

Which connective fits best: ‘car’ or ‘parce que’?

A challenge for both humans and LLMs

Louis Escouflaire¹, Joanna Blochowiak², Liesbeth Degand³, Marie-Catherine de Marneffe⁴

¹UCLouvain – louis.escouflaire@uclouvain.be

²UCLouvain – joanna.blochowiak@uclouvain.be

³UCLouvain – liesbeth.degand@uclouvain.be

⁴UCLouvain – marie-catherine.demarneffe@uclouvain.be

Abstract

We analyze how well human annotators, ChatGPT and a fine-tuned CamemBERT transformer model are able to predict words in French which differ subtly in meaning. We focus on *car* and *parce que*, two French connectives considered as near-synonymous, distinguished only by fine-grained syntactic, semantic and pragmatic features. We used a test set of 420 sentences from French news articles and SMS text messages containing the word *car* or *parce que*, which was masked and had to be predicted. Our results suggest that this task is particularly difficult both for native speakers of French and for large language models. However, we find that fine-tuning CamemBERT on a training corpus of 10,000 masked sentences containing *car* vs. *parce que* allows it to grasp the syntactic and semantic subtleties between the connectives, and to perform significantly better than the human annotators.

Keywords: Large Language Models, Annotation, Connectives

1. Introduction

Large language models (LLMs) have improved the state-of-the-art for most Natural Language Processing (NLP) tasks in recent years (Brown et al., 2020), and have pervaded the public sphere with the launch of ChatGPT (Berson and Berson, 2023). These models leverage advanced neural network architectures and are pre-trained on massive amounts of textual data, enabling them to capture complex linguistic patterns and contextual dependencies (Devlin et al., 2019). It is still unclear, however, whether those pre-trained transformer models are able to grasp subtle differences in meaning (Gao et al., 2021). In this paper, we compare how humans and large language models perform in handling such fine-grained differences through experiments focusing on two near-synonymous French causal connectives: *car* and *parce que* (‘because’). Beyond syntactic and semantic differences, the subtleties between those two connectives are related to their status regarding subjectivity: *parce que* has been described as a typically ‘objective’ discourse marker, whereas *car* has been acknowledged as typically ‘subjective’. Yet, ample linguistic work has questioned this division of labor (Nazarenko, 2000; Zufferey et al., 2018; Blochowiak et al., 2020).

In this work, we asked three native speakers of French to predict the masked connective (*car* or *parce que*) in a sample of 420 sentences extracted from news articles and text messages. The same task and the same test set were given as a prompt to ChatGPT. Finally, we fine-tuned the French state-of-the-art LLM CamemBERT (Martin et al., 2019) on a larger corpus of 10,000 sentences containing *car* or *parce que* to try to reach the highest accuracy in this specific task, and evaluate its performance on our test set. We find that predicting the presence of *car* or *parce que* based on a single sentence is a particularly difficult task both for humans and for LLMs, though fine-tuning CamemBERT allows it to uncover syntactic and semantic patterns in the

sentences and to handle the subtleties between *car* and *parce que* better (6.5%) than human native speakers of French.

2. The case of *car* vs. *parce que*

The French connectives *car* and *parce que* have been the object of detailed linguistic work since Groupe Lambda-L's (1975) seminal proposal to distinguish between uses of connectives like *car* to relate claims and arguments, as in sentence (1), from uses such as that of *parce que* to link causes and consequences that unfold in the world, as in sentence (2). More precisely, in sentence (2) the cause of Max's illness is made explicit, while in sentence (1) the speaker infers that Anne must be ill based on her paleness. According to Groupe Lambda-L, this semantic distinction follows directly from the different status between the two connectives: in sentence (1) the connective is a "speech act marker" (*marqueur d'acte de parole*) realizing a speech act of inference between two independent clauses, while in sentence (2) the connective functions as an "operator" (*opérateur*), construing a new causal proposition on the basis of two connected propositions.

- (1) Anne doit être malade, **car** elle est très pâle.
Anne must be ill, [car] she is very pale.
- (2) Max est tombé malade **parce qu'**il a attrapé froid.
Max got ill [parce que] he got cold.

In Groupe Lambda-L's (1975) paper, this distinction is further substantiated by the existence of syntactic differences between the uses of these two causal connectives. When it comes to cause-consequence relations, it is only *parce que* – but not *car* – that can undergo a number of syntactic transformations, such as clefting, as in sentence (3), modification by an adverb, as in sentence (4), or questioning.

- (3) C'est **parce que** Max a attrapé froid qu'il est tombé malade.
*C'est car Max a attrapé froid qu'il est tombé malade.
*It's [parce que/*car] Max got cold that he got ill.*
- (4) Max est tombé malade, probablement **parce qu'**il a attrapé froid.
*Max est tombé malade, probablement car il a attrapé froid.
*Max got ill, probably [parce que/*car] he got cold.*

In later work, this distinct behavior has been related to the semantic dependent vs. independent status of the connected clauses, allowing for different constructions (Debaisieux, 2004), like coordination, as in (5).

- (5) Max est tombé malade, **parce qu'**il a attrapé froid et qu'il ne s'est pas soigné.
*Max est tombé malade, car il a attrapé froid et car il ne s'est pas soigné.
*Max got ill because he got cold and [parce que/*car] he did not take care of himself.*

In more recent literature, Groupe-L's proposal is often reinterpreted within the cognitive approach that distinguishes between objective and subjective causality. Thus, the relation in sentence (2) is objective as it connects real-world events, while the relation in sentence (1) is

subjective as it relies on the speaker’s own reasoning. Considering these two categories, there is a prevailing trend in the literature to view *parce que* as the most versatile French causal connective, applicable to any type of relation (objective and subjective), while *car* is viewed a prototypical indicator of subjective relations (see, e.g. Degand and Pander Maat, 2003; Zufferey, 2012; among many others). Nevertheless, it is important to highlight that the findings regarding how speakers use and process the two connectives are not consistent. For instance, Zufferey (2012) observes that the two connectives can be swapped in various objective and subjective contexts when written, particularly among the younger native speakers of French. Importantly, certain studies have cast doubt on the subjective nature of *car* and indicate that it can actually be employed to convey objective relations as well (Nazarenko, 2000; Zufferey et al., 2018). These findings have been further supported in a recent study analyzing the use of *parce que* and *car* in news discourse and SMS text messages (Blochowiak et al., 2020). On the basis of several annotation tasks, results indicate that the connective *car* is used mainly to convey the message in a more objective way, and not to predominantly express subjectivity as expected.

In summary, the distinctions among the French causal connectives under scrutiny here are threefold. At the syntactic level, *car* cannot occur in specific syntactic configurations (such as clefting, adverbial modification, coordination), while *parce que* does not have such restrictions (even if linguistic intuitions of native speakers seem to evolve, especially in spoken language). At the semantic level, these syntactic patterns are shaped by the inherent semantic properties of the connectives: *parce que* functions as an operator, akin to other logical connectives, relating two propositions to create a new complex proposition, whereas *car* connects two independent propositions with their proper illocution. Finally, at the pragmatic level, both connectives seem to occur in subjective as well as objective contexts.

3. Data

3.1. Test set

SMS	<i>car</i>	<i>parce que</i>	<i>N</i>	Le Monde	<i>car</i>	<i>parce que</i>	<i>N</i>
subjective	45	60	105	subjective	50	51	101
objective	50	47	97	objective	56	51	107
ambiguous	2	1	3	ambiguous	1	6	7
<i>N</i>	97	108	205	<i>N</i>	107	108	215

Table 1. Distribution of sentences in the SMS and Le Monde portions of the test set.

To evaluate the accuracy of the annotators and LLMs, we use the corpus designed in (Blochowiak et al., 2020). It consists of 420 French pieces of text containing an occurrence of either *car* or *parce que*, manually annotated as objective or subjective. The texts belong either to the conversational or to the journalistic genre: they are Belgian SMS text messages collected in the *SMS4Science* project (Fairon et al., 2006), such as sentence (6), or excerpts of news articles from the French newspaper *Le Monde*, such as sentence (7). The distributions are well-balanced along the different dimensions (*car* vs. *parce que*, subjective vs. objective, Le Monde vs. SMS), as shown in Table 1.

(6) Coucou, juste pour te demander de m'envoyer un SMS quand tu viens **car** je dois passer à l'école, gros bisous à tantôt
*Hi, just to ask you to text me when you come over **because** I have to go to school, big kisses see you soon*

(7) Mais si elle est l'institution la plus respectée des Allemands, **c'est notamment parce qu'**elle apparaît comme le dernier recours de la population (qui peut la saisir directement) contre les puissants.
*But if it is the most respected institution among Germans, **it is especially because it** appears to be the last resort of the population (who can contact it directly) against the powerful.*

In each text, the context sentences located before or after the sentences containing *car* or *parce que* were removed. We systematically replaced each occurrence of *car* or *parce que* in the test set by a <MASK> token. The resulting corpus of 420 sentences contains 7,672 tokens in total: 33 tokens per sentence on average (35 in Le Monde against 29 in SMS).

In total, the corpus contains 34 occurrences of *parce que* with syntactic restrictions. Of these, 11 are restricted by clefts, 16 by adverbial modifications, and 6 by both. For example, sentence (7) contains *parce que* both restricted by a cleft and an adverb (*notamment*). In addition, we counted 111 connectives used alongside semantic coordinators between two dependent clauses: 4 occurrences of *parce que* introduced by a coordinating conjunction, and 107 sentences containing *car* (78 times) or *parce que* (29 times) directly preceded by a comma.

3.2. CamemBERT training corpus

As mentioned above, LLMs can achieve very good performance on some language tasks, especially when fine-tuned (i.e., trained on specific data to adapt them to particular tasks or domains). To fine-tune CamemBERT on the classification task of predicting *car* or *parce que* to fill a gap in a given sentence, we assembled a corpus of 10,000 sentences containing *car* or *parce que*. On the one hand, we extracted all 5,800 sentences containing one of the two connectives from a corpus of 40,000 articles published by four Belgian media (Le Soir, La Libre Belgique, L'Avenir and RTBF). On the other hand, we extracted 4,200 additional sentences containing the connectives, including alternate spellings of *parce que* (*parceque* or *pcq*) from the Belgian component of the *SMS4Science* corpus (text messages). By design, the corpus is perfectly balanced in terms of connectives: the newspaper section contains 2,900 sentences with *car* and 2,900 with *parce que*, and the SMS section contains 2,100 sentences with *car* and 2,100 with *parce que*.

This corpus is randomly split into a training and a development set, both balanced in terms of connectives and text genres, and containing respectively 8,600 and 1,400 sentences. The full corpus accounts for a total of 327,877 tokens. Sentences contain on average 33 tokens (36 in the newspaper corpus against 29 in SMS). Each occurrence of *car* or *parce que* in the training and development sets are replaced by a <MASK> token, and the sentences are shuffled using a random seed.

4. Experiments

4.1. Human annotation

To assess the ability of human readers to accurately predict whether a sentence should contain *car* or *parce que*, we carried out an annotation experiment with three (female) native speakers of French. They were 21, 28 and 51 years old.

They were asked to read the 420 sentences of our test set one by one, and to guess whether the <MASK> was hiding the word *car* or *parce que*. They were explicitly told to guess the connective right after reading the sentence fully (from start to end), to rely on their linguistic intuition, and to avoid overthinking. The annotation was conducted with a simple text editor, without direct supervision. (8) gives an example of a masked sentence, in which the <MASK> token hides the word *car* in the original sentence.

(8) Ça vous dit pas de mater le match ce soir <MASK> moi ce soir y a des gens qui viennent et je m'en tape !
How about watching the game tonight <MASK> I got people coming over, and honestly, I couldn't care less!

4.2. ChatGPT

We use the free version of ChatGPT, which relies on the transformer-based LLM GPT-3.5 (Brown et al., 2020) and can be accessed online.¹ First, through a process of trial-and-error prompt engineering, we iteratively designed a fitting prompt that can be fed as a new query to ChatGPT to receive the expected form of answer from the chatbot. The goal is to have ChatGPT complete the same task as the human annotators. Because ChatGPT tends to always choose the same connective after a random number of sentences, we forced it in the prompt to “try not to use the same word all the time, with as even a distribution as possible.” Given that the masked sentences are written in French, the prompt was also given in French to ChatGPT. The prompt and its translation are presented in Table 2.

Je vais te donner des phrases numérotées, 10 phrases à la fois. Pour chaque phrase, je veux que tu remplaces le <MASK> par un mot. Pour chaque phrase, tu as le choix entre deux mots pour remplacer le <MASK> : le mot "car" ou le mot "parce que". Choisis le mot approprié en fonction du contexte de la phrase dans laquelle tu dois remplacer le <MASK>. Essaie de ne pas utiliser toujours le même mot, avec une répartition aussi homogène que possible. Imprime ensuite le texte en surlignant le mot que tu as remplacé dans le texte, "car" ou "parce que". Ensuite, je veux que tu imprimes un tableau avec le numéro du texte que je t'ai donné et le connecteur que tu as choisi pour ce texte. Est-ce que tu as bien compris ?	I'm going to give you numbered sentences, 10 sentences at a time. For each sentence, I want you to replace the <MASK> with a word. For each sentence, you have the choice between two words to replace the <MASK>: the word "car" or the word "parce que". Choose the appropriate word based on the context of the sentence in which you have to replace the <MASK>. Try not to use the same word all the time, with as even a distribution as possible. Then print out the text, highlighting the word you've replaced in the text, "car" or "parce que". Next, I want you to print out a table with the number of the text I've given you and the connector you've chosen for this text. Do you understand?
--	---

Table 2. Prompt given to ChatGPT, with its English translation.

¹ <https://chat.openai.com/>, last accessed on 31 January 2024.

Once the chatbot replies that the task is understood, we present it masked sentences 1 to 10. ChatGPT replies with the filled sentences in order and a two-column table containing the ID of the sentence and the completed word (*car* or *parce que*). This operation is repeated 42 times, with a block of 10 sentences each time.

To evaluate the consistency of the decisions made by the chatbot, we carried out the same experiment three different times as new query threads, for a set of 100 sentences randomly selected from the test set. This allowed us to find that ChatGPT is not consistent in the predictions made across its different query threads.

4.3. Fine-tuned CamemBERT

We use the transformer model CamemBERT (Martin et al., 2020) in its basic and case-insensitive version, which contains 110 million parameters. CamemBERT is based on the architecture of the RoBERTa model (Liu et al., 2019) and has been pre-trained on the French section of the OSCAR corpus (138 GB of text). The bidirectional training of transformer models such as CamemBERT allows them to consider both left and right contexts, facilitating a deeper understanding of textual structure and word meanings within different discursive genres (Devlin et al., 2019). Despite their achievements, transformer models still face challenges in fully comprehending more complex and culturally nuanced meanings. Ongoing research endeavors focus mostly on enhancing model interpretability and mitigating biases in model predictions (Lyu et al., 2024).

In this experiment, we fine-tune CamemBERT on the *car* vs. *parce que* classification task using the training set of 8,600 masked sentences. The classification task consists in predicting whether the <MASK> token in a sentence hides *car* or *parce que*.

To test for domain-dependency, we also fine-tune the model with data from one genre only: one version fine-tuned only on 5,000 sentences from news articles, and another version fine-tuned only on 3,600 SMS sentences. All versions of CamemBERT are fine-tuned during two epochs with the same hyperparameters (learning rate = 2×10^{-5} ; batch size = 4).

5. Results

5.1. Human annotation

Inter-annotator agreement		Accuracy	Le Monde (N = 215)	SMS (N = 205)	Total (N = 420)
K(a1, a2)	0.068	annotator 1	49.3	51.2	50.2
K(a1, a3)	0.037	annotator 2	65.6	58.0	61.9
K(a2, a3)	0.131	annotator 3	56.3	59.5	57.9
		aggregated	61.9	58.5	60.2

Table 3. Human inter-annotator agreement (Cohen's K) and classification accuracy (%) on the different portions of the test set.

Table 3 presents the results of the human annotation. The inter-annotator agreement between the three participants is very low, ranging between 0.037 and 0.131 (Cohen's Kappa). When comparing the annotator's guesses to the original connectors in the 420 sentences, the highest

classification accuracy is 61.9% (annotator 2), while annotator 1 obtains a score that is very close to random distribution for a binary annotation task (50.2%). We also measure accuracy when aggregating all three annotators’ predictions, taking for each sentence the connective guessed by at least two annotators. This aggregated human accuracy reaches 60.2%. Overall, these results suggest that predicting the presence of *car* and *parce que* based on the context of a single sentence is a particularly difficult task for human readers.

5.2. ChatGPT consistency

Table 4 shows the agreement (Cohen’s Kappa) between the predictions made by the three query threads of ChatGPT for the same 100 sentences from the test set. The agreement is very low, showing that ChatGPT’s predictions are not consistent. Moreover, only 41 sentences out of 100 were given the same connective (*car* or *parce que*) by the three query threads. Because of its higher accuracy on the 100 first sentences, the first query thread was then used to predict the full 420 sentences of the test set (see Section 5.4).

ChatGPT inter-thread agreement		Accuracy	Sample (N = 100)
K(t1, t2)	0.189	thread 1	56.0
K(t1, t3)	0.023	thread 2	51.0
K(t2, t3)	0.139	thread 3	48.0
		aggregated	54.0

Table 4. Agreement between the ChatGPT query threads (Cohen’s *K*) and classification accuracy (%) on 100 sentences from the test set.

5.3. CamemBERT: impact of text genre

Table 5 presents the classification accuracy of the three versions of CamemBERT (fine-tuned on three different parts of the training corpus) on the test set. Not surprisingly, the most accurate version overall is the one fine-tuned on the entire training set (newspaper and SMS sentences), with 66.7% accuracy on the test set. When analyzing the results on Le Monde and SMS test sentences separately, a drop in accuracy of 9.2% can be observed, suggesting that the connectors in SMS are more difficult to classify for CamemBERT than those in Le Monde. This may be a consequence of the uneven distribution of newspaper and SMS sentences in the training set, and of the great disparity between the two discursive genres. Table 5 shows that both specialized models are clearly better at classifying sentences belonging to the genre on which they were fine-tuned. In the next sections, we only use the version of CamemBERT fine-tuned on the entire training set (news and SMS).

	CamemBERT (News)	CamemBERT (SMS)	CamemBERT (News + SMS)
Le Monde (N = 215)	69.8	48.8	71.2
SMS (N = 205)	58.0	65.9	62.0
Total (N = 420)	64.0	57.1	66.7

Table 5. Accuracy (%) of the different CamemBERT models (fine-tuned on news, on SMS, or on both) on the test set.

5.4. Overall results

Table 6 presents the accuracy of the aggregated annotators’ predictions, ChatGPT and CamemBERT (fine-tuned on news articles and SMS), on the whole test set and for different portions of the test set: Le Monde sentences, SMS sentences, sentences in which syntactic restrictions (clefts or adverbial modifications) prescriptively enforce the presence of *parce que*, and sentences in which the connective is preceded by a coordinator (conjunction or comma).

Overall, on the complete test set, CamemBERT achieves a significantly better accuracy than the human annotators and ChatGPT (Cochran’s $Q = 41.21$; $p < 0.001$).

	Annotators (aggregated)	ChatGPT	CamemBERT
Le Monde (N = 215)	69.8	54.6	71.2
SMS (N = 205)	58.0	60.6	62.0
Restricted (N = 34)	91.2	61.8	91.2
Coordinated (N = 110)	60.9	68.2	73.6
Total (N = 420)	60.2	57.6	66.7

Table 6. Classification accuracy (%) of human annotators, ChatGPT and CamemBERT on the different portions of the test set.

5.5. Sensitivity to syntactic and semantic restrictions

In Table 6, the results on the specific set of 34 sentences where *parce que* is expected because of syntactic restrictions show that both CamemBERT and the human annotators correctly predict *parce que* 31 times. Sentence (9) is one of two of those 34 sentences upon which CamemBERT and the human annotators do not agree: only annotator 1 and CamemBERT predicted *parce que*, while ChatGPT and annotators 2 and 3 predicted *car*. In comparison, in sentence (6) (see Section 3.1), which contains both a cleft and the adverb *notamment*, all models and annotators (except annotator 2) correctly predicted *parce que*.

(9) Ce lieu va rayonner sur toute l'Europe, **notamment parce que** Voltaire va s'impliquer personnellement dans les affaires Calas, Sirven, et surtout celle du chevalier de La Barre, accusé de sacrilège, condamné à avoir la langue coupée, à être décapité et brûlé vif.

*This place will shine throughout all of Europe, **especially because** Voltaire will become personally involved in the Calas and Sirven cases, and above all that of the Chevalier de La Barre, who was accused of sacrilege and sentenced to have his tongue cut out, beheaded and burned alive.*

Humans and the fine-tuned CamemBERT model seem to be sensitive to the constraints of clefts and adverbial modifications when predicting connectives (91.2% accuracy). On the other hand, ChatGPT (61.8%) seems to not rely on the expected use of *parce que* in those constructions.

We also looked at the 110 sentences of the test set in which *car* or *parce que* is used as a semantic coordinator between two clauses: when it is found after a coordinating conjunction, as in (10), or right after a comma, as in (11). This occurs in 78 sentences with *car*, and in 33 sentences with *parce que*. However, the accuracy results on those 110 sentences (see Table 6) show that both CamemBERT (73.6% accuracy) and ChatGPT (68.2%) are better than humans (60.2%) at predicting the correct coordinating connective. While all humans and models correctly predicted *parce que* after *et* ('and') in (10), the *car* in (11) was correctly predicted by CamemBERT and ChatGPT, but not by two out of three human annotators. These results may suggest that predicting a connective inside a single sentence (without further context) is a particularly difficult task for human readers, but that LLMs may rely on additional knowledge (such as semantic relationships between coordinators and connectives) acquired during pre-training or fine-tuning.

(10) Tiens-moi vite au courant si tu veux que je la fasse pour aller chercher le tendeur **et parce que** j'ai une autre voiture à faire en urgence...

*Let me know as soon as you can if you want me to do it to go and get the tensioner **and because** I've got another car to do in a hurry...*

(11) Et, de toute façon, je ne vois pas l'utilité de parler anglais, **car** je ne sais pas combien de temps je ferai les saisons.

*And, anyway, I don't see any point in speaking English, **because** I don't know how long I'll be doing the seasons.*

6. Conclusion

Using human annotators and LLMs (ChatGPT and CamemBERT) for predicting *car* or *parce que* in French, we find that distinguishing between these connectives with such subtle differences in meaning is a very difficult task, both for humans and LLMs, although fine-tuning a CamemBERT model allows it to handle these subtleties better than human native speakers of French. The public-available LLM-based chatbot ChatGPT showed weaker results than the human annotators, especially on sentences containing restrictions syntactically forcing the presence of one connective.

Further research avenues should focus on providing more context around the sentences, to investigate how additional semantic and pragmatic cues might influence humans and LLMs in this task. Gathering additional data with a larger panel of human annotators stratified across different age groups would further validate our results, as it has been shown that the age variable significantly influences the use of *parce que* and *car* in modern French (Blochowiak and Grisot, 2022). Additionally, our study could be extended to analyze how humans and (LLMs) handle similar words in other languages, such as the *want* vs. *omdat* pair in Dutch, where we observe a more stable semantic-pragmatic profile (Degand, 1996; Sanders and Spooren, 2009).

References

Berson, I. R., & Berson, M. J. (2023). The democratization of AI and its transformative potential in social studies education. *Social Education*, 87(2), 114-118.

- Blochowiak J. & Grisot C. (2022). The mapping between the subjective-objective divide and French causal connectives: Is it about the relation, the connectives or the speakers? In Maria Josep Cuenca & Liesbeth Degand (eds.), *Discourse Markers in Interaction: From Production to Comprehension*, 45–68 Berlin: De Gruyter.
- Blochowiak, J., Grisot, C., & Degand, L. (2020). What type of subjectivity lies behind French causal connectives? A corpus-based comparative investigation of *car* and *parce que*. *Glossa: a journal of general linguistics*, 5(1).
- 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.
- Debaisieux, J.-M. (2004). Les conjonctions de subordination: Mots grammaticaux ou mots de discours? Le cas de *parce que*. *Revue de Sémantique et Pragmatique*, 15–16, 51–67.
- Degand, L. (1996). Causation in Dutch and French: Interpersonal aspects. In Ruqaiya Hasan, Carmel Cloran & David Butt (Eds), *Functional Descriptions: Theory in Practice*. John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/cilt.121.08deg>
- Degand, L., & Pander Maat, H. (2003). A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In A. Verhagen & J. M. van de Weijer (Eds.), *Usage-based approaches to Dutch* (pp. 175–199). Lot.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers)*, 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- 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.
- Gao, S., Alawad, M., Young, M. T., Gounley, J., Schaefferkoetter, N., Yoon, H. J. et al. (2021). Limitations of transformers on clinical text classification. *IEEE journal of biomedical and health informatics*, 25(9), 3596–3607.
- Groupe Lambda-L. (1975). *Car, parce que, puisque*. *Revue Romane*, 10, 248–280.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv:1907.11692*.
- Lyu, Q., Apidianaki, M., & Callison-Burch, C. (2024). Towards faithful model explanation in nlp: A survey. *Computational Linguistics*, 1-70.
- 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*.
- Sanders, T. & Spooren, W. (2009). Causal categories in discourse. Converging evidence from language use. In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition*, 205–246. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110224429.205>
- Nazarenko, A. (2000). *La cause et son expression en Français*. Paris: Ophrys.
- Zufferey, S. (2012). “Car, parce que, puisque” revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics*, 44(2), 138–153.
- Zufferey, S., Mak, W., Verbrugge, S., & Sanders, T. (2018). Usage and processing of the French causal connectives ‘car’ and ‘parce que.’ *Journal of French Language Studies*, 28(1), 85–112.