiguity of the *grey zone*. The issue of the *grey zone* and *grey content* runs through the legislative discussions until the second reading in January 2020. Most of the time *grey content* is discussed in the context of regulation and the impact on freedom of speech, not on its definition. The debates show that this unresolved tension is a central point, but despite this the meta-discourse dwells on questions relating to regulatory distinction, where definition is a secondary consideration, necessary to be defined only to support the regulatory framework. Ultimately, there is no clear definition, and the matter is left to the judges, and the online platforms, to identify and determine. This unsatisfactory situation led to the censure.

Extracts 6, 7 and 8 also show stance markers within the meta-discourse, highlighting the conflicts and absence of common understanding of the parameters of the *grey zone*.

(F) Cet amendement vise à résoudre la question de la gestion des contenus gris, c'est-à-dire les contenus qui ne sont pas **manifestement** illicites et qui n'ont pas donné lieu à une décision de justice.

This amendment aims to resolve the issue of the management of grey content, i.e., content which is not **manifestly** illegal, and which has not given rise to a court decision.

E. Menard—**** *doc_20190619 *EC_19062019 *Lecture_1

(G) Mme la rapporteure a affirmé, en substance, que la proposition de loi ne permettra sans doute pas de tout régler, et qu'elle n'a pas vocation à **résorber la**

fameuse zone grise, qui sépare les contenus manifestement illicites de ceux qui sont sujets à interprétation. [...] Le vrai problème, madame la rapporteure, c'est que votre proposition de loi confie le pouvoir de retrait et d'appréciation des contenus aux plateformes, sans ménager une place au juge s'agissant de cette fameuse zone grise, que nous-mêmes ne parvenons pas à déterminer comme évidente.

Madam Rapporteur affirmed, in essence, that the Bill will undoubtedly not make it possible to resolve everything, and that it is not intended to **reduce the grey area, which separates the manifestly illicit contents from those that are subject to interpretation.** [...] The real problem, Madame la Rapporteure, is that your Bill entrusts the power to withdraw and assess content to the platforms, without leaving room for the judge when it comes to this specific grey area, which we ourselves fail to determine as obvious.

**** *doc_20190703 *SP_0307 *Lecture_1

(H) Quand il s'agit de propos manifestement haineux, il n'y a aucun problème; **tout le monde est d'accord pour demander leur retrait.** Mais comment la **plateforme appréhendera-t-elle la zone grise ?** Puisque ce ne sont pas des personnes physiques qui s'y attelleront, mais des algorithmes, comment ces derniers distingueront-ils **les propos clairement haineux des messages perturbants mais n'outrepasant pas les limites de la liberté d'expression**, telles qu'elles sont définies par la Cour européenne des droits de l'homme ?

When it comes to obvious hateful remarks, there is no problem; **everyone agrees to request their removal.** But **how will the platform approach the grey area?** Since it is not physical persons who will tackle it, but algorithms, how will the latter **distinguish between clearly hateful remarks and messages that are disturbing but do not overstep the limits of freedom of expression**, such as they are defined by the European Court of Human Rights?

La Raudière - **** *doc_20200114 *EC_14012020 *Reading_2

In extracts 6 and 7, the speakers express themselves in the first and second readings and define the *grey zone* negatively. In extract 6, the *grey zone* relates to unqualified semantic features. In extracts 7 and 8, the semantic feature *uncertainty* is present, opposing the certainty of "manifestly illicit content" for which "everyone agrees" concerning their removal. On the other hand, *grey content* is considered as "disturbing but does not exceed the limits of freedom of expression" and "subject to interpretation". In other words, it is not possible to define the *grey zone* outside the legal interpretative framework, which must operate on a case-by-case basis. These are the semantic features of normative vagueness and definitional impossibility that surrounding this notion. The transfer of the object of this normative vagueness to the control of the online platforms worries some parliamentarians on the grounds that the scope for online platforms to stray into the restriction of free speech was too great if the *problem of definition* remained unresolved.

Ultimately, the *grey zone* never moves beyond discussion within the parliamentary debates and is not resolved. It remains in the liminal, discursive space of the debates. Despite this failure to resolve, the debates moved beyond liminal discussion in respect of the *problem of regulation*. However, without resolution of precisely

what is to be legislated against, both the *problem of definition* and the *problem of regulation* remain as there no legislative clarity is established, and no consensus on the normative standards of individual online platforms.

As a result of the inability to move forward on the *question of definition*, in the extracts from the Second Reading the question of the *grey zone* is still discussed. In extract 9, Mme. Avia, the Deputy, “takes responsibility” for focusing on “manifestly illegal content”. She states her point with assurance. The opposing speaker in extract 10 retorts that leaving aside the *question of definition* where it now emerges that the intention is to exclude *grey zone* content from regulation, the Deputy is in “intellectual and psychological denial”. She criticizes the Deputy for “privatising the withdrawal decision” and transferring power to commercial entities:

- (I) Avia: Dernier point: **les contenus gris. J’assume pleinement** le fait de ne m’intéresser ici qu’à un sujet: les contenus manifestement illicites. On sait bien que ce texte ne réglera pas tout le problème, mais si l’on arrive déjà à supprimer ces contenus, internet sera un peu plus sain.

Avia: Last point: **gray content. I fully bear** the responsibility that I am only interested here in one subject: manifestly illegal content. We know that this text will not solve the whole problem, but if we already manage to delete this content, the internet will be a little healthier.

**** *doc_20200121 *SP_21012020 *Lecture_2

- (J) **La fameuse zone grise**, par exemple, me semble donner lieu à un véritable déni intellectuel et psychologique. **Je vous crois sincères** lorsque vous dites ne pas viser la zone grise mais celle-ci, par essence, ne saurait être « visée » ou non puisqu’elle présente un problème d’interprétation. Or vous confiez ce pouvoir d’interprétation – qui devrait revenir au juge – à des plateformes privées, d’où la **privatisation de la décision de retrait**.

The famous gray zone, for example, seems to me to give **rise to real intellectual and psychological denial**. I believe you are sincere when you say that you are not aiming for the gray area, but this, in essence, cannot be “targeted” or not since **it presents a problem of interpretation**. However, you entrust this power of interpretation – which should be up to the judge – to private platforms, hence **the privatization of the decision to withdraw**.

Dumas **** *doc_20200121 *SP_21012020 *Lecture_2

By extract 11, discussions of *grey content* draw out the implicit proposals from some parliamentarians to “locate” (define) the *grey zone* more precisely, and thus provide solutions to the *problem of definition*:

- (K) On s’imagine que l’on a affaire à une catégorie juridique qui tombe sous le sens, et qu’un modérateur privé sera à même de gérer les situations en vingt-quatre heures, mais les frontières de l’injure sont délicates et les zones grises sont larges. On peut citer le cas retentissant des caricatures de Mahomet; il y a douze ans, le tribunal correctionnel de Paris avait estimé en première instance que la qualification d’injure envers les musulmans pouvait être retenue s’agissant du

dessin montrant Mahomet avec une bombe dans son turban, même si Charlie Hebdo était par ailleurs relaxé au titre de sa bonne foi. Cette analyse sera ensuite infirmée par la cour d'appel.

We imagine that we are dealing with **an obvious legal category** and that a private moderator will be able to manage situations in twenty-four hours, but the borders of insult are delicate, and the gray areas are wide. We can cite the resounding case of the caricatures of Muhammad; twelve years ago, the Paris Criminal Court had considered in first instance that the qualification of insult to Muslims could be retained with regard to the drawing showing Muhammad with a bomb in his turban, even if Charlie Hebdo was otherwise released under the good faith of its editors. This analysis will then be reversed by the Court of Appeal.

Dumas **** *doc_20190703 *SP_0307 *Reading_1

The speaker, Dumas, connects the vagueness of the *grey zone* by citing an example of a court order regarding the Muhammed cartoons controversy. She justifies her point by highlighting that a fact can be qualified as an insult, but at the same time could not be condemned. Here, she seems to refer to the notion of intention, rather than the words themselves, and she also refers to an attack on a community dignity. Identifying facts and intent, and qualifying their perception and impact, however, are two separate things, and perhaps establishing the parameters of the *grey zone* would be helped by establishing whether intention and/or impact are relevant legal factors.

Finally, in extract 12, the speaker returns to the *problem of regulation*, highlighting the fact that online platforms are not equipped with suitable mechanisms and frameworks, both in terms of time and competence.

(L) [...] nous avons discuté de ces **zones grises**, ainsi que de la viralité des contenus, ce qui a donné lieu à un débat tout à fait intéressant. [...] je me suis interrogé, lors de la discussion générale, sur l'éventuelle absence d'un chaînon manquant. Si le texte traite en effet des contenus manifestement illicites, un doute subsiste quant à la suppression de **contenus seulement présumés tels**. Si la saisine du juge judiciaire sera possible, le problème dont nous parlons, **nous le savons bien, requiert une grande réactivité, et le temps judiciaire n'offre pas forcément l'immédiateté dont on a besoin en la matière**. [...] J'avais donc, au cours de la discussion générale, indiqué qu'il aurait peut-être été opportun d'avoir à consulter une **sorte d'autorité**, laquelle serait évidemment en charge des plateformes, tout en jouissant d'une **certaine indépendance**.

Je sais que le problème n'est pas simple du tout, mais une telle solution permettrait sans doute de progresser sur ces sujets et d'y apporter des réponses. L'idéal serait une recommandation rapide fondée sur le droit applicable : **les médias sociaux ne seraient ainsi plus seuls à décider**, avec le risque que cela comporte par rapport à des remises en cause de la liberté d'expression.

[...] we discussed these **gray areas**, as well as the virality of content, which gave rise to a very interesting debate. [...] I wondered, during the general discussion, about the possible absence of a **missing link**. If the text does indeed deal

with manifestly illegal content, a doubt remains as to the deletion of content **presumably illicit**. If the referral to the judicial judge will be possible, the problem we are talking about, **we know it well, requires a great effort and the judicial time does not necessarily offer the immediacy that we need in the matter**. (...) I had therefore, during the general discussion, indicated that it might have been appropriate to have to consult some **sort of authority**, which would obviously oversee the platforms, while enjoying a **certain independence**. I know that the problem is not simple at all, but such a solution would undoubtedly make it possible to progress on these subjects and to provide answers to them. The ideal would be a quick recommendation based on the **applicable law: social media would no longer be the only ones to decide**, with the risk that this entails in relation to challenges to freedom of expression.

Reiss; **** *doc_20190704 *SP_0407 *Lecture_1

The speaker, Reiss, replaces *grey zone*, qualifying it as meaning “presumably illicit”, referring to uncertainty and connecting it to a “missing link”. This statement is semantically interesting because it situates the *grey zone* in a continuously moving and evolving space. This link exists both semantically and structurally in the new regulation solution the speaker proposes. Thus, the answer provided by the deputy is articulated in the expression of a semantics of *space* (link) and of *time* (reactivity, judicial time, immediacy) and proposes an intermediary to arbitrate this type of content. This will, Reiss argues, be achieved by creating an “independent” body “in charge of the platforms”. The formulation and development of the meta-discourse on the notion of *grey zone* accompanies those of proposals to try to solve the *problems of regulation*.

6 Conclusions

Our analysis highlighted the key themes of the debates, and the respective prominence given to them. We also identified power dynamics within the debates by considering the discourses of the party in power and the opposition. Analysis of the semantic features reinforced our statistical findings, showing that the discourse of the party in power uses stance markers in relation to their focus on the matter of (bureaucratic) regulation, rather than the (moral) question of definition, despite the attempts of the opposition to discuss and define the *grey zone*.

The use of quali-quantitative methods enabled us to triangulate our findings, blending statistical evaluation, not evident from manual reading of the debates, with a closer stance and semantic examination at individual speaker level. Combined with our study of the meta-discursive activity, these analyses reveal the strategies of the parliamentarians. Our analysis of the distribution of the lexical variations within the debates revealed specific language choices of the political majority, and of the opposition. Analysis of the meta-discourse of the parliamentary debates provides source of furthering understanding of the legal trajectories. This analysis illuminated the *processes* by which the object’s codification, transmutation, and passage from a social problem to a legal notion is undertaken. This study did not, therefore, focus on the general structure of parliamentary speeches and interactions, instead it

considered the macro question of the *conditions of production* of these texts in the parliamentary sphere and the micro question of how the *object* of the discussion was discussed, formulated, (re)negotiated, defined, and qualified.

From a diachronic perspective the emergence, and then the disappearance, during the debates of the lexical variations concerning the definition online hate speech was evident. This was observed during the trajectory of the debates as they stray from consideration in cluster 1 of the *problem of definition* towards the considerations in clusters 2,3 and 4 of the *problem of regulation*, despite the attempts of the opposition shown in the extracts to re-introduce the unresolved issue of the *grey zone*. Although the opposition is not a homogeneous block, our analysis showed that the political majority used a specific discourse. The analysis showed the semantic conflicts, particularly evident in the differences identified in cluster 1 concerning the *problem of definition*. For example, the majority expressed *hate* dichotomously as both a concrete phenomenon, and a generic abstract concept. However, the opposition discussed and tackled *hate* from a more practical stance.

These analyses shed light on the fundamental issues faced by parliamentarians when implementing any new legislation that specifically confers state powers to limit private rights: The issue must first be defined before it can be regulated. In our analysis this was articulated in the discussions around how to delimit *grey content* and *who* should arbitrate the content. Neither question was adequately resolved. We observe that within the liminal space of the debates, their identification and qualification are closely linked, perhaps too closely to the system that will then be required to moderate and judge them.

In our example, we selected debates concerning a modern phenomenon, that of online content and its ability to pose both a societal and a political problem via ‘soft invasions’ which undermine democratic election processes. The nature of online content post-dates most worldwide attempts to legislate in order to control the limits of free speech. In our French example, the pre-existing law was the *Loi de 1889* [50]. Within the debates, the opposition argued against applying these laws to content whose intrinsic characteristics are instantaneity, and globalization. This problem refers to the complex, supra-jurisdictional nature of both the problem and any proposed solutions.

Taken together, our analyses interrogated the dynamics of both the dominant and the counter-discourses within the corpus. The study revealed how the meaning of hate speech was discussed, mediated, and ultimately suppressed, and how the position of individual jurisdictions, individual online platforms, and their individual and collective responsibilities were negotiated within the debates. The study also highlighted how legal norms were articulated and challenged by capturing the characteristics of the meta-discourse. Our analysis was grounded in the wider social context of the French attempt to address online hateful content in the *Loi Avia*. The Bill and subsequent *Loi* attempted to address issues such as (online or otherwise) insults, defamation, negationism (the revision of history to omit something that has happened), and revisionism (the revision of history to downplay the significance of certain events) on the basis that the existing law has been argued to be inadequate in most circumstance, not least in relation to online communication [51].

We conclude that analysis of discourse within the French parliamentary debates yielded insights into the ways parliamentarians speak both *directly* and *indirectly* about a topic whilst failing to adequately resolve either of the key themes we identified (the *problem of definition* and the *problem of regulation*). Given the *Loi Avia*, French parliamentary texts on the topic of hate speech offer a rich opportunity for such analysis. For any parliament, the regulatory framework must take account of the need for such a framework to transcend individual jurisdictional control, firstly because there is no supra-jurisdictional definition of hate speech, no supra-jurisdictional notion of the precise boundaries of where reporting becomes fake news or hate speech, and no unified understanding of the limits of freedom of speech. Secondly, the practicalities of regulating online content available to be viewed by citizens in a particular jurisdiction (in this example, France) transcends the ability of national law (in this case French law).

In terms of the way forward, following the censure of the *Loi Avia*, initiatives such as the *European Observatory of Online Hate* [52], funded by the Rights, Equalities and Citizenship programme of the European Union may assist. Such initiatives exist to create and publicise areas where hate speech and fake news are observed, and thus take a more incremental and empirically grounded approach to establishing a repository of cases which we can say with certainty fall outside the *grey zone*. The growing body of materials within such repositories offer new opportunities for analysis of the language decided to be online hate.

Acknowledgements The research leading to these results received funding from United Kingdom Research and Innovation (UKRI) under the Expanding Excellence in England (E3) fund Grant number: 28031.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Lecolle, M., M. Veniard, and O. Guérin. 2018. Pour une sémantique discursive: propositions et illustrations. *Langages* 210 (2): 35–54.
2. Rastier, F. 2015. Interpretative semantics. In *The Routledge handbook of semantics* (pp. 507–522). Routledge.
3. Hyland, K. 2018. *Meta-discourse: Exploring interaction in writing* (1st ed). Bloomsbury Academic; Bloomsbury Collections. <https://doi.org/10.5040/9781350063617>
4. European Court of Human Rights. Judgment. 1987. Mathieu-Mohin and Clerfayt v Belgium. Application No. 9267/81. Para. 47.
5. Steiger, D. 2022. Protecting democratic elections against online influence via “Fake News” and hate speech—The French *Loi Avia* and *Loi No.* 2018–1202, the German Network Enforcement Act and the

- EU's Digital Services Act in Light of the Right to Freedom of Expression in Schiedermaier, S., Schwarz, A. and Steiger, D. (Eds.). *Theory and practice of the European Convention on Human Rights*. Hart.
6. Council of Europe. (1950). *Convention for the protection of human rights and fundamental freedoms*. Rome. Article 10, p.12.
 7. ECtHR, Judgement, 4 December 2003, *Gündüz v Turkey*, Application No. 35071/97.
 8. ECtHR, Judgement, 5 December 2019, *Tagiyev and Huseynov v Azerbaijan*, Application No. 13274/08.
 9. Siegel, M.L. 1998. Hate speech, civil rights, and the Internet: The jurisdictional and human rights nightmare. *Albany Law Journal of Science & Technology* 9: 375.
 10. Sung-Yul Park, J., and M. Bucholtz. 2009. Public transcripts: Entextualization and linguistic representation in institutional contexts. *Text & Talk* 29–5 (2009): 485–502.
 11. Assemblée Nationale, Proposition de loi no 388, adoptée par l'Assemblée nationale, en nouvelle lecture, visant à lutter contre les contenus haineux sur internet, 13 Mai 2020, https://www.assemblee-nationale.fr/dyn/15/textes/I15t0388_texte-adoptee-seance#
 12. French Conseil Constitutionnel. Judgment. 2020. Décision n° 2020–801 DC—JORF n° 0156 du 25/06/2020, Texte 2 sur 181 du 18 Juin 2020. Decision No. 2020–801 DC of 18 June 2020 | Constitutional Council (conseil-constitutionnel.fr)
 13. Arsène, S., and C. Mabi. 2021. L'action publique au prisme de la gouvernamentalité numérique. *Reseaux* 225 (1): 9–22.
 14. Cardozo, J. 2009. What can be done against cyber hate—Freedom of speech versus hate speech in the council of Europe. *Journal of International and Comparative Law* 417: 17.
 15. Bleich, E. 2011. The rise of hate speech and hate crime laws in liberal democracies. *Journal of Ethnic and Migration Studies* 37 (6): 917–934.
 16. Alkiviadou, N. 2018. Hate Speech on social media networks: Towards a regulatory framework? *Information & Communications Technology Law* 28 (1): 19–35. <https://doi.org/10.1080/13600834.2018.1494417>.
 17. Bakalis, C. 2018. Rethinking cyber-hate laws. *Information & Communications Technology Law* 27 (1): 86–110. <https://doi.org/10.1080/13600834.2017.1393934>.
 18. Digital, Culture, Media and Sports Committee. 2018. *Disinformation and 'fake news': Interim Report*. Fifth Report of Session 2017–2019. HC363. Disinformation and 'fake news' (parliament.uk)
 19. Digital, Culture, Media and Sports Committee. 2019. *Disinformation and 'fake news': Final Report*. Eighth Report of Session 2017–2019. HC1791. Disinformation and 'fake news' (parliament.uk)
 20. Hildebrandt, M. 2020. *Law for computer scientists and other folk*. Oxford: Oxford University Press.
 21. Roberts, S. T. 2019. *Behind the screen: Content moderation in the shadows of social media*. In *Behind the Screen*. Yale University Press. <https://doi.org/10.12987/9780300245318>
 22. Banks, J. 2010. Regulating hate speech online. *International Review of Law, Computers & Technology*. 24 (3): 233–239. <https://doi.org/10.1080/13600869.2010.522323>.
 23. Sangsuban, K. 2013–2014. Balancing freedom of speech on the internet under international law. 39 N.C.J. *Int'l L. & Com. Reg.* 701
 24. Bleich, E. 2015. Freedom of expression versus racist hate speech: Explaining differences between high court regulations in the USA and Europe. In Maussen, M. and Grillo, R. (Eds.) *Regulation of Speech in Multicultural Societies*. Routledge.
 25. Horder, J. 2021. Online free speech and the suppression of false political claims. *Journal of International and Comparative Law* 8 (15): 2021.
 26. Droin, N. 2018. L'appréhension des discours de haine par les juridictions françaises : Entre travail d'orfèvre et numéro d'équilibriste. La Revue des droits de l'homme. *Revue du Centre de recherches et d'études sur les droits fondamentaux*, 14, Article 14. <https://doi.org/10.4000/revdh.4302>
 27. Masson, F.-X., and P. Mariatte. 2018. Pharos et la lutte contre le discours de haine sur internet. *Archives de politique criminelle* 40 (1): 127–142.
 28. Doyle, N. 2019. *Pervasive Racism or American Giants? The target of France's hate speech law*. Australian Institute of International Affairs.
 29. The grey zone: as discussed in the debates: e.g., Mme Frédérique Dumas, XVe législature, Session ordinaire de 2019–2020, Séance du mardi 21 janvier 2020, <https://www.assemblee-nationale.fr/dyn/15/comptes-rendus/seance/session-ordinaire-de-2019-2020/deuxieme-seance-du-mardi-21-janvier-2020>
 30. Council of Europe. 2022. *Recommendation CM/Rec(2022)16 of the Committee of Ministers to member States on combating hate speech* (Adopted by the Committee of Ministers on 20 May 2022 at the 132nd Session of the Committee of Ministers), <https://rm.coe.int/native/0900001680a67955>

31. Claire de Galember. Les trois corps du parlementaire. Un débat d'assemblée au prisme d'Alceste. Faire parler le parlement. Making Parliament Speak, Oct 2011, Paris, France. (halshs-00832729) <https://halshs.archives-ouvertes.fr/halshs-00832729>
32. Moor, P. 2022. Le droit et ses limites: Le juridique et le non-juridique. *International Journal for the Semiotics of Law* 35: 71–91. <https://doi.org/10.1007/s11196-020-09776-9>.
33. Moor, P. 2013. Les figures de l'ordre juridique dans les relations entre le droit et son environnement. *International Journal for the Semiotics of Law*. <https://doi.org/10.1007/s11196-013-9332-y>.
34. Eemerin, F. van. (ed). 2017. *Prototypical Argumentative Patterns: Exploring the relationship between argumentative discourse and institutional discourse*. John Benjamin Publishing.
35. Karsenti, B. 2014. Chapitre 19. Foucault et la parole de vérité des modernes. In *Usages de Foucault* (p. 319–334). Presses Universitaires de France. <https://www.cairn.info/usages-de-foucault--9782130621102-page-319.htm>
36. Chevallier, J. 2004. Science politique et science juridique. In O. P. Eric Darras (Éd.), *La science politique une et multiple* (pp. 41–59). L'Harmattan. <https://hal.archives-ouvertes.fr/hal-01735006>
37. Rastier, F. 2011. *Cartographie de la sémiotique actuelle* / *Mapping Current Semiotics*. Presses Universitaires de Liège-Sciences Humaine. La-semiotique-visuelle-grands-paradigmes-et-tendances-lourdes.pdf (researchgate.net)
38. Legislative dossier from which the corpus was constructed available at: https://www.assemblee-nationale.fr/dyn/15/dossiers/lutte_contre_haine_internet
39. Assemblée nationale. 2019. *Welcome to the english website of the French National Assembly, Passing a law*, https://www2.assemblee-nationale.fr/langues/welcome-to-the-english-website-of-the-french-national-assembly#node_9511
40. Loubère, L. and Ratinaud, P. 2014. *Documentation IRaMuTeQ 0.6 alpha 3 version 0.1*. documentation_19_02_2014.pdf (iramuteq.org)
41. Lamalle, C., Martínez, W., Fleury, S., Salem, A., Fracchiolla, B., Kuncova, A. & Maisondieu, A. 2003. Lexico3: Outils de statistique textuelle. Manuel d'utilisation. Lexi&co. Available from: <http://lexi-co.com/ressources/manuel-3.41.pdf>
42. Van Dijk, T. A. 2000. On the analysis of parliamentary debates on immigration. In Reisigl M. & Wodak, R. (Eds.). *The semiotics of racism. Approaches to Critical Discourse Analysis*. (pp. 85–104).
43. Wagener, A. 2019. Pragmatique discursive du témoignage numérique Sexisme ordinaire dans le tumblr «Paye ta fac». *Interfaces numériques*. <https://doi.org/10.25965/interfaces-numeriques>
44. Everitt, B.S., S. Landau, M. Leese, and D. Stahl. 2011. Hierarchical clustering. *Cluster Analysis*. 5 (71–110): 71–110.
45. Gries, S. 2013. *Statistics for Linguistics with R: A practical introduction*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110307474>
46. Lebart, L., Salem, A., & Berry, L. 1997. *Exploring textual data*. Springer Science & Business Media.
47. Bourdieu, P. 1986. La force du droit. *Actes de la Recherche en Sciences Sociales*. 64 (1): 3–19. <https://doi.org/10.3406/arss.1986.2332>.
48. Foucault, M. 2011. *L'Ordre du discours*. Leçon inaugurale au Collège de France prononcée le 2 décembre 1970, Éditions Gallimard, Paris, 1971.
49. Foucault, M. 1980. *Power/knowledge: Selected interviews and other writings 1972–1977*. New York: Pantheon.
50. Loi sur la liberté de la presse du 29 juillet 1881. Law of 29 July 1881 on freedom of the press - Légifrance (legifrance.gouv.fr)
51. Mouron, P. 2017. Le renforcement des moyens de lutte contre les discours de haine diffusés en ligne. 5.
52. *European Observatory of Online Hate*. Online: <https://eoooh.eu/>