

# The Roots of Doubt. Fine-tuning a BERT Model to Explore a Stylistic Phenomenon

Margherita Parigini<sup>1</sup>, Mike Kestemont<sup>2</sup>

<sup>1</sup>Université de Genève

<sup>2</sup>University of Antwerp

## Abstract

The narrative work of well-known Italian author Italo Calvino (1923-1985) features a phenomenon that literary critics refer to as “dubitative text”: this stylistic device consciously hinders the narrative progression of a story, by questioning its own content. We report on an attempt to model the presence of dubitative text in Calvino’s fictional oeuvre and examine whether this model can also be used to retrieve dubitative instances in his essayistic oeuvre. We hypothesize that precisely the category of the dubitative text yields interesting points of intersection between both writing modes. We fine-tuned a BERT model based on a manually annotated dataset and report inter-annotator scores. We situate our findings and model criticism in the current landscape of Calvino scholarship. While detecting dubitative text is challenging, our model provides fresh insights into the device’s surface features.

## Keywords

Italian literature, Italo Calvino, BERT, entity recognition

## 1. Introduction

Italo Calvino (1923-1985) has been one of the most important authors on the Italian literary scene in the twentieth century: he published some twenty volumes, including novels and collections, selling approximately 4 million copies during his 40-year career.<sup>1</sup> As previously emphasized by scholars, a central characteristic of his narratives is that they «are being made as they unravel: they deny themselves, biting their own tails. They fill the void of the blank page through the thematization of their dissolution» [48, p. 126][own translation]. The narrative progression of the text, in these cases, is invariably linked to a questioning of what was previously stated.<sup>2</sup> This mechanism only recently has become the subject of systematic research in the project *Atlante Calvino. Literature and Visualization*<sup>3</sup>, where it is called “dubitative text”.

---

CHR 2022: Computational Humanities Research Conference, December 12 – 14, 2022, Antwerp, Belgium

✉ margherita.parigini@unige.ch (M. Parigini); Mike.Kestemont@uantwerpen.be (M. Kestemont)

🌐 <http://mikekestemont.github.io/> (M. Kestemont)

🆔 0000-0002-0832-2202 (M. Parigini); 0000-0003-3590-693X (M. Kestemont)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>To further explore the extent to which Calvino was a successful writer [47, pp. 358-366].

<sup>2</sup>For an in-depth study of the role of *correctio* and *dubitatio* in Calvino’s writing [33][39].

<sup>3</sup>*Atlante Calvino. Literature and Visualization* (<https://atlantecalvino.unige.ch/?lang=en>) is a project funded by the Swiss National Science Foundation and directed by Professor Francesca Serra (University of Geneva) in collaboration with the DensityDesign Lab (Politecnico di Milano) that tries to apply Data Visualization techniques to address some research issues related of literary criticism. The itinerary of Doubt in the Atlante is dedicated to the dubitative text.

The project looks at the manifestation of the phenomenon in fiction only, without considering his essays production. Nevertheless, it is important to underline that Calvino, responding to a trend of the time, is leaning over the course of his career towards hybrid forms that mix the techniques of both his fictional and essayistic writing (e.g. [38, 18, 24]).<sup>4</sup> A recent study devoted to the author’s nonfiction production showed how he tends to express himself by adopting different «forms of perplexity» [6, p. 10][own translation]: in the essays, «the reader is accompanied by a subtle enunciative game relying on questions and stylistic devices of doubt and questioning» [6, p. 11][own translation]. Our hypothesis is that these “forms of perplexity” could be related to the dubitative text.

The dubitative phenomenon belongs to the fictional universe, yet it might have its roots within the essays, or vice versa it may have been able to propagate into nonfiction writing: however, it seems clear to us that a link exists. This research intends to explore the nature of this link and aims to assess what stylistic and structural components the dubitative text might have in common with the nonfiction genre. To investigate this hypothesis, we fine-tuned a BERT model to automatically detect the presence of dubitative text in a cross-genre setting. Afterwards we annotated a representative sample of texts in both genres for dubitativity (§2.2.)

One of the defining characteristics of the dubitative phenomenon is that affects “the main progression of the *story*” (§2.1.). The genre of the essay possesses a mobile definition in the various critical traditions, which, however, has some constants: in particular, this form of discourse «is based on the implicit prescription not to invent a fictional world to convey reflections on the world» [31, p. 152][own translation]. Consequently, trying to analyze the essays using the category of the dubitative text might seem like a stretch. However, we created a validation set of essays by collecting dubitative occurrences as if they were fictional texts, with two specific purposes: (a) field-testing the eventual limits of the dubitative category; (b) having a reference point for more thoughtful evaluation of the model results.

Instead of focusing on model performance, our aim was to derive interpretative insights into the dubitative text in both genres, in order to deepen our understanding as to which stylistic or structural elements are associated with this phenomenon.

## 2. An Ambiguous Category

### 2.1. Dubitative Text

The dubitative text is a category of analysis which draws inspiration on multiple linguistic and narrative dimensions of a text. It is a challenging notion to operationalize in a formal annotation framework [45]. We set up an annotation process with the ultimate aim of training a machine learning model that can automatically detect its presence in an unseen text. We focus specifically on a stylistic conceptualization of this label in the sense proposed by Herrmann who argue that «style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively» [25, p. 44].

---

<sup>4</sup>«Scholars such as Thomas Pavel have recognized in the mixture of narrative and nonfiction meditation a decisive character of the modern and contemporary novel (think of Marcel Proust, Thomas Mann, Robert Musil, Jorge Luis Borges, and, among more recent examples, Thomas Bernhard, Milan Kundera, or Enrique Vila-Matas)» [19, p. VII][own translation][40].

To consider a string of text as dubitative, we require that the progression of the story (or a cohesive section of it) must be based on the deconstruction of its content. The dubitative text is therefore characterized by the following linguistic features:

- expressions of epistemic modality [42, pp. 54-61];
- discourse connectives related to the argumentative genre [30];
- punctuation marks (parentheses, dashes, question marks, ellipsis);

These features serve as useful surface cues in a text to flag any dubitative occurrences in Calvino's writings. However, these characteristics in isolation do not automatically entail the presence of dubitativity – the first condition regarding the deconstruction of the narrative progress must still be met. Let us take an example.

#### UN-DUBITATIVE

«piccole increspature che si propagavano sott'acqua: forse il battere di coda d'un pesce, forse un bambino che riempiva un secchiello su una riva»

[small ripples that propagated underwater: perhaps the tapping of a fish's tail, perhaps a child filling a bucket on a shore] [own translation]

*Com'era grande il mare*, 1948

#### DUBITATIVE

«Una bestia si mosse in fondo a un cespo d'erice: forse una lepre, forse una volpe, forse un tedesco coricato tra gli arbusti che lo prendeva di mira »

[An animal moved at the back of a bush of heather; perhaps it was a hare, perhaps a fox, perhaps a German lying in the thickets keeping him covered]

*Paura sul sentiero*, 1946

**Figure 1:** Examples of un-dubitative and dubitative text: *Com'era grande il mare* [11, p. 855]; *Paura sul sentiero* [9, p. 250][7, p. 71].

The stylistic form adopted in the two passages is almost identical. A movement whose origins are unknown is described («piccole increspature», «una bestia si mosse»), and three hypotheses introduced by as many adverbs of doubt («forse») are presented. The difference between these two extremely similar texts lies in the way the respective plots are developed. *Com'era grande il mare* is a short story about a couple ascending a headland on a scorching summer day, who, blinded by the sun and mugginess, cannot always see clearly [11, p. 855]. *Paura sul sentiero* narrates the nighttime crossing of a forest by a partisan dispatcher: darkness, sleep and fear generate a deeply doubting attitude toward what surrounds the partisan, thus making the progression of the narration coincide with a concatenation of uncertainties [9, p. 250]. While in the case of the first short story we are dealing with a circumscribed visual difficulty, in the second the thematization of doubt is the pivot around which the narrative revolves. Determining which text can be considered dubitative is an intersubjective, interpretative act that needs a case-by-case inspection during a manual annotation process.

## 2.2. The Inter-Annotator Agreement

Two annotators have independently annotated a representative sample of text to be able to assess the feasibility of this annotation task. Each was asked to analyze a different set of texts, selecting the parts considered as dubitative.<sup>5</sup> Their work was subsequently compared to the group of annotations linked to the corresponding texts in the dubitative text dataset derived from the Atlante Calvino project (§3.1.). Annotator A was in charge of the two related novels *Il castello dei destini incrociati* and *La taverna dei destini incrociati*. Annotator B analyzed the collection *Palomar*. We decided to apply Cohen’s Kappa [14] [22] to measure the IAA.

**Table 1**

*Il castello dei destini incrociati* (1973). Length (words): 26,519.

	Number of annotations	Annotation average length
<b>Team Atlante Calvino</b>	353	15.45
<b>Annotator A</b>	256	22.16

**Table 2**

*Palomar* (1983). Length (words): 26,867.

	Number of annotations	Annotation average length
<b>Team Atlante Calvino</b>	403	14.25
<b>Annotator B</b>	135	20.81

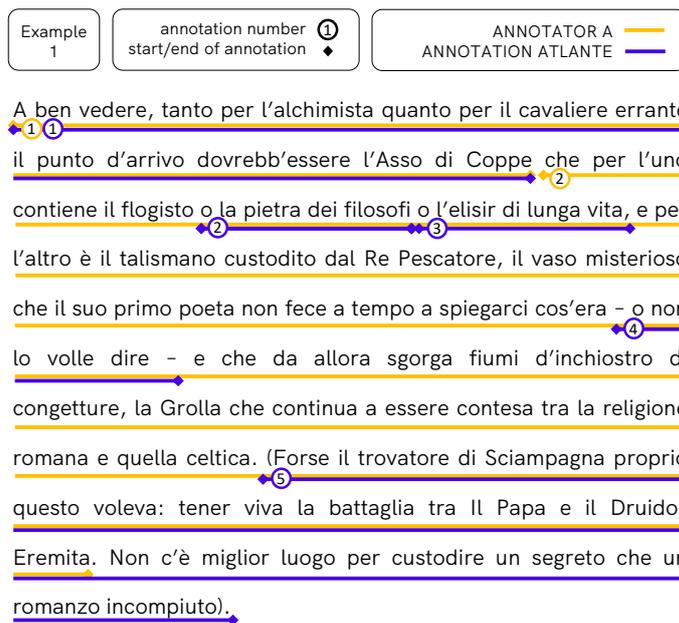
Annotator A obtained a score of 0.54 and Annotator B of 0.50. The difficulty of the task is illustrated by the weak agreement observed.<sup>6</sup> We were therefore aware from the outset that we had to manage our expectations regarding the performance of the automated model, deciding to «exploit disagreement between crowd workers as a signal, rather than try to eliminate it» [2, p. 23].<sup>7</sup>

Dubitative text is an elusive category that presents classification problems. We tried to understand what could be the causes of such a low score, analyzing the discrepancies: (a) within an area considered unanimously dubitative, the boundaries of occurrences are not always clear (cf. **Fig. 2**); (b) it seems to us that annotations diverge when confronted with sentences with a strong semantic component of uncertainty but without clear stylistic markers (cf. **Fig. 3**). Both of these problems later resurfaced in the results of the model, studying the false positives and the false negatives (§6.).

<sup>5</sup>To further explore the units of analysis used in content analysis [27, pp. 97-110].

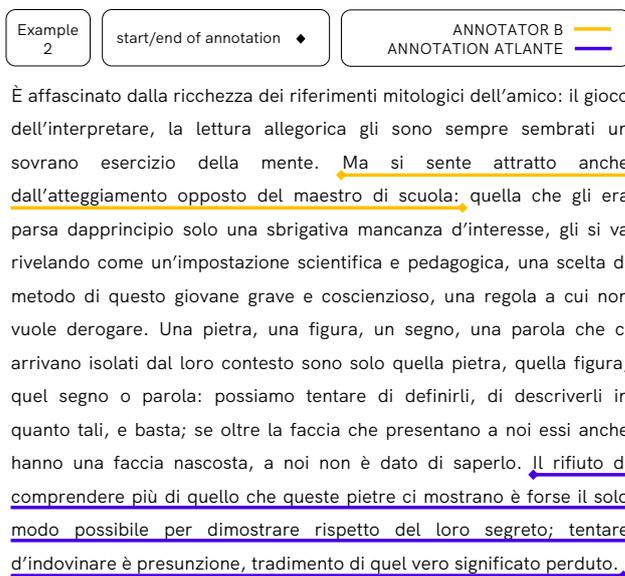
<sup>6</sup>As reference points for evaluating the scores obtained [32, 28].

<sup>7</sup>We do not want to claim that the «exploratory tool» produced, «even if wrong», is «intrinsically valuable because exploration is intrinsically valuable» [15, p. 602] Rather, we want to go beyond the numbers, as daunting as they may seem, to check their cognitive reliability. Actually, «Parsing f-scores, and precision and recall values enable a certain degree of understanding of the overall performance of the classifier and the function of the workflow as a whole, but this leaves aside many other post-classification metrics that can be directly applied to the underlying data model. These data can tell us much about our input datasets and the criteria by which the classifier made its classifications» [17].



If you look carefully, the destination for both the alchemist and the knight-errant should be the Ace of Cups which, for the one, contains phlogiston or the philosopher's stone or the elixir of long life, and for the other the talisman guarded by the Fisher King, the mysterious vessel whose first poet lacked time—or else was unwilling—to explain it to us; and thus, since then, rivers of ink have flown in conjectures about the Grail, still contended between the Roman religion and the Celtic. (Perhaps the Champagne troubadour wanted precisely this: to keep alive the battle between The Pope and the Druid-Hermit. There is no better place to keep a secret than in an unfinished novel.)

**Figure 2:** Excerpt from *Il castello dei destini incrociati* (1973) [10, pp. 583-584][13, p. 107].



He is fascinated by his friend's wealth of mythological references: the play of interpretation, allegorical readings, have always seemed to him a supreme exercise of the mind. But he feels attracted also by the opposite attitude of the schoolteacher: what had at first seemed only a brisk lack of interest is being revealed to him as a scholarly and pedagogical position, a methodological choice by this serious and conscientious young man, a rule from which he will not swerve. A stone, a figure, a sign, a word that reach us isolated from its context is only that stone, figure, sign or word: we can try to define them, to describe them as they are, and no more than that; whether, beside the face they show us, they also have a hidden face, it is not for us to know. The refusal to comprehend more than what the stones show us is perhaps the only way to evince respect for their secret; trying to guess is a presumption, a betrayal of that true, lost meaning.

**Figure 3:** Excerpt from *Serpenti e teschi* (1978) [10, pp. 955-956][8, pp. 117-118].

### 3. Data

#### 3.1. Dataset of the Dubitative Text

The annotated training dataset was obtained from the Atlante Calvino project. It is composed of almost 5,000 occurrences of dubitative text, derived from the fictional work published when Calvino was alive (205 short stories and 10 novels). Occurrences vary in length, including groups of words, sentences, or even whole paragraphs.<sup>8</sup>

#### 3.2. Corpus of Essays

Calvino produced numerous essays and articles throughout his career, about 400 of which constitute the two volumes edited by Mario Barenghi [12]. Nevertheless, we limited our study to the three volumes set up by the author himself as coherent ensembles: *Una pietra sopra. Discorsi di letteratura e società*, published in 1980 and containing 42 essays written between 1955 and 1978; *Collezione di sabbia*, published in 1984 and containing 23 essays written between 1980 and 1984; *Lezioni americane. Sei proposte per il prossimo millennio*, published posthumously in 1988, which collects 5 lectures that the author should have given at Harvard University in 1985-1986 as part of the Charles Eliot Norton Poetry Lectures. The volumes were purchased in electronic format and automatically converted to plain text.

#### 3.3. Validation Set

We set up two validation sets for the model validation: one consisting of fictional writings (and thus closer to the training material) and one consisting of essays (and thus closer to the target domain for our model) (§5.).

Five different short stories were selected ensuring that they were distributed chronologically across the entire time span of Calvino’s career and with a variable presence of the dubitative phenomenon. Ten essays were selected following the same guidelines.

**Table 3**

Validation set: fiction.

ID	title	year of publication	collection
S007	<i>Paura sul sentiero</i>	1946	<i>Ultimo viene il corvo</i>
S101	<i>L'avventura di un miope</i>	1958	<i>Gli amori difficili</i>
S142	<i>Meiosi</i>	1967	<i>Ti con zero</i>
S159	<i>Prima che tu dica "Pronto"</i>	1975	-
S190	<i>La spada del sole</i>	1983	<i>Palomar</i>

<sup>8</sup>The occurrences have been collected, using the guidelines mentioned above (§2.1.). For an accurate description of the structure of the dataset cf. DOUBT – DATASET 2 in <https://atlantecalvino.unige.ch/capta?lang=en> The dataset will be available once the PhD of its author, Margherita Parigini, is completed. Instead, the corpus data from which the IDs and collateral information were derived for this paper are accessible cf. [49].

**Table 4**  
Validation set: essay.

ID	title	year of publication	collection
E006	<i>Dialogo di due scrittori in crisi</i>	1961	<i>Una pietra sopra</i>
E011	<i>L'antitesi operaia</i>	1964	<i>Una pietra sopra</i>
E019	<i>Cibernetica e fantasmi</i>	1967-68	<i>Una pietra sopra</i>
E038	<i>La penna in prima persona</i>	1977	<i>Una pietra sopra</i>
E044	<i>Collezione di sabbia</i>	1974	<i>Collezione di sabbia</i>
E046	<i>Il viandante nella mappa</i>	1980	<i>Collezione di sabbia</i>
E070	<i>I mille giardini</i>	1984	<i>Collezione di sabbia</i>
E072	<i>La spada e le foglie</i>	1977	<i>Collezione di sabbia</i>
E078	<i>La foresta e gli dei</i>	1984	<i>Collezione di sabbia</i>
E082	<i>La leggerezza</i>	1988*	<i>Lezioni americane</i>

## 4. Fine-tuning BERT model

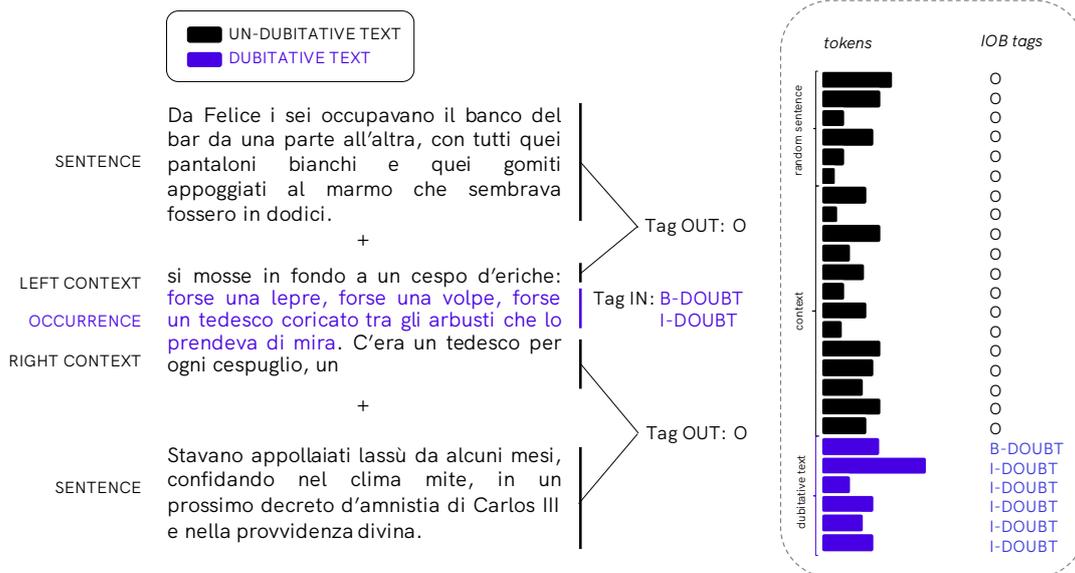
### 4.1. Research background

We fine-tuned a Transformer BERT to detect the presence of the dubitative text adopting a NER approach. BERT, which means Bidirectional Encoder Representations from Transformers [51], is «designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers» [16]. One of the aspects that make BERT such an effective tool is that it can be easily fine-tuned to accomplish a variety of specific tasks, «with only one additional output layer» [16]. To conduct our research, we used BERT as a sequence tagger. For the most part, we have reproduced [1].

### 4.2. Preprocessing Data

During preprocessing, an important initial choice related to the relative proportion of dubitative text presented to the model in comparison to the un-dubitative text; our corpus statistics suggest a serious class imbalance – the proportion of dubitative text has an average occurrence of only 20%. During training, we therefore artificially varied the relative proportion of dubitative text fed to the model as a hyperparameter [46, 36]. Regarding un-dubitative text, we decided to provide the model with (a) the context of each occurrence, using the sentences naturally coming before/after the dubitative ones, and (b) examples of non-dubitative sentences extracted from Calvino’s fictional corpus.

- (a) To determine the size of the context, we calculated the average number of words in relation to the total number of words in occurrences in the dataset, which correspond to 22;
- (b) We randomly extracted sentences from the fictional corpus, looking at the mean and variance of the distribution of dubitative occurrences in the dataset, for a total of 5000 sentences. We also took into account the possible overlap with dubitative occurrences and we made sure to avoid it;



**Figure 4:** Example of the analysis and tagging process carried out on the text.

To structure the files needed to fine-tune the BERT model, we interspersed a non-dubitative sentence before and after each group of tokens formed by context and occurrence.

Next, we were careful to stratify the data to form a training set and a test set, adopting the proportions 80% and 20%, respectively, without shuffling the tokens.

### 4.3. Training Steps

Our model was developed through numerous iterations to determine the optimal (hyper)parameters on the validation data, also in the light of the prior literature on the topic (e.g. [26, 35, 51]). Here, we briefly report on the main insights derived from this process.

**Table 5**  
Step 1.

M	nr interweave	doubt	no doubt	BERT	batch size	f1 best score
	1	15.5%	84.5%	base-multilingual-uncased	128	0.31%
	1	15.5%	84.5%	italian-xxl-cased	128	0.35%
1	2	24.4%	76.6%	italian-xxl-cased	256	0.49%

The BERT Lang Street platform [37], developed by a group of researchers within the Bocconi University of Milan, allowed us to identify the most suitable model to perform our type of task. After an initial attempt with the multilingual version, we then opted for the italian XXL cased model. The model trained with the starting distribution of dubitative text obtained an F1 of 0.49. Because of the fairly low Kappa’s score and the scores of previous experiences performed

on similar tasks [46], our expectations linked to the model were modest from the beginning. Next, we experimented with the proportion of the doubt/no doubt instances in the training data.

**Table 6**

Step 2.

M	nr interweave	doubt	no doubt	context size	un-dubitative sentences	f1 best score
	3	29%	71%	11+11	2000	0.54%
	-	46.3%	53.7%	11+11	-	0.64%
	2	34%	66%	-	2500	0.66%
	3	33%	67%	5+5	2000	0.56%
	6	22.3%	77.7%	25+25	1000	0.50%
	3	29%	71%	11+11	2300	0.53%

As soon as we manipulated the internal distribution of the data, moving away from the values established at the outset, the model went into overfitting.

**Table 7**

Step 3.

M	training score	BERT	best score
1	f1	italian-xxl-cased	0.49%
2	precision	italian-xxl-cased	0.51%
3	precision	italian-xxl-uncased	0.53%

We then took the trained model with the best score so far and we decided to change the training score from the traditional F1 to precision score, so that the model selected during fine-tuning would be as reliable as possible. In fact, our goal at this stage of the research was to obtain sufficiently accurate results to be able to interpret the model selection processes.

## 5. Results

### 5.1. Model Criticism

To analyze the results we decided to select the three models with the best scores in the fine-tuning phase, namely Model1 (M1), Model2 (M2) and Model3 (M3). We wanted to take an experimental approach, going beyond the numerical results that would lead us to choose a single model. We wanted to see if, by combining the different results, we could learn more about the dubitative text and how it was decoded.

At first, we noticed a correlation between the density ratio<sup>9</sup> [50] of the dubitative text and the scores of the validation set: this aspect is probably related to the strong imbalance in the

<sup>9</sup>Given two data samples  $x_1$  and  $x_2$  from unknown distributions, the density ratio is equal to:  $(x_1 + x_2)/x_1$ , where  $x_1$  in this case are the dubitative occurrences and  $x_2$  are the non-dubitative occurrences cf. [https://github.com/oxo-m/densratio\\_py](https://github.com/oxo-m/densratio_py).

**Table 8**

Average values of validation sets.

	VALIDATION SET: FICTION				VALIDATION SET: ESSAY		
	Model 1	Model 2	Model 3		Model 1	Model 2	Model 3
F1-score	0.48	0.49	0.46	F1-score	0.41	0.39	0.49
precision	0.65	0.52	0.61	precision	0.37	0.31	0.50
recall	0.40	0.50	0.40	recall	0.57	0.64	0.60
COLOR KEY	weak <0.39	moderate <0.59	strong <0.79				

**Table 9**

Values of validation sets.

ID	density ratio	VALIDATION SET: FICTION								
		Model 1			Model 2			Model 3		
		F1	precision	recall	F1	precision	recall	F1	precision	recall
S007	14.01	0.33	0.37	0.29	0.32	0.32	0.32	0.34	0.48	0.27
S101	7.45	0.43	0.80	0.30	0.47	0.46	0.48	0.33	0.56	0.24
S142	5.2	0.44	0.40	0.48	0.54	0.51	0.56	0.52	0.50	0.53
S159	9.08	0.71	0.84	0.62	0.49	0.39	0.64	0.65	0.68	0.62
S190	4.16	0.48	0.84	0.33	0.64	0.92	0.49	0.48	0.84	0.34

ID	density ratio	VALIDATION SET: ESSAY								
		Model 1			Model 2			Model 3		
		F1	precision	recall	F1	precision	recall	F1	precision	recall
E006	9.23	0.38	0.30	0.51	0.43	0.30	0.72	0.63	0.53	0.77
E011	8.00	0.48	0.39	0.64	0.43	0.30	0.74	0.52	0.42	0.69
E019	5.73	0.39	0.41	0.38	0.39	0.34	0.45	0.46	0.55	0.40
E038	6.89	0.34	0.25	0.53	0.51	0.43	0.62	0.53	0.57	0.50
E044	5.66	0.51	0.46	0.56	0.54	0.46	0.64	0.67	0.81	0.57
E046	47.87	0.18	0.10	0.91	0.17	0.09	1.00	0.28	0.16	1.00
E070	6.87	0.13	0.15	0.11	0.17	0.20	0.15	0.16	0.28	0.12
E072	6.63	0.52	0.45	0.62	0.31	0.28	0.34	0.39	0.45	0.34
E078	4.94	0.83	0.99	0.72	0.72	0.56	1.00	0.91	1.00	0.84
E082	22.87	0.33	0.21	0.76	0.24	0.14	0.70	0.36	0.24	0.75

COLOR KEY training scores	minimal <0.20	weak <0.39	moderate <0.59	strong <0.79	almost perfect <1.00
---------------------------	------------------	---------------	-------------------	-----------------	-------------------------

COLOR KEY density ratio	strong <0.5	moderate <0.7	weak <0.10	minimal <0.20
-------------------------	----------------	------------------	---------------	------------------

source data. Both models generally perform better when the presence of the phenomenon is frequent (e.g. S159), but their performance weakens in the most extreme phenomenon distribution situations: when its concrete presence is minor (e.g. S007), the models tend to predict more than necessary; in the same way, when confronted with a highly dubitative text, the models tend to categorize as dubitative those parts of the text with marked stylistic features (especially graphical marks and argumentative connectives), leaving out more ambiguous expressions of doubt (e.g. S142).

We also observed a general decrease of the scores once the models were applied to the essays.<sup>10</sup> Rather than being alarming, this finding allowed us to better understand a fundamental

<sup>10</sup>This could be explained with the literary genre of the training data: the dubitative text is a phenomenon that takes place in the narrative corpus; there is therefore a mismatch between the material used to train the models

aspect of the phenomenon which we have already had occasion to mention (§3.3.): the models numerically confirmed the difficulty in transporting this category into a realm other than fiction. However, our goal was not to verify the existence of the dubitative text in essays, but to explore the possible connections between the phenomenon and essayistic expression. An initial numerical data confirms the high compatibility between these two stylistic modes: in fact, we can easily notice an increase in the percentage of essay text recognized as dubitative compared to fictional text.

**Table 10**  
Percentage of text recognized as dubitative by the models.

VALIDATION SET: FICTION				VALIDATION SET: ESSAY			
titles	Model 1	Model 2	Model 3	titles	Model 1	Model 2	Model 3
S007	6%	7%	4%	E006	18%	26%	16%
S101	5%	14%	6%	E011	20%	31%	21%
S142	23%	21%	20%	E019	16%	23%	13%
S159	8%	18%	10%	E038	30%	21%	13%
S190	9%	13%	10%	E044	21%	25%	13%
				E046	18%	23%	6%
				E070	11%	10%	11%
				E072	21%	19%	17%
				E078	15%	36%	17%
				E082	16%	21%	14%
<b>COLOR KEY</b>	low <10%	medium <19%	high 20%>				

All models react with more greediness once they are applied to essays, recognizing a strong presence of the phenomenon. This first finding brought us to guess the existence of a point of contact between the category of the dubitative text and the essay genre. The essay is a text with a reasoning framework that attempts to approximate reality, weighing certain aspects of it [19, p. VII]. It is a «genre of inquiries and questions», «intimately doubtful» [21, p. 186, 197][own translation]. These characteristics are highly compatible with our definition of the dubitative phenomenon. Yet the essay, however hybrid and fluctuating in nature, belongs to the field of argumentation. Applying the model trained to recognize the dubitative text to essays allowed us to verify the presence of certain argumentative techniques that filter into narrative production, redirecting the development of the story.<sup>11</sup>

Dobson states that «determining the meaning of textual data requires both computational knowledge that can determine significance within the model and domain-specific contextual knowledge that can be applied to the understanding of these features»[17]. After evaluating the numerical performance of the models, we then made a second reading, studying the concrete results. A careful analysis of the errors made, both on the narrative and essay validation sets, allowed us:

- to verify the presence in the nonfiction genre of certain stylistic features related to the phenomenon;

and the type of text on which the models are applied. This apparent discrepancy is part of the experience, as our goal is precisely to use the models to understand which are stylistic features related to the argumentative genre may be reused by the dubitative text.

<sup>11</sup>The mix between nonfiction and fiction, though not through the prism of the dubitative text, has been reported in the field of Calvinian studies particularly with regard to *La giornata di uno scrutatore* (1963), «a kind of novel-essay, suspended between testimony and reflection»[3, p. 56][own translation].

- to identify stylistic and argumentative predispositions that open up new research directions.

## 5.2. Model Inspection

A manual inspection of the models' output indicated that certain surface features provided strong cues to the models to flag text as dubitative (cf. Appendix): (A.1) the presence of punctuation marks; (A.2) sentences with keywords related to modality. Additionally, a closer scrutiny of the false positives allowed us to identify two additional dubitativity markers which we named respectively cerebrality and figurality:

- (A.3) Cerebrality: verbs related with a brain activity such as *pensare* [to think] and *sapere* [to know];
- (A.4) Figurality: verbs *parere* [to seem], *guardare* [to look, to watch], *vedere* [to see] and adverb *come* [like];

The first aspect (A.3) suggests an intellectual distance between the narrated object and the act of writing. From the essay, the dubitative text assimilates the reasoning attitude. The essay is marked by a strong «experimental element», a type of «prose as a means of relating to the world» [19, p. VII][own translation]. The dubitative text thus seems to insert a first and foremost conceptual frontier into the narrative, anchoring the development of fiction to a reflexive logic. Narrative strategies are then hybridized to argumentative strategies based on approximation to eventual truth.

Regarding the second aspect (A.4), the role of the visual dimension has particular relevance for Calvino [4]. Thus, it is not surprising that the tendency to associate narrative objects based on a figurative logic would emerge in his work. However, it is interesting that this same logic is associated with the dubitative phenomenon, moreover in an essayistic dimension: the author uses comparison, metaphor, and analogy as conceptual bridges.<sup>12</sup> There is no unanimous agreement on the role that rhetorical figures should play in argumentation, because their presence does not correspond to the need for clarity associated with the genre: metaphor, for example, «is both true and false, guilty of ambiguity and categorical error» [43][own translation][44]. However, Fenoglio traveling «backwards along the chain of tradition in search of the masters of *essais* modern», identifies a «metamorphic drive at the origins of the genre» [21, p. 180][own translation]. The essay, we have said before, is «the quintessential questioning genre» [24, p. 31][own translation], marked by «the uncertainty, the contradictory, the swarming of opinions and points of view» [24, p. 33][own translation]. It is a type of text designed to willingly accommodate both visual and conceptual refraction. And the dubitative text has absorbed this characteristic, readily rehousing it in the realm of fiction, which is particularly predisposed to accommodate this kind of stylistic momentum [5].

---

<sup>12</sup>This aspect had been pointed out by Pier Vincenzo Mengaldo, speaking of the «nexus between analogy-based figurality and precision» [33, pp. 253-55][own translation].

## 6. Conclusion

«The opacity and ineffability of the text and the ethical demand to attend to it remain central to practices of literary interpretation today» [29, p. 371]. Love's "today" is now twelve years old, yet the issue continues to polarize our attention. The act of interpretation means «always wrestling with the text» [34, p. 33][own translation], a strenuous struggle aimed at taking it apart in an attempt to better understand the internal mechanisms that govern it. Currently, it seems that research efforts have shifted, focusing not so much on discovering the mechanisms, but rather on the principles that govern this struggle. Establishing incontrovertible and transparent rules is the new criterion by which to attribute value to an experience, relegating the act of interpretation to an entirely secondary level.

In the current research landscape of the Digital Humanities the debate resonates around the proper weight of the interpretive act of the researcher and, implicitly related, the role of error.<sup>13</sup> Da negatively judges the way that for CLS «misclassifications become object of interest, imprecisions become theory» [15, p. 602]. But one would have to wonder how error is perceived in the more traditional humanistic sphere. Literary criticism is not based on the revelation of absolute evidence, but on the demonstration of a point of view: «the field of argumentation is that of the plausible, the probable, insofar as the latter escapes the certainties of calculation» [41, p. 3][own translation].

The perspective in which this research fits does just what Da argues should not be done: we did not think of the «statistical tools» relying on their so-called «true functions» [15, pp. 619-620]. Rather, «from a humanist perspective, we might want to think of data models created from textual sources as alternative representations of supplied» [17].

The trained model BERT was a way of viewing a research problem of literary criticism in an innovative way. In this sense we tried, following Dobson's suggestion, a «making computational work interpretable» [17]. Through this experience, we learned that the point of contact between the dubitative mechanism, belonging to the universe of fiction, and the nonfiction genre is the latter's peculiarity of possessing the «lens of a double vision» [21, p. 196][own translation].<sup>14</sup> This characteristic of Calvino's nonfiction is translated into the two characteristics: figurality on the one hand and cerebrality on the other.

«The assumption of the essay is that the reader is reading things that are true, perhaps paradoxical and provocative, but all the more so as they should be taken literally and be understood in a dimension of reality. The assumption of the novel, on the other hand, is that, however veraciously taken from reality, the things being read are to be understood on a plane of fiction» [23, pp. 20-21][own translation]. The relationship between fiction and reality is thus put under tension by a technique such as the dubitative text, which applies certain essayistic features within

---

<sup>13</sup>We find the attention to this issue that has emerged in recent CoP linked to some prestigious publication venues particularly representative: Working on and with Categories for Text Analysis: Challenges and Findings from and for Digital Humanities Practices for DHQ; Reproducibility and Explainability in Digital Humanities for IJDH. Even among the topics of interest in the call for CHR22, two inherent items appeared: "development of empirical methods for humanities research" and "modeling bias, uncertainty, and conflicting interpretation in the humanities".

<sup>14</sup>«The essayist in effect contemplates the gaze of others, thus his is a refracted gaze [...] he does not bend to the evidence of things, he questions it, he subjects it to verification, that is, he looks at it from a foreshortened, unusual and oblique point of view» [21, p. 196][own translation].

a fictional universe: this phenomenon implicitly prompts us to consider what is said within the fiction as a truth to be proved. The dubitative text deceives the reader that a fictional truth can be reached by approximation, by adding and subtracting pieces of the different versions of a single story that multiplies before our eyes through doubt.

## 7. Future work

Through analysis of the results, we were able to see the strengths and weaknesses of the various models. By cross-referencing the three models with each other, it is possible to derive a new one in turn that could be more accurate in identifying dubitative occurrences. Once created, this synthesis will be applied to the totality of the essays to study where the two features of figurality and cerebrality appear and how they behave. At the same time, an analysis will be conducted on the totality of argumentative connectives in the essays, leaning on the work of [20], to understand which discourse relations are transported to the fictional domain by the dubitative text.

Thereafter, it is our intention to conduct new training by creating subsets of the starting dataset based on the structural types of the occurrences: in fact, one of the major difficulties when fine-tuning the models is related to the occurrences' uneven length. Our hope is that by refining the type of analysis, the model can improve its performance. Furthermore, in analyzing false negatives, we realized that some features seem particularly difficult to identify: (B.1) negations; (B.2) questions; (B.3) alternatives; (B.4) sentences that rephrase something stated earlier if taken in context, but once isolated possess no clear dubitative markers (cf. Appendix). We hope that the creation of training subsets in the dataset, based on the structural types of occurrences, can help us resolve these impasses in categorization.

## Acknowledgments

We first thank Paavo Van der Eecken (University of Antwerp) for following this research project from the very beginning, sharing his views on various conceptual junctures, and suggesting ingenious technical solutions. We are grateful to Valeria Cavalloro (Università per Stranieri di Siena) and Virginia Giustetto (Université de Genève) for their contribution as annotators. We lastly thank Emilien Schultz (Université de Paris) and Simon Gabay (Université de Genève) for constructive exchanges.

## References

- [1] W. Amamou. *How to Fine-Tune BERT Transformer with spaCy 3. A step-by-step guide on how to fine-tune BERT for NER*. 2021. URL: <https://towardsdatascience.com/how-to-fine-tune-bert-transformer-with-spacy-3-6a90bfe57647>.
- [2] L. Aroyo and C. Welty. "Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation". In: *AI Magazine* 36.1 (2015), pp. 15–24. DOI: 10.1609/aimag.v36i1.2564. URL: <https://ojs.aaai.org/index.php/aimagazine/article/view/2564>.

- [3] M. Barenghi. *Calvino*. Bologna: il Mulino, 2009.
- [4] M. Belpoliti. *L'occhio di Calvino*. Torino: Einaudi, 1997.
- [5] M. Bonhomme. "La figuralité comme événement de style : l'exemple de la métonymie". In: *Cahiers de Narratologie* 35 (2019). DOI: 10.4000/narratologie.9286.
- [6] S. Bozzola and C. D. Caprio. *Forme e figure della saggistica di Calvino*. Roma: Salerno Editrice, 2021.
- [7] I. Calvino. *Adam, One Afternoon*. London: Vintage books, 2000.
- [8] I. Calvino. *Mr. Palomar*. London: Vintage books, 1999.
- [9] I. Calvino. *Romanzi e racconti*. Vol. I. Milano: Mondadori, 1991.
- [10] I. Calvino. *Romanzi e racconti*. Vol. II. Milano: Mondadori, 1992.
- [11] I. Calvino. *Romanzi e racconti*. Vol. III. Milano: Mondadori, 1994.
- [12] I. Calvino. *Saggi*. Vol. II. Milano: Mondadori, 1995.
- [13] I. Calvino. *The Castle of Crossed Destinies*. London: Vintage books, 1998.
- [14] J. Cohen. "A coefficient of agreement for nominal scales". In: *Educational and Psychological Measurement* 20.1 (1960), pp. 37–46. DOI: 10.1177/001316446002000104.
- [15] N. Z. Da. "The Computational Case against Computational Literary Studies". In: *Critical Inquiry* 45.3 (2019), pp. 601–639. DOI: 10.1086/702594.
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. DOI: 10.48550/arxiv.1810.04805. URL: <https://arxiv.org/abs/1810.04805>.
- [17] J. Dobson. "Interpretable Outputs: Criteria for Machine Learning in the Humanities". In: *Dhq* 15.2 (2021). URL: <http://www.digitalhumanities.org/dhq/vol/15/2/000555/000555.html>.
- [18] S. Ercolino. *Il romanzo-saggio*. Milano: Bompiani, 2014.
- [19] S. C. Federico Bertoni and N. Rubbi. "I confini del saggio. Per un bilancio sui destini della forma saggistica". In: *Ticontre. Teoria Testo Traduzione* 9 (2018), pp. VII–xii.
- [20] A. Feltracco, E. Jezek, B. Magnini, and M. Stede. "LICO: A Lexicon of Italian Connectives". In: *Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it 2016)*. 2016. URL: <http://ceur-ws.org/Vol-1749/paper24.pdf>.
- [21] C. Fenoglio. "Euforia di un genere: il saggismo novecentesco". In: *Methodologica* 19.2 (2021). DOI: 10.6092/issn.1721-4777/10715.
- [22] G. Gagliardi. "Inter-Annotator Agreement in linguistica: una rassegna critica". In: *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*. 2018, pp. 206–212. URL: <http://ceur-ws.org/Vol-2253/paper28.pdf>.
- [23] M. D. Gesù. "Conversazione con Alfonso Berardinelli". In: *Il saggio critico. Spunti, proposte, riletture*. Palermo: Duepunti, 2007, pp. 20–21.
- [24] M. Graziano. *Oltre il romanzo. Racconto e pensiero in Musil e Svevo*. Roma: Carocci, 2013.

- [25] B. J. Herrmann, K. Dalen-Oskam, and C. Schöch. “Revisiting Style, a Key Concept in Literary Studies”. In: *Journal of Literary Theory* 9 (2015). DOI: 10.1515/jlt-2015-0003.
- [26] A. Kamsetty, K. Fricke, and R. Liaw. *Hyperparameter Optimization for Transformers: A guide*. 2020. URL: <https://medium.com/distributed-computing-with-ray/hyperparameter-optimization-for-transformers-a-guide-c4e32c6c989b>.
- [27] K. Krippendorff. *Content Analysis: An Introduction to Its Methodology*. 2nd. California: Sage Publication, 2004.
- [28] K. Krippendorff. “Reliability in Content Analysis: Some Common Misconceptions and Recommendations”. In: *Human Communication Research* 30.3 (2004), pp. 411–433. DOI: 10.1111/j.1468-2958.2004.tb00738.x.
- [29] H. Love. “Close but Not Deep: Literary Ethics and the Descriptive Turn”. In: *New Literary History* 41.2 (2010), pp. 371–391. URL: <https://www.jstor.org/stable/40983827>.
- [30] T. S. e. A. M. Manfred Stede. “Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives”. In: *Discours* 46 (2019). DOI: 10.4000/discours.10098.
- [31] L. Marchese. “È ancora possibile il romanzo-saggio?” In: *Ticontre. Teoria Testo Traduzione* 9 (2018), pp. 151–170. DOI: 10.1007/s10503-020-09523-1.
- [32] M. L. McHugh. “Interrater reliability: the kappa statistic”. In: *Biochem Med* 22.3 (2012), pp. 276–282. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>.
- [33] P. V. Mengaldo. “Aspetti della lingua di Calvino”. In: *La tradizione del Novecento. Terza Serie*. Torino: Einaudi, 1991, pp. 227–291.
- [34] F. Moretti. *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: Nottetempo, 2022.
- [35] J. Morris. *Does Model Size Matter? A Comparison of BERT and DistilBERT*. 2020. URL: <https://wandb.ai/jack-morris/david-vs-goliath/reports/Does-Model-Size-Matter-A-Comparison-of-BERT-and-DistilBERT--VmlldzoxMDUxNzU>.
- [36] T. Nguyen, D. Nguyen, and P. Rao. “Adaptive Name Entity Recognition under Highly Unbalanced Data”. In: *CoRR abs/2003.10296* (2020). arXiv: 2003.10296. URL: <https://arxiv.org/abs/2003.10296>.
- [37] D. Nozza, F. Bianchi, and D. Hovy. “What the [MASK]? Making Sense of Language-Specific BERT Models”. In: *CoRR abs/2003.02912* (2020). arXiv: 2003.02912. URL: <https://arxiv.org/abs/2003.02912>.
- [38] C. de Obaldia. *The Essayistic Spirit: literature, modern Criticism and the Essay*. Cotswolds: Clarendon Press, 1995.
- [39] N. Palmieri. “Il segno cancellato. Correzione e commento nella scrittura di Calvino”. In: *Studi sulla modernità* (1993), pp. 289–314.
- [40] T. Pavel. *The Lives of the Novel: A History*. Princeton: Princeton University Press, 2014.
- [41] C. Perelman and L. Olbrechts-Tyteca. *Trattato dell’argomentazione. La nuova retorica*. Torino: Einaudi, 1966.

- [42] P. Pietrandrea. *Epistemic modality. Functional properties and the italian system*. Amsterdam - Philadelphia: John Benjamins Publishing Company, 2005.
- [43] C. Plantin. “Un lieu pour les figures dans la théorie de l’argumentation”. In: *Argumentation et Analyse du Discours* (2009). DOI: 10.4000/aad.215.
- [44] L. V. Poppel. “The Study of Metaphor in Argumentation Theory”. In: *Argumentation* 35 (2021), pp. 177–208. DOI: 10.1007/s10503-020-09523-1.
- [45] A. Roque. “Towards a computational approach to literary text analysis”. In: *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Montréal, Canada: Association for Computational Linguistics, 2012, pp. 97–104. URL: <https://aclanthology.org/W12-2514>.
- [46] R. Ruiz-Dolz, J. Alemany, S. M. B. Heras, and A. García-Fornes. “Transformer-Based Models for Automatic Identification of Argument Relations: A Cross-Domain Evaluation”. In: *IEEE Intelligent Systems* 36.6 (2021), pp. 62–70. DOI: 10.1109/mis.2021.3073993.
- [47] F. Serra. *Calvino*. Roma: Salerno Editrice, 2006.
- [48] F. Serra. “La notte del morto nel paese nemico”. In: *Paragone* 61.90-91-92 (2010), pp. 125–134.
- [49] F. Serra, V. Cavalloro, V. Giustetto, M. Parigini, M. Mauri, T. Elli, Á. Briones, and B. Gobbo. *Corpus Atlante Calvino*. Yareta, 2022. DOI: 10.26037/yareta:4w2fzwd2j5atzzlhjcv4ntm.
- [50] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 2012. DOI: 10.1017/cbo9781139035613.
- [51] S. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. “Attention Is All You Need”. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.

## A. False Positives

Selection of examples to illustrate different types of false positives:

### A.1. Interpunction

S159: “(A pagarla anche a suono d’unità a peso d’oro [...] inevitabile.)”; S90: “Che cosa avviene [...] mondo?” (M2); S101: “- una faccia [...] faccia -” (M2); S142: “cosa intendo?” (M1, M2), “- perché intanto [...] la polpa -” (M1, M3, M2), “- impercettibilmente o smisuratamente diversa -” (M1), “- nella scelta che Priscilla fa di me” (M2); E006: “Ma a cosa ci porta il nostro continuo nervoso sfogliare giornali freschi d’inchiostro?” (M1, M2, M3), “Quale situazione migliore per avere un’idea complessiva del mondo?” (M1, M3); E011: “(questa suggestione [...] in poi)” (M1, M2, M3), “(almeno all’immaginazione)” (M1, M2, M3), “(non rare quando una coscienza empirica e spontanea matura prima di una coscienza riflessa e organizzata)” (M2, M3); E038: “Il poeta («la man che ci movea»)” (M1), “E l’arte?” (M1, M2, M3); E046: “(Popper dovrebbe esser contento)” (M1, M2), “(vera)” (M1, M2), “(A Venezia nel Seicento il Vestri disegna una carta delle correnti che ora le prospezioni via satellite compiute per determinare l’inquinamento della

Laguna confermano punto per punto)” (M1), “(E’ interessante [...] senza precedenti)” (M1, M2, M3), “(cioè da 1 a 86.400)” (M1, M2).

## A.2. Sentences with argumentative connectives

S007: “o tornare con la risposta” (M1), “quasi con uno sciacquo” (M1), “forse già al Carmo” (M1, M2, M3), “ma più si scava [...] sottile” (M2), “Ma era anche il suo compito [...] per il fieno” (M3), “eppure riconosceva il sentiero, le pietre, gli alberi, il muschio” (M3); S159: “o a metà” (M1), “o pensato” (M1), “o sentire” (M1), “oppure uno smorzato [...] operazione” (M3), “o ancora [...] del buio” (M3); S190: “o per meglio [...] a lui” (M3), “o autolesionista” (M3); S101: “quasi femminei” (M1), “ma quando poi [...] fortemente” (M2), “o troppo in fretta o troppo piano, senza libertà di movimenti, Amilcare doveva seguire la corrente o risalirla a fatica” (M1, M2, M3); S142: “cioè che di momento in momento io non sono più lo stesso io e Priscilla non è più la stessa Priscilla” (M1, M2, M3), “ma forse sollevarlo non [...] complicate” (M1, M3), “cioè per via di quel che si dice” (M1, M3); E011: “cioè a ribadire le proprie catene” (M1, M3), “quasi diremmo extrastorica, catastrofica” (M1, M2, M3), “ma anzi forse con la possibilità di capovolgere il rapporto tra i due termini” (M1, M2, M3), “anzi fa aggravare fino all’esplosione naturale” (M1, M2, M3); E019: “Ma questo non è che il primo gradino della grammatica e della sintassi narrativa;” (M2, M3); E044: “o una mescolanza cangiante di rosso bianco nero grigio che sull’etichetta porta un nome ancor più policromo” (M1, M2, M3), “forse in mezzo al deserto” (M1, M2, M3), “Eppure, chi ha avuto la costanza di portare avanti per anni questa raccolta sapeva quel che faceva, sapeva dove voleva arrivare” (M1, M2, M3).

## A.3. Cerebrality

S007: “Ma Binda ora pensava [...] sotterranei” (M1, M2); S159: “non so” (M2), “o pensato come in un delirio” (M2); S190: “Allora pensa: «Se io vedo e penso e nuoto [...] raggi” (M1), “A ben pensarci, una tale situazione non è nuova:” (M2, M3); S142: E019: “Vediamo di tentare un ragionamento opposto a quello che ho svolto finora: questo è sempre il sistema migliore per non restar prigioniero nella spirale dei propri pensieri” (M1); E038: “Talvolta io penso e immagino che tra gli uomini esiste una sola arte e scienza, e che questa sia il disegnare” (M1, M2, M3); E082: “Se pensiamo che questa perorazione per una vera fraternità universale è stata scritta quasi centocinquanta anni prima della Rivoluzione francese, vediamo come la lentezza della coscienza umana a uscire dal suo parochialismo antropocentrico può essere annullata in un istante dall’invenzione poetica” (M1, M2, M3), “al contrario penso che la razionalità più profonda implicita in ogni operazione letteraria vada cercata nelle necessità antropologiche a cui essa corrisponde.” (M1, M2, M3), “(penso naturalmente agli affascinanti studi di Francis Yates sulla filosofia occulta del Rinascimento e sui suoi echi nella letteratura)” (M2, M3).

## A.4. Figurality

S007: “A volte pareva a Binda...” (M1); “come una scimmia aggrappata al collo” (M2) “come enormi ragni sotterranei” (M1, M2); S159: “come in trance” (M2, M3) “come un’altra parte di me stesso cui corrispondono altre funzioni” (M2), “o pensato come in un delirio” (M2), “come un’altra parte di me stesso cui corrispondono altre funzioni” (M2), “come un organo della

mia persona” (M2); S101: “adesso magari faceva istintivamente per guardarle, ma subito gli **pareva** che scorressero via **come** un vento, senza dargli nessuna sensazione, e allora abbassava indifferente le palpebre” (M3), “adesso il poterle vedere [...] non più soltanto gli **pareva** un vederle ma già addirittura un possederle” (M2), “**come se** fosse la faccia tipica d’una categoria di persone a lui estranea” (M2), “**come** un rimorso” (M2), E006: “**come** un limbo innocente e funereo” (M2), “**come** se fossero stati rosi all’interno dalle termiti, appena gli s’avvicina la mano non ne resta che polvere” (M2), “ma ormai, anche su questo terreno, **pare** che non possa più crescere erba fresca” (M2, M3), E038: “Anziché il mondo **come** oggetto rappresentabile dall’arte e l’arte **come** rappresentazione del mondo, ci si apre un nuovo orizzonte in cui il mondo vissuto è visto **come** opera d’arte e l’arte propriamente detta **come** arte al secondo grado” (M1, M2, M3), “La linea **come** segno del movimento, **come** godimento del movimento, **come** paradosso del movimento” (M1), E082: “Non mi **pare** una forzatura connettere questa funzione sciamanica e stregonesca documentata dall’etnologia e dal folklore con l’immaginario letterario;” (M1, M3), “anzi, la leggerezza pensosa può far apparire la frivolezza **come** pesante e opaca.” (M2, M3), “una fuga di immagini, che è **come** un campionario delle bellezze del mondo” (M2).

## B. False Negatives

Selection of examples to illustrate different types of false negatives:

### B.1. Negations

S007: “**non** d’agrifoglio”, “**non** con acqua e rane”, “E arrivando **non** avrebbe...”; S159: “**non** è perché mi sia rimasto da dirti qualcosa d’indispensabile”, “**né** è la nostra intimità interrotta al momento della partenza che sono impaziente di ristabilire”; S190: “Tutto questo avviene **non** sul mare, **non** nel sole, – pensa il nuotatore Palomar”, “il sole **non** tramonta, il mare **non** ha quel colore, le forme sono quelle che la luce proietta nella retina”; S101: “(**Non** li metteva sempre [...] lontano)”, “**non** c’era dubbio [...] diverso”, “Ma **non** per i cambiamenti”, “ma un margine di dubbio che **non** fosse colui che credeva restava sempre”, “ma se veniva in qua adesso, **non** poteva esser lei che aveva fatto tutto il giro”; S142: “Andando avanti vedremo che **non** c’è niente di fatto apposta, che nessuno ha messo lì niente”, “quindi **non** si sa quanti io ci siano a monte dell’io che credo di essere io, e quante Priscilla a monte della Priscilla verso la quale io sto credendo di stare correndo”

### B.2. Questions

S159: “Sarà stato già tradotto in comandi ai selettori [...] centrali di transito?”; S190: “«E quello il solo dato non illusorio, comune a tutti, il buio?» si domanda il signor Palomar”, “dove finirebbe la spada?”, “Un principio unico e assoluto da cui prendono origine gli atti e le forme?”; S101: “Possibile che non l’avesse riconosciuto?”, “Ma com’era possibile scambiare Isa Maria per Gigina?”

### B.3. Alternatives

S007: “o l’avrebbe sentito a un rotolar di pietre che si metteva al suo fianco, a camminare assieme a lui in silenzio.”; S159: “**oppure** rivelarsi inaspettatamente attiva senz’aver dato prima alcun segno di vita”, “o quale altra città sia la tua”; S190: “depressivo o autolesionista”; S101: “o sono tutto”, “o non possono essere più niente”, “o li si segue giorno per giorno”, “**oppure** non si riesce più a entrarci”, “se era stato visto o no”; S142: “alla specie o all’ambiente o a noi due”, “o già vicini una volta per tutte”, “fusione [o mescolanza] o scambio”, “l’addensarsi in sciame di cellule-semi o il concentrato maturare di cellule-uova”, “impercettibilmente o smisuratamente”, “o è avvenuto o avverrà”

### B.4. Sentences without clear dubitative markers

S007: “Tanti lumi diversi, potevano essere, in marcia per tutti i sentieri di Tumena bassa”, “era il fischio convenuto dei tedeschi che serravano intorno a lui, ecco che un altro fischio gli rispondeva, era circondato!”; S190: “«È un omaggio speciale che il sole fa a me personalmente», è tentato di pensare il signor Palomar”, “«Tutti quelli che hanno occhi vedono il riflesso che li segue; l’illusione dei sensi e della mente ci tiene sempre tutti prigionieri»”, “«E quello il solo dato non illusorio, comune a tutti, il buio?» si domanda il signor Palomar”, “Tutto questo avviene non sul mare, non nel sole, – pensa il nuotatore Palomar”; S142: “il nostro patrimonio genetico, tra virgolette”, “E dicendo forma intendo tanto quella che si vede quanto quella che non si vede”, “il rapporto tra i soli elementi differenziali, perché quelli comuni si possono trascurare da una parte e dall’altra”, “e allora bisogna vedere se si tratta di quelli comuni”, “la somma dei caratteri dominanti del passato, il risultato d’una serie d’operazioni che davano sempre un numero maggiore di zero”, “Tutto quel che possiamo dire è che in certi punti e momenti quell’intervallo di vuoto che è la nostra presenza individuale viene sfiorata dall’onda che continua a rinnovare le combinazioni di molecole”