How to Annotate Emotions in Historical Italian Novels: a Case Study on *I Promessi Sposi*

Rachele Sprugnoli, Arianna Redaelli
Università di Parma, Viale D'Azeglio, 85, 43125 Parma, Italy
{rachele.sprugnoli, arianna.redaelli}@unipr.it

Abstract

This paper describes the annotation of a chapter taken from *I Promessi Sposi*, the most famous Italian novel of the 19th century written by Alessandro Manzoni, following 3 emotion classifications. The aim of this methodological paper is to understand: i) how the annotation procedure changes depending on the granularity of the classification, ii) how the different granularities impact the inter-annotator agreement, iii) which granularity allows good coverage of emotions, iv) if the chosen classifications are missing emotions that are important for historical literary texts. The opinion of non-experts is integrated in the present study through an online questionnaire. In addition, preliminary experiments are carried out using the new dataset as a test set to evaluate the performances of different approaches for emotion polarity detection and emotion classification respectively. Annotated data are released both as aggregated gold standard and with non-aggregated labels (that is labels before reconciliation between annotators) so to align with the perspectivist approach, that is an established practice in the Humanities and, more recently, also in NLP.

Keywords: annotation, emotion analysis, historical texts, Italian literature, 19th century Italian

1. Introduction

Emotion analysis is a task at the intersection of Natural Language Processing (NLP) and Affective Computing whose aim is to automatically recognize the emotions conveyed in a text. It is important to note that the concept of emotion is notoriously difficult to define (Scherer, 1984); for the purposes of this paper, we will use the word "emotion" as an umbrella term to encompass various affective states including all kinds of feelings, moods, attitudes, and behavioral responses.

Applications, domains and text genres considered in the emotion analysis task are extremely varied (Acheampong et al., 2020) and the organization of specific evaluation exercises in various languages demonstrates the growing interest of the NLP community towards the analysis of emotions (Mohammad et al., 2018; Plaza-del Arco et al., 2021; Araque et al., 2023). In this context, literary texts are less studied in NLP than, for example, social media posts but, on the contrary, the relationship between emotions and literary texts is of enormous interest in the field of Digital Humanities especially after the so-called affective-turn in literary studies (Keen, 2011). Therefore, emotion analysis is a task where a collaboration between the two communities can be extremely fruitful and beneficial for both. This paper[1] presents an example of such collaboration by describing the sentence-level emotion annotation of a chapter from a 19th century novel (for a total of 338 sentences and more than 9,000 tokens) according to 3 distinct classifications. The purpose of this paper is mostly methodological; instead of aiming for a large amount of data, in this phase we want to study in depth: i) how the annotation changes depending on the granularity of the classification, ii) how the different granularities impact the inter-annotator agreement, iii) which granularity allows good coverage of emotions, and iv) if the chosen classifications are missing emotions that are important for a literary text of the 19th century. To achieve these goals, a questionnaire was also created involving 45 anonymous non-experts.

The data of our study are from the final edition (1840-1842) of Alessandro Manzoni's *I Promessi Sposi* (*The Betrothed*). This novel is fundamental to both the history of Italian literature and the development of the Italian language, as it introduced a functional model of written literary language that closely mirrored common speech and was widely imitated by Italian authors, scholars and learners. Following the unification of Italy in 1861, the novel emerged as a symbol of national identity, and its prominence was particularly felt in the educational sector, where it was swiftly incorporated into the literary canon. This not only strengthened its status as a cornerstone of Italian literature but also positioned it as a practical model from which to learn Italian language and even derive grammatical norms to be taught in schools. However, over the years, this educational emphasis cast the novel in a somewhat gray, heavy, and static light for many students. This per-

[1] This paper is the result of the collaboration between the two authors. For the specific concerns of the Italian academic attribution system: Rachele Sprugnoli is responsible for Sections 2, 4, 5 and 6; Arianna Redaelli is responsible for Sections 1 and 3. Section 7 was collaboratively written by both authors.

LT4HALA 2024@LREC-COLING-2024, pages 105–115

ception stands in stark contrast to the novel's true nature, which is dynamic and original. Furthermore, Manzoni's meticulous exploration of the language of passions, underscored by a moral perspective (Maiolini et al., 2017), ensures the novel's emotional depth and variety. Such qualities, together with the intricate narrative and well-rounded characters, far from melodramatic stereotypes, not only affirm its status as a literary masterpiece, but also highlight its suitability for emotion analysis. In turn, emotion analysis can even serve as a mean to re-emphasize the novel's positive features, potentially revitalizing its perception in education and encouraging renewed appreciation among students.
From the data availability standpoint, the text of *I Promessi Sposi* is free from copyright, fully digitized, and available in a machine-readable and clean (that is without OCR errors) format. This format ensures seamless integration with computational tools with minimal manual intervention.
To sum up, our main contributions are as follows: i) an in-depth study on the annotation of emotions in an Italian historical literary text that, despite its critical significance to Italian literary history, has not previously been examined through NLP methods; ii) the development of a new dataset manually annotated with 3 emotion classifications of different granularity that is released with both aggregated and non-aggregated annotations; iii) the release of a new polarity lexicon derived from 19th-century Italian narrative texts.[2]

2. Related Work

Over the last few years, numerous datasets for emotion analysis have been developed following two main approaches. The first approach is based on the idea that emotions are innate, universal and limited in number, thus they can be classified using categorical labels, often borrowed from psychological theories, such as those of Ekman (Ekman, 1992) and Plutchik (Plutchik, 1980). On the contrary, in the second approach, emotions are represented by combining a small set of dimensions using continuous values. For example, Russell and Mehrabian (1974) identify valence (degree of pleasantness), arousal (degree of excitement) and dominance (degree to which a person feels in control of a situation) as the three fundamental dimensions for defining all emotions. From this theory derives the so-called VAD (Valence-Arousal-Dominance) model which serves as the foundation for both lexicons and annotated datasets, see among others (Buechel and Hahn, 2017; Mohammad, 2018). Both approaches have advantages and disadvantages: categorical classifications are intuitive to understand but use culture- and language-specific labels that are not actually universal, while dimensional models can describe feelings that would otherwise be difficult to label but are harder to interpret by humans. Therefore there are studies that aim not only to analyze the two approaches but also to unify them (Calvo and Mac Kim, 2013; Bostan and Klinger, 2018). However, in our work we have decided to adopt a discrete classification for its ease of interpretation because, as anticipated in Section 1, our aim is to create a resource easily accessible even to non-experts, humanities scholars and students first and foremost.
The main issue when dealing with the categorical approach is the choice of the classification to adopt. Together with works that borrow Ekman's 6 emotions[3] or Plutchik's 8 basic emotions[4], usually adding a label for neutral cases (Alm et al., 2005; Schuff et al., 2017; Öhman et al., 2020), there are also datasets that employ a much narrower or much broader set. For example, Grounded-Emotions is annotated only with `sadness` and `happiness` (Liu et al., 2017), while FEEL-IT with `anger`, `fear`, `sadness` and `joy` (Bianchi et al., 2021). On the contrary, the dataset of SemEval-2018 Task "Affect in Tweets" uses 11 emotions[5] (Mohammad et al., 2018) and Demszky et al. (Demszky et al., 2020) propose a taxonomy of 27 categories plus `neutral` (see Section 3.2 for the complete list). Although the various classifications are often applied to texts that are very different from each other (e.g., posts on social media, song lyrics, transcriptions of dialogues) in an indistinct manner, some works instead focus on how to find the most suitable taxonomy for the textual genre to be annotated. This is particularly important for literary texts where emotions tend to be complex, subtle and intertwined with narrative, aesthetic and cultural aspects. For example, for the annotation of historical German plays different annotation schemes have been tested (Schmidt et al., 2018), and then 13 hierarchically structured emotion concepts have been defined (Schmidt et al., 2021). On the other hand, in the Kāvi corpus, Punjabi poems are annotated following the concept of Navrasa, that distinguishes nine emotions, such as "shaanti" (meaning *peace*) and "raudra" (meaning *anger*), in order to better reflect Indian culture (Saini and Kaur, 2020). The survey papers by Kim and Klinger (2019) and Reb-

[2] All data presented in this paper are available in a GitHub repository: `https://github.com/RacheleSprugnoli/Emotion_Analysis_Manzoni`

[3] `Anger, disgust, fear, joy, sadness, surprise.`

[4] `Anger, disgust, fear, joy, sadness, surprise, trust, anticipation.`

[5] `Anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust.`

ora (2023), to which we refer for further details, well describe the broad and multifaceted panorama of emotion and sentiment analysis applications in the field of computational literary studies.
In the present work we decided not to uncritically adopt one classification but to try different taxonomies to identify the one that best suits our case study and that can be potentially applicable to other Italian novels as well. Furthermore, the annotated data produced in this work enriches the inventory of linguistic resources for emotion analysis available for Italian which, although always growing, is not as abundant as for other languages. Notable examples of recent Italian datasets in the field of emotion analysis are: FEEL-IT (tweets annotated with 4 emotions, see above), the EMit dataset (Araque et al., 2023) (tweets annotated with Plutchik's basic emotions plus `love` and `neutral`), MultiEmotions-it (Sprugnoli, 2020) (comments posted on YouTube and Facebook annotated with both Plutchik's basic and complex emotions) and AriEmozione (Zhang et al., 2022) (opera verses annotated with 6 emotions, namely, `love`, `joy`, `admiration`, `anger`, `sadness` and `fear`).

3. Data and Annotation

This Section describes the data used in our annotation and the workflow we followed giving details on the selected chapter and on the emotion taxonomies adopted.

3.1. Data Selection

Among the 38 chapters of the novel, chapter VIII appeared to be the most suitable one to start the annotation. Indeed, this chapter is particularly noteworthy for its structure, consisting of 5 macro-sequences: the failed marriage attempt in the house of the priest Don Abbondio, the failed kidnapping of Lucia (the female protagonist) by the *bravi* (hired assassins), the gathering of the crowd outside Don Abbondio's house at the tolling of the bell, the meeting of the bethroted and Lucia's mother (Agnese) with Fra Cristoforo (a monk) in a church, and the abandonment of the hometown. Given the profound diversity of the aforementioned themes, chapter VIII also shows a wide range of scenes and tones (moving between the extremes of Don Abbondio's sympathetic opening line and Lucia's final weeping), and a great stylistic-narrative variety (shifting from dialogue to vivid description, and finally to the lyrical depth of the *Addio ai Monti* [*Farewell to the mountains*]). Additionally, the many events of the chapter involve more than 15 characters, each one distinctly marked by his own linguistic features, gestures, and emotional states. Furthermore, chapter VIII is one of the longest in the novel (9.808 tokens, including punctuation) which allowed us to have a good amount of textual material to annotate.
For all these reasons, chapter VIII not only offers a microcosm of the novel's intricate emotional and linguistic features but also provides a comprehensive and varied dataset for emotion analysis. By focusing on this chapter, we were allowed to obtain a condensed and yet diverse representation of the emotional dynamics that permeate the entire novel of *I Promessi Sposi*.

negative	0.78
neutral	0.76
positive	0.57
mixed	0.46
overall	0.73

sadness	0.79
fear	0.75
anger	0.73
surprise	0.72
joy	0.69
neutral	0.57
anticipation	0.53
trust	0.53
disgust	0.44
overall	0.53

Table 1: Inter-annotator agreement in terms of Krippendorff's Alpha for emotion polarity annotation (on the left) and for the annotation of Plutchik's basic emotions (on the right).

3.2. Annotation Workflow

The annotation was carried out using a simple spreadsheet with a sentence per line in their original order.[6] Sentence splitting was performed manually because the automatic segmentation proved to be very challenging for the models currently available for Italian due to issues related to the novel's complex punctuation. For example, the text contains low quotation marks («») indicating direct speech spoken aloud, while the long dash (–) is used to delimit thought or muttered direct speech. These punctuation marks, and their so specific and diverse use, are not common in contemporary texts, thus systems are not trained to recognize them correctly. For example, an accuracy of 64% was registered with Stanza (Qi et al., 2020). At the end of the manual sentence splitting procedure, we obtained 338 sentences of different length (from 1 to 109 tokens).
Two annotators were involved in the annotation: one with a significant expertise in Manzoni's work but limited annotation experience, and the other being an experienced annotator with basic knowledge of Manzoni. The first 20 sentences were annotated collaboratively, while the remaining sentences were

[6] By "sentence" we mean a coherent set of words that conveys a complete thought and ends with a strong punctuation mark (e.g., full stop, question mark, or exclamation point), typically followed by a capital letter.

love	0.86	remorse	0.66	annoyance	0.42
curiosity	0.83	optimism	0.66	admiration	0.33
sadness	0.75	nervousness	0.65	relief	0.30
gratitude	0.74	embarrassment	0.59	caring	0.29
fear	0.73	joy	0.57	disappointment	0.15
anger	0.71	disapproval	0.55	approval	NEG
neutral	0.71	surprise	0.45	desire	NEG
disgust	0.67	confusion	0.42	realization	NEG
				overall	0.44

Table 2: Inter-annotator agreement in terms of Krippendorff's Alpha for the annotation using GoEmotions classification.

annotated independently by each annotator following 3 types of emotion classification.

The first classification takes into consideration the polarity of the emotions conveyed by the text. More specifically, emotion polarity is categorized into 4 classes: i) `positive` (meaning that positive emotions are clearly prevalent in the sentence), ii) `negative` (which means that negative emotions are clearly prevalent in the sentence), iii) `mixed` (which indicates that opposite emotions are expressed in the sentence and it is not possible to find a clearly prevailing emotion polarity), iv) `neutral` (to be used when no emotions are expressed in the sentence). This coarse-grained taxonomy requires a single-label annotation while the other two adopted classifications allow a multi-label annotation being Plutchik's basic emotions and the taxonomy proposed for the GoEmotions dataset (Demszky et al., 2020). The first consists of 8 labels (namely, `anger`, `fear`, `sadness`, `disgust`, `surprise`, `anticipation`, `trust`, and `joy`) plus `neutral`, whereas the second is made of 27 distinct emotion categories (`admiration`, `amusement`, `anger`, `annoyance`, `approval`, `caring`, `confusion`, `curiosity`, `desire`, `disappointment`, `disapproval`, `disgust`, `embarrassment`, `excitement`, `fear`, `gratitude`, `grief`, `joy`, `love`, `nervousness`, `optimism`, `pride`, `realization`, `relief`, `remorse`, `sadness`, and `surprise`) plus `neutral`.

The guidelines prescribed, for all 3 annotation types, to: i) evaluate both the lexicon used and the images evoked (for example through the use of rhetorical figures) in the sentence, ii) focus on the emotions expressed by the author, either directly (as the narrator present in the story) or indirectly (through the characters), and not on those perceived by the reader; iii) take into consideration the flow of the narrative also considering the previous sentences but not the ones that follow. Subsequently, for each classification, the individual labels were explained; for example, for the GoEmotions taxonomy the brief descriptions reported in the corresponding paper were taken (Demszky et al., 2020).

neutral	166	neutral	133
negative	129	anticipation	75
mixed	22	fear	68
positive	21	anger	52
		surprise	29
		sadness	25
		trust	24
		joy	11
		disgust	5

Table 3: Number of annotated labels after reconciliation: emotion polarity on the left and Plutchik's basic emotions on the right.

4. Data Analysis

This section presents details on the inter-annotator agreement (IAA) and on the dataset obtained after the reconciliation of disagreements.

4.1. Inter-Annotator Agreement

Tables 1 and 2 report the results of the IAA in terms of Krippendorff's alpha for each label and for each classification together with the overall score. Labels are ranked in descending order of agreement. The overall scores show a substantial agreement for emotion polarity annotation (0.73) and a moderate agreement for both the annotation of Plutchik's basic emotions (0.53) and the GoEmotions classification (0.44). Given the well-known high subjectivity of emotion annotation and the multi-label nature of two of the three used classifications, these results can be considered promising.

The IAA on single labels varies greatly: such wide variability is common in emotion annotation, as attested in several previous works, for example (Strapparava and Mihalcea, 2008; Schuff et al., 2017). In the emotion polarity annotation, the `negative` and `neutral` classes proved to be the easiest to annotate (0.78 and 0.76, respectively), followed by `positive` (0.57), whereas `mixed` was the most problematic (0.46). Although difficult to recognize, we think that the `mixed` class is important because it captures the complexity of the literary text. Eliminating that class would impoverish the annotation making it less interesting for humanities scholars. Among the Plutchik's basic emotions, the highest scores were achieved with three negative emotions (`sadness`, `fear`, `anger`) however, even in this case, the agreement is between substantial and moderate for all the labels. Moreover, 64% of the sentences have both the annotators agreeing on at least one emotion label. As for the GoEmotions taxonomy, 17 labels out of 24 have at least a moderate agreement but `approval`, `desire` and `realization` registered slightly negative values (-0,004, -0,007 and -0,007 respectively) indicating an inverse agreement, less than that expected by chance. Indeed, these 3 classes had been misinterpreted by an annotator who had never used them. However, in general, we note that the values are on average higher than those reported for the original English dataset. In addition, 81% of the sentences have the annotators agreeing on at least one emotion label.

4.2. Annotated Data after Consolidation

Disagreements were discussed and consolidated to obtain gold labels. Our consensus-building efforts was primarily centered on enhancing the annotation methodology itself, enabling us to adjust our guidelines and labels for clearer future annotations. Using Plutchik's and GoEmotions classifications, most of the sentences resulted with a single emotion label (77% for the former and 64% for the latter, respectively), followed by sentences with 2 labels (22% and 33%, respectively) while 3 emotions are a strong minority (1,5% and 3%, respectively).
Tables 3 and 4 present the number of labels for each classification after the reconciliation in descending order. The `neutral` class is always the most frequent: it makes up 49% of all the labels in the emotion polarity annotation, 31% in Plutchik's classification and 24% in the GoEmotions annotation. The fact that the number of neutral sentences is not constant is due to the greater annotation granularity allowed by the Plutchick's and GoEmotions classifications. Having much more detailed labels available, led annotators to be able to better specify emotional nuances, recognizing them more easily. In particular, what is annotated as `neutral` in the first classification is instead marked with an ambiguous emotion (namely, `surprise` and `anticipation` following the Plutchick's distinction, `realization`, `surprise`, `curiosity`, `confusion` in GoEmotions) in the others. For example, the first sentence of the chapter (the exclamation of a proper name), i.e., "– Carneade!" (EN: - *Carneades!*), is annotated as `neutral`, `surprise`, `surprise` respectively. Often, a sentence marked as `neutral` in the emotion polarity annotation is marked as `anticipation` following the Plutchik's annotation and as `nervousness` following GoEmotions taxonomy: this last label makes explicit the anxiety that underlies the expectation of an event disambiguating an ambiguous emotion. An example is given by the sentence "Entraron pian piano, in punta di piedi, rattenendo il respiro; e si nascosero dietro i due fratelli."[7]
Apart from the `neutral` class, there is a large disparity in terms of label frequency. Although a similar disparity is also present in the GoEmotions dataset, a very different distribution of emotions is noted due to the different nature of the texts considered. In fact, the most frequent labels in the English GoEmotions data are `admiration` and `approval` whereas in Manzoni's chapter negative and ambiguous emotions prevail. It is interesting to note that the strong presence of negative emotions in our data is also attested in other literary datasets, such as (Zhang et al., 2022) and (Schmidt et al., 2021), regardless the annotation scheme used.
To better understand the relationship between emotions across the three types of annotation, we calculated the correlation between emotion polarities and the classes of Plutchik and GoEmotions. More specifically, we converted emotion labels into their corresponding polarity value leaving out ambiguous emotions. For example, `anger` and `embarrassment` were mapped onto the `negative` class, whereas `joy` and `approval` onto the `positive` one. Annotations made of opposite emotions (as the third sentence in Table 7) were converted into the `mixed` class. We found a strong positive correlation both between the emotion polarity annotation and the Plutchik's classification (0.70) and between the emotion polarity annotation and the GoEmotions classification (0.76).

5. Preliminary Experiments

The small size of the dataset did not allow it to be used to train new models but was instead adopted as a test set. In particular, we tried two approaches for polarity detection:

- Lexicon-based: a score is computed for each sentence by summing the polarity values of the tokens as recorded in a polarity lexicon (see below for more details). `Positive` and `negative` labels are assigned to sentences with a score above or below zero, respectively. Instead, we assign the `neutral` label to sentences in which all words have a score of 0 and the `mixed` label when the positive and negative values balance each other resulting in a sum of 0.

[7] EN: *They came in slowly slowly, on tiptoe, holding their breath, and hid behind the two brothers.* (Manzoni, 2022)

neutral	111		caring	21		gratitude	5
nervousness	74		annoyance	15		remorse	5
fear	42		relief	13		joy	4
curiosity	33		sadness	13		love	4
disapproval	25		optimism	10		approval	3
anger	23		disappointment	7		admiration	2
surprise	23		desire	5		disgust	2
confusion	22		embarrassment	5		realization	2

Table 4: Number of annotated labels after reconciliation for the annotation using GoEmotions classification.

LEXICON-BASED: W-MAL					LEXICON-BASED: XIX Cent.					CROSS-LANGUAGE SYSTEM			
	P	R	F1			P	R	F1			P	R	F1
pos	0.08	0.76	0.14		pos	0.17	0.67	0.27		pos	0.22	0.67	0.33
neg	0.57	0.54	0.56		neg	0.67	0.56	0.61		neg	0.65	0.37	0.47
neu	0.85	0.07	0.12		neu	0.78	0.55	0.64		neu	0.63	0.75	0.68
mix	0.00	0.00	0.00		mix	0.22	0.32	0.26		mix	0	0	0
avg	**0.37**	**0.34**	**0.21**		**avg**	**0.46**	**0.52**	**0.45**		**avg**	**0.37**	**0.45**	**0.37**

Table 5: Results of emotion polarity detection in terms of precision (P), recall (R) and F1-measure (F1).

- Cross-lingual model: a zero-shot cross-language system (Sprugnoli et al., 2023) that classifies emotion polarity into the same 4 classes used in our annotation, trained on an English dataset of social media texts and fine-tuned on XLM-RoBERTa (Conneau et al., 2020).

As for emotion classification, we tested two off-the-shelves models:

- FEEL-IT (Bianchi et al., 2021): a monolingual emotion classification system, trained on Italian tweets, that identifies 4 emotions (`fear`, `joy`, `sadness`, `anger`). We evaluated this tool only on the 73 sentences annotated with these emotions.

- XLM-EMO (Bianchi et al., 2022): a multilingual emotion classification system, fine-tuned on XLM-RoBERTa, that identifies the same emotions as FEEL-IT. Also in this case, only 73 sentences were used for the evaluation being them annotated with `fear`, `joy`, `sadness` or `anger`.

For emotion polarity detection, the lexicon-based approach relied on two polarity lexicons. The first one, W-MAL (Vassallo et al., 2020), is based on contemporary Italian whereas the second was developed to be more representative of the lexical characteristics of 19th century Italian. For this reason, we downloaded[8] the narrative texts published in the period of interest (including *I Promessi Sposi*), listed the tokens in order of frequency and assigned a polarity value (i.e. -1 for negative polarity, +1 for positive polarity and 0 for neutral cases) to all the tokens with a frequency higher or equal to 5. The final lexicon is made of 18,885 entries with a strong majority of neutral tokens (69.1% of the total) and more negative entries (19.5% of the total) than positive ones (11.4% of the total). The IAA calculated on a randomly chosen subgroup consisting of 10% of the entries was substantial (Cohen's kappa = 0.76).

As reported in Table 5, the lexicon-based approach using this new lexicon achieved the best F1 (0.45, weighted macro-average F1 0.58, accuracy 54) and it is the only method capable of identifying sentences with mixed polarity, even if only 7 times out of 22. Performances on the `neutral` and `negative` classes are good but, on the contrary, they are low on `positive`. A similar pattern is registered for the cross-lingual model,[9] whereas with the W-MAL lexicon a good F1 is achieved only for the `negative` class.

As for emotion classification, Table 6 shows that the multilingual model performed better than the monolingual one obtaining a F1 of 0.47 (weighted macro-average F1 0.50, accuracy 0.49). However, precision and recall are non balanced, with the latter being higher than the former. For FEEL-IT the lowest performance was on `fear`, which is the least frequent emotion in the training corpus and the most difficult to recognize even in the experiments carried out by the system developers. Instead, in the case of XLM-EMO the lowest F1 was registered for `joy` for which the recall is perfect but the precision is very low.

These results confirm the need to create adequate

[8]http://www.bibliotecaitaliana.it/.

[9]Please note that these results are worse than those that the same system obtained both on Italian social media texts and on Opera verses written in 18th-century Italian.

FEEL-IT			
	P	R	F1
anger	0.62	0.56	0.59
fear	0.38	0.11	0.17
sadness	0.29	0.67	0.41
joy	0.14	0.33	0.20
macro avg	**0.36**	**0.42**	**0.34**

XLM-EMO			
	P	R	F1
anger	0.54	0.52	0.53
fear	0.71	0.36	0.48
sadness	0.50	0.60	0.55
joy	0.20	1.00	0.33
macro avg	**0.49**	**0.62**	**0.47**

Table 6: Results of emotion classification in terms of precision (P), recall (R) and F1-measure (F1).

resources for the development of new models suitable for the processing of historical literary texts.

6. Emotion Annotation Elicitation

An additional study involved non-experts through an online questionnaire (made with Google Form) circulated on social networks (namely, LinkedIn, Mastodon and X). We selected 21 sentences taken from chapter VIII (i.e., the same text annotated by experts). These sentences belong to three textual passages chosen for their structural and emotional differences in order to present a good variability without, however, making the questionnaire too long (consequently reducing the risk of non-completion by the participants). The first group of sentences describes the final agitated phases of the failed attempt at marriage between Renzo and Lucia; the second is a sequence of short direct speeches between the crowd who rushed to help Don Abbondio and the priest himself, who regretted having raised the alarm; the third passage reports Lucia's thoughts while, on board a boat, she sadly says goodbye to her beloved homeland. Instructions were as follows.[10] "*Your task is to tell us which emotions you think are expressed in each sentence. For each sentence you can report one or more emotions; we won't give you a list of emotions to choose from, but you can express yourself freely. The sentences are taken from chapter VIII of The Betrothed by Alessandro Manzoni (1840). Read one sentence at a time and indicate the emotions that you think are expressed and/or felt by the narrator or the characters. ATTENTION: not what you feel when reading the sentence. If you want to list multiple emotions, separate them with a comma; if you can't express the emotion with a single word, also describe it with a sentence or a phrase; if it seems to you that the text does not express any emotion, write NO.*" Under the instructions, the groups of sentences were presented in distinct sections so as to make it clear that they were separate units. We also collected some socio-demographic information: namely, age (i.e., under 18, between 18 and 29, between 30 and 50, over 60), self-perceived gender identity (i.e., male, female, other, I prefer not to specify) and level of education (i.e., high school diploma, bachelor's degree, master's degree, Phd).

In one week we collected 45 responses. In general, the most mentioned emotions for each sentence correspond to those identified by the experts (see Table 7)[11] but, not having given a predefined list of labels, we recorded a great lexical richness with the use of numerous synonyms and plesionyms. For example, *spavento* (fright), *timore* (dread), *angoscia* (anguish), *panico* (panic), *terrore* (terror), *orrore* (horror), *allarme* (alarm), *sgomento* (dismay) can be traced back to the `fear` label, while `anger` is expressed also with words such as *furia* (fury), *collera* (wrath), *ira* (rage), *odio* (hate), *aggressività* (aggression). This observation prompted us to enhance the guidelines by incorporating lists of synonyms into the descriptions of emotions, thereby clarifying that each label encompasses a range of emotional shades.

The analysis of the responses also highlighted the recurring emergence of some emotions, such as `resignation`, not present in the classifications used by experts; adding such labels could make the annotation more precise but their adoption must be carefully evaluated to avoid that the increase in labels leads to a decrease in agreement.

Finally, the responses were analyzed from the point of view of the socio-demographic characteristics of the participants. In particular, we studied the propensity to assign more than one emotion per sentence based on differences in age, gender and education level. The only statistically significant difference detected (with alpha = 0.05) is the one between males and females, with the latter indicating more emotions per sentence than the former.

[10]The original instructions were written in Italian.

[11]Translations of sentences in Table 7, taken from (Manzoni, 2022): i) *Having dropped the lamp he'd been holding, he used that hand to gag her with the cloth, almost suffocating her. And all the while he kept shouting at the top of his lungs, "Perpetua! Perpetua! Treachery! Help!"*; ii)*And saying this, he stepped back and closed the window once more.*; iii) *Farewell childhood home, where lost in private thoughts, she had learned to hear the difference between normal footsteps and the footsteps of the youth she awaited with a mysterious fear.*

Sentence	Polarity	Basic	GoEmotions	Questionnaire
E subito, lasciata cader la lucerna che teneva nell'altra mano, s'aiutò anche con quella a imbacuccarla col tappeto, che quasi la soffogava; e intanto gridava quanto n'aveva in canna: «Perpetua! Perpetua! tradimento! aiuto!»	negative	anger,fear	anger,confusion,fear	fear,anger
E, detto questo, si ritirò, e chiuse la finestra.	neutral	neutral	neutral	no
Addio, casa natìa, dove, sedendo, con un pensiero occulto, s'imparò a distinguere dal rumore de' passi comuni il rumore d'un passo aspettato con un misterioso timore.	mixed	sadness	love,sadness	nostalgia,sadness

Table 7: Examples taken from our data after reconciliation; the last column presents the two most mentioned emotions in the questionnaire. Please note that the answers to the questionnaire are translated into English from the original Italian. Sentence translation is provided in footnote 11.

7. Discussion and Conclusion

This paper describes, from a methodological point of view, the annotation of emotions in a chapter taken from *I Promessi Sposi*, the most famous Italian novel of the 19th century written by Alessandro Manzoni. The annotation was based on 3 different classifications with the final goal of finding the best taxonomy for a historical literary text, balancing the richness of the recognized emotional states and the feasibility of the annotation. During their work, annotators could add any suggestions or doubt in an ad-hoc field. Other useful suggestions came from a questionnaire, aimed at non-experts, that helped us improve the guidelines. The following issues emerge from the comments by both annotators and non-expert. Regarding emotion polarity, annotators felt the lack of a label to indicate sentences with ambiguous emotions. On the other hand, Plutchik's emotions were not always considered suitable because they were too generic: indeed, often the annotator chose an emotion going solely by exclusion (a repeated comment was "it seems to me that none of the other options are suitable"). Finally, the lack of a label to indicate resignation is reported when annotating with the GoEmotions taxonomy: this need was declared also by non-experts (see Section 6). An additional suggestion was to introduce a specific level of annotation for irony. This proposed layer aims to address the subtle use of irony in Manzoni's writing, a topic extensively analyzed by literary critics (see (Raimondi, 1990; Mancini, 2005) among others), and its correlation with the emotional features of the text. The feasibility of integrating this layer and its impact on inter-annotator agreement are subjects for further investigation.

In addition, preliminary experiments were carried out using the new dataset as a test set to evaluate off-the-shelves tools for emotion polarity detection and emotion classification respectively. In this context we developed a new lexicon created by assigning a polarity value to almost 19,000 tokens taken from 19th-century Italian narrative texts. The outcomes of the experiments underscore the necessity for developing systems tailored to process historical literary texts, which have very specific linguistic features.

Going back to the aims of the work listed in Section 1, we can summarize the results obtained by our study as follows. The presence of more detailed labels leads to a wider recognition of the different emotional nuances and a reduction in the number of neutral sentences. As expected, a greater granularity of the classifications is accompanied by a lower agreement: however, the majority of emotions have an IAA between substantial and moderate. The 27 classes of the GoEmotions taxonomy seem to be suitable for representing the complexity of the literary text but it is necessary to add a label to express resignation, and to refine the guidelines by adding more information to each emotion description (for example, providing a list of synonyms and plesionyms).

We release the annotated data both with aggregated and non-aggregated labels. The annotation elicited from the questionnaire are also available. Offering more than one perspective in identifying emotions allows this study to be in line with established practices in the Humanities. For example, a central assumption of contemporary literary theory is that facts, values, reason, and nature are constructs, not objective and immutable realities (Fischer, 1990). According to this theory, literary texts are not static entities but are open to multiple interpretations, each shaped by the unique perspective of the reader or critic. This notion suggests that a single text can offer a multitude of readings, each valid in its own right, and emphasizes the importance of understanding literature as a dynamic interplay between text and reader. This approach aligns perfectly with the perspectivist turn in the field of NLP (Cabitza et al., 2023), which we see as an interesting topic for future collaboration between NLP and DH scholars.

Future work extends in at least three directions. First, we want to expand the annotation to other chapters so that we have more data and can run new experiments and train new models. Secondly, we plan to apply the same type of annotation to other Italian novels to verify the degree of generalization of the proposed approach. Another interesting future study concerns the annotation of the emotions as elicited in the reader that would allow us to have two complementary points of view on the same text. This double approach (writer- and reader- oriented) is particularly suitable for literary texts, as demonstrated by recent reader response studies (Rebora, 2023; Pianzola et al., 2020), and further highlights the need to include multiple perspectives in the computational analysis of emotions.

8. Acknowledgements

Questa pubblicazione è stata realizzata da ricercatrice con contratto di ricerca cofinanziato dall'Unione europea - PON Ricerca e Innovazione 2014-2020 ai sensi dell'art. 24, comma 3, lett. a, della Legge 30 dicembre 2010, n. 240 e s.m.i. e del D.M. 10 agosto 2021 n. 1062. Questa ricerca è stata anche finanziata dall'Università degli Studi di Parma attraverso l'azione Bando di Ateneo 2022 per la ricerca co-finanziata dal MUR-Ministero dell'Università e della Ricerca - D.M. 737/2021 - PNR - PNRR - NextGenerationEU.

9. References

Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. 2020. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7):e12189.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Oscar Araque, Simona Frenda, Rachele Sprugnoli, Debora Nozza, and Viviana Patti. 2023. EMit at EVALITA 2023: Overview of the Categorical Emotion Detection in Italian Social Media Task. In *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023)*, pages 37–44, Parma, Italy. AILC - Associazione Italiana di Linguistica Computazionale, Accademia University Press.

Federico Bianchi, Debora Nozza, and Dirk Hovy. 2021. FEEL-IT: Emotion and sentiment classification for the Italian language. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 76–83, Online. Association for Computational Linguistics.

Federico Bianchi, Debora Nozza, and Dirk Hovy. 2022. XLM-EMO: Multilingual emotion prediction in social media text. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 195–203, Dublin, Ireland. Association for Computational Linguistics.

Laura-Ana-Maria Bostan and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain. Association for Computational Linguistics.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. volume 37, pages 6860–6868.

Rafael A. Calvo and Sunghwan Mac Kim. 2013. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, online. Association for Computational Linguistics.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th*

Annual Meeting of the Association for Computational Linguistics, pages 4040–4054, Online. Association for Computational Linguistics.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Michael Fischer. 1990. Perspectivism and literary theory today. *American Literary History*, 2(3):528–549.

Suzanne Keen. 2011. Introduction: Narrative and the Emotions. *Poetics Today*, 32(1):1–53.

Evgeny Kim and Roman Klinger. 2019. A survey on sentiment and emotion analysis for computational literary studies. Wolfenbüttel.

Vicki Liu, Carmen Banea, and Rada Mihalcea. 2017. Grounded emotions. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 477–483. IEEE.

Elena Maiolini et al. 2017. *Manzoni. Il linguaggio delle passioni*, volume 68. Franco Cesati.

Massimiliano Mancini. 2005. *Il romanzo dell'ironia. I Promessi Sposi nella critica*. Vecchiarelli.

Alessandro Manzoni. 2022. *The Betrothed: A Novel*. Modern Library. Translated by Michael F. Moore.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia. Association for Computational Linguistics.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. Association for Computational Linguistics.

Emily Öhman, Marc Pàmies, Kaisla Kajava, and Jörg Tiedemann. 2020. XED: A multilingual dataset for sentiment analysis and emotion detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6542–6552, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Federico Pianzola, Simone Rebora, and Gerhard Lauer. 2020. Wattpad as a resource for literary studies. quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. *PloS one*, 15(1):e0226708.

Flor Miriam Plaza-del Arco, Salud M Jiménez Zafra, Arturo Montejo Ráez, M. Dolores Molina González, Luis Alfonso Ureña López, and María Teresa Martín Valdivia. 2021. Overview of the EmoEvalEs task on emotion detection for Spanish at IberLEF 2021. Sociedad Española para el Procesamiento del Lenguaje Natural.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Ezio Raimondi. 1990. *La dissimulazione romanzesca: antropologia manzoniana*. Il Mulino.

Simone Rebora. 2023. Sentiment analysis in literary studies. a critical survey. *DHQ: Digital Humanities Quarterly*, 17(2).

James A Russell and Albert Mehrabian. 1974. Distinguishing anger and anxiety in terms of emotional response factors. *Journal of consulting and clinical psychology*, 42(1):79.

Jatinderkumar R Saini and Jasleen Kaur. 2020. Kāvi: An annotated corpus of punjabi poetry with emotion detection based on 'navrasa'. *Procedia Computer Science*, 167:1220–1229.

Klaus R Scherer. 1984. Emotion as a multicomponent process: A model and some cross-cultural data. Sage Publications, Inc.

Thomas Schmidt, Manuel Burghardt, and Katrin Dennerlein. 2018. Sentiment annotation of historic german plays: An empirical study on annotation behavior. In Sandra Kübler and Heike Zinsmeister, editors, *Proceedings of the Workshop on Annotation in Digital Humanities 2018 (annDH 2018)*, pages 47–52. Sofia, Bulgaria.

Thomas Schmidt, Katrin Dennerlein, and Christian Wolff. 2021. Towards a corpus of historical german plays with emotion annotations. In *3rd Conference on Language, Data and Knowledge (LDK 2021)*, pages 1–11. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger. 2017. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. In *Proceedings of the 8th Workshop on*

Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 13–23, Copenhagen, Denmark. Association for Computational Linguistics.

Rachele Sprugnoli. 2020. Multiemotions-it: A new dataset for opinion polarity and emotion analysis for italian. In *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020)*, pages 402–408. Accademia University Press.

Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti, and Giovanni Moretti. 2023. The Sentiment of Latin Poetry. Annotation and Automatic Analysis of the Odes of Horace. *IJCoL. Italian Journal of Computational Linguistics*, 9(9-1):53–71.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560.

Marco Vassallo, Giuliano Gabrieli, Valerio Basile, Cristina Bosco, et al. 2020. Polarity imbalance in lexicon-based sentiment analysis. In *Proceedings of the Seventh Italian Conference on Computational Linguistics*, pages 1–7. CEUR.

Shibingfeng Zhang, Francesco Fernicola, Federico Garcea, Paolo Bonora, and Alberto Barrón-Cedeño. 2022. Ariemozione 2.0: Identifying emotions in opera verses and arias. *IJCoL. Italian Journal of Computational Linguistics*, 8(8-2):7–26.