Chère maison or *maison chère*? Transformer-based Prediction of Adjective Placement in French

Eleni Metheniti ① ② **Tim Van de Cruys** ③

Wissam Kerkri ④ **Juliette Thuilier** ① **Nabil Hathout** ①

① CLLE-CNRS ② IRIT ③ UK Leuven
④ Université Toulouse - Jean Jaurès (UT2J) Université Toulouse - Paul Sabatier (UT3) Department of Linguistics Leuven.AI institute

firstname.lastname@{univ-tlse2.fr,irit.fr,kuleuven.be}

Abstract

In French, the placement of the adjective within a noun phrase is subject to variation: it can appear either before or after the noun. We conduct experiments to assess whether transformer-based language models are able to learn the adjective position in noun phrases in French—a position which depends on several linguistic factors. Prior findings have shown that transformer models are insensitive to permutated word order, but in this work, we show that finetuned models are successful at learning and selecting the correct position of the adjective. However, this success can be attributed to the process of finetuning rather than the linguistic knowledge acquired during pretraining, as evidenced by the low accuracy of experiments of classification that make use of pretrained embeddings. Comparing the finetuned models to the choices of native speakers (with a questionnaire), we notice that the models favor context and global syntactic roles, and are weaker with complex structures and fixed expressions.

1 Introduction

In French, the placement of the adjective is subject to a considerable amount of variation—a phenomenon that has been under close scrutiny among linguists. Generally speaking, adjective placement in anteposition or postposition is attributed to many intertwining, linguistic processes, rather than a few rigid grammatical rules. However, the order of the adjective can be crucial to the meaning of the noun phrase; in the titular example, *chère maison* means “dear house" but *maison chère* means “expensive house".

Meanwhile, natural language processing researchers investigate whether language models built by transformer architectures are able to capture some of the inner workings of human language during their learning process. So far, research has shown that the high performance of such models does not imply the understanding of basic concepts such as grammatical order because the transformer architecture is non-sequential by design.

We are exploring whether transformer-based language models are capable of perceiving the adjective's position in a sentence with regard to its head noun, with a variety of experiments. Our goal is not to set a new state-of-the-art, but to explore if and how this information on word order is learned and used in tandem with the contextual word embedding information. While previous work has shown that transformer models are insensitive to word order (Pham et al., 2021; Gupta et al., 2021), finetuned models were successful in classifying adjective position (Sinha et al., 2021b). We also tested variations of finetuning training sizes and the use of attention masks to hide either the context of the sentence or the head noun and adjective.

For most adjectives, classifying their position is a relatively easy decision based on frequency; to observe the models' underlying competencies in more complex cases, we carried out an error analysis and additional experiments and visualizations on the pretrained versions of the models. We also had the opportunity to conduct an experiment with native French speakers, to compare their choices in challenging cases of adjective placement to the models' predictions.

Our findings show that finetuned models are capable of learning word order and efficiently classifying it; this knowledge is fainter in pretrained embeddings, but some layers demonstrate some specialization. Finetuning a model helps to learn these variations in adjective position and very successfully select the correct one. CamemBERT models were more successful than FlauBERT models over all experiments and captured more positional information in the finetuned adjective embedding. However, all transformers models show weaknesses (to different degrees) in complex cases of adjective/noun dependent phrases and fixed expressions.

Findings of the Association for Computational Linguistics: EACL 2023, pages 1668–1683

2 Position of adjective in French noun phrases

While traditional grammar proposes that adjectives in French follow the noun, in a noun phrase, linguistic analysis supports that adjectives are mobile, i.e. can be placed in anteposition or postposition relative to the noun (Abeillé and Godard, 1999). However, most adjectives tend to appear in specific positions; adjectives that accept only anteposition, only postposition, and those that alternate position (Benzitoun, 2013). For example, ordinal adjectives in *-ième* (e.g. *troisième* 'third'), are almost always anteposed to the noun, the adjectives *exotique* 'exotic', *idéal* 'ideal', *populaire* 'popular', *moderne* 'modern', *géant* 'giant', *naturel* 'natural' are always postposed, and the adjectives *énorme* 'huge', *immense* 'immense', *superbe* 'superb' alternate between the two possible position (Larsson, 1994; Benzitoun, 2014).

The preferred position of an adjective depends on its features and frequency; for example, Benzitoun (2014) claims that the adjective *prochain* 'next' in plural form does not occur in postposition (based on corpora statistics), but the singular does. Wilmet (1980, 1981) calculated that the most frequent adjectives in a corpus of literary works tend to precede the noun. However, chromatic adjectives (e.g. *rouge* 'red') which are of high frequency are always postposed to nouns when not a part of a multi-word expression. Adjectives derived from nouns and adjectives have a very strong tendency to be postposed (Forsgren, 2016; Goes, 1999). Wilmet (1981) and Forsgren (1978) support that the length of the adjective affects its position; short adjectives (e.g. *bon* 'good', *beau* 'pretty') tend to antepose, while longer adjectives and derivatives can only be postposed.

Semantic factors may also affect the position of an adjective with respect to its head word. For example, adjectives with multiple meanings may have different meanings in different positions; e.g. *propre* when anteposed refers to possession 'own', but when postposed means 'clean' (Thuilier, 2013). Benzitoun (2014) also presents the concept of adjective-noun pairs, where the meaning of the noun influences the position of the adjective. These pairs differ from fixed expressions because it is possible to create a pair with a different order (and different meaning), while fixed expressions are lexicalized and do not allow the existence of a variation with a different meaning. For example, the lexicalized phrase *arts premiers* (where *premier* is postposed) has a very specific meaning ('arts of the non-Western world') compared to *premiers arts* 'first arts' where it used in its literal sense and is not a lexicalized phrase.

The presence of more dependents in the noun phrase also affects the position of the adjective. The presence of an adverbial modifier to the adjective may force the adjective phrase to postposition or increase the occurrence of the adjective in postposition, or at least allow more flexible positioning of the adjective phrase relative to the noun (Forsgren, 1978; Thuilier, 2013). A definitive case of postposition happens when an adjective has a multi-word modifier, e.g. a prepositional phrase (Thuilier, 2013). Postposition is also favored when there are multiple adjectives defining the noun. Thuilier (2013) also suggests that elements in the syntactic phrase are ordered by increasing length (known as *increasing* or *relative mass*). However, it may not apply to high-frequency adjectives such as *magnifique* 'magnificent' (Larsson, 1994).

3 Word order and transformer models

There has been extensive work on analyzing the syntactic and semantic capabilities of transformer models and their pretrained word embeddings, with positive and negative findings on the abilities of these models to capture linguistically salient word relations. In this review, we focus on word position and word order findings.

The addition of absolute word order (i.e. the sequential order of words) to the training process of contextual word embedding models has proven quite beneficial. Transformer models with bidirectional training, which captures adjacent word order, showed improvement compared to the original self-attention neural networks (Yang et al., 2019). Transformer models trained with masked language modeling, such as BERT and RoBERTa, are able to learn absolute word positions, but they also learn structural word positions (i.e. phrase position in hierarchical tree structures) and make use of them (Wang et al., 2019; Wang and Chen, 2020). Multiple experiments combine absolute and structural word positions to create better-informed and better-performing word embeddings (Wang et al., 2020; He et al., 2021; Chang et al., 2021; Wang et al., 2021).

However, experiments on already pretrained language models and shuffled word order tell a different story. Pham et al. (2021) conducted experiments on BERT-based models (BERT, RoBERTa, ALBERT) with GLUE classification tasks, and showed that tasks such as sentiment analysis were not affected by shuffled word order, except for the grammatical correctness task. O'Connor and Andreas (2021) conducted experiments on the effect that context variation has on transformer models' usable information, and discovered that word shuffling has a negative effect, whether the shuffling was implemented on short or long distances among words. Gupta et al. (2021) conducted similar experiments with GLUE tasks and observed that model performance was lower on shuffled word orders (in methods that render a sequence ungrammatical and incomprehensible to humans) but close enough to support that models rely more on embedding information rather than sequential context. Sinha et al. (2021b) confirm that pretrained language models are insensitive to word order in tasks of Natural Language Inference and show that, on some occasions, classification is successful only with certain (random) word order variations of an input sequence. They also conducted experiments on finetuned models and noted finetuning's positive influence on learning word order. Finetuning improved performance on tasks of inference and grammaticality as well (even with models pretrained with scrambled word order) (Sinha et al., 2021a). For French, Li et al. (2021) conducted experiments, on the transformer models' capacity to capture long-range object-verb agreement and word order (in one of their experiments). They observed that models performed worse with scrambled inputs, and increasingly worse, for increasingly complex relations.

4 Experiment 1: Finetuning and classification of adjective position

4.1 Methodology

Given the findings from previous work, highlighting the syntactic and semantic capacities of transformer models as well as also their weakness in learning word order, we want to test whether transformer models are able to classify the position of the adjective in a sentence.

In order to provide the two possible positions that the adjective may have in a noun phrase, we provide a pair of sentences as input: the first sentence of the input has the adjective always anteposed to the noun, and the second sentence always postposed. We label the two-sentence sequences with '0' if the first sentence is correct (i.e. the correct order is anteposition) and '1' if the second sentence is correct (i.e. postposition)—see example in Table 1. The sentences are separated by the specific end-of-sequence token of each model. With this task, we aim to observe if word order is insignificant to the models or if they are able to capture the preferred word order between two sentences with identical tokens and different word order. We finetuned the transformer models for 4 epochs based on the guidelines by Devlin et al. (2019) and McCormick and Ryan (2019) (see Section 4.2 for datasets and details).

We also run the same experiment with a one-sentence input, with the original sentence without any permutations. The models were finetuned for 4 epochs as well, with the original sentence and its label of ante-/postposition. This method is less informative, as the model is not aware of the different possible positions of the adjective, and will only predict correctness.

In order to further study the contribution of different tokens in the input sequence, we also finetuned the models with blocked attention to certain tokens; we used the attention mask, which is an array that instructs the model's self-attention mechanism to attend to specific tokens of the input sequence, by assigning 1s to the "visible" tokens and 0s to the "invisible" ones. In addition to the *default* setting of attending to all tokens, we tested a *pair* setting, in which all tokens are masked except for the adjective and its head noun, and a *context* setting, in which the adjective and noun are masked and all the other tokens are visible. Our goal is to observe whether the adjective-noun pair is significant enough to encapsulate their preferred positions or not, and whether the context contains (enough) information on preferred adjective-noun positions even without explicit information on the pair. We present a visualization of what an input sentence looks like in these settings in Table 2.

4.2 Datasets

We extracted sentences with correct adjective-noun pairs from two parsed corpora: the frWaC corpus (Baroni et al., 2009) and the French corpora of Universal Dependencies 2.9 (UD; Zeman et al., 2021), in different combinations[1].

[1]The list of corpora can be found at `https://universaldependencies.org/fr/`

On construit les éléments de plus **haut niveau**. ⇓	
SENTENCE	LABEL
On construit les éléments de plus haut niveau. </s>On construit les éléments de plus niveau haut.	0

Table 1: An example input of two sentences for the original sentence *On construit les éléments de plus haut niveau* 'We build the higher level elements'. We only shift the position of the adjective-noun pair in the noun phrase, without affecting any other elements of the phrase (e.g. the dependent adjective *plus*).

MASK	TOKENS	
Default	on construit les éléments de plus	haut niveau
Pair		haut niveau
Context	on construit les éléments de plus	

Table 2: Use of attention masks for the sentence: *On construit les éléments de plus haut niveau*. In this sentence, the adjective-noun pair is *haut niveau* (the adjective is before the noun). The label for all three inputs is [0]. For the double-sentence input, the same process will be followed for the second sentence of the input *On construit les éléments de plus niveau haut.*

We used all UD sentences and selected 120K relevant sentences from frWaC, with a 2/3 ratio of anteposition/postposition, which is roughly the ratio documented in the literature and the one that occurs in our corpora[2] –this ratio is also beneficial since anteposed adjectives are fewer but more frequent. However, we excluded the adjectives and words which were incorrectly parsed as adjectives, such as numerals, pronouns such as *autre*, *certain*, *chacun*, *quelque* which may have other linguistic functions than an adjective. In addition, we also excluded the adjectives and the nouns which were tokenized into subwords by the transformer model tokenizers, in order to create the attention mask described in Section 4.1.

The sentences of the two datasets were combined and used in various ways. In one setting, we trained the model only with frWaC, and used the UD sentences as an additional test set. In another one, we added a subset of the UD sentences to the train set and tested on the rest of UD; we also finetuned the model just with the (significantly) smaller UD dataset. When applicable, we tested both with frWaC and UD sentences. The size of the datasets is presented in Table 3.

[2]Measured on 1M frWaC sentences and the entire UD corpora.

Dataset	**Train**	**Val.**	**frWaC test set**	**UD test set (entire)**	**UD test set (part)**
frWaC	76,164	7,672	7,740	19,437	5,151
frWaC + UD	91,615	7,672	7,740	-	5,151
UD	13,905	1,546	7,740	-	5,151

Table 3: Dataset sizes for the finetuned models.

4.3 Transformer models

We used two monolingual French transformer-based models, available from the HuggingFace Python library (Wolf et al., 2020), CamemBERT (Martin et al., 2020) and FlauBERT (Le et al., 2020). CamemBERT is the pioneering monolingual French model and is built based on the RoBERTa architecture and trained on monolingual data. Experiments showed its advantage on traditional NLP tasks over multilingual transformer models. The authors also highlight the base version's high performance with a fraction of the size of the large version. FlauBERT is a monolingual French BERT-based model trained with multiple, heterogeneous corpora and a more extensive tokenization procedure. It has been shown to (slightly) outperform CamemBERT on French benchmark tasks.

4.4 Baselines

The simplest baseline we can establish is based on frequency in our corpus: we assign each adjective a label of ante-/postposition based on its most frequent position in the training set. We also performed classification with more classical NLP methods, namely a logistic regression model on bag-of-words, implemented with `scikit-learn` (Pedregosa et al., 2011), and a CNN-based classifier, more sensitive to word order, implemented with PyTorch (Paszke et al., 2019).

4.5 Results

The results for the two-sentence input experiment can be found in Table 4 (and for the one-input in Table 7 in Appendix A). We can already observe that frequency yields a quite high accuracy, the bigger the training set is and the smaller the test set is. The CNN classifier is very successful when the training set is large enough. Therefore, it comes as no surprise that the finetuned transformer models made very few mistakes, with the overall accuracy being close to 100%. The results were consistently high, even when testing with a different dataset (frWaC and UD). However, with a much

Model	frWaC train			frWaC+UD train		UD train	
	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test
`camembert-base`	**0.99**	**0.93**	**0.93**	**0.99**	**0.99**	0.93	0.95
`camembert-large`	**0.99**	0.91	**0.93**	**0.99**	**0.99**	**0.98**	*0.66*
`flaubert-small-cased`	**0.99**	0.90	0.90	**0.99**	**0.99**	0.62	*0.66*
`flaubert-base-cased`	**0.99**	0.90	0.87	**0.99**	0.97	0.96	**0.96**
`flaubert-base-uncased`	**0.99**	0.90	0.91	**0.99**	**0.99**	0.95	0.95
`flaubert-large-cased`	**0.99**	**0.93**	0.88	**0.99**	**0.99**	0.91	0.87
Position frequency	0.91	0.77	0.93	0.91	0.94	0.45	0.62
Logistic Regression	*0.45*	*0.68*	*0.66*	*0.45*	0.65	0.82	0.87
CNN	0.94	0.48	0.94	0.96	0.95	0.55	0.72

Table 4: Classification results for the finetuned models and baselines, with the different training and test sets. Values in *italics* indicate that the model completely failed to classify.

Model	Attention mask: hidden context							Attention mask: hidden adj + noun						
	frWaC train			frWaC+UD train		UD train		frWaC train			frWaC+UD train		UD train	
	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test
`camembert-base`	**0.99**	**0.80**	**0.83**	**0.99**	**0.99**	0.78	0.83	**0.99**	0.45	0.57	**0.99**	0.98	0.45	0.66
`camembert-large`	0.98	0.76	0.76	0.98	0.99	**0.87**	**0.91**	*0.45*	*0.66*	*0.66*	*0.45*	*0.66*	0.45	0.63
`flaubert-small-cased`	*0.45*	*0.68*	*0.68*	*0.45*	*0.66*	*0.45*	*0.66*	0.99	0.52	0.52	0.99	0.98	0.47	0.64
`flaubert-base-cased`	*0.45*	*0.68*	*0.68*	*0.45*	*0.66*	*0.45*	*0.66*	**0.99**	0.47	0.47	**0.99**	**0.99**	**0.58**	**0.68**
`flaubert-base-uncased`	*0.45*	*0.68*	*0.68*	*0.45*	*0.66*	*0.45*	*0.66*	**0.99**	**0.61**	**0.61**	**0.99**	**0.99**	0.47	0.62
`flaubert-large-cased`	*0.45*	*0.68*	*0.68*	*0.45*	*0.66*	*0.45*	*0.66*	**0.99**	0.54	0.54	**0.99**	**0.99**	0.50	0.64

Table 5: Classification results of the finetuned models with attention masks. Values in *italics* indicate that the model completely failed to classify.

smaller training set, results were slightly lower (as expected; finetuning guidelines recommend a training set of at least 100K inputs). In comparison, the accuracy of the one-sentence finetuning experiment is 11-12% lower, which is even lower than the frequency-based baseline and the CNN classifier.

The results of the experiments with attention masks are presented in Table 5 (and Table 8 in Appendix A for the one-input finetuning). In these experiments, the models' attention mechanism had access to only certain tokens. When attention was only allowed to the adjective and noun pair, the Flaubert models were unable to classify, while the Camembert models showed equally outstanding performance with the frWaC sentences (but lower performance with the UD test sets). Meanwhile, masking the adjective and noun pair and only allowing attention to the rest of the sequence was surprisingly successful for the finetuned models with the larger training sets (except for `camembert-large`), reaching similar accuracies to those of the no-mask finetuned models. For the one-input finetuning experiment, we notice that, for the masked context scenario, performance rose drastically only for CamemBERT models and only in the frWaC domain, while the Flaubert models were again unsuccessful. For the masked adjective-pair scenario and for the UD domain, the performance is significantly lower.

4.6 Qualitative analysis

In most cases, the models make very few mistakes, which are not consistent among models. Moreover, the models are very confident in their choices, assigning high probabilities to all predictions (see Figure 1 in Appendix).

Focusing on the frWaC training set with the UD dataset as the test set, we notice that most of the sentences that were mislabeled are ones where the adjective could possibly be in a different position, with a different meaning than the original one (i.e. the utterance remains grammatical when the adjective-noun order is reversed). For example, the sentence *Une école a ouvert dans une **ancienne église** en 1950* 'A school opened in a former church' remains correct with *ancienne* postposed to the noun, but the meaning of the adjective becomes 'old'. The context provided by the sentence is not sufficient to decipher the actual meaning, and native French speakers agree that both sentences are grammatical. On the other hand, mistakes in the classification of sentences such as *Les **créations sensuelles**, modernes et orientales se font remarquer* 'The sensual, modern and oriental creations stand out' uncover the models' shallow perception of syntactic relations –these mistakes were, however, very rare. Finally, we notice a few badly-parsed and badly-formed sentences in the dataset, which were not enough to warrant a redesign, but were confusing to the models.

5 Experiment 2: Pre-existing knowledge in pretrained embeddings

The previous experiment shows that the finetuned transformer models are quite successful in classifying the adjective's position when asked to distinguish between two possible positions. The following part of this research aims to observe whether this capability is given by the finetuning, or whether the pretrained models had already learned enough information on the adjective's preferred position, with regard to its context.

5.1 Classification with adjective embeddings

The layers of a transformer model specialize in creating different dynamic word embeddings, which capture and interact with a word's context in a different way than the previous layer. Therefore, the adjective embedding might contain the syntactic, contextual, and semantic information that determine its position with regard to the noun. We extracted the word embeddings for the adjective of each sentence, per layer, and we trained a simple logistic regression model –built in the same way as in Section 4.4. We used the frWaC training set and tested on the frWaC test set and on the entire UD dataset.

The results of the classification for the two test sets can be seen in Figure 2 in Appendix C. The classification results for the frWaC test set are quite low –close to being non-classifiable– except for the `flaubert_base_uncased` model, which unexpectedly reached 97% accuracy on the last layer. Results for the UD test set were more unpredictable, with a few layers of *camembert-base* reaching a very high accuracy, but the final layer having the lowest accuracy. On the other hand, the `flaubert` models had a progressively better performance, but they are not as good as their finetuned counterparts nor as the baselines.

5.2 Adjective [MASK] probabilities with Masked Language Models

Pretrained models can predict the tokens that can fill a masked position in a sequence. We use this method to retrieve the probability that the models have assigned to the adjective in the sentence, specifically in the position it was found in. We make use of the sentences of the frWaC test set. (These probabilities are presented in Figure 3 in Appendix D.) We observe that overall the models assigned higher probabilities to anteposed adjectives being in anteposition, than to postposed adjectives in postposition; apart from the stricter linguistic constraints for anteposed adjectives, this could also be due to the fact that transformer models favor token frequency, and most of the most frequent adjectives in French are anteposed, while postposition harbors far more adjectives. Additionally, we observe that CamemBERT models give higher probabilities in the predictions of both anteposed and postposed adjectives.

When we shifted the [MASK] position from its original place to the opposite one, and asked the models to assign the adjective's probability in the "wrong" position, the probability of the adjectives was close to zero for at least 85% of the cases, even for anteposed adjectives which are more versatile.

6 Experiment 3: Human vs Transformers judgments of adjective order

6.1 Methodology and Dataset

We had the opportunity to carry out an additional experiment on adjective word order, in which we studied how native French speakers and the finetuned transformer models dealt with challenging cases of adjective position, caused by structural or semantic idiosyncrasies. As observed in the previous experiments, the models demonstrated weaknesses in cases of adjacent adjectives that did not belong to the noun phrase, and their choices did not always align with the original sentence in cases of semantic ambiguity relating to the adjective position. These cases cannot always be coined as errors, since native speakers may also make similar choices whether intentionally (e.g. different comprehension of context) or unintentionally (e.g. haste, lack of attention).

The structure of the experiment is the same as in Experiment 1, where speakers and the finetuned models were presented with a sentence containing a noun-adjective pairing, and its variation having the target adjective in the opposite position. Regardless of the original order, each sentence pair of the two positions was presented in the order of anteposition-postposition. We created 89 prompt sentences, written by a native French speaker or extracted and modified from frWaC, and evaluated by French speakers. The full dataset can be found in Appendix F. The sentences are split into four categories based on the type of relations that the adjective has with the noun, or the context of the sentence:

1. *Presence of adjective/noun dependent*: The only categorical constraint that governs the position of the adjective in French is the presence of a dependent to the adjective, which forces the position of the adjective to postposition. However, if the dependent is to the noun, the position of the adjective is not restricted. We included sentences with the same adjectives and dependents either to the adjective or the noun.

2. *Fixed expressions*: Adjectives in fixed expressions will always have a fixed position in this specific context and meaning. Apart from sentences with fixed expressions we selected, we added sentences with the adjectives found in those expressions, but not in restrictive structures.

3. *Structural persistence*: Speakers are sensitive and tend to reuse repeating syntactic constructions (*syntactic priming*, (Branigan et al., 1995)). The presence of a noun phrase with an adjective in a certain position may influence the processing of the next noun phrase, especially if it contains the same adjective. We want to test the extent of this effect on native speakers and our models.

4. *Blocked and mobile adjectives*: In this category, we are including adjectives which are (almost) always found in postposition, and adjectives with free position depending on the meaning (*propre*, *ancien*). This category serves both as a control group, but could also provide unexpected results.

6.2 Questionnaire diffusion

While the finetuned models received all sentences as a test set, we divided the prompt sentences in 3 questionnaires, ensuring that there is equal distribution of the four categories in each. The participants were asked to select the sentence that sounded “most natural" to them, out of the two position variations. In order to eliminate outliers or non-native speakers of French, at the start of each questionnaire we asked for input of first language, and to confirm that they were native speakers of French (and also to acclimatize the participants with the experiment) there was a mini-tutorial with two sentence pairs which could not be mistaken by French speakers. The questionnaire was built with LimeSurvey[3] and distributed to French university students and French locals. Out of the 71 participants who completed the questionnaire and were not considered outliers, each version of the questionnaire had 22-25 participants, i.e. each sentence pair was evaluated by at least 22 speakers.

[3] https://www.limesurvey.org/

Model	Micro avg.	Macro avg.
`camembert-base`	0.3326	0.1629
`camembert-large`	**0.5801**	**0.4673**
`flaubert_small_cased`	**0.6014**	0.3711
`flaubert_base_cased`	0.433	0.3446
`flaubert_base_uncased`	0.5192	0.3298
`flaubert_large_cased`	0.3688	0.3554

Table 6: Correlation between the average choice of the speakers and each model’s output. Micro-averaged is aggregating all sentences regardless of category while macro-average is category-sensitive.

6.3 Quantitative and Qualitative Results

We calculated the average selection over all speakers and used this as the baseline to make judgments for our models. In Table 6 we are presenting the Pearson correlation between the speakers’ and the models’ choices, in order to see which of the models was closer to the behavior of the speakers. The model that achieved the highest micro- and macro-averaged correlation was `camembert-large`, although `flaubert_small_cased` model was slightly better at micro-averaged correlation – an interesting finding, since this model is created for debugging purposes and its results are unreliable. The `camembert-base` and `flaubert_large_cased` models showed the lowest correlations, and all models except for `camembert-large` did not show a strong positive correlation (>0.4) in the macro-averaged correlation.

We also examined the speakers’ decisions and the models’ predictions per category and performed error analysis. For the *presence of adjective/noun dependent* category, the speakers preferred longer adjectives in postposition, even when the dependent phrase was attached to the noun: for example, the speakers unanimously chose the postposed variation of the sentence *Ils vivent une différente relation sans amour.* “They lived a different relationship without love." and so did most of the models. However, for shorter adjectives, the speakers chose anteposition when there was a noun dependent and postposition when there was an adjective dependent. The models however did not present a uniform behavior, with some models mostly preferring postposition (camembert-large) or anteposition (flaubert-large-cased), while the more successful ones made mistakes on the shorter adjectives.

In the *fixed expressions* category, the speakers naturally did not make any mistakes on the fixed expressions, and were able to differentiate between the fixed and the free position of the same adjective in different contexts. However, the models made several mistakes on very common fixed expressions, e.g. *la grasse matinée* “the morning of sleeping in", but were not mistaken on expressions with a short adjective, e.g. *bénéfice net* “net benefit" (i.e. the short adjective was not anteposed, while its variations in non-fixed phrases are commonly anteposed). In the category of *structural persistence*, the speakers were able to make their choices for the adjective position despite being primed by a previous noun phrase with the opposite adjective position, e.g. they preferred the variation *Il lui a offert des volumineuses plantes à fleurs volumineuses.* “He offered them voluminous plants with voluminous flowers." for the noun phrase *fleurs volumineuses*. However, all the models predicted anteposition, and this could have been affected by the adjectives being in the same wordform. Finally, in the *blocked/mobile* adjectives category, the speakers did not make any inexplicable choices, and always preferred postposition for the postposed adjectives (e.g. chromatic) and both positions for the mobile adjectives (despite the length). The only model which made mistakes on the postposed adjectives was `flaubert-large-cased`, while the other models made very few mistakes on mobile adjectives–decisions which are to some extend acceptable, since the meaning may be different but still grammatical.

7 Discussion

Previous work on exploring transformer models has supported that their success in NLP tasks is heavily based on their vast training data and efficient learning of frequencies. Our experiments, compared to a frequency-based uninformed baseline, show that there are more complex operations in play. Transformers were more efficient than sequential-order-learning neural networks, and were in fact able to differentiate between two sentences with identical tokens and slightly different word order. Finetuning is more efficient with a larger training corpus and different domains, but can still be successful with a smaller dataset if necessary.

When the models' attention mechanism only has access to the context, and not to the adjective-noun pair itself, they were still quite capable of classifying adjective position even without attending to it. This observation is consistent with the linguistic description that supports that adjective position is also determined by context and not solely by the noun phrase. However, the fact that CamemBERT models were extremely successful in identifying position without the use of context, while Flaubert models failed completely, is caused by the models' different architectures and choices in the way the tokens are handled. In our more detailed experiments, we saw that CamemBERT models assign an overall higher probability to adjectives, regardless of their position, and that, at least for the UD dataset, the adjective embeddings were, in some layers, very informed on the preferred word position. This knowledge is correlated to the learned contextual word embeddings, rather than the word itself, as we observed a lack of semantic similarity in the visualization.

Regarding the models' mistakes in the testing phase, they were either caused by low-frequency adjectives, bad parsing, or ambiguous meaning which may be grammatical and acceptable in both adjective positions. However, comparing the models to human performance showed their true strengths and weaknesses; when they are successful, the models tend to follow a more rigid syntactic structure and favor postposition, as it is the most frequent adjective position over all adjectives. They showed severe problems in recognizing some fixed expressions, and were more easily swayed than humans by being primed with the same adjective. In cases where both positions were possible, they usually preferred the more “traditional" postposition. These findings may demonstrate that the models base their decisions on adjectives more on frequency rather than the syntactic and semantic information of a particular adjective, and are impervious to factors that affect speakers' decisions such as length, difficulty of processing with regard to cognitive load, and substantial or subtle semantic differences.

8 Conclusion

In this work, we aimed to study the capabilities of transformer-based language models in understanding word order, specifically the order of adjectives in a noun phrase in French. Our findings, for pretrained models, confirmed previous ones which claimed that these models are agnostic to word position. However, the process of finetuning and classification with two variations of the sentence (one correct and one with permutated adjective order) was very successful, which proves that the models are capable of learning and becoming sensitive to word order. Concerning the use of attention masks, the CamemBERT models were very capable of classifying word order by only attending to the adjective and noun, while for the Flaubert models it was impossible. The differences between the two architectures were also reflected in our study of the pretrained word embeddings and the adjective probabilities, where we noticed that CamemBERT's adjective embeddings were better informed. The adjective embeddings themselves, for all models, seem to contain more contextual than word-specific information, which makes different iterations of an adjective differ from each other. In our experiment comparing native speakers to the models' preferences, we observed that the models showed weakness in structures with dependents, fixed expressions, and priming, and reverted to the grammatically-established postposition more than humans. Therefore, the models' understanding of the position relies both on context and on shallow syntactic roles, but is lacking semantic nuances. We also observed that the information on position is specialized in some layers –and easily learned via finetuning.

Limitations

This work has been conducted in the French language, due to the available language resources and transformer models in this high-resource, in addition to the authors' adept knowledge of the language and its linguistic properties. We decided to focus on the specific phenomenon of adjective placement because it offers the possibility to study the models' sensitivity to word order on pairs with one grammatical and one ungrammatical sentence, but also with pairs where both sentences were grammatical. The finetuning of the transformers models, especially of the large versions, was made possible with the use of a server with GPU clusters, provided by our institution.

References

Anne Abeillé and Danielle Godard. 1999. La position de l'adjectif épithète en français: le poids des mots. *Recherches linguistiques de Vincennes*, (28):9–32.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.

Christophe Benzitoun. 2013. Adjectifs épithètes alternants en français parlé: premiers résultats. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, (29).

Christophe Benzitoun. 2014. La place de l'adjectif épithète en français: ce que nous apprennent les corpus oraux. In *SHS Web of Conferences*, volume 8, pages 2333–2348. EDP Sciences.

Holly P Branigan, Martin J Pickering, Simon P Liversedge, Andrew J Stewart, and Thomas P Urbach. 1995. Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24(6):489–506.

Tyler Chang, Yifan Xu, Weijian Xu, and Zhuowen Tu. 2021. Convolutions and self-attention: Re-interpreting relative positions in pre-trained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4322–4333, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Mats Forsgren. 1978. *La place de l'adjectif épithète en français contemporain: étude quantitative et sémantique*. Ph.D. thesis, Acta Universitatis Upsaliensis.

Mats Forsgren. 2016. La place de l'adjectif épithète. *Encyclopédie grammaticale du français (online)*. Accessed on Jan 04, 2022.

Jan Goes. 1999. *L'adjectif: entre nom et verbe*, volume 777. De Boeck Supérieur.

Ashim Gupta, Giorgi Kvernadze, and Vivek Srikumar. 2021. Bert & family eat word salad: Experiments with text understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12946–12954.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention. In *International Conference on Learning Representations*.

Björn Larsson. 1994. *La place et le sens des adjectifs épithètes de valorisation positive: Une étude de 113 adjectifs d'emploi fréquent dans la langue du tourisme et dans d'autres types de prose non-littéraire*. Lund University Press.

Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoit Crabbé, Laurent Besacier, and Didier Schwab. 2020. FlauBERT: Unsupervised language model pre-training for French. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France. European Language Resources Association.

Bingzhi Li, Guillaume Wisniewski, and Benoit Crabbé. 2021. Are Transformers a modern version of ELIZA? Observations on French object verb agreement. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4599–4610, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Chris McCormick and Nick Ryan. 2019. BERT Fine-Tuning Tutorial with PyTorch. `https://mccormickml.com/2019/07/22/BERT-fine-tuning/`. Retrieved January 24, 2021.

Joe O'Connor and Jacob Andreas. 2021. What context features can transformer language models use? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 851–864, Online. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Thang Pham, Trung Bui, Long Mai, and Anh Nguyen. 2021. Out of order: How important is the sequential order of words in a sentence in natural language understanding tasks? In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1145–1160.

Koustuv Sinha, Robin Jia, Dieuwke Hupkes, Joelle Pineau, Adina Williams, and Douwe Kiela. 2021a. Masked language modeling and the distributional hypothesis: Order word matters pre-training for little. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2888–2913, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Koustuv Sinha, Prasanna Parthasarathi, Joelle Pineau, and Adina Williams. 2021b. UnNatural Language Inference. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7329–7346, Online. Association for Computational Linguistics.

Juliette Thuilier. 2013. Syntaxe du français parlé vs. écrit: le cas de la position de l'adjectif épithète par rapport au nom. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, (29).

Benyou Wang, Lifeng Shang, Christina Lioma, Xin Jiang, Hao Yang, Qun Liu, and Jakob Grue Simonsen. 2021. On position embeddings in {bert}. In *International Conference on Learning Representations*.

Benyou Wang, Donghao Zhao, Christina Lioma, Qiuchi Li, Peng Zhang, and Jakob Grue Simonsen. 2020. Encoding word order in complex embeddings. In *International Conference on Learning Representations*.

Xing Wang, Zhaopeng Tu, Longyue Wang, and Shuming Shi. 2019. Self-attention with structural position representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1403–1409, Hong Kong, China. Association for Computational Linguistics.

Yu-An Wang and Yun-Nung Chen. 2020. What Do Position Embeddings Learn? An Empirical Study of Pre-Trained Language Model Positional Encoding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6840–6849, Online. Association for Computational Linguistics.

Marc Wilmet. 1980. Antéposition et postposition de l'épithète qualificative en français contemporain. *Travaux de linguistique*, 7:179–201.

Marc Wilmet. 1981. La place de l'épithète qualificative en français contemporain. etude grammaticale et stylistique. *Revue de Linguistique Romane Lyon*, (177-178):17–73.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. pages 38–45.

Baosong Yang, Longyue Wang, Derek F. Wong, Lidia S. Chao, and Zhaopeng Tu. 2019. Assessing the ability of self-attention networks to learn word order. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3635–3644, Florence, Italy. Association for Computational Linguistics.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arıcan, Hórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Deniz Baran Aslan, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkaður Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Lauren Cassidy, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Mihaela Cristescu, Philemon Daniel, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Jannatul Ferdaousi, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Kaoru Ito, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Krister Lindén, Nikola Ljubešić, Olga Loginova, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, Lorena Martín-Rodríguez, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhevich, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayọ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Dmitry Sichinava, Janine Siewert, Einar Freyr Sigurðsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Shafi Sourov, Carolyn Spadine, Rachele Sprugnoli, Steinhór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Sri Hartati Wijono, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Anna Zhuravleva, and Rayan Ziane. 2021. Universal dependencies 2.9. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

A Results for finetuning with one sentence input

Model	frWaC train			frWaC+UD train		UD train	
	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test
camembert-base	**0.89**	0.8	0.8	**0.89**	**0.87**	**0.84**	**0.87**
camembert-large	**0.89**	0.8	0.8	**0.89**	**0.87**	**0.84**	**0.87**
flaubert-small-cased	**0.88**	0.81	0.81	**0.88**	**0.87**	0.84	0.85
flaubert-base-cased	**0.89**	0.81	0.81	**0.89**	**0.87**	0.82	**0.87**
flaubert-base-uncased	**0.89**	**0.82**	**0.82**	0.88	**0.87**	0.82	0.87
flaubert-large-cased	**0.89**	0.81	0.81	**0.89**	**0.87**	0.83	0.87
Logistic Regression	*0.45*	*0.68*	*0.66*	*0.45*	0.65	*0.45*	0.65
CNN				0.8	0.84	0.68	0.79

Table 7: Classification results for finetuning models and baselines, with only one sentence as input, with our different training and test sets. Values in *italics* indicate that the model failed completely to classify

Model	Attention mask: hidden context						
	frWaC train			frWaC+UD train		UD train	
	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test
camembert-base	**0.99**	**0.99**	**0.99**	0.8	0.8		
camembert-large	0.97	0.98	0.98	**0.97**	**0.98**		
flaubert-small-cased	*0.45*	*0.68*	*0.68*	*0.76*	*0.79*	*0.45*	*0.66*
flaubert-base-cased	*0.45*	*0.68*	*0.68*	*0.8*	*0.8*	*0.45*	*0.66*
flaubert-base-uncased	*0.45*	*0.68*	*0.68*	*0.8*	*0.8*	*0.45*	*0.66*
flaubert-large-cased	*0.45*	*0.68*	*0.68*	*0.45*	*0.66*	*0.45*	*0.66*

Model	Attention mask: hidden adj + noun						
	frWaC train			frWaC+UD train		UD train	
	frWaC	UD-full	UD-test	frWaC	UD-test	frWaC	UD-test
camembert-base	0.79	0.77	0.77	0.79	0.89	0.67	0.82
camembert-large	*0.45*	*0.66*	*0.66*			*0.45*	*0.66*
flaubert-small-cased	0.76	0.75	0.75	0.76	0.82	0.59	0.74
flaubert-base-cased	0.8	0.69	0.69			**0.7**	**0.86**
flaubert-base-uncased	0.81	0.76	0.76			0.7	0.86
flaubert-large-cased	**0.82**	**0.79**	**0.79**			0.69	0.83

Table 8: Classification results of finetuning models with only one sentence as input and with attention masks. Values in *italics* indicate that the model failed completely to classify.

B Probabilities of predicted labels during classification

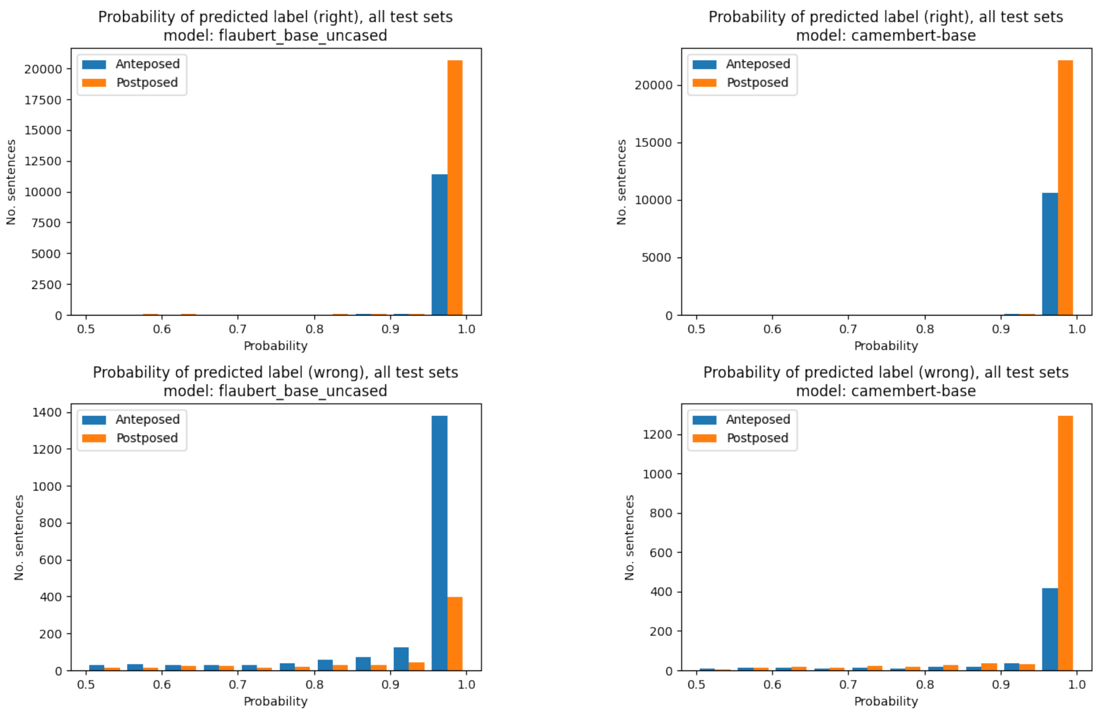

Figure 1: The probability of predicted labels, for wrong and correct predictions, from the frWaC train set and both test sets.

C Classification based on the adjective's pretrained embedding, with logistic regression

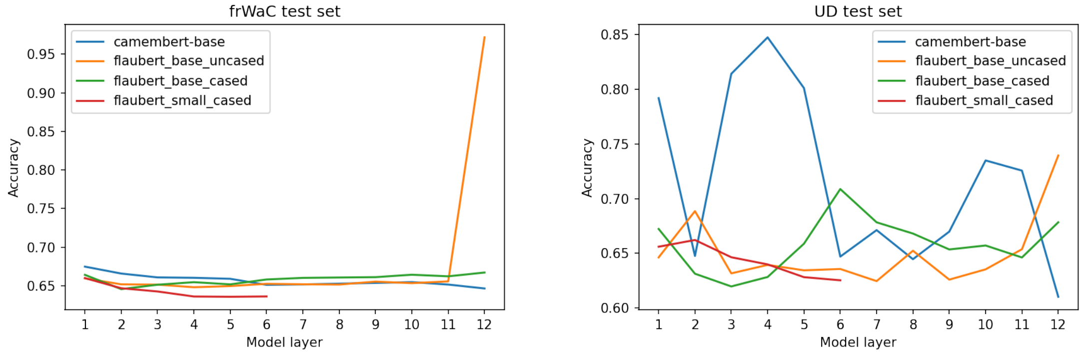

Figure 2: Logistic regression accuracy trained with layer-specific adjective embeddings, with our two large test sets.

D Adjective [MASK] probabilities with Masked Language Models

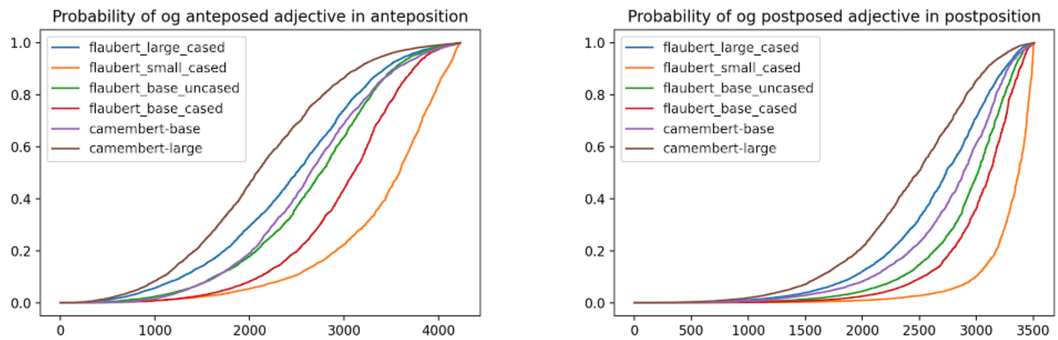

Figure 3: The assigned probability of each adjective instance, when placed in its original position which has been masked (anteposition/postposition), for each model.

E Visualizing adjective pretrained embeddings per layer

We extract layer-specific embeddings of some transformer models, and use them to visualize static embeddings by reducing their dimensions and plotting them on a 2-dimensional space, in order to observe their closest neighbors and possible clusters or patterns emerge. We selected a few frequent adjectives from the literature, either with a preferred position or without: *grand*, *petit* for always-anteposed, *naturel* for always-postposed, *ancien* for ambivalent. All the embeddings of each adjective (from the different sentences it appeared in) were used and plotted per layer.

We reduced the embeddings' dimensional with t-distributed Stochastic Neighbor Embedding (t-SNE) from `scikit-learn` and plotted with `matplotlib`. Some of the plots are presented in Figure 4. Our intuition was that the anteposed and postposed adjectives would form clusters. However, we could not observe discernible clusters in any of the data –the closest being for some early layers, for some adjectives, and for complex word forms rather than the base ones.

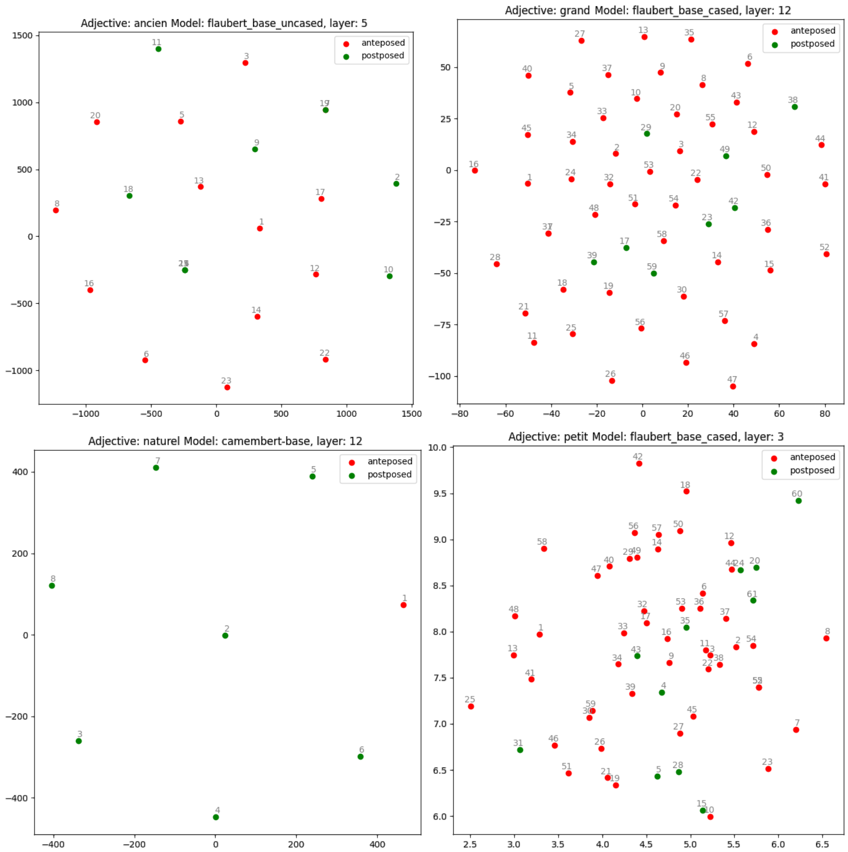

Figure 4: Embedding projections for base-form adjectives *ancien* 'old', *grand* 'large', *naturel* 'natural', *petit* 'small' –from various layers and models. The numbers correspond to the sentence id.

F Dataset for questionnaire (with English translations)

We have annotated in italics the sentences for which the French speakers preferred anteposition.

Anteposition	**Postposition**	**Translation**
Ces fiers époux attendent avec impatience le jour J.	*Ces époux fiers attendent avec impatience le jour J.*	*These proud spouses are eagerly awaiting the go time.*
Cette fière équipe de travail se hâte de présenter son projet.	*Cette équipe fière de travail se hâte de présenter son projet.*	*This proud work team is eager to present its project.*
Cette longue saison de football a été intense.	*Cette saison longue de football a été intense.*	*This long football season has been intense.*
Elle connait ce fier artiste depuis des années.	*Elle connait cet artiste fier depuis des années.*	*She has known this proud artist for years.*
Il a écrit un long article de linguistique.	*Il a écrit un article long de linguistique.*	*He wrote a long article on linguistics.*
Ils ont emprunté un long chemin sans visibilité.	*Ils ont emprunté un chemin long sans visibilité.*	*They took a long path without visibility.*
J'ai lu un long roman comme je les aime.	*J'ai lu un roman long comme je les aime.*	*I read a long novel as I like them.*
Les fiers ouvriers déjeunent actuellement.	*Les ouvriers fiers déjeunent actuellement.*	*The proud workers are currently lunching.*
Ma tante est une fière cuisinière de renom.	*Ma tante est une cuisinière fière de renom.*	*My aunt is a proud cook of renown.*
Elle a participé à un long séminaire de quelques jours.	*Elle a participé à un séminaire long de quelques jours.*	*She participated in a seminar lasting a few days.*
Il a écrit un long article de 50 pages.	*Il a écrit un article long de 50 pages.*	*He wrote a 50 page long article.*
Ils ont emprunté un long chemin de plusieurs kilomètres.	*Ils ont emprunté un chemin long de plusieurs kilomètres.*	*They took a path several kilometers long.*
J'ai lu un long roman de plusieurs tomes.	*J'ai lu un roman long de plusieurs tomes.*	*I read a novel several volumes long.*
Elle annote un différent segment de 32 caractères.	Elle annote un segment différent de 32 caractères.	She annotates a different segment of 32 characters.
Ils vivent une différente relation sans amour.	Ils vivent une relation différente sans amour.	They live a different relationship without love.
L'architecte a construit une différente maison dans le sud.	L'architecte a construit une maison différente dans le sud.	The architect built a different house in the south.
Tu as acheté un différent cahier pour dessiner.	Tu as acheté un cahier différent pour dessiner.	You bought a different notebook to draw.
Vous avez couru un différent marathon toujours populaire.	Vous avez couru un marathon différent toujours populaire.	You ran a different, ever-popular marathon.
Ces fiers époux de leurs préparatifs attendent avec impatience.	Ces époux fiers de leurs préparatifs attendent avec impatience.	These spouses proud of their preparations are waiting impatiently.
Cette fière équipe de son projet se hâte de le présenter.	Cette équipe fière de son projet se hâte de le présenter.	This team, proud of its project, is eager to present it.
Cette longue saison de 4 mois a été intense.	Cette saison longue de 4 mois a été intense.	This 4 month long season has been intense.
Elle annote un différent segment du précédent.	Elle annote un segment différent du précédent.	It annotates a different segment from the previous one.
Elle connait ce fier artiste de sa création.	Elle connait cet artiste fier de sa création.	She knows this artist who is proud of his creation.
Ils vivent une différente relation de la suivante.	Ils vivent une relation différente de la suivante.	They live a different relationship than the following one.
L'architecte a construit une différente maison de celle prévue.	L'architecte a construit une maison différente de celle prévue.	The architect built a different house than planned.
Les fiers ouvriers de leur avancement s'accordent une pause.	Les ouvriers fiers de leur avancement s'accordent une pause.	The workers, proud of their advancement, take a break.
Ma tante est une fière cuisinière de ses talents.	Ma tante est une cuisinière fière de ses talents.	My aunt is a cook proud of her talent.
Tu as acheté un différent cahier du sien.	Tu as acheté un cahier différent du sien.	You bought a notebook different from his.
Vous avez couru un différent marathon de celui de Toulouse.	Vous avez couru un marathon différent de celui de Toulouse.	You ran a different marathon than that of Toulouse.

Table 9: Sentences in the *Presence of adjective/noun dependent* category.

Anteposition	Postposition	Translation
Dimanche, ils ont pu faire la grasse matinée.	*Dimanche, ils ont pu faire la matinée grasse.*	*On Sunday, they were able to sleep in.*
Elle a écrit un vibrant hommage pour sa mère décédée.	*Elle a écrit un hommage vibrant pour sa mère décédée.*	*She wrote a moving tribute for her late mother.*
Elle aime la grasse matinée du lundi.	*Elle aime la matinée grasse du lundi.*	*She loves sleeping in on Mondays.*
Il a passé une dure semaine.	*Il a passé une semaine dure.*	*He's had a tough week.*
Il admet son net avantage sur les autres.	*Il admet son avantage net sur les autres.*	*He admits his clear advantage over others.*
Il ne retient pas ses diverses leçons.	*Il ne retient pas ses leçons diverses.*	*He does not retain his various lessons.*
Ils ont rendu un vibrant hommage à ce digne soldat.	*Ils ont rendu un hommage vibrant à ce digne soldat.*	*They paid a vibrant tribute to this worthy soldier.*
J'avais des doubles objectifs précis.	*J'avais des objectifs doubles précis.*	*I had specific dual objectives.*
Nous effectuons diverses expériences.	*Nous effectuons des expériences diverses.*	*We perform various experiments.*
Elle a fait un net bénéfice ce mois-ci.	*Elle a fait un bénéfice net ce mois-ci.*	*She made a net profit this month.*
Depuis la mort de son hamster, il a le dur cœur.	Depuis la mot de son hamster, il a le cœur dur.	Since the death of his hamster, he has had a hard heart.
Depuis la mort de son hamster, il a une dure vie.	Depuis la mort de son hamster, il a une vie dure.	Since the death of his hamster, he has had a hard life.
Dimanche, ils ont mangé des gras plats.	Dimanche, ils ont mangé des plats gras.	On Sunday, they ate fatty dishes.
Elle essaiera par elle-même pour en avoir le net cœur.	Elle essaiera par elle-même pour en avoir le cœur net.	She will try on her own to find out for sure.
Elle n'aime pas laver la grasse boîte.	Elle n'aime pas laver la boîte grasse.	She doesn't like to wash the greasy box.
Il est adepte de divers faits.	Il est adepte de faits divers.	He is adept at various facts.
Il n'a pas accepté sa défaite, il a le dur cœur.	Il n'a pas accepté sa défaite, il a le cœur dur.	He did not accept his defeat, he has a hard heart.
Ils ont acheté un vibrant fauteuil pour leur salon.	Ils ont acheté un fauteuil vibrant pour leur salon.	They bought a vibrating armchair for their living room.
J'ai mis les doubles bouchées pour arriver à temps.	J'ai mis les bouchées doubles pour arriver à temps.	I worked hard to get there on time.
Nous suivons les divers faits à la télévision.	Nous suivons les faits divers à la télévision.	We follow the news on television.
Vous avez mis les doubles bouchées pour terminer.	Vous avez mis les bouchées doubles pour terminer.	You worked hard to finish.

Table 10: Sentences in the *Fixed expressions* category.

Anteposition	Postposition	Translation
A nouvelle année, nouveaux dynamismes pour cette entreprise.	*A nouvelle année, dynamismes nouveaux pour cette entreprise.*	*A new year, new dynamics for this company.*
Fabuleux amis, fabuleux camarades : l'ennemi n'est pas à l'intérieur !	*Fabuleux amis, camarades fabuleux : l'ennemi n'est pas à l'intérieur !*	*Fabulous friends, fabulous comrades: the enemy is not within!*
J'ai aimé le concept : bonne ambiance, bonne musique, les gens sont contents.	*J'ai aimé le concept : bonne ambiance, musique bonne, les gens sont contents.*	*I liked the concept: good atmosphere, good music, people are happy.*
Ce document vise à expliquer le déficit véritable, la véritable dette dans son ensemble.	*Ce document vise à expliquer le déficit véritable, la dette véritable dans son ensemble.*	*This document aims to explain the real deficit, the real debt as a whole.*
Nous avons adopté pour des stratégies communes, actions communes et positions communes.	*Nous avons adopté pour des stratégies communes, actions communes et communes positions.*	*We have adopted for common strategies, common actions and common positions.*
Avec la merveilleuse sélection et de merveilleux essais, ils ont trouvé les résultats qu'ils cherchaient.	*Avec la merveilleuse sélection et des essais merveilleux, ils ont trouvé les résultats qu'ils cherchaient.*	*With the wonderful selection and wonderful testing, they found the results they were looking for.*
Il lui a offert des volumineuses plantes à volumineuses fleurs.	Il lui a offert des volumineuses plantes à fleurs volumineuses.	He gave her bulky plants with bulky flowers.
Je suis d'accord avec eux : à événement exceptionnel, exceptionnel dispositif.	Je suis d'accord avec eux : à événement exceptionnel, dispositif exceptionnel.	I agree with them: for an exceptional event, an exceptional device.
Cette année, ils préparent un diplôme professionnel en professionnel lycée.	Cette année, ils préparent un diplôme professionnel en lycée professionnel.	This year, they are preparing a professional diploma in vocational high school.
Concernant la protection des données personnelles, aucune personnelle information n'est collectée.	Concernant la protection des données personnelles, aucune information personnelle n'est collectée.	Regarding the protection of personal data, no personal information is collected.
Elle a procédé à l'étude de quelques instruments pitoyables et pitoyables illusions.	Elle a procédé à l'étude de quelques instruments pitoyables et illusions pitoyables.	She proceeded to study some pitiful instruments and pitiful illusions.
Ce bâtiment n'a pas changé depuis sa construction : lumineuses couleurs, lumineux lampadaires.	Ce bâtiment n'a pas changé depuis sa construction : lumineuses couleurs, lampadaires lumineux.	This building has not changed since its construction: bright colors, bright streetlights.

Table 11: Sentences in the *Structural persistence* category.

Anteposition	Postposition	Translation
Elle préfère son propre pantalon à celui de sa soeur.	*Elle préfère son pantalon propre à celui de sa sœur.*	*She prefers her own pants to her sister's.*
Nous nous sommes rejoins autour d'un chaleureux repas.	*Nous nous sommes rejoins autour d'un repas chaleureux.*	*We came together for a hearty meal.*
Tu m'as fait part de ta fabuleuse idée.	*Tu m'as fait part de ton idée fabuleuse.*	*You told me about your fabulous idea.*
Cet ancien fer n'est plus utilisé.	*Ce fer ancien n'est plus utilisé.*	*This old iron is no longer used.*
C'était un fabuleux voyage que nous avons organisé.	C'était un voyage fabuleux que nous avons organisé.	It was a fabulous trip that we organized.
Ce chaleureux accueil m'a fait chaud au cœur.	Cet accueil chaleureux m'a fait chaud au cœur.	This warm welcome warmed my heart.
Ce légendaire récit me tourmente chaque jour.	Ce récit légendaire me tourmente chaque jour.	This legendary tale torments me every day.
Ce puéril discours lui a porté préjudice.	Ce discours puéril lui a porté préjudice.	This childish speech harmed him.
Cette fermière entreprise n'est plus aussi familiale que dans le temps.	Cette entreprise fermière n'est plus aussi familiale que dans le temps.	This farm business is no longer as family-run as it used to be.
Cette jaune chaise est très tendance.	Cette chaise jaune est très tendance.	This yellow chair is very trendy.
Cette puérile plaisanterie ne l'a pas fait rire.	Cette plaisanterie puérile ne l'a pas fait rire.	This childish joke did not make him laugh.
Elle m'a fourni la volumineuse archive.	Elle m'a fourni l'archive volumineuse.	She provided me with the voluminous archive.
Il m'a apporté une bleue gourde.	Il m'a apporté une gourde bleue.	He brought me a blue water bottle.
Il mange des roses bonbons.	Il mange des bonbons roses.	He eats pink candies.
Ils n'ont pas pu télécharger le volumineux fichier.	Ils n'ont pas pu télécharger le fichier volumineux.	They were unable to download the large file.
J'ai écrit sur une bleue feuille.	J'ai écrit sur une feuille bleue.	I wrote on a blue sheet.
La jaune trousse contient ses feutres.	La trousse jaune contient ses feutres.	The yellow pencil case contains her markers.
La pétrolière industrie ne m'attire pas du tout.	L'industrie pétrolière ne m'attire pas du tout.	The oil industry does not appeal to me at all.
Le ferroviaire transport est voué à s'étendre.	Le transport ferroviaire est voué à s'étendre.	Rail transport is destined to expand.
Le ministériel arrêté a confirmé les mesures prises.	L'arrêté ministériel a confirmé les mesures prises.	The ministerial decree confirmed the measures taken.
Les filles ont opté pour une mauve couverture.	Les filles ont opté pour une couverture mauve.	The girls opted for a purple blanket.
Leur financière situation s'aggrave de jour en jour.	Leur situation financière s'aggrave de jour en jour.	Their financial situation is getting worse day by day.
Ma sœur porte des mauve lunettes.	Ma sœur porte des lunettes mauve.	My sister wears purple glasses.
Mon bureau est décoré d'un vert panier.	Mon bureau est décoré d'un panier vert.	My office is decorated with a green basket.
Sa rose poubelle lui plait énormément.	Sa poubelle rose lui plait énormément.	His pink trash can pleases him enormously.
Son doudou est une verte peluche.	Son doudou est une peluche verte.	His cuddly toy is a green plush.
Elle a acheté un vibrant jouet pour son fils.	Elle a acheté un jouet vibrant pour son fils.	She bought a vibrating toy for her son.

Table 12: Sentences in the *Blocked and mobile adjectives* category.