Engagement and Non-Engagement: Two Notions at the Core of an Annotation Schema of Enunciative Strategies

Cyril Bruneau
Laboratoire MoDyCo
Université Paris Nanterre
c.bruneau@parisnanterre.fr

Delphine Battistelli
Laboratoire MoDyCo
Université Paris Nanterre
dbattist@parisnanterre.fr

Abstract

This study provides an annotation schema of a wide range of enunciative strategies underlying every enunciation process by which an enunciator actualizes a predicative content. We show that most of these enunciative strategies involve the enunciator in a relationship of *Engagement* (concerned with the notions of truth value and axiological/appreciative value) or *Non-Engagement* toward a stated predicative content. Our approach takes place in the French enunciative framework rooted in the work of Bally (1932). We explicitly compare our approach with that of Appraisal theory (Martin and White, 2003). We also illustrate the applications of our schema with a manual annotation experiment conducted on a corpus of French history textbooks. This experiment reveals interesting diachronic variations in the enunciator's modes of *Engagement* and *Non-Engagement*.

1 Introduction

In the lineage of what is called the enunciative approach of language notably rooted in the work of Bally (1932), we are interested in the various strategies of actualization of a predicative content that are mobilized by an enunciator when producing an utterance. A predicative content such as *be blue(the car)*, comprising a predicate (*be blue*) and its argument (*the car*), can be mobilized in an utterance in multiple ways (see ex. (1a) to (1g)), suggesting different actualizations and attitudes with regard to the same predicative content.

(1a) The car is blue
(1b) The car might be blue
(1c) I'm glad the car is blue
(1d) The car should not be blue
(1e) Is the car blue?
(1f) Do you know that the car is blue?
(1g) I heard the car is blue

The characterization of these enunciative operations is not addressed in NLP as a standalone problem, although several tasks make use of notions which are part of it, such as modality. We propose an operational schema for NLP of the full range of enunciative operations. Our schema is based on (Desclés, 2009), with several additions which are made explicit in Section 3. The characterization of these enunciative operations provides an analytical framework for observing and quantifying enunciative profiles across diverse corpora. Furthermore, enriching a corpus with such semantic and enunciative information can prove crucial for automatic detection tasks in NLP related to the enunciator's engagement, such as hate/toxic speech detection, opinion mining, ideological content analysis, uncertainty analysis, and more. After clarifying our theoretical framework in Section 2, we present a comprehensive annotation schema for enunciative operations in Section 3. Section 4 details the results of a manual annotation experiment conducted on a corpus using this schema.

2 The notion of engagement: a recurring issue in literature

2.1 Related works

The question of how an enunciator validates a predicative content and/or positions in relation to it has been approached in various ways, directly or not concerned with the notion of engagement. Works on **commitment** (see De Brabanter and Dendale (2008) for an extended presentation) examine the beliefs of a speaker that can be inferred from their discourse, aiming to determine whether an event (term covering a conceptual notion close to that of predicative content) is presented as *actual*, *non-actual*, or *uncertain*. These studies suggest that such inferences are not limited to assertions, the primary mode of commitment (De Marneffe et al.,

2019). For example, the utterance (1e), formulated as a question, leaves room for doubt regarding the truth value of the predicative content *be blue (the car)*, unlike example (1f), which — though also phrased as a question — demonstrates speaker commitment inferable from the verb "know" (Jiang and de Marneffe, 2019).

The enunciator's commitment can also be linked to the notion of **factuality** when it is related to the truth value of the predicative content (i.e. when presented as factual). In fact, Jiang and de Marneffe (2019) use commitment and factuality interchangeably. Works on factuality e.g. (Saurí and Pustejovsky, 2009) propose annotation schemas covering (i) the enunciator's certainty regarding an event's truth, (ii) its possibility (as in ex. (1b)), or (iii) its probability.

Other related notions are **modality** (see for example in NLP (Pyatkin et al., 2021)) and **evidentiality** (see for ex. (Su et al., 2010)). They seek to capture respectively the attitude of the enunciator toward their content, and the nature of the source of the information. Modality may convey judgments of the enunciator regarding for instance the uncertainty of the predicative content (as in (1b)), while evidentiality may distantiate the enunciator from the information by highlighting its source (1g). These notions thus play a role regarding the enunciator's commitment we can infer.

The notions of **stance** and **sentiment** as described in NLP might also be intuitively linked to this question. Stance detection aims at assessing the enunciator's favorability toward a predetermined target (*in favor* or *against* the target which is not necessarily mentioned in the predicative content) (Mohammad et al., 2016). More recent works rely on the SDQC tagset (*Support, Deny, Query, Comment*) (Gorrell et al., 2019; Evrard et al., 2020). Similarly, sentiment analysis aims to detect the polarity (*positive* as in (1c), *negative* as in (1d), *neutral*) of the enunciator's opinion toward an explicitly mentioned target (e.g., a person, an organization), sometimes including the identification of the target (Nakov et al., 2016).

Closer to our approach is the **Appraisal theory** (Martin and White, 2003), which meticulously classifies evaluative language into categories such as *Attitude*, *Engagement* and *Graduation*. This theory is *"concerned with the interpersonal in language, with the subjective presence of writers/speakers in texts as they adopt stances towards* *both the material they present and those with whom they communicate."*(ibid., p.1). The Appraisal theory models the evaluative operations available to an enunciator. These operations are not limited to the truth value of the statement but also refer to axiological and appreciative dimensions. We provide an overview of the Appraisal framework in Figure 5 in the appendix, to facilitate comparison with our approach.

2.2 Contributions and Distinctions

In this study, we provide a new typology for NLP of the enunciative operations an enunciator may deploy in order to actualize their predicative content into an utterance, following the conceptualization initiated in (Desclés, 2009), himself in the lineage of Bally's works. Desclés (2009) designates this set of operations by the term *"prise en charge"* (De Brabanter and Dendale, 2008; Coltier et al., 2009). At the core of these enunciator's operations lies the notion of **Engagement**, defined as the enunciator's capacity to engage or disengage with either the truth value or the axiological/appreciative value of the predicative content they articulate. What we define as engagement is close to the aforementioned notion of speaker commitment, except that its scope is not limited to (non)factual events. Our notion of engagement also coexists with the opposing notion of **Disengagement**, a common operation that enables the communication of uncertain or distanced information. The notion of **Non-Engagement** (a default category implying neither engagement nor disengagement) is also central to this schema. It captures the enunciator's apparent neutrality. The central articulation of these three notions enables contrasted analyses of the enunciative strategies in diverse corpora, as presented in Section 4, and constitutes an original contribution for NLP. Moreover, our definition of the enunciator's engagement relies on explicit linguistic markers. Thus, unlike works on commitment, the simple declaration of an event (presented as factual) without explicit markers of engagement (assertion, certainty...) will be classified as *Non-Engagement*.

Our approach also differs from works on stance and sentiment as we do not seek to capture the polarity of the utterance, nor do we focus on an explicit target within (or outside) the predicative content. Consequently, a strongly favorable or unfavorable opinion would be treated as strong engage-

2

ment, with no consideration for polarity or target, while a weakly favorable opinion accompanied by uncertainty might indicate disengagement. Stance detection and sentiment analysis are therefore very different tasks from ours, but they can be seen as complementary in certain use cases.

We share certain notions presented in Appraisal theory, which relate to the annotation schema we propose. The category *Attitude* partially aligns with our *Appreciative modality* category, as explained in section 3.2. The *Engagement* category, which aims to describe the enunciator's stance towards the positions referenced in the discourse, encompasses notions related to both engagement and potential disengagement, which are described in our approach under the categories of *Validation*, *Modality*, and *Representation of speech* (see section 3.2). However, significant differences can be outlined between Appraisal theory and our approach:

1. We do not consider the notions captured by the *Graduation* category in Appraisal theory as modifiers, but as indicators that characterize the categories of engagement we describe, particularly by clarifying the concepts of certainty/uncertainty and negotiability/non-negotiability.

2. Our conception of engagement only focuses on the predicative content expressed by the enunciator. Consequently, we do not consider the evaluation that is made of an interlocutor, or the enunciator of a discourse to which the primary enunciator refers.

3. Unlike stance detection approaches, we do not assume that an enunciator's engagement is necessarily in favor of or against a specific target. Additionally, we do not incorporate the notion of engagement *Polarity* into our schema.

4. As a result of 3, we propose the central concept of *Non-Engagement*, which describes a form of neutrality adopted by the enunciator. This concept enables us to differentiate between utterances presented as negotiable and other types of utterances that exhibit explicit engagement or disengagement. We believe this notion of Non-Engagement is crucial for analyzing certain corpora, as demonstrated in Section 4.

Modality and *Evidentiality* find their place in our typology among operations involving enunciator engagement / disengagement (with evidentiality overlapping both *Epistemic modality* and *Plausibility*, as detailed in the description of the *Plausibility* category) (see Section 3.2).

3 Annotation schema

3.1 Global view

Figure 1 represents the enunciative operations which can be mobilized by an enunciator. This typology is inspired by the one proposed in (Desclés, 2009), which has since been revised: (i) We added the *Representation of speech* categories as they are presented in (Authier-Revuz, 2020) in order to refine the category formerly referred to as *Reported Enunciation*; (ii) we added the notion of *Deictic Anchoring*; (iii) we added the *Appreciative Modality* present in (Desclés, 2009) but not in the final typology, as we do not want to limit the analysis to the truth value of the predicative content; and (iv) we changed the layout in order to make clear which categories fall under the scope of *Engagement/Disengagement*.

The primary enunciator (i.e. the initiator of the utterance act) is denoted as "E" on the left side of the schema. Each category represents an operation through which the enunciator actualizes a predicative content (that is, transforms a raw predicative content into a concrete utterance: see examples (1a) to (1g)). This actualization may entail a degree of validation of the content, which is precisely what this typology seeks to capture. Four ways - not necessarily exclusive - of actualizing a predicative content are distinguished: on the one hand, *Non-Engagement* and *Engagement / Disengagement* which are mutually exclusive; on the other hand, *Contextual frame of reference* and *Deictic anchoring* which can be added to one of the previous ones. These operations are analyzed at the clause level (independent clauses, which may be juxtaposed or coordinated). Each category can be associated to a clause within the utterance, and multiple categories may describe a single clause, except for a few categories excluding each other, which are presented in Section 3.2. The resulting manual annotations can thus be used to train a multilabel text classifier.
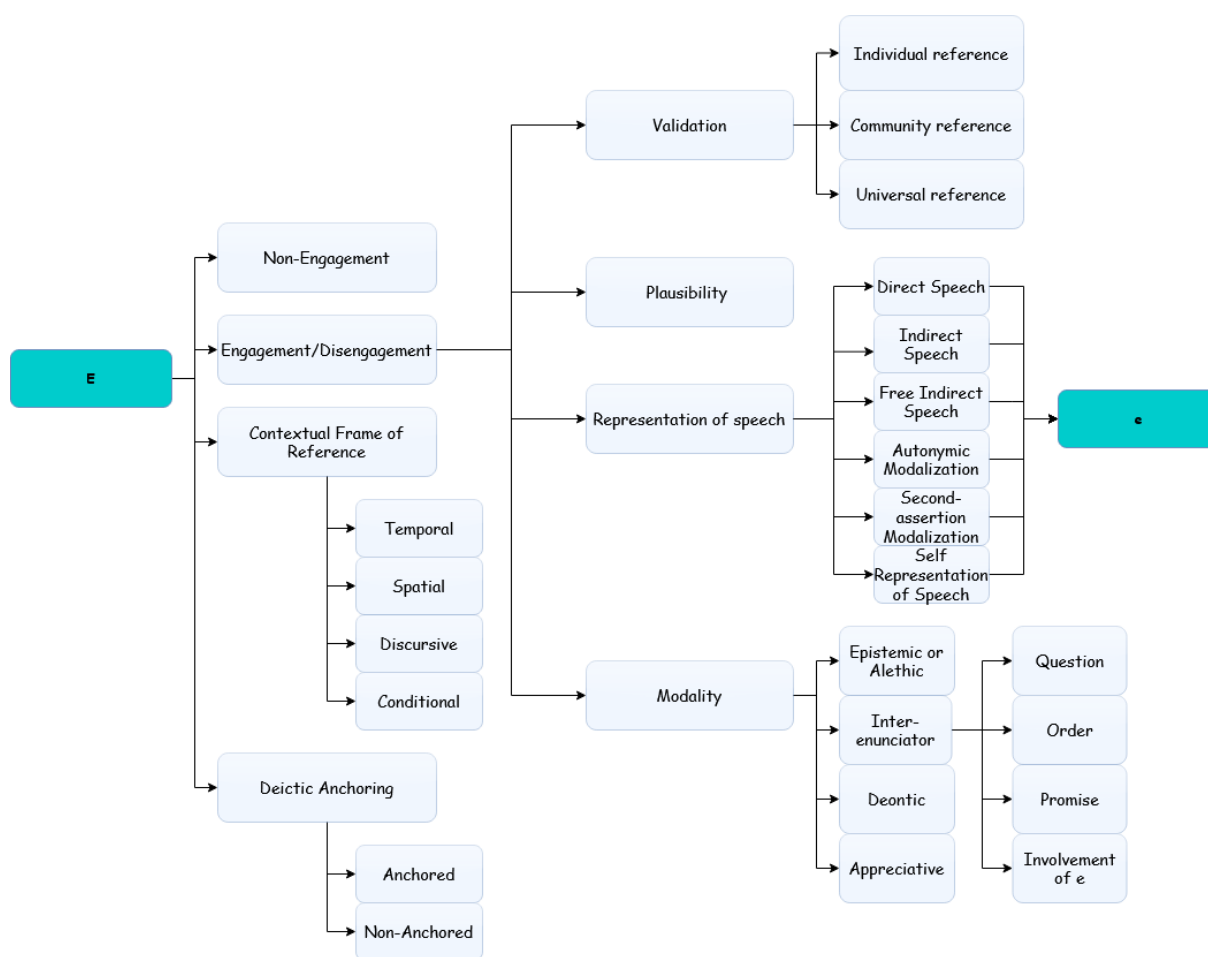
3

Figure 1: Annotation schema for enunciative operations[1]

3.2 Detailed description of the categories

We now detail the categories of our enunciative operations typology, with examples drawn from a corpus of French textbooks, which we present in Section 4.1. The categories under *Engagement / Disengagement* constitute the main part of the schema.

- The *Non-Engagement* category corresponds to the default mode of utterance: the enunciator does not express any particular engagement or disengagement, and their statement is perceived as *"negotiable"* with the interlocutor (Desclés, 2009) (see example (1a)). This category is often identified by the absence of markers from other categories. It is described in opposition to (and is thus incompatible with) any category classified under *Engagement / Disengagement*.

- *Contextual frame of reference* describes a type of utterance in which the truth value of the enunciator's predicative content is conditioned by an explicit context, which may be *Discursive*, *Conditional*, *Spatial*, or *Temporal*. These contextual frames of reference often serve to clarify the scope of the predicative content and are not considered markers of the enunciator's engagement or disengagement. Example (2) illustrates a temporal context via the adverbial *"from the late 15th century onward"*, example (3) a spatial context via *"in the capital"*.

(2) Comment l'Europe s'ouvre-t-elle sur le monde à partir de la fin du XVe siècle ? / How did Europe open itself to the world from the late 15th century onward?

(3) et commença dans la capitale la plus épouvantable tuerie dont nos annales fassent mention. / and began in the capital the most terrible slaughter recorded in our annals.

---

[1]Arrows between categories indicate a subdivision.

4

- The *Deictic anchoring* category is divided into two subcategories: *Anchored* and *Non-anchored*. It characterizes how the predicative content is related to the enunciative situation. *Anchored* clauses contain markers of the enunciator's presence and/or anchors that indicate a direct relationship with the enunciative context, such as the pronoun *"our"* in example (3). These markers do not signal the enunciator's engagement or disengagement.

- The *Validation* of the enunciator's predicative content is related to statements formulated as assertions: the enunciator presents their predicative content as *"non-negotiable"*. These utterances strongly reflect the enunciator's engagement regarding the truth value of the content. Three types of validation may be distinguished, depending on their referential framework - that is, the origin of the assertion -: the enunciator as an individual (*Individual*) (example (4) due to the adjective *"true"*), the enunciator as part of a broader community (*Community*) (example (3) where *"our"* stands for the French community), or a universal idea adopted by the enunciator (*Universal*, example (5)). Validation characterizes operations identified within the Appraisal framework as *monoglossic* engagement, or *heteroglossic* when associated with the notion of dialogic contraction, which can be linked to the criterion of non-negotiability.

(4) Ils substituèrent le culte de la déesse Raison à celui du vrai Dieu / They replaced the worship of the true God with that of the goddess Reason.

(5) L'envie s'attache toujours aux grands talents. / Envy always clings to great talents.

- *Plausibility* describes a mediated type of utterance in which the enunciator formulates a plausible hypothesis triggered by observed evidence (which is shared by co-enunciators) and/or by an inference from shared knowledge (example (6)). The inferential and hypothetical nature of the predicative content leads to a partial disengagement of the enunciator regarding the truth value: the statement is presented as *"negotiable"* (4). *Plausibility* only relies on abductive inference, which should not be confused with deductive inference (Desclés and Guentchéva, 2024). The former can

be considered as a specific case of mediativity: it aims to infer the cause of an objectively observed situation. The latter seeks to deduce a consequence from an observed situation and corresponds, from an enunciative perspective, to what we describe under the category *Epistemic modality*.

(6) Le siège de la Rochelle, où périrent plus de quarante mille catholiques, fut une preuve que le parti calviniste n'avait rien perdu de sa puissance. / The siege of La Rochelle, where over forty thousand Catholics perished, proved that the Calvinist faction had lost none of its power.

- The macro-category *Modality* encompasses all the attitudes an enunciator may adopt toward their predicative content. Except for the *Appreciative* and *Deontic* subcategories, *Modality* encompasses evaluations categorized within the Appraisal framework under *Expand*, which reflects a dialogic expansion. *Epistemic (or Alethic)* presents a predicative content within a framework of uncertainty. It constitutes a partial disengagement by the enunciator (ex. (7), (1b)). *Inter-enunciator* modality applies to *Questions*, *Orders*, and *Promises*, which require linking predicative content to another enunciator in order to acquire a truth value, as well as the involvement of the co-enunciator in the statement. Example (2) also illustrates a question. These two types of modality reflect the enunciator's partial disengagement regarding the truth value of the clause. However, promises may convey a judgment or intention of the enunciator on an axiological level, thus indicating engagement on the axiological value. *Deontic modality* characterizes a judgment expressed by the enunciator based on external codes and rules, such as institutional norms. It reflects the enunciator's engagement on the axiological value while simultaneously disengaging from the truth, as the predicative content does not receive a truth value (8). *Appreciative modality* (which includes bouletic, axiological, and appreciative dimensions) expresses an individual judgment by the enunciator, as well as their engagement on the axiological value, as in (3), (11). This category can be related to the *Attitude* defined in the Appraisal framework, whether it concerns affects, judg-

ments, or appreciations. These modalities are strong evidence of the enunciator's processes of engagement or disengagement.

(7) Il aurait dit à cette occasion "Paris vaut bien une messe ". / He reportedly said on that occasion "Paris is well worth a mass".

(8) Le sanctuaire de l' école doit être préservé des passions intéressées et des luttes stériles des partis. / The sanctuary of the school must be preserved from self-interested passions and the sterile struggles of partisan factions.

- The *Representation of speech* category encompasses all forms of reported speech, as well as *Self-Representation of speech* as developed by Authier-Revuz (2020). The six proposed categories of reported speech are distinguished by the type of enunciative anchoring employed by the primary and secondary enunciators ("unified" or "dissociated" anchoring, ibid.) and by the status of the reported speech (as the "source of speech" or as the "object of speech", ibid.). The six categories are as follows: *Direct speech*, *Indirect speech*, *Free indirect speech*, *Autonymic Modalization* (the act of borrowing lexical elements from another speech, as in example (9)), *Second-assertion modalization* (the act of reporting the source of an utterance as in (10)), and *Self-representation of speech* (the act of representing one's own speech in the process of being produced), as in (11) with *"I say it with regret"*. These six categories demonstrate the enunciator's engagement regarding the fact that another enunciator has uttered the reported speech. However, this initial engagement may be attenuated by a modalization of the reporting act, possibly leading to partial disengagement (see ex. (7)). The enunciator's engagement toward the predicative content of the reported speech is more difficult to evaluate. Although some categories of reported speech appear to favor a distancing between the primary enunciator and the reported speech (*Direct speech*, *Autonymic modalization*), particularly because they involve distinct enunciative anchoring, they do not constitute sufficient evidence to determine the primary enunciator's engagement. *Self-representation of speech* (11) is a special case, describing a mode of utterance in which the enunciator underlines their own speech, thereby confirming potential engagement or disengagement toward either value within the self-quoted speech. It is worth noting that a reflexive loop may emerge between the secondary enunciator ("e") and the primary ("E") at the origin of the utterance. Indeed, the reported speech may eventually be qualified by all the enunciative operations applicable to "E". As it describes various ways of representing another enunciator's speech, this category can be compared to the evaluations under the *Heteroglossic* category of Appraisal theory, although we do not adopt the distinction between *Contract* and *Expand* in the description of our subcategories, as our approach does not focus on the enunciator's stance.

(9) Pour construire la basilique Saint-Pierre-de-Rome, le pape vend des "indulgences". / To fund the construction of St. Peter's Basilica in Rome, the Pope sells "indulgences."

(10) Selon l'Église catholique, elles pardonnent les péchés. / According to the Catholic Church, they forgive sins.

(11) Et là, dans sa fureur, je le dis à regret, c'est le seul crime de ce héros, mais il est affreux, il fit massacrer 3000 personnes. / And there, in his fury, I say it with regret, for it is this hero's only crime, yet an awful one, he had 3,000 people slaughtered.

The schema incorporates a total of 24 distinct annotation labels. Except for a few incompatibilities between these labels (*Anchored* necessarily excludes *Non-anchored*, *Non-Engagement* excludes *Engagement / Disengagement*, and *Inter-enunciator modality* excludes Non-anchored) most of these enunciative mechanisms can apply additively to the same clause. An annotation guide (Bruneau and Battistelli, 2024) in French provides a more detailed description of the different categories, along with annotation examples and frequently encountered linguistic markers for each. In the following section, we present the results of a manual annotation of a corpus based on this schema, in order to illustrate the proportions of these enunciative categories that enunciators may employ in a specific context.

4 Exploring some manual annotations

4.1 Corpus description

In this section, we present the results of the manual annotation of a collected French corpus, comprising 858 clauses extracted from history lessons in eight textbooks. Four of them (referred to as "ancient") were published in the 19th century. They were referenced by the MoDoAp project[2], which draws on a larger corpus of digitized textbooks hosted by the French online library Gallica[3]. In addition, four textbooks published in the 21st century ("modern") were collected to support diachronic comparison in this study. The clauses are evenly distributed across 3 themes: (i) Christopher Columbus and the Age of Discovery, (ii) Martin Luther and the Reformation, and (iii) the French Revolution. The clauses composing this corpus are independent clauses (see Section 3.1). We designed a tool for the automatic segmentation of the corpus into independent clauses. Based on the syntactic analysis provided by Stanza (Qi et al., 2020), this tool implements a heuristic that consists in identifying verbs (or adjectives or nouns within verbal constructions) that have a syntactic dependency of the *conjunction* or *parataxis* type and are syntactically linked to a verbal (or adjectival or nominal within a verbal construction) root.

The complete annotated corpus is available online[4].

History textbooks constitute a specific textual genre, as the enunciator's engagement toward the truth and factuality of the described events is of particular interest, and as they are not axiologically neutral. Moreover, variations may exist among textbooks depending on factors such as educational level and the specific editorial competition of certain periods, which can be observed through the analysis of enunciative operations and the enunciator's engagement.

4.2 Annotation process

The annotation was conducted by two experts in enunciative linguistics. For each independent clause, a label (0 or 1) was assigned to each category in the annotation schema. Initially, a provisional annotation guide was provided to the annotators, detailing the different categories and rele-

---

Category	IAA	Category	IAA
Non-Engagement	0.83	Appreciative	0.90
Val Individual	0.93	IE Question	1.0
Val Community	0.84	IE Order	1.0
Val Universal	1.0	IE Promise	0.0
Context Temporal	0.97	IE involv. e	0.97
Context Spatial	0.91	Plausibility	1.0
Context Conditional	1.0	Direct Speech	0.96
Context Discursive	1.0	Indirect Speech	0.88
Anchored	0.90	Free Indirect Speech	
Non-anchored	0.92	Second-assert Modal	0.86
Epistemic	0.92	Autonymic Modal.	0.88
Deontic	0.96	Self repr. of Speech	1.0

Figure 2: IAA scores for each category

vant linguistic markers for French. A preliminary inter-annotator agreement score was computed on a larger corpus of 1000 clauses, to identify the categories leading to the most disagreement. A discussion between annotators was held to highlight sources of inconsistency, followed by a second round of annotation, leading to the inter-annotator agreement (IAA) results presented in Figure 2, with the categories related to *Engagement / Disengagement* highlighted in blue. Cohen's kappa (Cohen, 1960) was used to calculate agreement between the two annotators, applied in a binary manner for each category. During the preliminary annotation process, various difficulties emerged, leading to revisions of the annotation guidelines, including:

1. The need to distinguish between a spatio-temporal information contextually added to the clause (*Contextual frame of reference*) and that which is inherently part of the content. Spatio-temporal frames of reference were eventually considered as being exclusively adverbials, in opposition to spatio-temporal information playing the role of grammatical subjects or object complements (as the ones underlined in examples (12) and (13) respectively, which would not be annotated as *Contextual frame of reference*):

   (12) <u>Rome</u> est devenue la ville la plus puissante / <u>Rome</u> has become the most powerful city

   (13) Cette armée a conquis <u>des régions de plus en plus lointaines</u> / This army has conquered <u>increasingly distant regions</u>
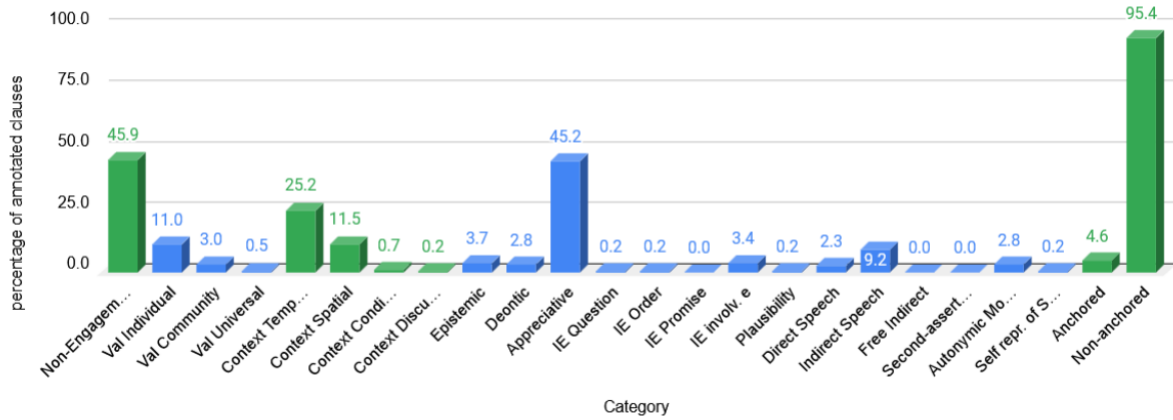
7

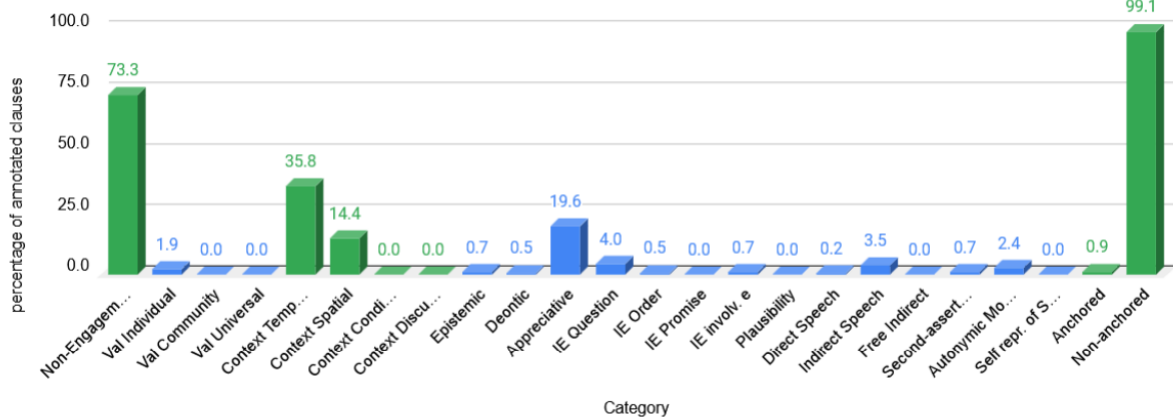Figure 3: Proportions of the annotated categories - Ancient corpus

Figure 4: Proportions of the annotated categories - Modern corpus

2. The necessity of differentiating adjectives that trigger a subjective evaluation from the enunciator (*Appreciative modality*, see the ones underlined in example (14)) from more "objective" adjectives (15). To implement this distinction, we relied on the division between "classifying" and "non-classifying" adjectives established in (Maingueneau, 2020).

(14) Louis XV fut un roi <u>vicieux</u>, <u>égoïste</u> / Louis XV was a <u>vicious</u>, <u>selfish</u> king

(15) Les <u>principales</u> villes du royaume devinrent le théâtre de scènes <u>analogues</u> / The kingdom's <u>major</u> cities became the stage for <u>similar</u> scenes;

3. The wide variety of possible linguistic markers for *Appreciative modality*.

After the IAA calculations, a gold standard corpus was established at the intersection of the two sets of manual annotations provided by the annotators, comprising the clauses reflecting a full agreement. 142 clauses were removed from the 1000 initially annotated, leading to the 858-clause corpus we describe. We adopted this conservative gold standard, based on strict intersection, in order to ensure maximal reliability of the annotated corpus. This choice also preserves the original distribution of disagreements, which can be analyzed separately in future work to better understand the sources of annotation variability.

8

4.3 Annotation results

Figure 3 presents the proportions of manually annotated clauses for each category in the ancient textbooks subcorpus, while Figure 4 focuses on the modern textbooks. Blue categories are the ones under *Engagement / Disengagement* in the typology, in contrast with the green categories. Among the 24 categories shown in the annotation schema, only *Promises* and *Free indirect speech* have not been encountered in the overall corpus.

The two temporal series primarily differ in their proportions of *Non-Engagement* (73.3% for modern textbooks vs. 45.9% for ancient), which is explained by the greater presence of *Appreciative modality* (implying engagement) in ancient textbooks (45.2% compared to 19.6%). Individual enunciator engagement regarding truth-value (*Validation - Individual*) is significantly higher in ancient textbooks (11% of the clauses vs. 1.9% of the modern ones), although the majority of enunciator engagement across the entire corpus relies on *Appreciative modality*. *Non-anchored* clauses are strongly favored in this textual genre. *Spatial* and *Temporal* contexts are relatively frequent in both series, as expected in history lessons.

5 Conclusion

The annotation schema we propose offers a framework for capturing the wide range of enunciative strategies underlying every enunciation process by which an enunciator actualizes a predicative content. This schema focuses on two important global categories relevant for describing the ways an enunciator positions in relation to the truth value and the axiological value of a predicative content: *Engagement* - notion most frequently used in works that refer to Appraisal theory (e.g. (Zeng et al., 2024)) - and *Non-Engagement* - notably absent from Appraisal approach-. When a lot of *Non-Engagement* textual units are highlighted in a text (or in a corpus of texts), it underlines a phenomenon in which the enunciator "fades away". We illustrated this phenomenon here by comparing history lessons from two corpora. More generally, our frame of analysis allows for the description of enunciative profiles and compare them between diverse corpora. Future work may extend this approach to other genres and corpora, and leverage the schema's potential for training NLP models in tasks such as toxic speech detection, as in (Battistelli et al., 2023) and ideological content analysis.

References

Jacqueline Authier-Revuz. 2020. *La représentation du discours autre: principes pour une description*. De Gruyter.

Charles Bally. 1932. *Linguistique générale et Linguistique française, par Charles Bally*. Impr. des Presses universitaires le France.

Delphine Battistelli, Valentina Dragos, and Jade Mekki. 2023. Annotating social data with speaker/user engagement. illustration on online hate characterization in french. In *International Conference on Computing and Communication Networks*, pages 317–330. Springer.

Cyril Bruneau and Delphine Battistelli. 2024. Guide d'annotation manuelle de la prise en charge énonciative - Version V2 (Février 2024). Https://hal.science/hal-04541398v1/file/$Guide_{Annotation_PEC_V}2_02 - 2024.pdf$.

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

Danielle Coltier, Patrick Dendale, and Philippe De Brabanter. 2009. La notion de prise en charge: mise en perspective. *Langue française*, (2):3–27.

Philippe De Brabanter and Patrick Dendale. 2008. Commitment: The term and the notions. *Belgian Journal of Linguistics*, 22(1):1–14.

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.

Jean-Pierre Desclés and Zlatka Guentchéva. 2024. Évidentialité, médiativité, modalité épistémique: une approche énonciative. In *SHS Web of Conferences*, volume 191.

Marc Evrard, Rémi Uro, Nicolas Hervé, and Béatrice Mazoyer. 2020. French tweet corpus for automatic stance detection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6317–6322.

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. Semeval-2019 task 7: Rumoureval 2019: Determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation: NAACL HLT 2019*, pages 845–854. Association for Computational Linguistics.

Nanjiang Jiang and Marie-Catherine de Marneffe. 2019. Do you know that florence is packed with visitors? evaluating state-of-the-art models of speaker commitment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4208–4213.

Dominique Maingueneau. 2020. Chapitre 6. classifiance et non-classifiance. *Lettres Sup*, 2:107–132.

James R Martin and Peter R White. 2003. *The language of evaluation*, volume 2. Springer.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, San Diego, California. Association for Computational Linguistics.

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.

Qi Su, Chu-Ren Huang, and Helen Kaiyun Chen. 2010. Evidentiality for text trustworthiness detection. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, pages 10–17.

Jiamei Zeng, Min Dong, and Alex Chengyu Fang. 2024. Annotating evaluative language: Challenges and solutions in applying appraisal theory. In *Proceedings of the 20th Joint ACL-ISO Workshop on Interoperable Semantic Annotation@ LREC-COLING 2024*, pages 144–151.
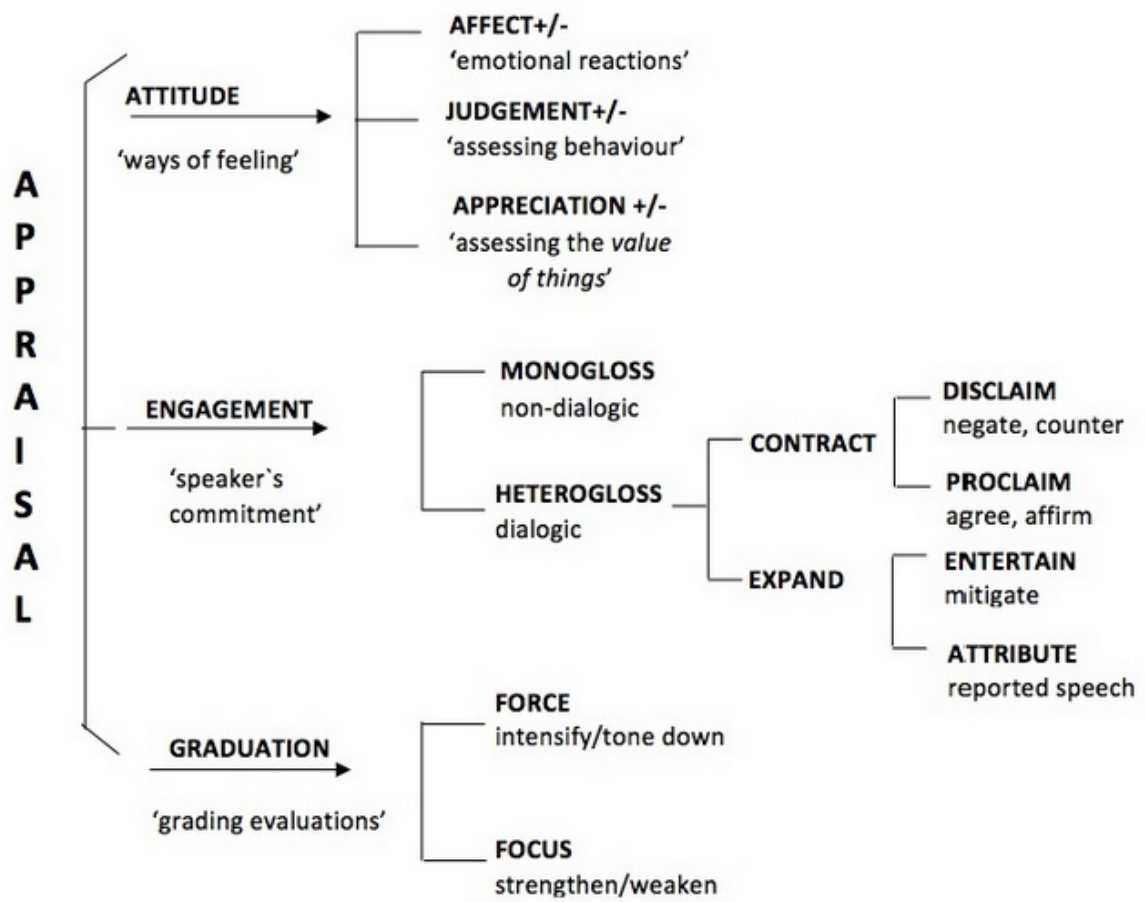
Figure 5: Appraisal Framework, adapted from (Martin and White, 2003, p. 38)