# Modality over Contemporary History Schoolbooks in French

**Anonymous IWCS submission**
The style files are borrowed from ACL-IJCNLP 2021

## Abstract

This paper investigates the analysis of modality in French. We use an annotation schema encompassing epistemic, deontic, appreciative, and inter-enunciators modalities. Using a fine-tuned CamemBERT model for multilabel classification, we achieve a strong weighted F1-score of 0.92, despite challenges with class imbalance. A classification experiment using this model on contemporary textbooks from two major French publishers reveals significant prevalence of appreciative and inter-enunciators modalities, with notable increases in appreciative judgments at higher educational levels.

## 1 Introduction

Understanding how historical knowledge is conveyed in school textbooks requires close attention to the processes through which facts are presented. The notion of modality provides a valuable framework for analyzing these processes because this semantic category relies on the enunciator's engagement or disengagement toward the truth or the axiological value of the content. In this study, we investigate how modality can be automatically detected in history lessons extracted from French textbooks. After a brief review focusing on several studies directly or indirectly interested in detecting modality in section 2, we present in section 3 an annotation schema of modality and its application on a dataset of French history textbooks. Section 4 presents a classifier of 4 categories of modality, by fine-tuning a transformer-based encoder with this dataset. We then present the results of this model when applied to a larger and contrastive dataset of textbooks issued by two French publishers.

## 2 Related work

As advances in NLP increasingly focus on capturing nuanced aspects of language, modality detection represents a crucial component in tasks related to enunciator's subjectivity such as opinion, commitment, or factivity analysis. (Benamara et al., 2012) examine how negation and modality interact with opinion expression, showing the complex interplay between these notions. (De Marneffe et al., 2019) investigate speaker commitment through the CommitmentBank, focusing on clause-embedding predicates embedded under an entailment canceling operator such as epistemic modality. (Saurí and Pustejovsky, 2009) partly relies on epistemic modality to assert the factivity of an expressed event. Modality detection can also be seen as a standalone task, as in (Pyatkin et al., 2021), who distinguish possible, plausible, and desirable events as seen by the enunciator.

Outside the field of NLP, some manually conducted studies focus more specifically on the analysis of modality in textbooks. (McCabe, 2004) compares the modality expressed in English and Spanish history textbooks and shows how the writers express uncertainty (epistemic modality) or, on the contrary, an *"appearance of objectivity"* (ibid). The study also shows that more recent historical topics in textbooks lead to less uncertainty. Several other studies rely on modality to analyze subjectivity or ideology in textbooks. (Suciati and Sugiarto, 2021) highlights links between interpersonal meaning conveyed by the enunciator and the use of probability and obligation modalization, in persuasive texts extracted from Indonesian 11th grade textbooks. (Tahir et al., 2021) shows how epistemic and deontic modalities are employed to embed ideological viewpoints (Islamic ideology in this case) in Pakistani English language textbooks.

Our approach follows the enunciative linguistic tradition which conceives modality as part of a broader typology of enunciative operations centered on the notion of the enunciator's engagement with their discourse. This perspective moves be-

yond traditional NLP modal approaches by encompassing both epistemic and axiological evaluation. This theoretical framework is well adapted for analyzing educational discourse, where authors must deal with intricate relationships between factual content, institutional authority, and pedagogical objectives.

## 3 Manual annotation of modality

### 3.1 Annotation scheme

Given that modality analysis in enunciative linguistics traditionally focuses on the propositional content (Bally, 1932; Desclés, 2009), we choose to annotate modality in discourse at the clause level (independant clauses, which may be juxtaposed or coordinates, see section 3.2). Following (Desclés, 2009), our schema comprises 4 modalities, which are annotation labels that will be used in the automatic detection experiment described in Section 4. We now describe these four categories using examples from our French history textbooks dataset introduced in Section 3.2:

- Epistemic Modality reflects the enunciator's uncertainty regarding the truth of the propositional content. This uncertainty (or possibility or probability) is obviously expressed. Example (1) is annotated as Epistemic due to the verb "could lead".

(1) *L'établissement d'un tel accord pourrait ensuite conduire à une entente, / The establishment of such an agreement could then lead to an understanding,*

- Deontic Modality expresses a judgement from the enunciator regarding "what should be" or "what must be", and the potential "self-imposed obligation of the enunciator to actualize it" (Desclés, 2009). See example (2) and the use of the verb "must".

(2) *Pour Hitler, pourquoi faut-il écarter les Juifs des Aryens ? / According to Hitler, why must Jews be separated from Aryans ?*

- Appreciative Modality encompasses evaluative judgments, which can be defined as appreciative, bouletic, or axiological (3). These judgements do not convey uncertainty on the truth value, but are sometimes expressed in a non-actualized form.

(3) *, ils ont réduit notre peuple à la plus noire misère. / , they have reduced our people to the darkest misery.*

- Inter-Enunciators Modality describes the act of linking a propositional content with an inter-locutor. These enunciations *"require, for their analysis, recourse to the co-enunciator to receive a determinate value."* (Desclés, 2009) Inter-Enunciators modality applies to orders (4), questions (2), promises, or apostrophic markers that involve the interlocutor in the discourse.

(4) *Rédige un texte pour résumer ce qu'est une ville sumérienne. / Write a text summarizing what a Sumerian city is.*

Those different modalities are not mutually exclusive and may co-occur within the same textual unit. The automatic annotation of the four modality categories presented in section 4 can thus be framed as a multilabel classification task.

### 3.2 Dataset and annotation process

We have composed a first corpus (referred to as CM858) which consists of history lessons extracted from textbooks in French. It is made of 858 clauses (our annotation level) drawn from 8 textbooks, evenly divided into two temporal series (4 published in the 19th century and 4 in the 21st century).The sentences in these lessons were segmented into independent clauses (juxtaposed and coordinates) using a custom heuristic, which identifies verbs (or adjectives/nouns in verbal constructions) with syntactic dependencies of the conjunction or parataxis type, linked syntactically to a verbal (or adjectival/nominal, in verbal constructions) root. This processing was implemented using the syntactic analysis provided by Stanza (Qi et al., 2020).

An initial dataset of 1000 clauses was manually annotated by two experts of the notion of modality, following specific guidelines. An inter-annotators agreement score was calculated using Cohen's kappa (Cohen, 1960). Scores are shown in table 1.

| | IAA |
|---|---|
| Epistemic | 0.92 |
| Deontic | 0.96 |
| Appreciative | 0.9 |
| Inter-enunc. | 0.81 |

Table 1: Inter-annotator agreement scores by label (Cohen's kappa)

A ground truth dataset was then exctracted, removing the 142 annotated clauses leading to partial or full disagreement, resulting in the CM858 dataset. Figure 1 shows the proportions of the man-

ually annotated modality categories in CM858. The proportions reveal a high class imbalance, with appreciative modality being significantly overrepresented.
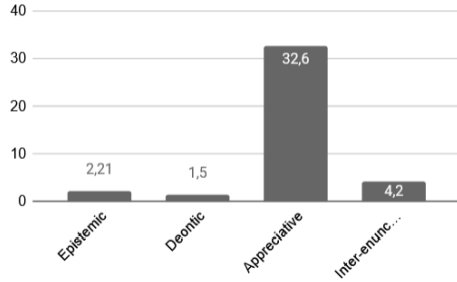


Figure 1: Proportions of modal categories - CM858

## 4   Analysis of history textbooks

### 4.1   Methodology

We have composed a dataset of 2 types of contemporary history textbooks, all from 2016: 4 were issued by the French publisher Hatier and 4 by the French publisher Magnard, each textbook belonging to one of the four educational levels of the French middle school system (6e, 5e, 4e, and 3e, corresponding to 6th through 9th grade in the U.S system). The resulting average number of clauses per textbook is 2563.75 for Magnard and 4136.5 for Hatier. The sentences in these lessons were segmented into independent clauses, using the heuristic applied on CM858 and described in section 3.2. In order to automatically annotate this dataset, we fine-tuned a multilabel text classification model on the CM858 corpus introduced in Section 3. The model architecture is based on CamemBERT (Martin et al., 2019), a transformer-based encoder specifically designed for NLP tasks in French A sigmoid activation function is used for multilabel classification, in a 5-fold cross-validation. To mitigate overfitting, early stopping is applied with a patience threshold of 3 epochs. In order to address the class imbalance identified in Section 3, we apply a weighted cross-entropy loss, with weights set inversely proportional to the class frequencies in the training data. The training is performed with a batch size of 8, using the HuggingFace Transformers library (Wolf et al., 2020). A test dataset of 100 manually annotated instances was created, reproducing the high class imbalance of the training dataset. The evaluation of the model is shown in table 2.

|  | P | R | F1 | Sup. |
|---|---|---|---|---|
| Epistemic | 1 | 0.5 | 0.67 | 2 |
| Deontic | 1 | 1 | 1 | 2 |
| Appreciative | 0.97 | 0.90 | 0.94 | 41 |
| Inter-enunc. | 1 | 0.80 | 0.89 | 5 |
| micro avg | 0.98 | 0.88 | 0.93 | 50 |
| macro avg | 0.99 | 0.80 | 0.87 | 50 |
| weighted avg | 0.98 | 0.88 | 0.92 | 50 |
| samples avg | 0.40 | 0.40 | 0.40 | 50 |

Table 2: Evaluation of the multilabel modality classification model

While the model achieves strong overall performance, with a weighted F1-score of 0.92, indicating effective classification on the majority class (appreciative) and good generalization, the macro-averaged recall is lower (0.801), reflecting difficulties in detecting less frequent categories such as epistemic and inter-enunciators. While the deontic and inter-enunciators classes show high precision, their limited support reduces the reliability of these scores. Notably, the low sample-based F1-score (0.400) suggests that the model may underpredict multiple labels per instance.
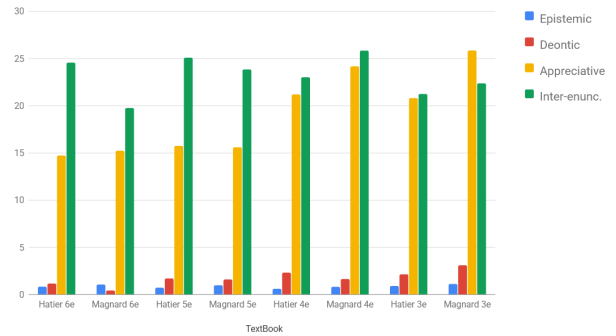
### 4.2   Results



Figure 2: Modalities over history textbooks

Figure 2 shows the proportions of clauses that were automatically annotated by the model according to the 4 modality categories. The distribution of the categories across history lessons reveals a strong prevalence of appreciative and inter-enunciators modalities. The latter is mostly explained by the large amount of questions and imperative verbs following each lesson, in order to evaluate the student's comprehension. The small variation of the proportions of Inter-enunciators

modality (from 19% to 26%) in the textbooks does not seem to be related to the educational level nor the publisher. It is worth noting that these evaluative questions may also contain appreciative (5) or deontic (6) modalities.

(5) *D'après vos connaissances, pourquoi Hitler veut-il embrigader la population ? / According to your knowledge, why does Hitler want to indoctrinate the population ?* (Hatier 3e)

(6) *Enfin il faut définir les notions clés à l'aide du cours. / Finally, you have to define the key concepts using the course.* (Magnard 3e)

Epistemic and deontic modalities are marginal. The expressed uncertainty may notably originate from the author and frame historical veracity (7). It may also originate from another enunciator, within the context of reported speeches in documents along the lesson, as shown in example (8) produced by a quoted Hiroshima survivor.

(7) *, les Dix commandements auraient été remis à Moïse sur le mont Sinaï. / , the Ten Commandments would have been given to Moses on Mount Sinai.* (Magnard 6e)

(8) *Il devait y avoir dans cette partie autour de nous une vingtaine de milliers de morts et entre 60 et 70 mille blessés. / There must have been in this area around us about twenty thousand dead and between 60 and 70 thousand wounded.* (Hatier 3e)

Interestingly, appreciative and deontic modalities (indicating axiological judgements) are more represented in the two higher educational levels (4e and 3e) for both publishers. This may be linked to the nature of the historical content covered at these levels.

The four graphs of figure 3 present a comparative analysis of the percentage of history textbook content devoted to various topics across the two French publishers. Overall, both publishers show a high degree of alignment in the thematic structure, indicating a shared adherence to national educational standards, except for the 'Mediterranean exchanges' chapter in Grade 5e which is not covered by Magnard. However, as students progress through the levels (4e and 3e particularly), there is a clear increase in politically and morally charged themes such as colonization, slavery, totalitarian regimes, and decolonization, which suppose stronger value judgments. While many topics are equally treated , some notable differences emerge: Hatier devotes significantly more attention to the French Revolution and Industrial Revolution in grade 4e, and to
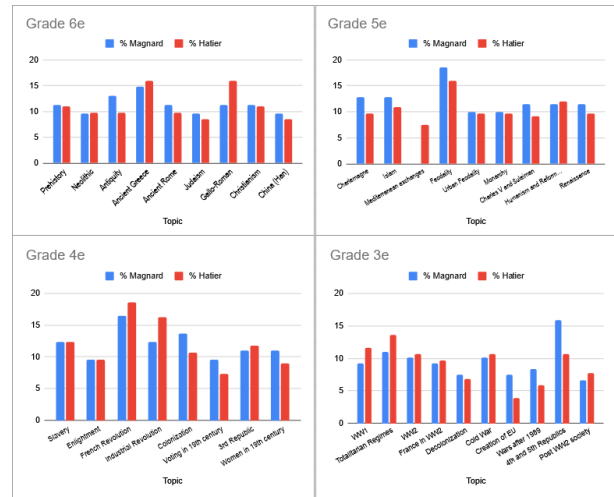


Figure 3: Pages distribution by topic for each grade and each publisher

Totalitarian Regimes in grade 3e, whereas Magnard allocates more space to the national topic of the 4th and 5th french republics (which focuses on the rights and duties of citizens).

The strong similarity between the proportions allocated to different historical themes gives no clue to explain the slight differences in modality between the two publishers. However, the evolution of themes across grade levels appears to explain the increase for both publishers in appreciative and deontic modalities related to value judgments in grades 4e and 3e.

As students's cognitive maturity increases, textbooks can introduce more evaluative and interpretive discourse. This shift may also reflect a didactic strategy to prepare students for final exams, where critical reasoning and argumentative writing are expected.

## 5 Conclusion

Our fine-tuned CamemBERT model demonstrates the feasibility of automated modality detection with a weighted F1-score of 0.92 indicating effective classification capability. The analysis reveals a predominance of appreciative modality, and an increase in axiological judgments at higher educational levels (4e and 3e). Expanding the analysis to include diachronic comparisons could reveal how modal strategies in textbooks have evolved over time.

# References

Charles Bally. 1932. Linguistique française et linguistique générale. *Leroux, Paris*.

Farah Benamara, Baptiste Chardon, Yannick Mathieu, Vladimir Popescu, and Nicholas Asher. 2012. How do negation and modality impact on opinions? In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 10–18.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Marie-Catherine De Marneffe, Mandy Simons, and Judith Tonhauser. 2019. The commitmentbank: Investigating projection in naturally occurring discourse. In *proceedings of Sinn und Bedeutung*, volume 23, pages 107–124.

Jean-Pierre Desclés. 2009. Prise en charge, engagement et désengagement. *Langue française*, 162(2):29–53.

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de La Clergerie, Djamé Seddah, and Benoît Sagot. 2019. Camembert: a tasty french language model. *arXiv preprint arXiv:1911.03894*.

Anne McCabe. 2004. Mood and modality in spanish and english history textbooks: The construction of authority. *Text & Talk*, 24(1):1–29.

Valentina Pyatkin, Shoval Sadde, Aynat Rubinstein, Paul Portner, and Reut Tsarfaty. 2021. The possible, the plausible, and the desirable: Event-based modality detection for language processing. *arXiv preprint arXiv:2106.08037*.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Roser Saurí and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language resources and evaluation*, 43:227–268.

Hani Suciati and Bambang Ruby Sugiarto. 2021. Modality realized on hortatory exposition texts used in senior high school grade xi textbook. *Journal of English Education Program (JEEP)*, 8(1):35–42.

Adnan Tahir, Rashid Mahmood, and Afzal ul Haque. 2021. Portrayal of islamic ideology: Modality analysis of ptb english language textbooks. *International Journal of Linguistics and Culture*, 2(1):1–17.

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. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.