

# Grammatical Gender and Politics: A Comparison of French and English in Political Discourse

Ellina Zhang<sup>1</sup> and Julia Deng

<sup>1</sup>Saint Theresa of Lisieux CHS, Richmond Hill, Ontario L4C 0E8, Canada

## SUMMARY

Grammatical gender systems are prevalent across many languages, and when comparing French and English the existence of this system becomes a strong distinction. There have been studies that attribute assigned grammatical gender with the ability to influence conceptualization (attributing gender attributes) of all nouns, thus affecting people's thoughts on a grand scale. We hypothesized that due to the influence of a grammatical gender system, French political discourse would have a large difference between the number of masculine and feminine nouns used. Specifically, we predicted there would be a larger ratio of feminine to masculine nouns in French political discourse than in non-political discourse when compared to English discourse. Through linguistic analysis of gendered nouns in French political writing, we found that there is a clear difference between the number of feminine versus masculine nouns, signaling a preference for a more "effeminate" language. This preference can be attributed towards the utilization of the conceptualization of nouns in swaying public opinion. We examined the ratio of complex words (two syllables or more) to the total number of words across both English and French political and non-political writing. When comparing these four categories, we found a significant increase in complexity from non-political to political writing in English, all of which were more complex than French writing. This suggests that, due to a lack of a grammatical gender system, English resorts to grand but vague language that hinders delivery and clarity in political discourse.

## INTRODUCTION

In what ways does the existence or lack of a grammatical gender system in a language influence political discourse? As stated by Roman Jakobson, "languages differ essentially in what they must convey and not in what they may convey" (1). That is, languages are distinctly unique based on conceptual or formal properties that they do or do not share. For example, in English the word "friend" does not give the gender of the friend, whereas in French "ami" or "amie" would immediately make the distinction ("ami" is male, and "amie" is female). As suggested in many works, such differences in obligatory

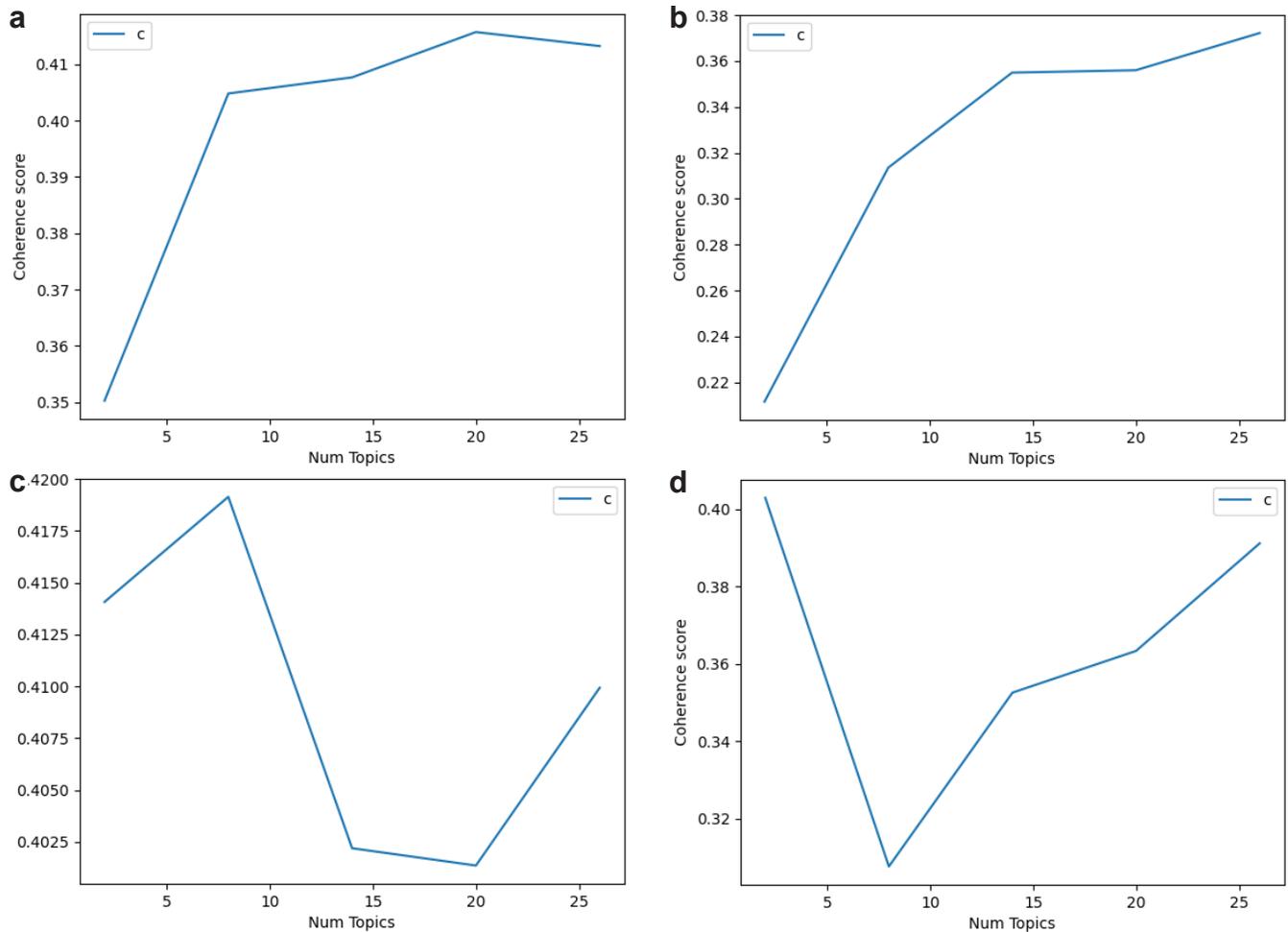
expression may imply that speakers of English and French would pay more or less attention to different lexical details (2,3). For example, French speakers may pay more attention to the sex of referents than English speakers. By extension, French speakers may have a tendency to think of objects in the world as more masculine or feminine on the basis of the word's grammatical gender (2). But how prevalent are these effects? And how do these differences in language affect the thoughts and writings of political discourse?

## Grammatical Gender

Grammatical gender is a system in most Romance languages where nouns are divided into classes based on articles and determiners (4). Languages like French possess a grammatical gender system that is split between male and female (masculine et feminine). Of course, English, being considered a Germanic language with Romantic influences, is a notable exception to this rule.

There are many different types of grammatical gender systems; however, the ones that garner an interest in research towards the effect of language on cognition usually share several properties. First, the system must have masculine and feminine genders (in French this is the le, la division) barring a few isolated exceptions. Second, the majority of nouns must fall into the semantic residue, meaning assignment is based on a word's phonological rather than semantic properties. For example, in French, almost all words that end in "-ion" are feminine. Third, a number of words in the semantic residue are already assigned a grammatical gender. Languages that fulfill these criteria provide the opportunity to ask questions such as: can the conceptualization of a noun gain gendered properties due to the assigned grammatical gender? If that is the case, then grammatical gender could exhibit a broad and pervasive effect on concept formation since almost all nouns require an assigned gender. This could then result in a broad range of concepts acquiring gendered properties that would otherwise be absent (5).

In regard to the effects of grammatical gender on cognition, many studies have determined a link between the gender assignment of an object and people's thinking towards the same object. In one study, they showed that 1. people do include gender in their conceptual representations of inanimate objects, and 2. the grammatical gender assigned to objects in the given language strongly influences people's ideas regarding the "gender" of an object (5). In particular, a study where participants were given proper names for objects



**Figure 1. Coherence Scores for English and French political and non-political discourse.**

a) English political coherence score. b) French political coherence score. c) French non-political coherence score. d) English non-political coherence score. Coherence score evaluates the quality of a given topic in terms of its coherence to a human.

(e.g., an apple may have been called “Patrick”) and were tested on their memory for these object name pairs later in the experiment, showed that when the proper name’s gender (i.e., Patrick) did not coincide with the assigned gender of the object, participants who spoke languages with a grammatical gender system had a much harder time remembering the proper names (3). In contrast, when the proper names did coincide with the grammatical gender, the participants with a grammatical gendered language remembered much more easily (3). This demonstrates that a person’s cognition, in this case their memory, is affected by the grammatical gender system under which their language operates.

How might people’s representations of objects be affected by the grammatical gender of their labels? One possibility is that in order to efficiently learn the grammatical gender of a noun, people focus on some property of that noun’s referent that may distinguish it as masculine or feminine. For example, in French the word for “sun” is “soleil”, which is masculine. One trying to remember this gender assignment may think of the sun in terms of what are perceived as stereotypically masculine properties, like powerful and threatening. In

comparison, in French the word for “moon” is “lune” which is feminine, meaning one might think of the moon as graceful and passive (stereotypical female qualities). Even after the grammatical genders of nouns are memorized, language may influence thought during “thinking for speaking” (6). Languages can force their speakers to attend to the genders associated with objects by making them grammatically obligatory. In the case of French, speakers need to refer to nouns with the gendered definite articles “le” and “la”, refer to objects using gendered pronouns (e.g., in French “hat” is “chapeau”, which is masculine, and we would have to refer to it as “il”, meaning “he”), and alter adjectives or even verbs to agree in gender with the nouns (e.g., describing a feminine noun as “beautiful” is “belle”, whereas masculine is “beau”). This need to constantly make the distinction between feminine objects and masculine objects throughout the language may lead people to selectively attend to that object’s masculine or feminine qualities (5).

This constant distinction between feminine and masculine qualities for gendered nouns creates implicit bias. In the study conducted in “How Language Affects Thought in

a Connectionist Model”, it was shown that the description of a given item was more male- or female-oriented depending on the grammatical gender of the item in that language. Additionally, this result persisted and was still highly significant even with extreme verbal interference. This finding supports the view that linguistic information helps shape semantic representations (7). Furthermore, in another study, there was evidence that gender information is central in people’s mental representations, meaning the assigned gender of an object creates a preconceived representation based on feminine and masculine descriptors alone. This effect is quite pervasive, as it has the potential to affect anything that could be named by a noun (4).

### Thinking in Politics

We chose to analyze political discourse in particular because political discourse and the persuasion it requires can be difficult. Having different ideologies and a competition for scarce resources gives political actors reasons to mistrust one another. However, persuasion requires a basis of trust, or else a listener is not inclined to change their beliefs if they think the speaker’s words are not to be trusted. The factors that can generate this basis for trust, persuasion and the possibility of learning, are based on the usages of specific language and information shortcuts in the form of heuristics (8).

Thus, the influence of grammatical gender is very useful in politics as the information shortcuts (connections between nouns and concepts) it creates are a large factor of political thinking. These shortcuts, or heuristics, are the basis on which citizens formulate opinions on politics and policies. By using a heuristic regarding a specific viewpoint, a citizen can attribute their position to be closer to one group than another — in effect, using these tendencies towards groups and ideas as information regarding their position. In regard to politics, most citizens face conditions of limited information and by any serious standard even relatively well-informed people fall short of being truly well informed (8). Therefore, heuristics, which can sway people based on more than fact, are indispensable to any citizen trying to make a well-informed judgement. Accordingly, the presence of existing heuristics in the makeup of a language becomes an invaluable tool to politicians who may use these heuristics to influence public decisions. Furthermore, due to its lack of a grammatical gender system and the accompanying heuristics, in political discourse English is at a distinct disadvantage (politicians lack communication tools, which also hinders the understanding of listeners), and would have to compensate for this through more complex speech. By not having information shortcuts built into the language by a grammatical gender system, English discourse would need to elaborate on details that would already be expressed in French discourse. Specifically, English discourse may require the usage of words that have much more specific definitions, generally longer words requiring more syllables. We refer to

the presence of these complex words (defined as two or more syllables) to be “political jargon” and estimate there is a much higher occurrence of these words in English than in French and more in political discourse than in non-political discourse.

This study hypothesizes that due to the influence of a grammatical gender system French will have a large difference between the number of masculine and feminine nouns used. This hypothesis was tested through the usage of topic-modeling to identify political and non-political articles in French and English. Then, the ratio of feminine to masculine nouns was compared in French political and non-political articles. The final results were subject to a t-test and modeled using a 2x2 ANOVA model. Overall, this study’s results implied that French politicians might unknowingly have a tendency towards feminine nouns, and thus that a grammatical gender system’s existence influences political discourse.

### RESULTS

In analyzing the vocabulary used in French versus English political discourse, this study used linguistic modeling in Python to determine which words are the most common in both areas. The linguistic model took into account topic parameters, such as “politics,” and divided the collected words into a specified number of topic distributions. Using “politics” as a parameter keyword, this study analyzed about 2000 articles in both English and French in the categories of political and non-political writings. In order to optimize the semantic similarity between keywords, this study analyzed the coherence score (measure of semantic similarity) in each model based on the number of topic distributions, as seen in Figure 1. This figure depicts the overall coherence score for English political and non-political discourse and French political and non-political discourse, respectively, compared to the number of topics that the words were grouped into. Using the number of topic distributions with the greatest coherence scores, these models led to finalized linguistic models that were further analyzed. Each of four models generated the 30 most salient terms across the four categories, and these words are depicted in Table 1.

In the analysis of the most common words across all topics in French political writing, this study looked at the occurrence of feminine nouns in relation to masculine nouns. This analysis was done on the French political and French non-political models, finding the average ratio of feminine to masculine nouns across the most salient terms in the articles. There was a percentage difference of 61.4% to 38.6% between feminine nouns and masculine nouns, respectively, in French political writing, compared to the 57.4% to 42.6% ratio in French non-political writing (Figure 2). French political writing contained more feminine nouns compared to non-political writing ( $t = 1.66$ ,  $p = 0.0489$ ). Thus, there exists an implicit or explicit choice in favor of some nouns over others based on assigned gender, and are apparent in the ratio of feminine to masculine nouns (Figure 2).

French Political	French Non-Political	English Political	English Non-Political
Politique	Pay	Trump	Player
Crise	Enfant	People	Sport
Pay	Tous	Government	Club
Opposition	Patient	Say	Game
Mouvement	Projet	Former	Play
Publique	Autre	Economy	People
Affaire	Personne	Covid	Community
Partis	Contre	Money	Money
President	Toute	Cumming	Receive
Point	Developpement	Pay	Case
Election	Leur	Report	Project
Semble	Economie	Economic	Fund
Tour	Plan	Crisis	Virus
Tous	Parent	Company	Report
Question	Etude	Health	Covid
Place	Risque	Official	Say
Responsable	Donc	Test	Sign
Possible	Place	Party	Test
Ensemble	Faire	Financial	Team
Comment	Entreprise	Virus	Fan
Face	Face	Market	Allow
Solution	Non	State	Close
Leur	Comme	Pandemic	Hold
Fait	Question	Case	Government
Demo	Encore	Political	Season
Debat	Fait	Business	Health
Discour	Traitement	Climate	Education
Situation	Ver	Labour	News
Service	Peuvent	Police	School
Avant	Impact	Use	Life

**Table 1. Salient Term Comparison.** The 30 most salient terms across English and French, Political and non-Political; same colored terms indicate the repeated terms.

Furthermore, when we analyzed the complexity of the four categories (English political, English non-political, French political, French non-political) we see from Figure 3) that English political writing has an average complexity score of around 0.392 while English non-political writing has an average of 0.148, a 153% increase. Additionally, the French political writing has an average of about 0.126 and the French non-political writing has an average of 0.115 — both lower than the English writings and a smaller increase of 9.5% (Figure 3). Thus, English on average has more complex words than French, and English political writing sees a dramatic increase in complexity.

Furthermore, in analyzing the amount of “political jargon” present in the two languages and across political versus non-political discourse, we computed a “complexity” score for each article and then used a 2x2 ANOVA to analyze the proportion of complex language in each condition (Table 2 and 3). This study defined a complex word as requiring two or more morphemes. (9) Furthermore, a complexity score is the ratio of complex words to overall word count in an article. In Figure 3, we see the mean complexity score for each of the four categories (English political, English non-political, French political, French non-political). In Table 2 there is the table depicting the mean and standard deviation for each

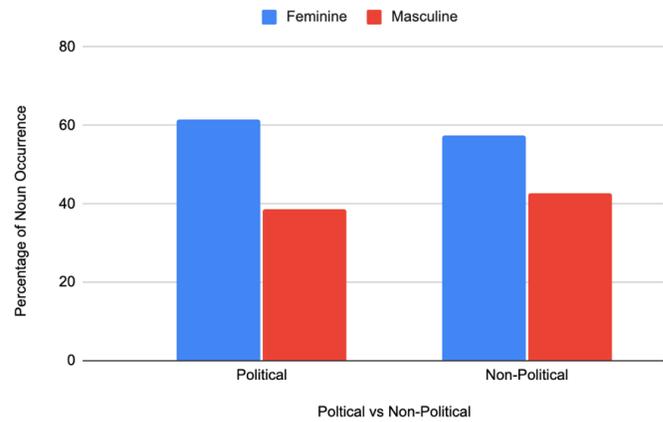
combination of the groups of the independent variables. In addition, the table provides “total rows”, which allows the means and standard deviations for groups only split by one independent variable, or none at all, to be known. Finally, Table 3 depicts the statistical significance of the 2x2 ANOVA. The particular rows we were interested in are the "Political", "Language" and "Political\*Language" rows. These rows informed us whether the independent variables (the "Political" and "Language" rows) and their interaction (the "Political\*Language" row) had a statistically significant effect on the dependent variable, "complexity score". For the “Political\*Language” row we can see from the "p-value" column that we do not have a statistically significant interaction with  $p < 0.05$ . Thus, we did not need to consider the interaction effect between the two variables.

### DISCUSSION

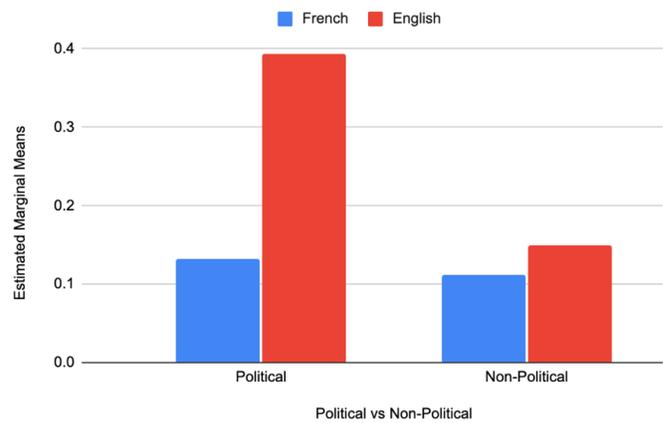
This study analyzed the occurrence of words across French and English media sources in political writings, in order to ascertain whether the grammatical gender system present in French and not in English could influence the effectiveness of political writing. The analysis of French language in political writing demonstrated that there was (1) a preference towards feminine nouns over masculine nouns and thus an (2) implicit bias towards one type of noun over another based on grammatical gender assignment. We hypothesized that this vocabulary choice was due to the lack of heuristics from a lack of grammatical gender system in the English language, thus leading to a need to compensate with jargon that the public can recognize but that may not necessarily convey much clear information. Overall, this may suggest that languages that have grammatical gender systems are more effective in political writing and discourse than those without it. As a next step, this study would analyze whether French political discourse is truly “more” effective than English politics based on a variety of parameters (i.e., public trust in government, political participation, etc.). Another avenue of study would also be to analyze why feminine nouns are seemingly preferred over masculine nouns, whether for societal, historical, or political reasons.

To summarize, this study investigated the effect of a grammatical gender system on political discourse by analyzing the ratio of feminine to masculine nouns in French political versus non-political writing. Additionally, we compared the language complexity of English and French discourse to support the hypothesis of English requiring more “jargon” due to not having the heuristics that grammatical gender provides in French.

These results supported our hypothesis regarding how a grammatical gender system in French will have an influence on political discourse, specifically in the ratio of feminine to masculine nouns. One possible explanation is that the political vocabulary used is not designed for the purpose of putting language to existing reality, but rather in order to perform a specific political function: putting existing reality in the service



**Figure 2: Gendered Noun Occurrence.** Masculine vs Feminine noun occurrence in French Political writing versus French Non-Political Writing.



**Figure 3: Estimated Marginal Means of Complexity.** 2x2 ANOVA (English vs. French) vs. (Political vs. Non-Political) marginal complexity score.

Language	Mean	Std. Deviation	N
Non-Political			
English	0.148	0.030	890
French	0.115	0.026	200
Total	0.142	0.032	1090
Political			
English	0.392	0.038	890
French	0.126	0.023	200
Total	0.343	0.109	1090
Total			
English	0.270	0.127	1780
French	0.121	0.025	400
Total	0.243	0.129	2180

**Table 2: Descriptive Statistics.** 2x2 ANOVA (English vs. French) vs. (Political vs. Non-Political) Descriptive Statistics (Dependent Variable: Complexity). N = sample size.

Source	Type of III Sum of Squares	df	Mean Square	F	p-value
Corrected Model	33.79	3	11.26	10393.48	0.0
Intercept	49.92	1	49.92	46058.59	0.0
Political	5.30	1	5.30	4890.21	0.0
Language	7.28	1	7.28	6718.57	0.0
Political*Language	4.44	1	4.44	4100.11	0.0
Error	2.36	2176	.001		
Total	164.59	2180			

**Table 3: Tests of Between-Subjects Effects.** 2x2 ANOVA (English vs. French) vs. (Political vs. Non-Political) Between Subjects Effects (Dependent Variable: Complexity ); R Squared = .935 (Adjusted R Squared = .935). Abbreviations: df = degrees of freedom, F = variation between sample means.

of politics through language. This political vocabulary varies from country to country, from language to language. Political vocabulary is distinguished by its pragmatic focus, by the fact that it does not necessarily reflect any existing situation, but rather represents an idea that a speaker is trying to influence. The most defining trait of political vocabulary is its “increased sensitivity to changes in life and state” (10), meaning political vocabulary would be strongly concerned with the nuances of public opinion and how to take advantage of existing preconceptions. Thus, the existence of heuristics within the grammatical gender system in French would be implicitly or explicitly appealing to users of political vocabulary (10).

In comparison, this study hypothesized that English, which lacks a grammatical gender system, would be at a disadvantage in politics and would have to resort to grander political terms in order to compensate. In the analysis of English political writings, when the most used words are compared against the most used words in other subjects of discourse, there is a large increase in political jargon, such as “democracy”, “freedom”, “patriotic”, etc. There is also an increase in Latin or Greek words such as “ameliorate”, “expedite”, etc. This is an overuse of grand language that is not present in our French analysis.

### Limitations

This study has potential limitations. The data comparison of feminine to masculine verbs in French political writing versus French non-political writing yielded that French political writing has a higher ratio of feminine nouns. However, this does not take into account the fact that many “political” nouns (i.e., la démocratie, la politique, etc.) are inherently feminine. Furthermore, we predict that the presence of a grammatical gender system uses heuristics to influence public opinion, leading to the significant difference in feminine and masculine nouns. However, although we studied the variance in complexity between languages as well as the ratio of feminine to masculine nouns, we did not quantitatively measure the influence of the use of feminine language on public opinion.

### Further Discussion

The existence of such patterns in political writing is in line with George Orwell’s criticism on the degradation of politics in the English Language. Orwell points out that modern English is full of bad habits which impair clear thinking and political writing. The writer usually has a meaning they cannot fully express to the larger public and, in attempting to be coherent, ends up with a mixture of vagueness characteristic of modern English political writing. As stated previously, the topics raised in politics are seldom well understood by the public, and when confronted with this uncertainty, English writers turn towards the abstract and prose becomes less choice words and more an amalgamation of random phrases (11).

The increase of Latin and Greek words in the analysis follows Orwell’s argument of how these words represent the vagueness that proliferates political writing. By utilizing Latin or Greek roots, it is much easier to make up words like “deregionalize” rather than think of an English word that fully conveys one’s meaning. The general result is an increase in vagueness and slovenliness, leading to a lack of clear understanding for readers and listeners. Furthermore, in the cross-analysis between English in politics and English in other topics, the increase in “political words” is another sign of sloppiness in English political writing. The words “democracy”, “freedom”, and “patriotic” appear commonly despite having several different meanings that cannot be reconciled. In the case of “democracy”, there is an inherent barrier against a proper definition as it is seen as universal praise for a regime to be “democratic”. The existence of a variety of governments and policies all under the wing of “democracy” makes the term completely vague and mostly meaningless when attempting to convey a singular idea. However, towards a general public who lack proper understanding of political theory and with a lack of heuristics provided by assigned gender as in French, only overused jargon such as “democracy” would result in some form of understanding (11).

### MATERIALS AND METHODS

The methodology of this study was the linguistic analysis

of English and French media sources through Gensim Topic Modeling in Python. Taking a large data set of English and French political articles dated from January 1st, 2020 to July 21st, 2020, we used the Latent Dirichlet Allocation (LDA) from the Gensim package along with the MALLET implementation. The English articles were pulled from the Guardian newspaper with an application programming interface (API) key, and the French articles were pulled from NewsAPI. All the articles we considered were writing pieces. The parameter used for the “politics” articles was “politic” while the “non-politics” articles used the exclusion of this parameter. In order to maintain equal sample size for each category, this study took the first 500 articles generated by the API. MALLET has an efficient implementation of LDA and it is known to run topic segregation. The results for French language were then re-analyzed based on the number of feminine nouns versus masculine nouns. The two English language models were cross-analyzed for differences in vocabulary.

The code for the LDA model takes the raw data from online articles through an API. The data was read in json form so that it was synthesizable by the LDA. In order to avoid common conjunctions or articles such as “and” or “the” appearing in the LDA’s output, English and French stop words were imported and removed from the text. Using the regular expressions package, empty spaces and phrases such as emails and websites are removed. In order to further clean up the text using Gensim’s simple processes, each sentence was tokenized into a list of words and punctuation was removed. The two main inputs to the LDA topic model were the dictionary and the corpus. Gensim created a unique id for each word in the document (corpus). The produced corpus is a mapping of (word\_id, word\_frequency).

The LDA considered each document as a collection of topics in various proportions with each topic having a collection of keywords in various proportions. Once the model was created and trained, it was provided with the number of topics and it rearranged the topic distribution within the documents and keywords in order to receive a good composition of topic-keywords distribution. In order to tune the number of segregation topics, the following parameters were considered: 1) Quality of text processing, 2) Variety of topics, 3) Choice of topic modeling algorithm, and 4) Number of topic divisions.

One particular indicator of the effectiveness of a topic model that this study considered is the topic coherence. Topic coherence evaluates the quality of a given topic in terms of its coherence to a human. This means after a human learns topics from the same collection of data how coherent their learned topics were in terms of their interpretability and association with a single overarching concept. The topic coherence of the model can vary based on the number of topic divisions made, therefore the study graphs the coherence values across a number of topic divisions to determine the optimal number of divisions (12). Coherence values are measured based on a one-set segmentation of the top words and an

indirect confirmation measure using normalized pointwise mutual information and the cosinus similarity. The measure retrieves the co-occurrence count for a given set of words; the counts are used to then calculate the NPMI of every top word to every other top word, resulting in a set of vectors. The one-set segmentation calculator calculates the similarity between these vectors using the cosinus similarity measure. The coherence value is then the mean of these similarities (13).

First, we used topic modeling to identify which articles were actually talking about politics vs non-politics in both French and English, and analyzed the most salient words in each topic breakdown. Second, we computed the proportion of feminine nouns in all the political and non-political French articles (proper nouns like “Trump” were removed from such a calculation) and used a t-test to determine that the proportion of feminine nouns is significantly higher in French political articles than non-political articles. Then using Python’s Pyphen and NLTK libraries we computed the proportion of complex words (words with 2 or more syllables) and used a 2 x 2 ANOVA (fully crossed design: French vs English x political vs non-political articles) to analyze the proportion of complex language in each condition.

## REFERENCES

1. Jakobson, Roman. “On Linguistic Aspects of Translation.” *On Translation*, 1959, edited by Reuben Arthur Brower, Harvard University Press, 2013, pp. 232-239., doi:10.4159/harvard.9780674731615.c18.
2. Vigliocco et al. *Grammatical Gender and Meaning*. 2004, doi.org/http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.448.5201&rep=rep1&type=pdf.
3. Boroditsky, L., Schmidt, L., Phillips, W. “Sex, Syntax, and Semantics.” *Language in Mind*, The MIT Press, 2003. DOI.org (Crossref), doi:10.7551/mitpress/4117.003.0010.
4. Speed, Laura J., and Asifa Majid. “Linguistic Features of Fragrances: The Role of Grammatical Gender and Gender Associations.” *Attention, Perception, & Psychophysics*, vol. 81, no. 6, Aug. 2019, pp. 2063–77. DOI.org (Crossref), doi:10.3758/s13414-019-01729-0.
5. Semenuks et al. “Effects of Grammatical Gender on Object Description.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2011.
6. Slobin, D. “Reevaluating Linguistic Relativity.” *Language in Mind*, 2003. doi:10.7551/mitpress/4117.003.0022.
7. Boroditsky, L. “How Language Affects Thought in a Connectionist Model.” *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2007, doi.org/https://escholarship.org/uc/item/7q47j690.
8. Lupia, Arthur, Mathew D. McCubbins, Samuel L. Popkin, A Lupia, M D McCubbins, and S L Popkin. *Reconsidering the Rational Public: Cognition, Heuristics, and Mass Opinion*. In *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*, 155–59. Cambridge:

Cambridge University Press, 2000.

9. Denning, Keith, et al. English Vocabulary Elements. 2nd Ed., Oxford University Press, 2007, [www.worldcat.org/title/english-vocabulary-elements/oclc/70699031](http://www.worldcat.org/title/english-vocabulary-elements/oclc/70699031).
10. Djachy, Ketevan, and Mariam Pareshishvili. "The Peculiarities of the Political Vocabulary in French and Georgian Languages." Theory and Practice in Language Studies, vol. 4, no. 9, Sept. 2014, pp. 1786–91. DOI.org (Crossref), doi:10.4304/tpls.4.9.1786-1791.
11. Orwell, George. Politics and the English Language. London: Renard Press, 2021.
12. David, Newman. Automatic Evaluation of Topic Coherence. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.
13. M. Röder, A. Both, and A. Hinneburg. Exploring the Space of Topic Coherence Measures. In Proceedings of the eighth International Conference on Web Search and Data Mining, 2015.

**Article submitted:** November 6, 2020

**Article accepted:** January 30, 2021

**Article published:** July 7, 2021

**Copyright:** © 2021 Zhang and Deng. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.