Multilingual and cross-cultural automatic modelling and analysis of argumentation structures in political debates

Cecilia Graiff

ALMAnaCH, Inria Paris

Abstract

This paper describes the topic of a Ph.D. thesis focusing on the automatic modelling and analysis of argumentation structures in a multilingual corpus of political data. More specifically, it merges argumentation theory and natural language processing by leveraging an argumentation framework to help argument retrieval by machine learning models. Furthermore, this thesis focuses on analyzing differences in argument structures and in their logic, thus assessing correlations to cultures and political trends.

1 Introduction

The modelling of argumentation structures in political debates falls between the realms of Natural Language Processing (NLP), argumentation theory, and political sciences. This multidisciplinary Ph.D. thesis aims to analyse argumentation structures and styles in a multilingual corpus of political data by leveraging both argumentation theory and NLP. Thus, we define the following research questions (RQ):

1. **RQ 1**: Is the merging of formal and propositional argumentation and NLP a valuable strategy to increase the performance of machine learning (ML) and large language models (LLMs) in analysing argumentation structures? Even though some studies rely on logical formalizations of fallacious arguments [Helwe et al., 2024, Sourati et al., 2023] or analyze the correlation between LLMs and argumentative reasoning [Castagna et al., 2024], most previous works about argument mining (AM) [Lawrence and Reed, 2020] focus on linguistical features, thus leaving this question unanswered.

2. **RQ 2**: How can the robustness of argument extraction algorithms be improved to deal with multilingual data, as well as with intra- and intercultural biases? Arguments proposed by different socio-cultural groups can vary greatly [Maurer et al., 2024]. Only few studies were conducted on linguistic variation in argumentation [Ruiz-Dolz et al., 2024]. Due to the inherent difficulty of generalizing in argument mining, this research question carries extreme importance.

3. **RQ 3**: Are political trends reflected by the argumentative strategies of politicians? Some studies analyze the connection between language style and populism [Decadri and Boussalis, 2020] [Kittel, 2025]. As this thesis will try to assess the the evolution of the quality of political argumentation, it may help improve our understanding of ideological and affective polarisation.

The first necessary step to be able to perform the above mentioned researches is the creation of an annotated multilingual corpus of political texts. In the next section, we will start by describing the data collection and argument extraction process, and we will then focus on the planned future work.

2 Method

Data

The used dataset currently includes US English and French, and will be expanded to German (Austria and Germany), Italian, Spanish, and UK English. The dataset for US English is ElecDeb60to20 [Goffredo et al., 2023], a collection of presidential debates in the US from 1960 to 2020 annotated for argumentative components and relations. The lack of annotated datasets in other languages highlighted the need to create one. The French corpus is composed by the transcripts of 8 publicly available presidential debates from 1974 to 2022. When a transcript already existed, it was scraped; alternatively, the video was transcribed and diarised with WhisperX [Bain et al., 2023], which provides fast automatic speech recognition (70x real time with large-v2) and was chosen over Whisper because it includes speaker diarisation. Moreover, the transcripts of French presidential speeches from 1958 until 2024 were scraped from the official website of the Élysée. The next section presents our formalization of arguments and their features, followed by a description of the annotation process for the unlabeled datasets.

Argument formalization

Similarly to [Goffredo et al., 2023], we define arguments as a set of propositions, divided in claims and premises (argumentative components), where the premises have the function of supporting or attacking the claims (argumentative relations). Because we consider argument components and relations based on their internal logical structure, we rely on propositional logic to define the information that we want to extract:

1. A set of arguments A, composed of a set of premises P and a set of claims C. We define claims and premises so that $\forall p \in P(\exists c \in C(p \rightarrow c))$. We thus allow the existence of claims that are not supported by premises, but not of premises that do not support any claim. We also choose to not limit the amount of accepted claims to only one, to allow more complex argument structures.
2. A support relation between premise p_1 and claim c_1, so that $p_1 \rightarrow c_1$.
3. An attack relation between premise p_1 and claim c_1, so that $p_1 \rightarrow \neg c_1$.

Because premises can be combined in several ways to build an argument, we include the relation type in the annotation scheme, based on [Peldszus et al., 2016]. We include 4 possible types of support relations: simple support ($p \rightarrow c$), convergent support ($p_1 \vee ... \vee p_n \rightarrow c$) occurring when all premises independently support a claim, linked support ($p_1 \wedge ... \wedge p_n \rightarrow c$) occurring when all premises are necessary to support or attack a claim, and serial support ($p_1 \rightarrow ... \rightarrow p_n \rightarrow c$) occurring when premises are interrelated and jointly support the claim. We also include 2 possible types of attack relations: the rebutter, defined as $p_3 \rightarrow \neg p_1$ in case of attack against a premise, and $p_3 \rightarrow \neg c_1$ in case of attack against a claim, and the undercutter ($p_3 \rightarrow \neg(p_1 \rightarrow c_1)$), occurring when the relation itself is attacked.

For each premise, we also started annotating the type of argument it builds, based on the 14 argument schemes introduced by [Walton et al., 2012] extended by the two schemes introduced in [Walton and Hansen, 2013]. We add this information to the premises in order to overcome the limitations of most argument schemes, which are too simple and strict for natural language arguments. Due to the high reasoning complexity of this task, we initially concentrated on components and relationships, with plans to leverage LLMs later for full task completion. To deal with the subjectivity of the task, the results will be manually checked, and an Inter-Annotator Agreement (IAA) metric will be chosen for evaluation.

Semiautomatic annotation workflow

To conduct preliminary experiments, we use the INCEpTION platform [Klie et al., 2018] to manually

annotate a portion of the dataset, namely the first (Estaing-Mitterrand, 1981), middle (Chirac-Jospin, 1995), and last debate (Macron-LePen, 2022) from the French presidential election. We annotate 84870 tokens following the Beginning-Inside-Outside (BIO) scheme [Ramshaw and Marcus, 1995], which marks tokens as beginning, inside, or outside an argument component.

To overcome the problem of annotation cost for the whole dataset, we propose a semiautomatic approach. Following [Yeginbergen et al., 2024], we start by automatically translating ElecDeb60to20 [Goffredo et al., 2022] into the 4 target languages (French, German, Spanish, and Italian) with OpusMT [Tiedemann and Thottingal, 2020]. The subsequent step is the projection of annotations, performed with SimAlign [Sabet et al., 2021]. This operation results in a multilingual dataset that we use for model fine-tuning. We initially relied on several BERT-based models, among which ModernBERT [Warner et al., 2025], chosen for its ability to process longer context windows of up to 8,192 tokens. Due to insatisfactory results, we decided to adopt a token-level approach instead, and fine-tune a BERT-based model with a CRF as last layer, starting from the implementation by [Goffredo et al., 2024]. The fine-tuned model is then tested on the annotated portion of the target dataset, and the results are iteratively improved following an active learning (AL) process.

Future steps

In the second main step of the project, we will focus on the analysis of argumentation structures leveraging an argumentation framework (AF) to be adapted to political debates. We orient our definition of AF to the ASPIC$^+$ framework [Modgil and Prakken, 2014]:

1. *An argumentation framework is a triple* $\mathcal{A} = (\mathcal{L}, \mathcal{R}, n)$*, where:*

- $\mathcal{L}$ *is a logical language closed under negation* ($\neg$)*.*
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a set of strict* ($\mathcal{R}_s$) *and defeasible* ($\mathcal{R}_d$) *inference rules*
- n *is a partial function such that* $n : \mathcal{R}_d \to \mathcal{L}$*.*

Defeasible inference rules are particularly suited to natural language arguments, which often do not follow strict logical rules. The extraction of a knowledge base also is necessary to define the acceptability of the arguments and the analysis of the cultural and diachronic differences:

2. *A Knowledge Base (KB) is a set* $\mathcal{K} \subseteq \mathcal{L}$ *consisting of two disjoint subsets* $\mathcal{K}_n$ *(the axioms) and* $\mathcal{K}_p$ *(the ordinary premises). [Modgil and Prakken, 2014]*

To make this step clearer, we propose the following example, extracted from the debate between Valéry Giscard d'Estaing and François Mitterrand (1981):

> [...] mais la Constitution dit que le Gouvernement est responsable devant le Parlement. Si vous agissez ainsi que vous dites, c'est-à-dire si votre Gouvernement ne va pas devant le Parlement, il ne pourra rien faire [...]. Nous n'avons pas, en France, un système tel qu'un Président de la République [...] puisse nommer un Gouvernement qui ne rende de comptes à personne et qui prenne des décisions importantes. [...] Si, donc, il ne va pas devant le Parlement, il ne pourra [...] qu'expédier les affaires courantes. [...] J'ai lu dans vos déclarations [...] que vous aviez l'intention de prendre tout de suite un certain nombre de mesures. [...] Ce sont des décisions qu'un Gouvernement chargé des affaires courantes ne peut pas prendre ! Donc, la procédure que vous indiquez n'est pas réalisable.

Giscard d'Estaing argues against Mitterrand by relying on the French Constitution, which in the scheme by [Walton et al., 2012] constitutes an argument from (legal) authority. In this example, the KB can be built from the cited law and formalised as follows:

$$\forall x \, (A(x) \wedge \neg B(x)) \tag{a}$$

$$\neg B(x) \to (C(x) \wedge \neg D(x)) \tag{b}$$

In natural language terms, we could say that for all x, if x is a government ($A(x)$), then x cannot not present itself to the Assembly ($\neg B(x)$). If x does not

go in front of the Assembly, then he will be able to only take care of current affairs ($C(x)$) and not to fully govern ($\neg D(x)$). Giscard d'Estaing then leverages a *modus ponens* and a deductive syllogism to build his argument starting from these premises. Because the KB is extracted from the text, we are not sure about its reliability. Especially in political discourse, it is not uncommon that an argument purposely relies on false grounds, e.g. as means for propaganda. Thus, we leverage subjective logic for argument formalization, as it allows modeling subjective opinions and states of uncertainty. Subjective logic conditional deduction is formalized as follows: $\omega_{C\|P} = \omega_P \odot \omega_{C|P}$, where $\odot$ denotes the general conditional deduction operator for subjective opinions, and $\omega_{C|P} = \{\omega_{Y|x_i} \mid i = 1, \ldots, k\}$ is a set of $k = |X|$ different opinions conditioned on each $x_i \in X$ respectively [Josang, 2009]. Our analyses will thus rely on the extracted arguments, a formalisation of the argument scheme, and a knowledge base. Particular emphasis will be placed on the dynamic nature of the KB, which is partly composed by implicit or common-sense facts [Becker et al., 2020]. We also consider leveraging existing cultural commonsense knowledge datasets and methodologies, such as CANDLE [Nguyen et al., 2023] or MANGO [Nguyen et al., 2024]. While the planned employ of argumentation graphs and ML models and algorithms will be further developed during the course of my Ph.D, we aim at leveraging the reasoning capabilities of LLMs, thus combining logic and NLP and delivering significant technical contributions to argumentation analysis by developing robust algorithms. As this thesis seeks to be interdisciplinary, its contributions to computational social sciences should also be mentioned. Its modeling of argumentative structures across corpora spanning a broad time range could allow a diachronic analysis of polarization. We also consider introducing novel quantitative metrics for the analysis of radicality, taking into account diachronic and geographic differences. The analysis of gender roles could also be explored, with particular attention to the challenge posed by the imbalanced dataset, as female leaders are significantly underrepresented compared to male leaders.

3 Discussion

While several resources already exist for the analysis of political data, a multilingual political dataset annotated for argumentation structures fills a gap in current research. The high subjectivity of the task will be taken into account to further improve the annotation process, as in [Helwe et al., 2024]. In the next months, we will continue the implementation and deal with annotation biases in its results. Later, we will analyze the argumentation structures with argument graphs and ML models. While some research was already conducted on the translation of explicit premises and claims into logical argumentation [Ben-Naim et al., 2024] and a few works on AM in political debates partly leveraged argument components [Goffredo et al., 2023], this Ph.D. thesis aims at leveraging logical reasoning and translating into practice an argumentation framework.

4 Conclusion

This paper describes the methodology developed for a recently started Ph.D. thesis on the automatic modeling of argumentation structures in political data. The first contribution is planned to be a semiautomatic annotation process that delivers a labeled multilingual corpus of political discourse, an useful and not yet existent resource. We define an annotation scheme that formalizes argument components and support and attack relations based on [Peldszus et al., 2016]. We add argument schemes based on [Walton et al., 2012] and [Walton and Hansen, 2013]. Annotation biases and errors will be manually checked and possibily limited by involving more than one human annotators and computing IAA. Furthermore, we rely on an argumentation framework oriented to the ASPIC$^+$ structure [Modgil and Prakken, 2014] and a knowledge base for the analysis of different argumentation structures. This will allow us to benchmark the performances of LLMs for AM when logical thinking is involved, rather than only linguistical features [Lawrence and Reed, 2020]. Moreover, it will bring contributions to political sciences, by unveiling patterns related to specific political situations.

Acknowledgements

This work in this paper is carried in the ALMAnaCH team of Inria Paris (Institut national de recherche en informatique et en automatique), with a collaboration with SciencesPo Paris. I would like to thank my advisors Benoît Sagot and Chloé Clavel at Inria, and Emiliano Grossman at the Centre de données sociopolitiques (CDSP) and Centre d'études européennes et de politique comparée (CEE) of SciencesPo.

References

[Bain et al., 2023] Bain, M., Huh, J., Han, T., and Zisserman, A. (2023). Whisperx: Time-accurate speech transcription of long-form audio. *INTERSPEECH 2023*.

[Becker et al., 2020] Becker, M., Hulpuş, I., Opitz, J., Paul, D., Kobbe, J., Stuckenschmidt, H., and Frank, A. (2020). Explaining arguments with background knowledge. *Datenbank Spektrum*, 20:131–141.

[Ben-Naim et al., 2024] Ben-Naim, J., David, V., and Hunter, A. (2024). Understanding enthymemes in argument maps: Bridging argument mining and logic-based argumentation.

[Castagna et al., 2024] Castagna, F., Sassoon, I., and Parsons, S. (2024). Can formal argumentative reasoning enhance llms performances? *ArXiv*, abs/2405.13036.

[Decadri and Boussalis, 2020] Decadri, S. and Boussalis, C. (2020). Populism, party membership, and language complexity in the italian chamber of deputies. *Journal of Elections, Public Opinion and Parties*, 30(4):484–503.

[Goffredo et al., 2024] Goffredo, P., Cabrio, E., Villata, S., Haddadan, S., and Torres Sanchez, J. (2024). Disputool 2.0: A modular architecture for multi-layer argumentative analysis of political debates. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13):16431–16433.

[Goffredo et al., 2023] Goffredo, P., Chaves, M., Villata, S., and Cabrio, E. (2023). Argument-based detection and classification of fallacies in political debates. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11101–11112, Singapore. Association for Computational Linguistics.

[Goffredo et al., 2022] Goffredo, P., Haddadan, S., Vorakitphan, V., Cabrio, E., and Villata, S. (2022). Fallacious argument classification in political debates. In Raedt, L. D., editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4143–4149. International Joint Conferences on Artificial Intelligence Organization. Main Track.

[Helwe et al., 2024] Helwe, C., Calamai, T., Paris, P.-H., Clavel, C., and Suchanek, F. (2024). MAFALDA: A benchmark and comprehensive study of fallacy detection and classification. In Duh, K., Gomez, H., and Bethard, S., editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4810–4845, Mexico City, Mexico. Association for Computational Linguistics.

[Josang, 2009] Josang, A. (2009). Conditional reasoning with subjective logic. *Mult.-Valued Log. Soft Comput.*, 15.

[Kittel, 2025] Kittel, R. C. (2025). Simply speaking? language complexity among (non-)populist actors in parliamentary debates. *Government and Opposition*, page 1–31.

[Klie et al., 2018] Klie, J.-C., Bugert, M., Boullosa, B., Eckart de Castilho, R., and Gurevych, I. (2018). The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In Zhao, D., editor, *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

[Lawrence and Reed, 2020] Lawrence, J. and Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

[Maurer et al., 2024] Maurer, M., Romberg, J., Reuver, M., Weldekiros, N., and Lapesa, G. (2024). GESIS-DSM at PerpectiveArg2024: A matter of style? socio-cultural differences in argumentation. In Ajjour, Y., Bar-Haim, R., El Baff, R., Liu, Z., and Skitalinskaya, G., editors, *Proceedings of the 11th Workshop on Argument Mining (ArgMining 2024)*, pages 169–181, Bangkok, Thailand. Association for Computational Linguistics.

[Modgil and Prakken, 2014] Modgil, S. and Prakken, H. (2014). The aspic + framework for structured argumentation: a tutorial. *Argument and Computation*, 5(1):31–62.

[Nguyen et al., 2023] Nguyen, T.-P., Razniewski, S., Varde, A., and Weikum, G. (2023). Extracting cultural commonsense knowledge at scale. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 1907–1917, New York, NY, USA. Association for Computing Machinery.

[Nguyen et al., 2024] Nguyen, T.-P., Razniewski, S., and Weikum, G. (2024). Cultural commonsense knowledge for intercultural dialogues. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, CIKM '24, page 1774–1784, New York, NY, USA. Association for Computing Machinery.

[Peldszus et al., 2016] Peldszus, A., Warzecha, S., and Stede, M. (2016). Annotation guidelines for argumentation structure. *complete*.

[Ramshaw and Marcus, 1995] Ramshaw, L. and Marcus, M. (1995). Text chunking using transformation-based learning. In *Third Workshop on Very Large Corpora*.

[Ruiz-Dolz et al., 2024] Ruiz-Dolz, R., Chiu, C.-J., Chen, C.-C., Kando, N., and Chen, H.-H. (2024). Learning strategies for robust argument mining: An analysis of variations in language and domain. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10286–10292, Torino, Italia. ELRA and ICCL.

[Sabet et al., 2021] Sabet, M. J., Dufter, P., Yvon, F., and Schütze, H. (2021). Simalign: High quality word alignments without parallel training data using static and contextualized embeddings.

[Sourati et al., 2023] Sourati, Z., Venkatesh, V. P. P., Deshpande, D., Rawlani, H., Ilievski, F., Ân Sandlin, H., and Mermoud, A. (2023). Robust and explainable identification of logical fallacies in natural language arguments.

[Tiedemann and Thottingal, 2020] Tiedemann, J. and Thottingal, S. (2020). Opus-mt – building open translation services for the world. In Martins [et al.], A., editor, *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Switzerland. European Association for Machine Translation. Annual Conference of the European Association for Machine Translation , EAMT2020 ; Conference date: 03-11-2020 Through 05-11-2020.

[Walton and Hansen, 2013] Walton, D. and Hansen, H. (2013). Arguments from fairness and misplaced priorities in political argumentation. *Journal of Politics and Law*, 6(3):78–94.

[Walton et al., 2012] Walton, D., Reed, C., and Macagno, F. (2012). *Argumentation Schemes*. Cambridge University Press.

[Warner et al., 2025] Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Adams, G. T., Howard, J., and Poli, I. (2025). Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T., editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*

(Volume 1: Long Papers), pages 2526–2547, Vienna, Austria. Association for Computational Linguistics.

[Yeginbergen et al., 2024] Yeginbergen, A., Oronoz, M., and Agerri, R. (2024). Argument mining in data scarce settings: Cross-lingual transfer and few-shot techniques. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11687–11699, Bangkok, Thailand. Association for Computational Linguistics.