

Argumentation through Discourse Relations and Subjectivity: Introducing FreCaDiS, a French Multi-Genre Corpus

Joanna Blochowiak^{1,2}, Cristina Grisot²

¹Université catholique de Louvain, ²Universität Zürich,
Institut Langage et Communication, Louvain-la-Neuve, Belgium; CLARIN-CH, Zürich, Switzerland
{joanna.blochowiak, cristina.grisot}@uzh.ch

Abstract

This paper addresses a crucial yet understudied issue in argumentation studies: the distinction between explanations and justifications, and their interaction with subjectivity. Building on insights from Bex and Walton (2016), who highlight the importance of not conflating explanations with arguments, we propose a corpus-based approach to operationalize this distinction in French. We present FreCaDiS (**F**rench **C**orpus of **C**ausal **C**onnectives, **D**iscourse **R**elations, and **S**ubjectivity), a novel corpus of French texts annotated for explanatory and justificatory discourse relations and their perceived subjectivity. FreCaDiS comprises excerpts of 2–3 sentences drawn from five distinct genres—SMS, online discussions, blogs, press, and contemporary literature—spanning informal to formal registers. Specifically, we focus on sentences introduced by the connectives *parce que* and *car* (“because”) and annotate them along two dimensions: (i) discourse relation (explanation vs. justification) and (ii) subjectivity (subjective vs. objective). The corpus was annotated by three independent human annotators using complementary approaches: a holistic, an intuitive method for subjectivity and a guided, operationalized method for discourse relations. FreCaDiS provides a rich resource for the study of argumentation, causal discourse, causal connectives, and subjective interpretation in French and can support future work in computational argument mining, discourse analysis, and NLP applications.

Keywords: explanation, justification, subjectivity, causal connectives, argumentation, human annotations, different text genres, French

1. Introduction

The automatic identification of argument components and structures has become a central topic in NLP. Recent work has explored various architectures and methods for argument structure learning (Wei et al., 2024), domain adaptation and robustness (Ruiz-Dolz et al., 2024), and mining arguments in less conventional domains (Liu et al., 2024). These approaches have advanced the state of the art but often treat arguments as a homogeneous category, typically focusing on claims and premises without distinguishing whether a premise functions as an explanation or a justification.

Research in argumentation mining has proposed different frameworks for automatically discovering argumentative structures in texts. These approaches generally aim to identify all arguments in a document, determine their relations, and represent their overall structure—often as trees or graphs (e.g., Marcu 2000; Mochales Palau & Moens 2009; Buckingham Shum et al. 2003). Such work draws on linguistic, rhetorical, and computational theories to detect propositions that function argumentatively, without assessing their logical validity. However, these broad frameworks tend to overlook the finer-grained linguistic cues that distinguish between argumentative and non-argumentative discourse relations.

By contrast, recent linguistic and discourse-analytic studies highlight the need to align computational models with such distinctions. For

example, Ding et al. (2024) link argumentative structure to cohesion in student essays, showing that rhetorical and discourse cues carry important information for argument mining.

Our work builds on this insight by focusing on a specific subset of discourse relations—those introduced by causal connectives, such as the French *parce que* and *car*—and by operationalizing the distinction between explanatory and justificatory uses. Unlike most argumentation mining approaches that seek to reconstruct entire argumentative networks, our approach targets a single, linguistically grounded pattern: discourse relations expressed with causal connectives that are potential indicators of argumentation. Beyond discourse relations, we also annotate subjectivity, indicating whether each discourse segment is perceived by humans as subjective or objective.

This dual focus enables us to address two closely related questions: (i) how to empirically distinguish between arguments and explanations in authentic discourse, and (ii) how subjectivity interacts with these distinctions. To address these questions, we construct a new French corpus, FreCaDiS, annotated with these features across five different text genres.

The rest of the paper is organized as follows. We first discuss the importance of distinguishing between arguments and explanations, with a focus on discourse relations and causal connectives (Section 2). Next, we examine the complexity of the notion of subjectivity and

present the rationale for our holistic approach to it (Section 3). We then describe our annotation studies (Section 4). Finally, we outline the main conclusions and contributions of our research (Sections 5 and 6).

2. Arguments, Discourse Relations, and Causal Connectives

The distinction between arguments and explanations has long been recognized as an important issue in the study of discourse, reasoning, and argumentation (van Eemeren et al., 1996; Walton, 2004; Bex and Walton, 2016). As Bex and Walton (2016: 55) rightly observe: “it would be a fundamental error to criticize an argument as falling short of standards for a rational argument, when what was put forward was actually an explanation.” Yet, despite its theoretical importance, this distinction remains insufficiently operationalized in NLP tasks: different automatic approaches to arguments detection have different level of granularity (sentence-, two sentences-, paragraph-level, etc.) but most of them tend to treat explanatory and justificatory structures uniformly, often subsuming both under broad categories of argumentative structures (Mochales Palau and Moens, 2011; Lawrence and Reed, 2020; Lippi and Torroni, 2016).

Disentangling explanations and justifications is challenging because they both rely on similar linguistic markers (such as causal connectives), and there are usually few additional cues that can help distinguish them. Yet, their conflation is problematic precisely because one of them—explanation—is not strictly argumentative, whereas the other—justification—typically is. To address this gap, the present study proposes a systematic way to distinguish between explanatory and justificatory uses in French, focusing specifically on clauses introduced by the causal connectives *parce que* and *car* (“because”). While these connectives are generally grouped under the umbrella of causal connectives in linguistic and discourse studies (Sweetser, 1990; Sanders et al., 1992; inter alia), their discourse function can vary and goes well beyond mere cause. Specifically, in explanatory uses, the subordinate clause presents an external cause for an event whereas in justificatory uses, it conveys a reason linked to the speaker’s stance, intention, or evaluation, thus aligning more closely with argumentative discourse (Blochowiak et al., 2020). Although it has been argued that *car* tends to favor justificatory uses, whereas *parce que* is more often associated with explanatory ones (e.g., Groupe Lambda-L, 1975 and subsequent works in various traditions), both connectives can technically introduce either type of clause. In contemporary French, particularly among younger speakers, their distribution

appears to be less strictly specified and more interchangeable (Author1 and Author2 2022).

The examples below illustrate how the same causal connectives, *parce que* and *car*, can introduce clauses that serve these different pragmatic functions in discourse.

- (1) a. Le chat est monté sur la table *car/parce qu’il* sentait le poulet.
‘The cat climbed onto the table *because* it smelled the chicken.’
b. Le chat peut rester sur la table *car/parce que* c’est son anniversaire.
‘The cat can stay on the table *because* it’s its birthday.’
- (2) a. Le chien aboie *car/parce que* le facteur est là.
‘The dog is barking *because* the mailman is here.’
b. Le facteur est là *car/parce que* le chien aboie.
‘The mailman is here *because* the dog is barking.’
- (3) a. Je suis tombé de la chaise *car/parce qu’elle* s’est cassée.
‘I fell off the chair *because* it broke.’
b. Je suis tombé de la chaise *car/parce que* je voulais faire rire tout le monde.
‘I fell off the chair *because* I wanted to make everyone laugh.’

In (1a), the subordinate clause introduced by the connective provides an external cause—the smell of the chicken—that explains why the cat climbed onto the table. By contrast, in (1b), the connective introduces a justificatory reason: the speaker provides a normative rationale (because it’s its birthday), not an actual cause. Similarly, in (2a), the causal connective expresses a straightforward causal relation between the mailman’s arrival and the barking of the dog. Interestingly, in (2b) a simple causal interpretation is not possible: the dog’s barking does not somehow mysteriously cause the mailman’s arrival. Rather, in (2b) the clause introduced by the connective provides a justification for the belief expressed in the first clause—the speaker infers the mailman’s arrival from the barking of the dog. Finally, in (3a), the connective introduces a physical explanation (the chair broke), whereas in (3b), it signals an intentional or motivational justification: the speaker fell deliberately to make others laugh.

Even though the sentences above are constructed examples, they reflect genuine interpretative ambiguities in natural discourse. In authentic texts, identifying whether a sentence containing a causal connective is explanatory or justificatory often requires context-sensitive interpretation, making annotation and automatic processing non-trivial. This is precisely why distinguishing between explanation and justification is crucial in argument mining: without such a distinction, systems risk misclassifying non-argumentative explanations as arguments, thereby introducing noise into downstream reasoning, stance detection, and discourse

structure modeling (Cabrio and Villata, 2018; Green, 2018).

3. Subjectivity and Its Perception

Another dimension that has received increasing attention is the role of subjectivity in arguments and their perception. Wachsmuth et al. (2024) argue that judgments about argument quality are inherently subjective and context-dependent, especially in the era of instruction-following large language models. Traditional methods for creating such corpora typically focus on detecting subjective lexical items (Wiebe and Riloff, 2005; Riloff and Wiebe, 2003), often derived from predefined lexicons (Das and Sagnika, 2020; Yu and Hatzivassiloglou, 2003; Villena-Román et al., 2015). However, these lexicon-based strategies face well-known limitations, as they tend to depend heavily on domain- and language-specific features (Pang and Lee, 2004).

Several recent resources and annotation efforts have tackled subjectivity more explicitly at the sentence-level, such as the NewsSD-ENG corpus (Antici et al., 2024), which provides subjectivity annotations for English news articles. In their guidelines, they provided annotators with a general definition of subjectivity, according to which: “A sentence is considered subjective when it is based on—or influenced by—personal feelings, tastes, or opinions. Otherwise, the sentence is considered objective.” A series of more specific indications has also been provided, such as: “A sentence is subjective if it explicitly reports the personal opinion of its author.” These definitions perfectly align with the traditional position on the topic of subjectivity, such as the classical one by Benveniste, who sees it as: “the capacity of the speaker to posit himself as subject” (Benveniste, 1966).

However, this type of approach seems to focus on only one aspect of subjectivity, namely by treating subjectivity as equivalent to expressing opinions, while equating objectivity with stating facts. In this view, explanations, which concern facts, are considered inherently objective, whereas justifications, which concern opinions, are seen as subjective. This position is not wrong, but it overlooks an important point: at another level, subjectivity also influences how humans interpret discourse relations more broadly. Specifically, the explanation of a fact can be perceived as more objective or more subjective, while the justification of an opinion can likewise be interpreted as more or less subjective, depending on the choice of words, tone of voice, stance, genre, context, background knowledge, and many other factors. For instance, our example (1a), repeated here as (4a) for convenience, is a simple explanation of a fact that can clearly be considered an objective explanation. However, its slight modification in (4b) does not change the type of discourse

relation it expresses—it remains an explanation. What changes is the use of the expressive word *bloody*, which makes it more subjective. In our approach, this would be classified as a subjective explanation.

- (4) a. The cat climbed onto the table because it smelled the chicken.
- b. The bloody cat climbed onto the table because it smelled the chicken.

Like explanations, justifications can also vary in their perceived degree of subjectivity. Some of them provide a basis for a claim in a more subjective way, some in a more objective way, as in the examples below.

- (5) a. The mailman is here because the dog is barking.
- b. The mailman is here because this neurotic dog is barking again.

In (5a), the justification relies on an external and observable event: the dog is barking. The speaker uses it to support the inference that the mailman is present. The inferential link between the dog’s barking and the mailman’s arrival is presented in a neutral way, without any additional evaluative language. This is therefore an instance of objective justification. In (5b), the type of discourse relation remains unchanged—the utterance still provides a justification for the conclusion that the mailman is here. What changes is the introduction of evaluative language through the phrase *this neurotic dog* and the attitudinal marker *again*. These elements convey the speaker’s stance and emotional attitude toward the dog, which makes the justification more subjectively framed.

In sum, the key for our study is to approach subjectivity in a more nuanced way—going beyond the simple divide between fact description and opinion expression—and to examine how it intersects with explanatory discourse relations (which are not inherently argumentative) and justificatory discourse relations (which often are), as well as how these relations are perceived in terms of subjectivity.

4. Resource Development

A central contribution of this research is the development of a new corpus FreCaDiS: a French corpus of causal connectives annotated for discourse relations and subjectivity. The FreCaDiS corpus consists of five text genres in French, ranging from informal texts such as SMS and online discussions to more formal ones such as blogs and press, as well as contemporary literature. This variation is relevant because different genres are associated with distinct degrees of subjectivity, discourse conventions, and argumentative strategies, which allows us to observe how explanatory and justificatory

relations are realized and perceived across contexts.

Informal genres such as SMS and online discussions are typically rich in personal stance, implicit reasoning, and evaluative language. Blogs, situated between informal and formal registers, often blend personal opinions with factual information, offering hybrid contexts where explanations and justifications can co-occur. Press texts, by contrast, generally favor explicit and structured argumentation on the one hand and are expected to display more objective explanatory and justificatory patterns. Finally, literary texts provide stylistically and narratively rich examples, including non-prototypical expressions of stance. This cross-genre design enables us to investigate how explanation, and justification are distributed across these diverse communicative contexts and how they interact with subjectivity.

4.1 Corpus

The corpus consists of a total of 1081 excerpts (52527 words): 200 from blogs, 220 from contemporary literature, 221 from web discussions, 205 from SMS and 215 from written press. The excerpts were randomly collected through search using “*car*” and “*parce que*” as keywords in Sketch Engine (blogs and web discussions), the UCL Corpus (contemporary literature), the Belgian SMS Corpus (Cougnon 2012) and the Le Monde Corpus (year 2012). The excerpts were then randomly selected among all search results in each of the listed sources with the aim to have half of the excerpts with *car* and half with *parce que*.

Each excerpt consists of 1–2 sentences containing either *parce que* or *car*. Each excerpt was annotated along two dimensions: (i) subjectivity (subjective vs. objective) in Annotation Study 1 and (ii) discourse relation (explanation vs. justification) in Annotation Study 2. The annotation was carried out by three human judges, who independently annotated each excerpt for both features. This dual annotation design enables us to examine not only the distribution of explanatory and justificatory relations across different genres but also how subjectivity interacts with these discourse functions.

4.2 Annotation Study 1

In Annotation Study 1, three independent judges were recruited to annotate the corpus excerpts. All annotators were female native speakers of

French with a background in linguistics¹ and were compensated for their participation. The study followed a non-guided annotation approach (Author 1 et al., 2020), which is based on the assumption that listeners and readers routinely—and often implicitly—evaluate the degree of subjectivity in the information they process.

Annotators were presented with excerpts containing causal connectives and were asked to decide whether the information conveyed by the sentence appeared more subjective or more objective. Importantly, no explicit definitional criteria for subjectivity or objectivity were provided. This deliberate choice was intended to activate the annotators’ intuitive interpretive layer, allowing for a more natural and holistic assessment of subjectivity.

Annotators were therefore instructed to rely on their overall impression and contextual understanding of each sentence, rather than on predefined diagnostic features. This design makes it possible to capture a gradient and context-sensitive notion of subjectivity that is closer to how it is perceived in everyday language use.

4.3 Annotation Study 2

In Annotation Study 2, we implemented a guided annotation approach to classify sentences containing the causal connectives *parce que* and *car* as instances of either explanation or justification. The same three annotators who participated in Study 1 were recruited for this second task. After a two-week interval, they were re-contacted and introduced to the new annotation procedure. To minimize potential carry-over effects, annotators were asked to delete all materials related to the first study from their devices prior to beginning the new task.

In contrast to the intuitive, non-guided approach used in Annotation Study 1, the Annotation Study 2 provided explicit operational definitions and clear guidelines to support consistent annotation. Annotators were given the following definitions. Explanation answers the question of why or how something happened, or why something is the case. It specifies the cause(s) of an event or state of affairs. Justification addresses why a claim, decision, or action is valid, correct, desirable, or appropriate. It provides reasons, evidence, or arguments to support an assertion, belief, or stance.

Following this presentation, annotators were provided with detailed written guidelines defining

¹ As pointed out by an anonymous reviewer, argument mining is often influenced by the reader’s disciplinary background. This raises the question of whether one would expect different agreement patterns if the task were performed by annotators with different backgrounds. We tested this assumption in a pilot study by recruiting seven annotators via Prolific (2 male, 5

female; native speakers of French; age range 20–51; age at completion of formal education ranging from 19 to 26; with diverse disciplinary backgrounds). We obtained similar results and agreement rates, suggesting that task performance is not substantially affected by individual differences among annotators.

the two discourse relations and were given the opportunity to ask clarifying questions to ensure a shared understanding of the criteria.

Based on these definitions, the annotators' task was to determine, for each sentence containing the causal connective *parce que* or *car*, whether it represented an explanatory or justificatory discourse relation.

4.4 Results

Inter-annotator agreement was measured with Cohen's Kappa, which corrects for agreement that could happen by chance. Its scores range from less than 0 (poor agreement) to 0.8-1 (almost perfect agreement), with intermediary values between 0.01-0.20 as slight agreement, 0.21-0.4 (fair agreement), 0.41-0.6 (moderate agreement), 0.61-0.80 (substantial agreement) (Landis & Koch, 1977). Nevertheless, more recently Grisot (2017) demonstrates using the examples of several types of semantic and pragmatic features and several annotators that inter-annotator agreement scores are strongly influenced by the nature of the linguistic information being annotated: semantic (encoded) information tends to yield higher agreement rates, whereas pragmatic (context-dependent) information often results in lower agreement. Grisot's approach offers a quantitative method for assessing linguistic meaning beyond theoretical intuition.

4.4.1 Discourse Relations

The average agreement among the three annotators was of 70% (i.e., 69% for the first pair of annotators, 66% for the second pair and 75% for the third pair). This average agreement rate corresponds to an average kappa score of 0.27 (i.e., 0.27 for the first pair of annotators, 0.2 for the second pair of annotators and 0.36 for the third pair). These low scores were thus expected due to the pragmatic and context-dependent nature of distinguishing between explanations and justifications, and for this reason, the annotations are accepted as reliable. Furthermore, the categorization of each corpus excerpt as explanation or justification corresponded to the categorization made by the majority (i.e., 2 out of 3, and by 3 out of 3 annotators). As such, no corpus excerpt was discarded from further analysis.

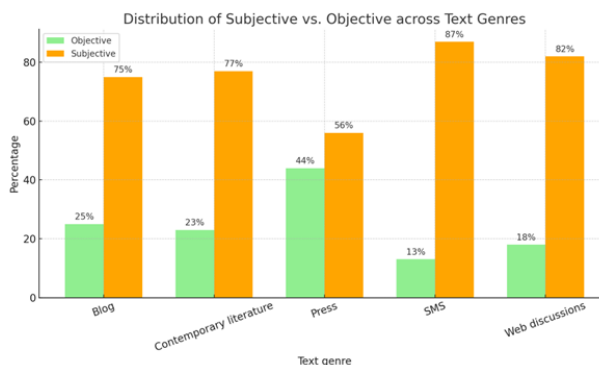
The corpus consists of 832 justifications (77%) and 249 explanations (23%). In this set of data, justifications and explanations are distributed as follows in the five corpora:

Figure 1 Frequency of subjective and objective relations in the five corpora

The results show an overall pattern across the five corpora according to which there are less explanations than justifications. Nevertheless, the

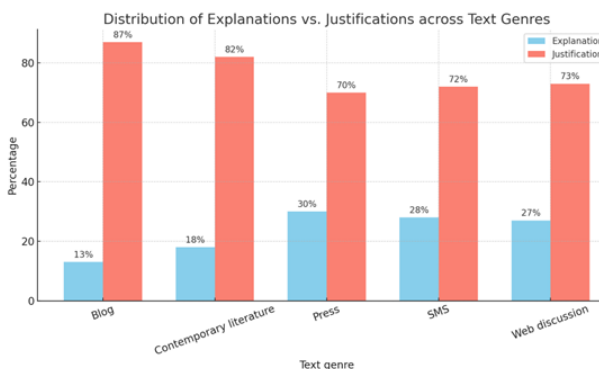
frequency of explanations is significantly lower in blogs than in press ($\chi^2(4) = 27.83, p < .001$).

4.4.2 Subjectivity



The average agreement among the three annotators was of 70% (i.e., 69% for the first pair of annotators, 66% for the second pair and 75% for the third pair). This average agreement rate corresponds to an average kappa score of 0.27 (i.e., 0.27 for the first pair of annotators, 0.2 for the second pair of annotators and 0.36 for the third pair). As before, these low scores were thus expected due to the pragmatic and context-dependent nature of subjectivity. Furthermore, the categorization of each corpus excerpt as subjective or objective corresponded to the categorization made by the majority.

The corpus consists of 814 (75.4%) excerpts perceived as subjective and 266 perceived as objective (24.6%). In this set of data, subjective and objective relations are distributed as follows



in the five corpora:

Figure 2 Frequency of justifications and explanations in the five corpora

The results reveal an overall pattern according to which there are more subjective relations than objective ones in blogs, contemporary literature, SMS and web discussions. In contrast, in press data, the frequency of objective relations is significantly higher than in the other four corpora, whereas in SMS the frequency of objective relations is significantly lower than in the other corpora ($\chi^2(4) = 62.9, p < .001$).

4.4.3 At the Intersection of Discourse Relations and Subjectivity

The first step in our analysis is to examine how, in general, discourse relations correlate with the intuitive perception of subjectivity. More specifically, we investigate the question of how explanations and justifications are perceived as more subjective or more objective in each of the five corpora.

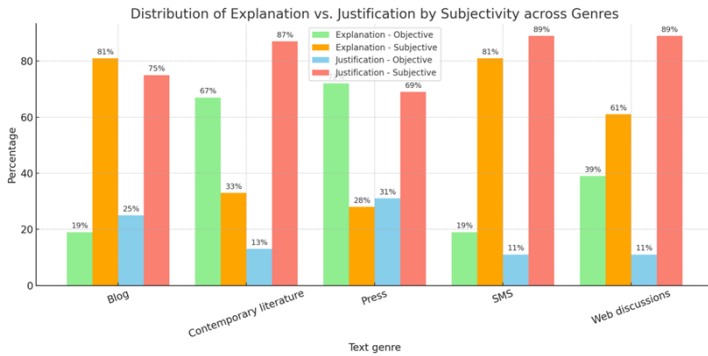


Figure 3 Frequency of explanations and justifications perceived as more subjective or more objective in the five corpora

The distribution of subjective and objective discourse segments reveals a clear predominance of subjectivity across all registers, with particularly high proportions in SMS (87%), web discussions (82%), blogs (75%), and contemporary literature (77%). Even in press texts, where objectivity is expected to play a central role, more than half of the segments (56%) are subjective.

These findings suggest that subjective stance is not limited to overtly personal or informal registers but is a pervasive feature of discourse more generally. The variation across registers reflects their communicative functions: interactive and expressive genres naturally foreground subjective expression, while more informational genres balance subjective and objective content to a greater degree.

5. Conclusions

This paper introduced FreCaDiS, a new French corpus designed to explore the intersection between causal discourse relations, argumentation, and subjectivity across five text genres. By focusing on the causal connectives *parce que* and *car*, we operationalized the theoretical distinction between explanation and justification, and complemented it with human judgments of subjectivity using both guided and non-guided annotation procedures. This dual approach provides an empirically grounded resource for investigating how subjectivity interacts with argumentative structures in natural language.

6. Applications

Beyond its theoretical relevance for the study of discourse and argumentation, FreCaDiS offers a wide range of applications for natural language processing (NLP) and AI research. By combining discourse relations (explanation vs. justification) with explicit subjectivity annotations across five distinct text genres, the corpus provides a valuable resource for advancing computational modeling of argumentation and subjectivity in natural language.

From an applied perspective, FreCaDiS can support several key NLP tasks. In stance detection and subjectivity analysis, its explicit encoding of subjective versus objective discourse enables the training and evaluation of models that more accurately identify subjective statements, opinions, or evaluative language. It also facilitates research on rhetoric and stance-taking, allowing models to capture subtle linguistic cues of perspective and bias. In discourse relation classification, the corpus operationalizes a fine-grained distinction between explanatory and justificatory uses of causal relations—an aspect often neglected in existing resources. This opens new possibilities for language models that distinguish between argumentative and non-argumentative causal relations, improving downstream applications such as text understanding, summarization, and question answering. In future research, this dataset could be used to create a contrastive benchmark to examine whether instruction-following LLMs exhibit patterns similar to those of standard NLP classifiers and human annotators.

This could be done in a manner similar to Escoufflaire et al. (2024), who assessed the performance of both humans and LLMs—specifically the state-of-the-art models at the time, GPT-3.5 (Brown et al., 2020) and a French fine-tuned CamemBERT transformer (Martin et al., 2019)—in choosing between *parce que* and *car* after these connectives had been removed from the original excerpts (with minimal context, i.e., a single sentence). In their study, they used a test set of 420 French segments containing either *car* or *parce que*, drawn from two genres: journalistic texts (*Le Monde*) and SMS messages from the SMS4science corpus (Fairon et al., 2006). The target connective was masked and surrounding context was removed. CamemBERT was fine-tuned on 10,000 additional sentences (5,000 per connective) extracted from SMS4science messages and Belgian news articles. Across three connective prediction experiments, human native speakers showed low accuracy and agreement, GPT-3.5 performed inconsistently and generally worse than humans, and the fine-tuned CamemBERT model achieved the best results (66.7% accuracy). Text genre consistently

affected the performance of both humans and models.

With respect to FreCaDiS, because it spans five genres ranging from informal (SMS, web discussions) to formal (press, blogs, literature), it can contribute to register- and genre-aware NLP and AI. Specifically, it can enable the development of systems that adapt to different communicative contexts, improving performance in tasks such as genre classification, style transfer, and domain adaptation. While the text genre distinction does not appear to strongly affect human annotators when distinguishing between explanation and justification, an open question is whether this also holds for computational models, or whether text genre-specific models will remain necessary for SSH-related tasks. As we saw, previous work suggests that text genre can indeed be an important differentiating factor in certain settings, such as connective selection task (see Escoufflaire et al., 2024).

Finally, FreCaDiS annotations on subjectivity as a binary feature crossed with discourse relations (such as explanations and justifications) make it highly relevant for sentiment and opinion mining, offering a richer basis for modeling nuanced evaluative language beyond simple polarity distinctions. Another methodological approach would be to use a Likert scale for subjectivity in order to provide more granular data, which could better support LLMs in nuanced stance detection tasks. In our study, however, we opted for a coarse-grained annotation scheme (*rather objective* vs. *rather subjective*) for two main reasons. First, our goal was to capture annotators' intuitive and holistic perception of subjectivity rather than to model subtle degrees of stance. We assume that finer distinctions are difficult to apply consistently without extensive guidelines and calibration, which could increase inter-annotator variability and reduce reliability. Second, agreement rates for a multi-point scale would likely have been even lower than for a binary feature, making the additional annotation burden disproportionate to the expected benefit.

7. Limitations

The corpus also presents a number of limitations. Since the task involves inherently subjective judgments, inter-annotator agreement is modest—an expected outcome in pragmatic annotation settings, but one that limits FreCaDiS's immediate suitability for high-stakes NLP benchmarking or fully automated, unsupervised model training. We acknowledge this constraint and view the present study as an initial step toward more robust modeling of subjectivity and discourse-based argumentation. In future work, we plan to strengthen the dataset by incorporating expert annotations, refining annotation

guidelines, and introducing a supervising adjudicator to handle ambiguous or contested cases. These measures aim to improve annotation reliability while preserving the richness of intuitive human interpretation that characterizes subjectivity and discourse phenomena.

Furthermore, FreCaDiS is at this stage a research-oriented corpus that lacks a fully validated pipeline for how these findings can be integrated into broader SSH research infrastructures.

8. Acknowledgments

The work was supported by grant no. 34821046 from the Belgian FNRS, awarded to the first author.

9. Bibliographical References

- Antici, F., Galassi, A., Ruggeri, F., Korre, K., Muti, A., Bardi, A., et al. (2024, May 24). A Corpus for Sentence-level Subjectivity Detection on English News Articles. arXiv. <https://doi.org/10.48550/arXiv.2305.18034>
- Benveniste, E. (1966). *Problèmes de linguistique générale*. Paris: Gallimard.
- Bex, F., & Walton, D. (2016). Combining explanation and argumentation in dialogue. *Argument & Computation*, 7(1), 55–68. <https://doi.org/10.3233/AAC-160001>
- Blochowiak J., Grisot C., & Degand L. (2020). What type of subjectivity lies behind French causal connectives? A comparative study of French causal connectives *car* and *parce que* from a corpus-based perspective. *Glossa: A Journal of General Linguistics*, 5(1), 50. DOI: <http://doi.org/10.5334/gjgl.1077>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Buckingham Shum, S., Uren, V., Li, G., Domingue, J., & Motta, E. (2003). Visualizing internetworked argumentation. In *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 185-204). London: Springer London.
- Cabrio, E., & Villata, S. (2018). Five years of argument mining: A data-driven analysis. In *IJCAI* (Vol. 18, pp. 5427–5433). <https://www.ijcai.org/proceedings/2018/0766.pdf>.
- Cougnon, L.-A. (2015). *Langage et sms: Une étude internationale des pratiques actuelles*. Presses universitaires de Louvain.
- Das, N., & Sagnika, S. (2020). A Subjectivity Detection-Based Approach to Sentiment Analysis. In D. Swain, P. K. Pattnaik, & P. K. Gupta (Eds.), *Machine Learning and Information Processing* (Vol. 1101, pp. 149–160). Singapore: Springer Singapore.

- https://doi.org/10.1007/978-981-15-1884-3_14
Ding, Y., Kashefi, O., Somasundaran, S. & Horbach, A. (2024) When Argumentation Meets Cohesion: Enhancing Automatic Feedback in Student Writing. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 17513-17524, Turin, Italy. ELRA and ICCL.
- Eemeren, F. H. van, & Grootendorst, R. (2016). *Argumentation, Communication, and Fallacies: A Pragma-dialectical Perspective*. New York: Routledge.
- Escoufflaire L., Blochowiak J., Degand L., & de Marneffe, M-C. (2024). "Which connective fits best: 'car' or 'parce que'? A challenge for both humans and LLMs." In: *JADT 2024 Proceedings: 17th International Conference on Statistical Analysis of Textual Data*, JADT: Brussels 2024 <http://hdl.handle.net/2078.1/288770>
- Fairon, C., Klein, J., Paumier, S. (2006). Le corpus SMS pour la science. *Base de données de 30.000 SMS et logiciels de consultation*. Louvain-la-Neuve, Presses universitaires de Louvain.
- Green, A. (2018). *Creativity in the distance: The neurocognition of semantically distant relational thinking and reasoning*. The Cambridge handbook of the neuroscience of creativity, 25, 363–381.
- Grisot, C. (2017). A quantitative approach to conceptual, procedural and pragmatic meaning: Evidence from inter-annotator agreement. *Journal of Pragmatics*, 117, 245–263.
- Lambda-I, G. (1975). Car, parce que, puisque. *Revue romane*, 10(2), 248–280.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
- Lawrence, J., & Reed, C. (2020). Argument mining: A survey. *Computational linguistics*, 45(4), 765–818.
- Lippi, M., & Torroni, P. (2016). Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology*, 16(2), 1–25. <https://doi.org/10.1145/2850417>
- Liu, H., Xu, W., Lin, X., Huo, J., Chen, H., & Zhao, Z. (2024, May 30). AntCritic: Argument Mining for Free-Form and Visually-Rich Financial Comments. arXiv. <https://doi.org/10.48550/arXiv.2208.09612>
- Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É. V., Seddah, D. & Sagot, B. (2019). CamemBERT: a tasty French language model. *arXiv:1911.03894*.
- Mochales, R., & Moens, M.-F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19(1), 1–22. <https://doi.org/10.1007/s10506-010-9104-x>
- Pang, B., & Lee, L. (2004, September 29). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. arXiv. <https://doi.org/10.48550/arXiv.cs/0409058>
- Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* (pp. 105–112).
- Román, J. V., Cámara, E. M., Morera, J. G., & Zafra, S. M. J. (2015). Tass 2014-the challenge of aspect-based sentiment analysis. *Procesamiento del Lenguaje Natural*, 54, 61–68.
- Ruiz-Dolz, R., Chen, C. C., Kando, N., & Chen, H. H. (2024). Learning strategies for robust argument mining: An analysis of variations in language and domain. In *Joint 30th International Conference on Computational Linguistics and 14th International Conference on Language Resources and Evaluation, LREC-COLING 2024* (pp. 10286–10292). European Language Resources Association (ELRA).
- Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1), 1–35. <https://doi.org/10.1080/01638539209544800>
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (Vol. 54). Cambridge University Press.
- Van Eemeren, F. H., Grootendorst, R., Jackson, S., & Jacobs, S. (1997). Argumentation. In *Discourse as structure and process: Discourse studies: A multidisciplinary introduction* (pp. 208–229). SAGE Publications Ltd.
- Wachsmuth, H., Lapesa, G., Cabrio, E., Lauscher, A., Park, J., Vecchi, E. M., et al. (2024, March 24). Argument Quality Assessment in the Age of Instruction-Following Large Language Models. arXiv. <https://doi.org/10.48550/arXiv.2403.16084>
- Walton, D. (2004). A new dialectical theory of explanation. *Philosophical Explorations*, 7(1), 71–89. <https://doi.org/10.1080/1386979032000186863>
- Wei, S., Li, M., Sun, J., Davis J. & Moens, M.-F. (2024). DMON: A Simple Yet Effective Approach for Argument Structure Learning. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 5109-5118, Turin, Italy. ELRA and ICCL.
- Wiebe, J., & Riloff, E. (2005). Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (Vol. 3406, pp. 486–497). Berlin, Heidelberg: Springer Berlin Heidelberg.

https://doi.org/10.1007/978-3-540-30586-6_53
Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* (pp. 129–136).