RAGE: Roman And Greek Emotions

Frederick Riemenschneider, Jonathan Geiger, Thomas Kuhn-Treichel, Anette Frank
Department of Computational Linguistics Department of Classical Philology
Heidelberg University, Germany
{riemenschneider, frank}@cl.uni-heidelberg.de

Abstract

The study of emotions in ancient Greek and Latin literature has largely been qualitative, relying on close reading, while existing computational methods often focus on coarse-grained sentiment polarity, which limits their use for nuanced literary analysis. To bridge this gap, we present RAGE (Roman And Greek Emotions), a new corpus of approximately 100 000 words of annotated classical literature spanning multiple genres and authors. Our multi-layered annotation framework, inspired by semantic role labeling, is designed for fine-grained analysis, capturing not only the emotion itself but also its `experiencer`, `cause`, and `target`. We adopt a nuanced emotion taxonomy and enrich each emotion instance with additional layers for intensity, explicitness, and negation. To facilitate comparative analysis, characters are linked to Wikidata or a local ontology. We demonstrate the utility of our corpus through corpus-level exploratory analyses and an in-depth case study. RAGE and its accompanying guidelines provide a valuable resource for applying quantitative methods to the study of emotions in classical texts.

Keywords: Emotion Analysis, Classical Literature, Ancient Greek, Latin, Distant Reading, Semantic Roles

1. Introduction

Μῆνιν ("wrath") is the first word of Homer's *Iliad*: the driving force behind the epic with which the history of ancient Greek and Latin literature begins. Following the "emotional turn" in the humanities, the study of emotions has become a vibrant field within classical philology, yet this work remains largely qualitative, relying on close readings of individual texts. At the same time, computational methods for emotion analysis are maturing in NLP (Plaza-del Arco et al., 2024), but their application to the nuanced landscapes of classical literature remains in its early stages. Initial efforts in this area have focused on coarse-grained sentiment polarity (Sprugnoli et al., 2024). While valuable for their intended scope, our work is motivated by different research goals that require a more fine-grained analysis.

Guided by the perspective of literary scholars, we present RAGE (Roman And Greek Emotions), a new corpus built on a multi-layered annotation framework designed to capture the complexity of literary emotion. Our approach draws inspiration from the semantic roles of emotions proposed by Kim and Klinger (2018), identifying not only *what specific emotion* is felt, but also *who feels it*, *what caused it*, and *towards whom it is directed*. We adapt and extend this model for the complexities of classical literature. Importantly, we move beyond the relatively coarse taxonomies of Ekman and Friesen (1971) or Plutchik (1982), adopting the more nuanced 27-category set from GoEmotions (Demszky et al., 2020). In addition, we enrich each emotion instance with multiple descriptive layers, including its *intensity, explicitness, negation*, and whether it is *merely hypothetical* or *realized*. Finally, to enable large-scale character analysis, we link each annotated entity to Wikidata[1] where possible, or to a project-specific ontology, making it possible to systematically compare the emotional profile of a character, such as Medea, across different works.

Grounded in this rich annotation scheme, the aims and contributions of our paper are as follows:

(i) We introduce RAGE, a new resource of approximately 100 000 words of emotion-annotated classical literature. This undertaking represents a significant increase in both scale and annotation complexity compared to existing corpora in this domain, spanning multiple authors, genres, and historical periods.

(ii) We present a multi-layered annotation scheme tailored to emotions in literary texts, which captures `cues`, `experiencers`, `causes`, `targets`, emotion qualifiers (intensity, explicitness, negation, realization) and entity linking, and release guidelines to enable reuse for similar projects on historical languages.

(iii) We demonstrate the utility of our corpus with several corpus-level exploratory analyses and an in-depth case study that showcase the variety of fine-grained literary analyses it enables.

2. Related Work

2.1. Emotions in Classical Philology

The "emotional turn" of the late 1990s sparked a wave of research into emotions across the humanities, including cultural and literary studies; several handbooks document this trend (Koppenfels and

[1] https://www.wikidata.org/.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 947–957
11–16 May 2026.

Zumbusch, 2016; Kappelhoff et al., 2019; Pritzker et al., 2020; Schnell, 2021; Scarantino, 2025). In recent years, the interest has also reached classical literature. Numerous monographs and edited volumes on emotions in Greek and Latin literature and philosophy have been published, including such programmatic titles as *Emotions in the Classical World* (Cairns and Nelis, 2017), *A Cultural History of the Emotions in Antiquity* (Cairns, 2019) or *Emotions and Narrative in Ancient Literature and Beyond* (Bakker et al., 2022). Most of these works rely on close readings of individual texts. By contrast, quantitative approaches, which allow for comparison between larger bodies of text, are still rare.

2.2. Emotions in NLP

The computational analysis of emotions is a wide-ranging field in NLP, with applications from opinion analysis to psychotherapy (Na et al., 2025). A noted challenge within this field is the lack of standardized terminology (Plaza-del Arco et al., 2024). The present study focuses on the application of emotion analysis to a corpus of classical literature.

Our approach is informed by several key works in the subfield of emotion analysis for literary texts. The primary inspiration for our annotation scheme is the work of Kim and Klinger (2018), who propose a method for annotating the semantic roles of emotions: `cue`, `experiencer`, `cause`, and `target`. Our focus on tracking the emotions of individual characters is further motivated by the findings of Vishnubhotla et al. (2024). Their research shows that distinguishing between the emotional arcs of the narration and those of specific characters provides a more accurate analysis, which supports our character-centric methodology.

The selection of an emotion taxonomy is a key methodological decision. Many studies rely on a small set of basic emotions, e.g., six emotions proposed by Ekman and Friesen (1971), or the set of eight emotions established by Plutchik (1982). We have adopted the more fine-grained taxonomy from GoEmotions, which provides 27 emotion categories plus a neutral class (Demszky et al., 2020). For a comprehensive overview of this field, we refer to a recent survey by Plaza-del Arco et al. (2024).

2.3. Emotions at the Intersection of NLP and Classical Philology

Several recent efforts document a growing interest in the automated analysis of emotions in classical literature, with work spanning both Latin and Greek.

Sprugnoli et al. (2022) present a first analysis of sentiment in Latin poetry, which examines Horace's *Odes* by having annotators label sentences as `positive`, `negative`, `neutral`, or `mixed`. Similarly, Pavlopoulos et al. (2022) perform sentiment classification on the first book of the *Iliad*. For their main analysis, they use the same four categories, but also include a subtask where annotators could specify emotions in a free-text field. The inclusion of "Emotion Polarity Detection" in the EvaLatin 2024 Shared Task (Sprugnoli et al., 2024) further highlights the community's interest in this area. Moving beyond sentiment polarity, Picca and Pavlopoulos (2024) analyze ten books of the *Iliad* using an eight-category emotion model.

While these approaches are valuable for gauging overall sentiment and basic emotions, they are not designed to answer more granular literary questions. For many analyses, expanding from a `positive/negative` polarity to a richer emotional vocabulary is essential. For instance, a key goal of our project is to enable systematic comparisons of characters, such as Medea's `anger` in contrast to her `sadness` across different works. Such a comparison requires not only a finer-grained emotion set but also knowing who feels the emotion, what caused it and towards whom or what it is directed. Datasets focused on sentence-level polarity do not provide this structured, fine-grained information. Our work introduces an annotation framework that is suited to characterize distinct emotions with their associated `experiencers`, `causes`, and `targets` as well as qualifiers of emotion presence, intensity and subtlety.

3. From Research Questions to Design Principles

The design of our annotation scheme is motivated from the perspective of a literary scholar. A central goal of computational literary studies is to enable new forms of inquiry that are difficult to conduct through traditional close reading alone. For instance, a scholar might wish to study how the emotional profile of a shared character like Medea differs across the Greek and Latin literary traditions, or even more specifically, between the portrayals in Euripides and Seneca. One might ask: Are there systematic differences in the emotions experienced by gods versus those experienced by humans within a specific author's work? Do male and female characters exhibit distinct patterns of emotions? How is emotional agency distributed?

Answering such questions imposes a clear set of requirements on an annotation framework. Hence, we derive the following core design principles for emotion annotation from our scholarly desiderata:

1. It must be character-centric and contextual. To track emotional agency or compare characters, the framework must move beyond simple sentence-level classification. It must identify not only the

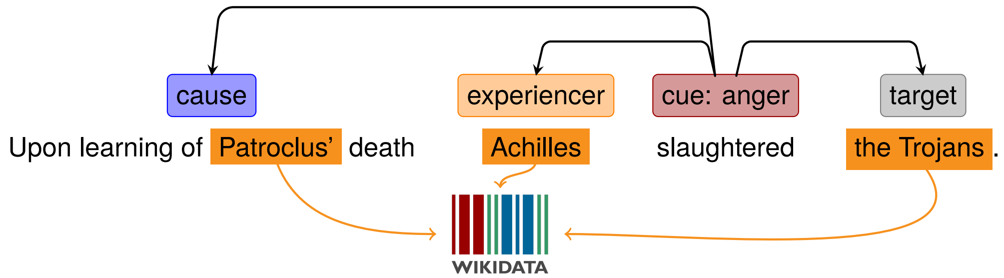


Figure 1: Annotation of an emotion with its roles and entities.

character `experiencing` an emotion but also the surrounding context, including the `cause` of the emotion and its `target`.

2. It must support a fine-grained emotional vocabulary. A simple `positive/negative` polarity is too rough a signal for literary analysis. The scheme must support an expressive range of distinct emotions to allow for meaningful comparisons, such as distinguishing a character's `sadness` from their `anger`.

3. It must distinguish presence from mention. Emotions in literature are not always directly experienced. To ensure analytical accuracy, the framework must be able to differentiate a present, felt emotion from one that is negated ("was not angry"), hypothetical ("if he were angry..."), or mentioned as an abstract concept ("anger is a powerful force").

4. It must account for subtlety and intensity. We need to distinguish an emotion's internal intensity from the subtlety of its textual expression. These are not necessarily correlated: a character may feel an *intense* emotion even when the textual cues remain highly *subtle*. Methodologically, this separation enables multi-faceted analysis: The annotation of overt emotions establishes a "common ground" where high agreement can be achieved, yielding a subset of the data suitable for "more objective" analyses. Conversely, annotations of subtle emotions, while more subjective, allow us to access the literary subtext which can reveal much of a character's psychology.

5. It must enable entity-driven comparisons. Comparing groups like "gods" vs. "humans" or "men" vs. "women" requires structured character information. The scheme should therefore include a mechanism to identify characters as entities and link them to an ontology that contains this categorical information.

4. Annotation Scheme

Our annotation scheme is tailored to the design principles presented above and was developed iteratively: we, the project leaders, conducted a pilot study to test and refine the label sets and guidelines before commencing the main annotation effort.

To allow for a character-centric, contextualized annotation, we begin by identifying the `cue`: the minimal span of text that signals an emotion. The `cue` is then linked to any recoverable roles – the `experiencer`, the `target`, and the `cause` – but none of these roles are required, and any may be implicit. Specific distinctions (e.g., negation or realization) are captured by attributes described below. Our scheme differs from Kim and Klinger (2018) insofar as we assign the roles (`cue`, `experiencer`, `target`, `cause`) directly to the spans rather than to the relations between spans. Emotion and role annotations are visualized in fig. 1.

The pilot annotation confirmed that a smaller set of 6 to 8 emotions (cf. section 2.2) was insufficient for our aims. Thus, we decided to rely on the emotion taxonomy presented by Demszky et al. (2020), which features 27 emotions (cf. fig. 2). Annotators pick one (or more) emotion(s) for each `cue`.

To distinguish an emotion's *presence* from its mere *mention*, we annotate its realization status using two binary flags. The first, `emotion is negated`, marks emotions that are explicitly negated. The second, `emotion is realized`, differentiates actually experienced emotions (`yes`) from unrealized ones, such as those that are hypothetical or mentioned as an abstract concept (`no`).

Two further checkboxes allow annotators to rate emotions for their *intensity* and *subtlety*: `emotion is explicit` is used to distinguish cases in which the cue includes an emotion term like "fear" or "afraid" from those in which the emotion is expressed in a more subtle way. As for an emotion's `intensity`, annotators can choose from `intense`, `moderate` and `mild`.

Two final checkboxes are necessary for cases in which it is not possible to mark an experiencer in the text: `experiencer is narrator` when the emotion is felt by the work's narrator, `experi-`

encer is speaker when it is felt by the speaker of a direct speech. This enables us to link cues to named entities even if the experiencer is not verbally represented in the text.

To support comparative analyses, we use a second annotation layer in which all animate entities within text spans identified in the initial annotation phase (including those not mentioned by name but only through a pronoun or a verb form) are linked to Wikidata.

5. Constructing RAGE

5.1. Corpus Composition

For our corpus, we selected a representative and balanced sample of texts for comparative analysis, which we structured along *three primary axes*: language (Greek and Latin), genre, and historical period, creating two language partitions of nearly equal size, each containing approximately $50\,000$ words. The selected texts span more than a millennium of literary history, from Homer to late antique authors like Procopius and Claudian, and cover four genres (Epic, Drama, Oratory, Historiography).

Within this balanced structure, we deliberately selected texts that create opportunities for controlled comparison. For instance, a researcher can trace the Medea narrative across Greek epic (Apollonius of Rhodes) and tragedy (Euripides), and then compare it directly with its Latin counterparts. To complement this, we also selected passages that appeared particularly promising for emotion analysis, aiming to create a dataset with a high density of relevant phenomena. The texts were sourced from the Perseus Digital Library (Crane). The full composition of the corpus is detailed in table 1.

5.2. Annotation Process

Annotation of the RAGE corpus comprised two distinct phases: a primary phase for detailed emotion annotation, and a subsequent, more rapid phase for entity identification and linking. This work was carried out by a team of 17 members with domain expertise in classical philology, primarily at the advanced bachelor's and master's level, along with three senior academics. All annotation tasks were conducted within the web-based annotation platform INCEpTION (Klie et al.). To facilitate rapid comprehension and ensure accurate interpretation, annotators were presented with the original Greek or Latin text alongside a modern translation.

Annotator Training and Quality Control. Our annotation protocol included several measures to ensure annotation quality and consistency. Annotators were paid at a fixed rate to encourage thoroughness over speed. For tasks involving multiple annotators, we regularly shuffled pairings to mitigate the development of team-specific bias. All annotators participated in an interactive introductory session where we presented the annotation guidelines, demonstrated the functionality of INCEpTION and discussed the project goals.

While the annotation guidelines were the main reference, we set up a group chat for quick, day-to-day questions and held regular group meetings to discuss difficult passages and clarify rules. We met weekly at the start of the project, and less often as annotators became more familiar with the task.

Phase 1: Emotion Annotation. The most substantial project phase was the detailed annotation of emotions. For this task, each text was assigned to two annotators who worked independently. After this independent annotation phase, each pair of annotators met for a curation session to produce a single, adjudicated version. We retained all three versions for each text – the two independent annotations and the curated one – for subsequent analysis.

Phase 2: Entity Identification and Linking. The second phase focused on identifying and linking entities to an ontology. This task was performed by a single annotator per text and was limited to the spans previously marked for emotion. This constrained scope made the process considerably faster and more straightforward than the emotion annotation.

Our approach here differs significantly from standard Named Entity Recognition, as we aim to identify all entities involved in an emotional event, to analyze their roles. This requires that we annotate any textual evidence of their presence, not only names ("Medea") but also pronouns ("she"). Crucially for pro-drop languages like Greek and Latin, we also annotate the inflected verb itself if an emotion role is left implicit. For example, in a phrase like *timet* ("she is afraid"), the verb *timet* is annotated as the textual anchor for the implicit experiencer.

To streamline this work, we used Anthropic's Claude Sonnet (v3.7; upgraded to v4.0 when available)[2] to generate recommendations for entities and their identifiers. The annotator reviewed these suggestions, accepting, correcting, rejecting or adding entities that the system missed. Whenever possible, entities were linked to their Wikidata ID. For characters or concepts not present in Wikidata, we created an entry in a local ontology.

[2] https://www.anthropic.com/.

	Greek Author & Text	#words	Latin Author & Text	#words
Epic	Homer, *Iliad* 3	3249	Virgil, *Aeneid* 4	4555
	Homer, *Odyssey* 5	3623		
	Apollonius of Rhodes, *Argonautica* 3.1–575	3942	Valerius Flaccus, *Argonautica* 7	4439
	Nonnus, *Dionysiaca* 1	3230	Claudian, *De raptu Proserpinae* 1	1765
Drama	Euripides, *Medea*	8043	Seneca, *Medea*	5525
	Menander, *Dyskolos*	6337	Terence, *Adelphoe*	8168
Oratory	Lysias, *Oratio* 3	2178	Cicero, *Philippicae* 8	2994
	Lucian, *Abdicatus*	4816	Apuleius, *Apologia* 90–102	3020
Historiography	Herodotus, *Histories* 1.1–55	6450	Livy, *Ab urbe condita* 1.7–31	7386
	Procopius, *Secret History* 4–10	8099	Sallust, *Catilina*	10 667
Σ		49 967		48 519

Table 1: Selected Greek and Latin texts with their word counts.

6. Data Statistics and Annotation Evaluation

General Statistics. In total, combining the work of both annotators and the final curation, we collected 21 219 `cue` annotations, 10 247 for `experiencers`, 9593 for `targets`, and 3232 for `causes`. The curated data contains 8661 `cue`, 4198 `experiencer`, 4184 `target`, and 1385 `cause` instances. Figure 2 presents a breakdown of the emotion distribution and associated qualifiers. The statistics confirm the utility of our rich emotion set, as all emotions were applied with meaningful frequency.

Evaluation of Roles. We evaluate the agreement on the identified role spans (`cue`, `experiencer`, `target`, `cause`) between independent annotators in table 2, using F_1-scores for both "exact match," which requires identical character spans, and "overlap match," which counts any overlap between spans as a match. Since the roles `experiencer`, `target`, and `cause` are dependent on a `cue`, we calculate their agreement conditioned on matched `cues`. This ensures we are only comparing roles that belong to the same annotated event.

Agreement for emotion `cues` reaches an F_1-score of 49.46 for exact matches. For dependent roles evaluated on these matched `cues`, agreement is highest for `experiencer` (56.98), followed by `target` (41.90), and is lowest for `cause` (25.16). These results are within the expected range for such a subjective annotation task and hold up well against similar efforts, such as the work by Kim and Klinger (2018), though we note that our setup is not directly comparable.

Evaluation of Emotions. In table 3, we assess the consistency of annotators assigning specific emotion labels to the matched `cue` spans. Just as with dependent roles, this comparison is only meaningful if the annotations refer to the same event; thus, we compute agreement conditioned on matched `cues`. Since our scheme permits multiple

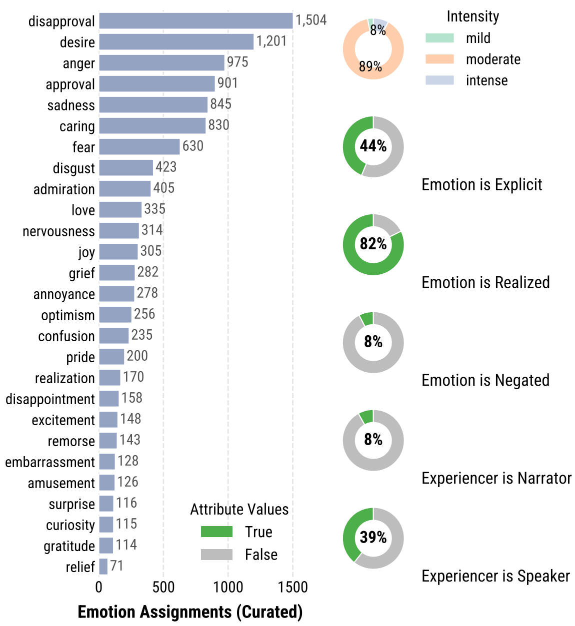


Figure 2: Emotion frequencies and attribute distributions of the curated annotations.

	Exact		Overlap	
	F_1	Support	F_1	Support
`cue`	49.46	6713	57.20	6713
`experiencer`	56.98	1700	60.82	1883
`target`	41.90	1198	50.46	1373
`cause`	25.16	497	35.16	547

Table 2: Annotator–annotator agreement (A–B) on role spans. "Exact": identical spans; "Overlap": any positive span overlap. Support = TP+FN (reference count; here, Annotator B). `experiencer`, `target`, `cause`: evaluated within matched cues.

emotion labels for a single `cue`, we treat the labels as a set and report several agreement metrics.

On the set of exactly matched `cues`, the exact set F_1-score is 48.89. This score rises substantially under more lenient conditions: Jaccard similarity is 61.66, relaxed set F_1, which requires only one

	Exact	Overlap
Exact Set	48.89	46.19
Jaccard	61.66	59.15
Relaxed Set	76.28	74.00
N	2976	3442
Single-Label Acc.	77.10	74.90
Single-Label N	1751	1964

Table 3: Emotion agreement conditioned on `cue` matches (A–B). Since annotators may assign multiple emotions to the same `cue`, we treat each `cue's` emotions as a set and compute set-based agreement. “Exact Set” requires set equality; “Jaccard” measures set overlap; “Relaxed Set” requires any shared label. “Exact” and “Overlap” refer to the `cue`-matching tier. “N” denotes number of matched cue pairs; “Single-Label Acc.” is computed on pairs where A and B chose exactly one label.

shared label between annotators, reaches 76.28. For cases where both annotators chose exactly one emotion, single-label accuracy is 77.10. These results indicate that while perfect agreement on the complete set of nuanced emotions is challenging, annotators show strong consistency in identifying the primary emotion of an event. The main source of disagreement appears to be the inclusion of secondary emotions. This trend also holds for the larger set of overlap-matched cues, and may turn out to be a feature rather than a weakness.

Discussion. Emotion annotation is an inherently subjective task, and it is widely reported in the literature that achieving high inter-annotator agreement can be difficult (Du and Hoste, 2025). Our moderate span agreement reflects a phenomenon also observed by Kim and Klinger (2018): even if annotators agree on an emotion taking place, they may annotate different spans due to different foci. Our annotators also reported this phenomenon.

Despite this, we see comparatively high emotion agreement between annotators. We partially attribute this to the 27 concrete emotion categories, which seem to guide annotators to appropriate classes, rather than forcing them to compose a feeling from broader options. At the same time, this granularity can lead to divergence for related emotions (e.g., `grief` vs. `sadness`).

While this higher level of detail inevitably introduces more dimensions for potential disagreement, it also makes the resulting resource flexible and can reveal subjective interpretations. Researchers can still adopt a coarser view, by, e.g., grouping the emotions into broader categories.

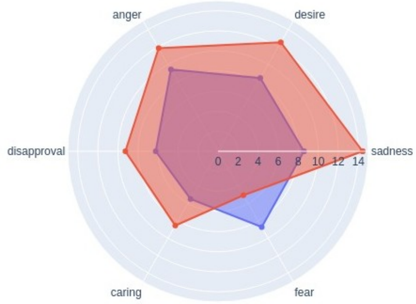


Figure 3: Percentage of the most frequent emotions in tragedy (red) and epic (blue).

7. Analyses and Interpretations

The RAGE corpus enables a “distant reading” (Moretti, 2013) approach that can enrich literary interpretation with a quantitative perspective in both research and teaching. On a broad scale, it allows for comparative inquiries across languages, genres, and character groups, revealing high-level patterns that can provoke new research questions. On a more focused scale, it supports nuanced comparisons of individual works and characters. We illustrate these applications through corpus-level exploratory analyses and a focused case study using the curated annotations.

7.1. Corpus-Level Exploratory Analyses

In this section, we show examples of how one can exploit our corpus design, which allows us to make useful comparisons, even if the corpus is not large enough to disregard specifics of each work it contains.

Emotion Frequencies. Genres may have different emotional profiles: As shown in fig. 3, the tragedies in our collection are disproportionately dominated by a few emotions that account for the majority of annotations. While the six most frequent emotions in epic and tragedy are the same, they occupy different ranks. `Fear` and `anger` are much more prominent in epic poetry, as they lend themselves more to external conflicts than `sadness` and `desire`, which in turn seem more fitting for inner conflicts and are thus more typical of tragedy.

Cue Density. How often emotions occur in a given text or in a portion of our corpus can be measured by a simple criterion: the number of `cue` spans per 100 words (without stop word filtering) gives some insights.

That tragedies are richer in emotions than epics seems plausible, even though the observed difference is starker than may be expected; partly this

Latin	Greek	Tragedy	Epic
6.8	10.3	14.4	6.2

Table 4: Frequency of `cues` per 100 words.

is due to the specific difference between Euripides and Seneca (section 7.2).

Interestingly, Greek texts – according to our data – contain considerably more emotions overall.The reason for this is not obvious and could be an interesting starting point for a more qualitative approach.

Explicitness. The difference in cue density may be related to the explicitness of emotions. In Greek texts, the annotators marked 32.8% of the emotions as explicit while in Latin texts the value is 59.5% – the ratio of explicit and implicit emotions is basically reversed. Thus, the *explicit* cue density in Latin is actually higher than in Greek (4 `cues` vs. 3 `cues` in 100 words). Does this point to a Roman literary style that is just less subtle? Or does the Greek language, with its richer morphology and vocabulary, lend itself more to implicit cues? Again, there is no obvious answer, but the numbers are bound to provoke the curiosity of classical scholars.

Emotional Similarity. An effective way to compare the emotion profiles of different works – specific works, authors, genres or languages –, is to calculate an "emotional similarity score." For this, we use the frequency of each emotion as one dimension in a vector to calculate pairwise cosine similarity between works.

It seems particularly interesting to look at works featuring Medea (table 5), as differences in the distribution of emotions are less likely to be caused by the plot. The analysis reveals that proximity in language, genre, and time period makes for a greater similarity score: The similarity between the two Latin authors, who are closer to each other in time, is higher than the one between the two Greek authors; the pairs with different language *and* genre have the lowest similarity.

Note also that the two tragedies are more similar to each other than the two epics, which is to be expected as they share a common plot, whereas the epics only share a common theme.[3]

Emotions in Relation to Entity Groups. Using the entity annotation allows us to analyze the roles of characters within a work or across works. By linking entity mentions to Wikidata, it is possible to aggregate not only characters, but also different categories they belong to. For example, one can investigate the difference in emotions and roles between humans and gods, as listed in table 6.[4]

	Greek		**Latin**	
	Tragedy 5th c. BCE Euripides	**Epic 3rd c. BCE Ap. Rhod.**	**Tragedy 1st c. CE Seneca**	**Epic 1st c. CE Val. Fl.**
Euripides	1	0.85	0.85	0.81
Ap. Rhod.		1	0.78	0.82
Seneca			1	0.95
Val. Fl.				1

Table 5: Emotion similarities among Medea works.

Em. `experienced` by **Humans** and **Gods**			**Em.** `targeted` at **Humans** and **Gods**		
disapproval	7.17	2.08	disapproval	20.06	12.73
desire	11.12	13.54	desire	9.58	5.45
anger	9.19	13.54	anger	8.67	5.45
caring	6.77	10.42	caring	11.09	9.09
approval	6.29	3.12	approval	9.88	10.91
fear	7.74	6.25	fear	3.33	7.27
admiration	3.22	5.21	admiration	6.45	16.36
gratitude	1.77	1.04	gratitude	2.02	5.45
disappointment	1.29	5.21	disappointment	1.21	3.64
embarrassment	1.45	3.12	embarrassment	0.20	1.82

Table 6: Select emotions of entity categories.

The emotion distribution suggests that gods experience a range of emotions not too dissimilar to people. We do, however, discern a few differences of note that point to gods being characterized as emotionally involved protagonists rather than detached spectators of human affairs:

Both `anger` and `caring` are considerably more frequently experienced by gods than humans, while gods exhibit less `disapproval` *and* `approval`. Gods appear to feel slightly more `desire` and `joy`, considerably more `disappointment`, `embarrassment`, and even `admiration`.

As to emotions felt towards gods, the data aligns with their distinct status: They are only rarely the objects of `disapproval`, `desire`, or `anger`, but are strongly associated with `admiration`, and, to a lesser degree, with `fear`, `gratitude`, `annoyance`, and `disappointment`.

We can cautiously conclude that within our corpus, deities are characterized as being authority figures to the outside but rather human on the inside.

7.2. Case Study: Medea Between Sadness and Anger

Apart from global evaluations, our corpus enables comparisons between individual works. It includes

[3] Overall, pairwise similarities within the Medea subset are relatively high compared to those across the full corpus; Procopius and Valerius Flaccus, e.g., only have a score of 0.40. The median score of pairwise similarities across our corpus is 0.78.

[4] Selected as "human" were those entities whose Wikidata property `P31` contained the word "character," "human," or "profession," as "gods" those entities whose `P31` category contained the word "deity" or "god."

(parts of) four texts dedicated to the mythological characters Medea and Jason, in two literary genres, with one Greek and one Latin representative each (see table 5). We chose the two tragedies for a case study, since they are fully annotated and represent the same series of events. Thus, any difference in the annotations can reflect different literary characteristics of the works and their authors – rather than just different plots. Indeed, our study reveals striking differences. First, cue density is much higher in Euripides' *Medea* (19.6 `cues` per 100 words) than in Seneca (6.7).

This is surprising, given that both works are usually considered to be steeped in emotions, and can therefore fuel the discussion about the peculiarities of the two plays. Here, different factors may play a role: Euripides' *Medea* contains significantly more changes of speaker, resulting in a higher frequency of references to different characters' emotions, even if the individual speakers are not more emotional. Seneca, on the other hand, was a Stoic philosopher and parts of his play reflect this philosophical background, in which emotions were generally considered dangerous.

Another difference is the relative frequency of individual emotions (fig. 4). In Euripides, the most frequent emotion is `sadness` (14.77%), followed by `desire` (12.51%) and `anger` (9.97%); in Seneca, `anger` (21.84%) is most frequent, while `sadness` ranks third (joint with `desire`, 12.66%) after `fear` (13.40%). In Euripides, the five most frequent overall emotions are identical with the five most frequent emotions attributed to Medea as an `experiencer`. In Seneca, this only applies to the highest ranked emotion, whereas the second, `fear`, is only ranked fifth in the list of Medea's emotions. This illustrates that the two Medeas, though committing similar deeds on plot level, have a different psychological profile. Seneca reworked the Euripidean model with focus on `anger`, which occupied center stage in Stoic philosophy and is the topic of one of Seneca's treatises (see e.g. Guastella, 2001). This reworking includes a more active heroine, as shown by the frequency of `desire`, whose determination causes `fear` in others.

The annotations also support observations regarding the development of the plot and the characters. Figure 5 shows the distribution of Medea's emotions across Euripides' tragedy. The dominant emotion `sadness` comes to an abrupt and almost complete end in the final section of the play – the part where Medea learns that she has succeeded in poisoning Jason's new bride Glauce and her father, and proceeds to kill her own children; at this point, Medea steps out of her inferior position, which is dominated by `sadness` and gains self-efficacy, which is accompanied by `anger` and `disapproval`. Seneca's Medea seems to start in the emotional state that Euripides' Medea reaches at the end of the tragedy. Future research could compare the quantitative perspective on the Medea texts with traditional literary interpretations such as Cairns (2021) and Lima (2021); besides, the graphs and numbers could be discussed in classes on the texts in question, be it in classics or in comparative literature.

8. Conclusion

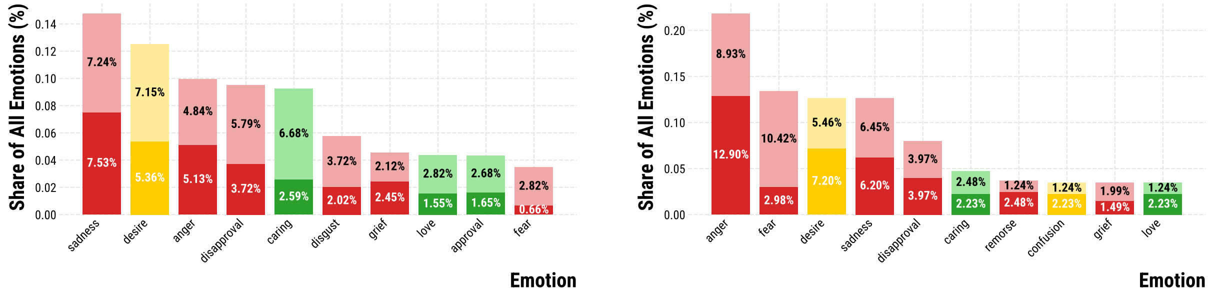
We introduce RAGE, a new corpus for fine-grained analysis of emotions in classical literature containing approximately $100\,000$ words. By creating a multi-layered annotation framework that captures not only 27 emotions but also their `experiencers`, `causes`, and `targets`, we bridge the gap between traditional qualitative literary studies and coarse-grained computational analysis.

We demonstrate some of the possibilities that RAGE enables through corpus-level exploratory analyses and a case study. On a more macroscopic level, the data reveals promising patterns that can be a starting point for deeper analyses, such as emotion frequencies and explicitness across languages and genres or different emotional profiles for gods and humans within the corpus. Our case study on the character of Medea in the tragedies of Euripides and Seneca quantitatively supports the observation that while both plays share a similar plot, they present markedly different psychological portraits, with Euripides' Medea defined by `sadness` and Seneca's by `anger`.

The data also offers a foundation for comparison with traditional literary interpretations. By making this resource available, we hope to encourage further development in this area and foster a productive dialogue between computational methods and the study of emotion in classical texts.

Limitations

Annotating emotions is not straightforward, nor is assessing such annotations. The mere existence of an emotion within a literary work is not always unambiguous; neither is whether an `experiencer`, a `target` or a `cause` exists in the environment of the `cue`. The exact string of words of each element may not be obvious either. And the intensity of the emotion is to a degree subjective as well. All these factors contribute to a relatively low inter-annotator agreement; and even though we anticipated this and therefore had established a curation phase, the outcome still cannot be marked as correct or incorrect. While the resulting annotations are plausible, different groups may come to somewhat different results. The final curated data represents a plausible and well-considered interpretation rather than



(a) Euripides, *Medea*: all vs. Medea as `experiencer`.

(b) Seneca, *Medea*: all vs. Medea as `experiencer`.

Figure 4: Emotion distributions (top 10) in Euripides' and Seneca's *Medea*.

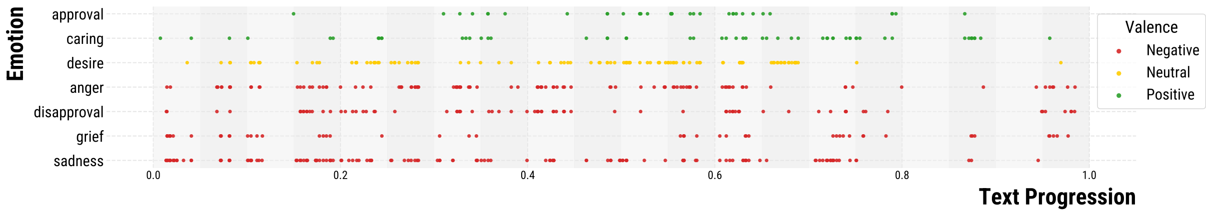


Figure 5: Medea's emotion progression in Euripides' *Medea* (seven selected emotions).

a single, absolute ground truth.

A second set of limitations relates to the size and composition of the corpus. Although RAGE is substantial for this domain, its scale is necessarily constrained by the cost-intensive nature of fine-grained annotation, which demands significant time and domain expertise. This size is insufficient to support broad generalizations across ancient literature. Consequently, statistical findings are highly sensitive to the specific texts included. The deliberate inclusion of works centered on the Medea narrative, for instance, disproportionately influences the dataset and skews any analysis of female characters, as other texts in the corpus feature them less prominently. Therefore, all findings should be understood as specific to the works analyzed, not as definitive trends across ancient languages or genres.

Ethics Statement

We do not foresee ethical risks. The corpus consists of selections from ancient Greek and Latin texts that are openly available as digital editions for scholarly study, and our annotations are created and released for scholarly research and teaching purposes.

All annotations were produced by trained annotators with domain expertise in classical philology. Annotators were compensated by the hour under standard university employment contracts. Participation consisted solely of textual labeling work; no sensitive personal information about annotators or third parties was collected.

Ancient texts inevitably reflect historical worldviews and biases – e.g., regarding gender, ethnicity, social status, and violence – that may be inaccurate or objectionable by modern standards. We preserve such content as part of the historical record because it is essential for scholarly inquiry into antiquity.

Acknowledgments

We are greatly indebted to Nina Stahl for her continuous support and invaluable ideas throughout the entirety of this project. We extend our sincere thanks to all of our hard-working annotators, whose dedication made the creation of this corpus possible. We also thank the anonymous reviewers for their constructive comments and suggestions. This work was funded as part of the Excellence Strategy of the Federal Government and the Länder.

9. Bibliographical References

Mathieu de Bakker, Baukje van den Berg, and Jacqueline Klooster, editors. 2022. *Emotions and Narrative in Ancient Literature and Beyond*. Brill, Leiden, Boston.

Douglas Cairns. 2021. The dynamics of emotion in Euripides' Medea. *Greece and Rome*, 68(1):8–26.

Douglas L. Cairns, editor. 2019. *A Cultural History of the Emotions in Antiquity*. The cultural histories series. Bloomsbury Academic, London, New York, Oxford, New Delhi, Sydney.

Douglas L. Cairns and Damien Nelis, editors. 2017. *Emotions in the classical world*. Franz Steiner Verlag, Stuttgart.

Gregory R. Crane, editor. *Perseus Digital Library*. Tufts University.

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.

Quanqi Du and Veronique Hoste. 2025. Another approach to agreement measurement and prediction with emotion annotations. In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, pages 87–102, Vienna, Austria. Association for Computational Linguistics.

Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129.

Gianni Guastella. 2001. Virgo, coniunx, mater: the wrath of Seneca's Medea. *Classical antiquity*, 20(2):197–220.

Hermann Kappelhoff, Jan-Hendrik Bakels, Hauke Lehmann, and Christina Schmitt, editors. 2019. *Emotionen*. J.B. Metzler Verlag, Berlin.

Evgeny Kim and Roman Klinger. 2018. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1345–1359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

Martin von Koppenfels and Cornelia Zumbusch, editors. 2016. *Handbuch Literatur & Emotionen*. De Gruyter, Berlin, Boston.

John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. Greek-bert: The greeks visiting sesame street. In *11th Hellenic Conference on Artificial Intelligence*, SETN 2020, page 110–117, New York, NY, USA. Association for Computing Machinery.

Paulo Alexandre Lima. 2021. Monstrous emotions in Seneca's Medea. *Greece and Rome*, 68(1):76–96.

Franco Moretti. 2013. *Distant reading*. Verso, London, New York.

Hongbin Na, Yining Hua, Zimu Wang, Tao Shen, Beibei Yu, Lilin Wang, Wei Wang, John Torous, and Ling Chen. 2025. A survey of large language models in psychotherapy: Current landscape and future directions. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 7362–7376, Vienna, Austria. Association for Computational Linguistics.

John Pavlopoulos, Alexandros Xenos, and Davide Picca. 2022. Sentiment analysis of Homeric text: The 1st book of Iliad. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7071–7077, Marseille, France. European Language Resources Association.

Davide Picca and John Pavlopoulos. 2024. Deciphering emotional landscapes in the Iliad: A novel French-annotated dataset for emotion recognition. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4462–4467, Torino, Italia. ELRA and ICCL.

Flor Miriam Plaza-del Arco, Alba A. Cercas Curry, Amanda Cercas Curry, and Dirk Hovy. 2024. Emotion analysis in NLP: Trends, gaps and roadmap for future directions. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5696–5710, Torino, Italia. ELRA and ICCL.

Robert Plutchik. 1982. A psychoevolutionary theory of emotions. *Social Science Information*, 21(4-5):529–553.

Sonya E. Pritzker, Janina Fenigsen, and James M. Wilce, editors. 2020. *The Routledge Handbook of Language and Emotion*. Routledge, Taylor & Francis Group, London, New York.

Andrea Scarantino, editor. 2025. *Emotion Theory: The Routledge Comprehensive Guide*. Routledge, Taylor & Francis Group, New York, London.

Rüdiger Schnell. 2021. *Histories of Emotion*. De Gruyter, Berlin, Boston.

Rachele Sprugnoli, Federica Iurescia, and Marco Passarotti. 2024. Overview of the EvaLatin 2024 evaluation campaign. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, pages 190–197, Torino, Italia. ELRA and ICCL.

Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti, and Giovanni Moretti. 2022. Sentiment analysis of latin poetry: First experiments on the odes of horace. In *Proceedings of the Eighth Italian Conference on Computational Linguistics (CliC-It 2021)*. Accademia University Press.

Rachele Sprugnoli and Arianna Redaelli. 2024. How to annotate emotions in historical Italian novels: A case study on I promessi sposi. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, pages 105–115, Torino, Italia. ELRA and ICCL.

Krishnapriya Vishnubhotla, Adam Hammond, Graeme Hirst, and Saif Mohammad. 2024. The emotion dynamics of literary novels. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2557–2574, Bangkok, Thailand. Association for Computational Linguistics.

10. Language Resource References

Crane, Gregory R. *Perseus Digital Library*. Tufts University. PID https://www.perseus.tufts.edu/.

Klie, Jan-Christoph and Bugert, Michael and Boullosa, Beto and Eckart de Castilho, Richard and Gurevych, Iryna. *INCEpTION*. UKP Lab, TU Darmstadt. PID https://inception-project.github.io/.