


Encoding Imagism? Measuring Literary Imageability, Visuality and Concreteness via Multimodal Word Embeddings

Yuri Bizzoni¹ 
Pascale Feldkamp¹ 
Kristoffer L. Nielbo¹ 

1. Center for Humanities Computing, Aarhus University , Aarhus, Denmark.

Citation

Yuri Bizzoni, Pascale Feldkamp, and Kristoffer L. Nielbo (2025). "Encoding Imagism? Measuring Literary Imageability, Visuality and Concreteness via Multimodal Word Embeddings". In: *CCLS2025 Conference Preprints 4* (1). [10.26083/tuprints-00030154](https://doi.org/10.26083/tuprints-00030154)

Date published 2025-06-17

Date accepted 2025-04-17

Date received 2025-02-07

Keywords

imageability, concreteness, visuality, embeddings, imagism, multimodal embeddings

License

CC BY 4.0 

Note

This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 4th Annual Conference of Computational Literary Studies at Krakow, Poland, in July 2025. Please check jcls.io for the final journal version.

Abstract.

This paper addresses the challenge of measuring "imageability" in literary texts – a concept from psycholinguistics that describes how words evoke sensory experiences. Imageability is connected to literature's immersive quality, but current methods face limitations due to vague definitions, poor lexical coverage, and difficulty scaling to sentence-level (/literary) analysis. To tackle this, we propose a data-driven approach using a multimodal model, alongside traditional word embeddings, to quantify imageability more effectively. Through three experiments: 1) a word-level analysis, 2) sentence-level comparison, and 3) a case study contrasting imagist and love poems – we test whether embeddings created with a multimodal model capture imageability, and the related features concreteness and visuality. We assess the extent to which multimodal embeddings capture imageability in literary texts, while considering compositionality and the multimodal nature of literary imagery.

1. Introduction

"The boy took the old army blanket off the bed and spread it over the back of the chair and over the old man's shoulders." – *The Old Man and the Sea*, Ernest Hemingway

Ernest Hemingway's prose is famously sparse, but conjures vivid mental images: Simple actions and objects – no florid descriptions, no overt emotional cues – yet the scene is immediately present, affective and immersive. One might say that the strength of literary texts lies in their *imageability*.

The concept of imageability originates in psycholinguistics, where it describes the ease with which words evoke sensory experiences (Paivio et al. 1968). However, when speaking of the imageability of literary texts, we are going beyond individual words, and rather touching upon implicit and evocative strategies, including imagery, narrated perception, and the overall immersiveness or experientiality of the text – strategies that have long been held to increase the appeal of texts and enhance the reading experience (Ellen J. Esrock 1994; Sharma Paudyal 2023).

These strategies are related. We can define imagery as language use that appeals to our senses – creating mental images (Lacey and Lawson 2013; Sharma Paudyal 2023). For example, Burroway (1987) notes that a certain use of nouns that evoke

sensory images and of verbs that represent visual actions makes writing “come alive”.¹⁸ This effect aligns with the broader concept of literary experientiality (Fludernik 1996).¹⁹ Experientiality describes how narratives prompt embodied engagement through a “quasi-mimetic evocation of real-life experience”, drawing on knowledge readers have acquired from their physical presence in the world. Psychological and reader-response studies further support this link, associating imagery with the intensity of emotional responses in reading and feelings of embodiment (Blackwell 2020; Goetz et al. 1993; Martínez 2024).²⁰⁻²⁵ By measuring imageability, we can gain insights into how texts evoke embodied responses, and better understand the immersive qualities that shape reader engagement.²⁶⁻²⁷

When measuring the imageability of literary texts, studies typically rely on dictionaries developed in psycholinguistic research that assign imageability or concreteness scores to words (Feldkamp et al. 2024; J. T. Kao and Jurafsky 2015). However, the use of such word-based dictionaries presents issues. We identify three main limitations in current computational approaches to imageability: conceptual vagueness, poor lexical coverage, and lack of compositionality at the sentence level. First, imageability itself is not a straightforward property but tends to be inherently vague, with human judgments often diverging.¹ It spans literal and figurative language, object descriptions, and metaphorical expressions that rely on shared cognitive schemas – not merely knowledge of the visual world. The precise nature of what imageability encodes remains debated, and it is often conflated with related constructs such as visualness and concreteness. While some studies suggest that imageability primarily reflects visual features of language (Ellis 1991), it strongly correlates with concreteness and visuality (Brysbaert et al. 2014), making it difficult to isolate as a distinct linguistic property. To test the relationship to these related concepts, we would need to compare dictionaries for imageability, visuality and concreteness.²

A second main issue is that dictionary-based scores struggle with sentence-level imageability. The accuracy of aggregated imageability, visuality, or concreteness scores at the sentence level is often poor (Verma et al. 2023), which may stem from the limited coverage of dictionaries and the questionable assumption that averaging word-level scores yields a meaningful sentence-level representation. A third issue is that dictionary-based approaches fail to account for compositionality – phrases such as “She painted a dark picture” and “She painted a picture in the dark” have different imageability despite containing similar words.²

Regarding the latter issue, recent advances in natural language processing (NLP) have attempted to quantify imageability using multimodal models that integrate textual and visual information Verma et al. 2023; S. Wu and Smith 2023. However, approaches such as gauging imageability through text-to-image generation output show uneven relation to human judgement, especially for literary texts (S. Wu and Smith 2023).

In sum, existing methods for computing imageability scores face three main challenges:

1. In Verma et al. (2023), inter-annotator agreement for sentence-level imageability ratings was 0.45 (Krippendorff’s α). Note that this was for non-literary texts, where we might expect literary texts to effect an even lower agreement, which seems to be the case in annotation tasks for other concepts (Feldkamp et al. 2024).

2. Also, note that most dictionaries assign imageability scores at the lemma level, abstracting from the word-form level. Working from lemmas means that the variations in word forms – such as ‘painted’ vs ‘paint’ – have the same imageability score, even though differences in tense and part-of-speech category may evoke a different intensity and, in theory, a different set of sensory associations for human readers.

(1) the vague conceptualization of imageability, (2) the limited lexical coverage of dictionary-based approaches, and (3) the difficulty of generalizing from word- or phrase-level scores to sentence-level imageability. These issues are relevant for literary texts: imageability is a core stylistic and aesthetic device; and since literature constructs immersive sensory experiences through language alone, it is an ideal domain for testing computational models of imageability. Unlike instructional or descriptive texts, literary language frequently employs figurative expressions, ellipsis, and symbolic imagery—requiring more nuanced tools to capture sensory evocativeness at the sentence or paragraph level.

To address these limitations, we propose a data-driven, scalable approach that moves beyond static dictionary-based methods. Given the impact of imageability and concreteness on immersivity – and, by extension, reader appreciation – we explore automatic assessment techniques based on text representations. Prior work has demonstrated the visual knowledge of text-only models (Sharma et al. 2024), while recent advances in multimodal models (Radford et al. 2021) offer new opportunities for capturing the visual dimension of language.

Examining the shape of both text-based and multimodal embeddings, we test their ability to approximate imageability, concreteness, and/or visuality scores.

Specifically, we evaluate their efficacy in characterizing literary texts through three experiments:

- Word-level analysis: We assess the relationship between human imageability scores and metrics of multimodal word embeddings for dictionary entries of the imageability dictionary.
- Sentence-level analysis: We compare dictionary-derived scores with metrics of multimodal sentence embeddings for literary texts.
- Literary case study: We examine the discriminatory power of these embedding-based metrics in distinguishing between text types where imageability is expected to differ: imagist poems versus love poems.³

By systematically evaluating these approaches, we aim to develop a more robust framework for measuring imageability in literary texts—one that accounts for compositionality, sentence structure, and the multimodal nature of literary imagery.

While our study does not include human annotations, it represents an initial computational exploration aimed at (1) testing the relationship between imageability and related constructs such as concreteness and visuality, and (2) evaluating the potential of embedding-based metrics to model imageability beyond static, word-level ratings.

3. In this third experiment, we use imagist poems as a testbed to probe whether multimodal embeddings capture stylistic and sensory variation. This should not be taken to imply a direct equivalence between imageability and imagism, nor a reduction of poetic imagery to literal visual representation. Nonetheless, the historical emphasis of imagist poetry on economy, concreteness, and sensory immediacy makes it a useful comparative corpus for our purposes.

2. Related Works 93

2.1 Imageability in literary texts 94

The evocation of mental imagery in literary texts has been a debated topic in literary and psychological scholarship (Kuzmičová 2014). Despite its prominence in early 20th century literary movements like Imagism, where the emphasis was placed on clear, visual language and the rejection of abstraction (Pound 1913), the role of imagery and the imageability of literature was often overlooked in structuralist and New Criticist frameworks, prioritizing linguistic networks and meaning-making (Ellen J. Esrock 1994). However, in recent years, the concept of imageability has regained attention, with both literary scholars and psychologists increasingly examining its role in reader response (Kuzmičová 2014; Magyari et al. 2020; Martínez 2024; Sharma Paudyal 2023).

Imageability, defined as the ability of a text to evoke sensory experiences and mental imagery, is closely linked to the heightened emotional responses that images can provoke (Goetz et al. 1993) and the embodied nature of reading experiences (Martínez 2024). Literary passages that employ concrete, sensory language—those that engage the senses without explicit emotional cues—have been shown to elicit emotional responses from readers. For instance, Hemingway’s minimalist style (see our example above), which uses stark imagery without overt emotional direction, is perceived as emotionally charged by human readers, despite being classified as neutral by automatic sentiment analysis systems (Feldkamp et al. 2024). Furthermore, the evocation of interoceptive or physiological states can activate a reader’s embodied experience (Martínez 2024), while concrete language enhances emotional engagement and heightens suspense (Auracher and Bosch 2016).

While imagery, concreteness and imageability have long been concepts employed in literary analysis (Ellen J. Esrock 1994; Sharma Paudyal 2023), computational literary history studies have further shown how quantifying imageability can be employed to characterize certain literary texts. For instance, studies of poetry have shown that Imagism, with its focus on direct, visual language, is associated with higher levels of imageability (J. T. Kao and Jurafsky 2015). Additionally, a historical shift toward more concrete and imageable language in poetry has been observed, suggesting a gradual evolution of literary style over time (Ibid.).

However, quantifying imageability – along with related concepts like concreteness and visuality – is a challenge for computational literary analysis. These concepts are often defined and operationalized differently across domains. For instance, in more communicative texts, such as journalism, imageability is often tied directly to sensory or visual representations, where literary language frequently employs imagery in more abstract or symbolic ways, and may use it more strategically and with greater nuance. The concept of “implicit” expression – “show, don’t tell” – is particularly significant in literature, where imagery is often used to evoke affect without explicitly naming it (Feldkamp et al. 2024), and literary scholarship frequently use terms like ‘evocative’ or ‘understatement’ to describe authorial styles (Strychacz 2002, Daoshan and Shuo 2014), further emphasizing the subtlety of literary imageability as a strategic tool. Furthermore, literary studies have made efforts to distinguish between conceptually different types of

imagery (Kuzmičová 2014). 136

This implicit nature of literary expression poses additional challenges for computational methods that rely on standardized lexical resources. Its subtleties may not align with the operational definitions of imageability found in existing dictionaries or lexicons. 137
138
139

2.2 Dictionaries for imageability 140

The terms imageability, concreteness, and visuality are often used interchangeably, though they capture distinct but overlapping dimensions of language. While concreteness typically refers to the degree to which a word denotes a tangible entity, imageability extends beyond the purely visual to include mental representations, including interoceptive states (Dellantonio et al. 2014). 141
142
143
144
145

In literary studies, these constructs have been applied in different ways. For example, J. Kao and Jurafsky (2012) use concreteness – or its reverse, abstractness – to assess literary imagery, while J. T. Kao and Jurafsky (2015) measure “concrete imagery” through a combination of object-word frequency, abstract-word frequency, and dictionary-based concreteness and imageability scores. 146
147
148
149
150

Even when focusing specifically on imageability, various resources have been developed, beginning in the 1960s with early psycholinguistic studies (Paivio et al. 1968). One of the first large-scale lexicons, the MRC database, was compiled in the 1980s Coltheart 1981 and remains widely used despite its limited scope. More recent and expansive dictionaries have been developed, such as the 40,000-lemma concreteness lexicon by Brysbaert et al. (2014), which significantly surpasses the 4,800 lemmas found in the MRC lexicon. However, coverage across different dimensions remains uneven, with imageability, visuality, and concreteness lexicons varying in size and consistency. 151
152
153
154
155
156
157
158

2.3 Development of models for imageability 159

Recently, the visual aspect of imageability has gained significant attention in Natural Language Processing (NLP), particularly in the context of text-to-image models like DALL-E. These models rely on dual processing of text and images yet struggle when dealing with long-form text containing spans of non-visual content. As a result, visualness has been proposed as a useful metric for characterizing the prompt prior to generation, with the goal of improving the accuracy of text-to-image generation models Chen et al. 2025; Verma et al. 2023. 160
161
162
163
164
165
166

For instance, Verma et al. (2023) introduces a binary classification task distinguishing imageable from non-imageable sentences to enhance prompt characterization before image generation. Similarly, Chen et al. (2025) explores the role of visualness in guiding the generation process, aiming to refine model outputs. Apart from augmented image generation, identifying imageable text might also have further downstream application such as on the fly visuals (Liu et al. 2023) and image-assisted video navigation (Zhao et al. 2019). However, while such binary classification is practical for generation and other tasks, is not as useful for describing nuanced data, where we ideally want to maintain a level of granularity: gauging more or less imageable text. 167
168
169
170
171
172
173
174
175

To address the shortcomings of existing imageability dictionaries, S. Wu and Smith 176

Construct	Scope	Relation to imageability
<i>Concreteness</i>	Tangibility of a word’s referent	<i>Partial overlap.</i> Concrete words tend to be imageable, but abstract items such as “whirlwind” can evoke vivid scenes; conversely “road” is concrete yet often yields weak imagery in isolation.
<i>Visualness</i>	Strength of <i>visual</i> associations	<i>Proper subset.</i> Imageability spans <i>all</i> modalities (auditory, tactile, olfactory, ...), not vision alone.
<i>Imagery</i> (literary studies)	Textual clusters of sensory details	<i>Complementary.</i> Imagery is a textual feature; imageability is the reader-side potential.

Table 1: How imageability relates to creating mental constructs.

(2023) propose methods that incorporate sentence compositionality, aiming to better capture the nuances of how imageability evolves across different sentence structures. While their work shows promise in addressing fixedness in representations, a critical challenge remains: many texts can evoke strong mental imagery in readers without these images being strongly encoded in a culturally shared or visual sense. For instance, creative and poetic language can provoke vivid imagery that is not directly tied to shared or commonsense visual representations. When the consistency of generated images is used as a proxy for imageability (S. Wu and Smith 2023), this may actually measure the stability of a text’s representation rather than its inherent imageability. Conversely, a non-visual passage may still elicit a text-to-image model to generate superficially coherent images. With high uncertainty, text-to-image models often generate visually similar images (e.g., images of actual text) that may appear coherent but do not necessarily align with any human reader’s mental image of the text.⁴

Finally, methods for gauging imageability may show variability across genres. S. Wu and Smith (2023) finds an insignificant correlation between their imageability measure and human assessments of poem lines, yet a significant correlation for news sentences. This suggests that the effectiveness of imageability metrics may depend on the genre and its inherent stylistic and thematic characteristics.

3. What we mean by *imageability*

We follow the psycholinguistic tradition in defining imageability as the ease with which a linguistic expression evokes sensory representations in the mind of a typical reader, but we extend the *expression* from individual words to any contiguous span of text.

Given a reader r and a text span t , the imageability $I(t, r)$ is the *subjective vividness* of the multi-sensory mental imagery spontaneously elicited by t . In group studies we use the expected value $I(t) = \mathbb{E}_r[I(t, r)]$.

Two implications follow. First, imageability is a *psychological potential*, approximated by behavioural data or cognitively motivated proxies. Second, it tends to be **compositional**: the vividness of “he smoked a crooked, emerald-green cigarette” might not be a linear sum of the scores for its component words.

4. See examples of such visually similar but disjunct images in Verma et al. (2023), Fig. 5.

Level	Typical operationalisation	Limitations
<i>Word</i>	Psycholinguistic norms from 25–40 k-entry lexica (MRC, Lancaster, BLP)	Coverage gaps for literary vocabulary; ignores syntax and context.
<i>Sentence/line</i>	Human ratings (rare) or context-aware embeddings (this work)	Ratings costly; embeddings need interpretability.
<i>Whole text</i>	Aggregations (mean, max, entropy) over sentences	Sensitive to length and genre; reliant on robust lower-level scores.

Table 2: Granularity choices when measuring imageability.

Ideally, we would move from the *word* to the *sentence* level without sacrificing scalability. 206
 Because imageability is, by definition, a reader experience, ultimate confirmation demands sentence- or passage-level ratings. The present work should therefore be read 207
 as a *bridging effort*: dictionary-validated, embedding-based metrics ready for human 208
 calibration.⁵ 209
 210

4. Resources 211

4.1 Dictionaries 212

For Experiment I, we utilize lexicon-based resources to analyze the imageability of words, 213
 primarily relying on the **MRC Psycholinguistic Imageability Lexicon** (Coltheart 1981). 214
 This lexicon comprises 4,828 lemmas that have been rated for their capacity to evoke 215
 mental imagery. Additionally, we resort to two other well-established resources: the 216
Concreteness Lexicon Brysbaert et al. 2014, which assigns ratings to words based on 217
 their *perceived tangibility and sensory grounding*, and the **Lancaster Sensorimotor Norms** 218
 (Lynott et al. 2020), which provide detailed *modality-specific perceptual ratings* (e.g., visual, 219
 auditory, tactile associations). As the lexica of the three resources largely overlap, we can 220
 systematically compare how they conceptualize and quantify imageability, concreteness, 221
 and sensory experience. Given that previous research has noted a strong correlation 222
 between imageability and concreteness, but also some key distinctions between them 223
 (Paivio et al. 1968), our analysis seeks to clarify the extent to which dictionary-based 224
 imageability measures capture cognitive and perceptual properties distinct from general 225
word concreteness and *modality-specific sensory attributes*. 226

- The **MRC Psycholinguistic Imageability Lexicon** (Coltheart 1981). One of the 227
 earliest large-scale resources for word imageability, this lexicon contains 4,828 228
 lemmas, each rated based on the extent to which they evoke *mental imagery*. The rat- 229
 ings were collected from human participants, making it an empirically grounded 230
 resource for word-level imageability. We compare this resource with later expan- 231
 sions and refinements: (i) **Cortese and Fugett’s Imageability Ratings** (here, Imag. 232
 C) (Cortese and Fugett 2004): an updated version that increases the coverage 233
 of imageability scores and refines earlier ratings. (ii) **Reilly and Kean’s Formal 234
 Distinctiveness Model** (here, Imag. R) (Reilly and Kean 2007): a lexicon that in- 235
 tegrates and updates multiple prior resources, including the MRC, while filtering 236

5. The final experiment of this paper already uses implicit human judgments by distinguishing two different literary genres.

out words with mid-range imageability ratings to focus on words that are strongly imageable or non-imageable. 237
238

- **The Visuality Lexicon of the Lancaster Sensorimotor Norms** (Lynott et al. 2020). 239
The Lancaster Sensorimotor Norms provide modality-specific sensory ratings 240
(e.g., visual, auditory, tactile, and motor associations) for 39,707 English words. 241
In our experiments, we use the *visuality* scores specifically. Unlike general image- 242
ability, *visuality* captures the extent to which a word evokes a visual percept. This 243
distinction is important because some words may be highly imageable but not 244
strongly visual (e.g., “fragrance” or “melody”). Comparing *visuality* to image- 245
ability allows us to examine how modality-specific sensory experience aligns with 246
broader notions of literary imagery . 247
- **The Concreteness Lexicon** (Brysbaert et al. 2014). This dataset provides con- 248
creteness ratings for 40,000 words, where concreteness is defined as the degree 249
to which a word refers to a tangible, physical entity. While concreteness and 250
imageability are often correlated, they are not identical concepts: some abstract 251
words (e.g., *freedom*) might be highly imageable due to their symbolic richness, 252
while some concrete words (e.g. *rock*) may elicit limited mental imagery despite 253
being physically tangible. By including concreteness as a comparative measure, 254
we assess how word-level concreteness and imageability interact, particularly in 255
literary contexts. 256

4.2 Literary texts 257

For Experiments II and III, we use full sentences from literary texts. Moving from single 258
words to entire sentences enables an assessment of how imageability, concreteness, and 259
visuality manifest in context. 260

For Experiment II, the dataset includes two modernist novels alongside a large-scale 261
corpus of fiction: 262

- *The Old Man and the Sea* by Ernest Hemingway (1952). This novel is characterized 263
by concise, concrete descriptions and a direct, unembellished prose, making it an 264
ideal candidate to evaluate imageability in an economical (yet vivid) narrative 265
style. 266
- *Mrs. Dalloway* by Virginia Woolf (1925). In contrast, Woolf’s novel employs 267
stream-of-consciousness narration, featuring long, fluid sentences that foreground 268
subjective perception with immersive sensory detail. Its contrast with Heming- 269
way’s style allows us to test whether opposite stylistic techniques correlate with 270
distinct levels of imageability. 271
- Sentences from the Chicago Corpus (1880–2000). A diverse dataset of 9,000 sen- 272
tences randomly sampled from 9,000 different novels, to ensure a broad coverage 273
of stylistic and historical variation in fiction. The corpus from which the sentences 274
are sampled includes works ranging from canonical literature to lesser-known 275
fiction, providing a representative snapshot of Anglophone prose writing across 276

the 19th and 20th centuries.⁶ For further details, see Bizzoni et al. (2024) and Y. Wu et al. (2024).

All sentences were tokenized using the SpaCy’s NLP library⁷. The inclusion of both single-author novels and a large, multi-author corpus allows us to assess how imageability varies both within and between different literary styles. In this context, Hemingway and Woolf serve as *controlled case studies* for contrasting narrative techniques: Hemingway’s prose is marked by concise, concrete descriptions of low abstraction, whereas Woolf’s stream-of-consciousness style favors immersive, introspective, but often highly imageable narration. These stylistic differences provide a useful basis for testing whether imageability metrics capture differences in literary technique. Their well-known status also makes them accessible and interpretable examples. Meanwhile, the Chicago Corpus, composed of diverse works spanning more than a century, offers a *broadly representative dataset* that enables generalization beyond the idiosyncrasies of individual authors.

	sentences	year	type
Hemingway	1,928	1952	prose
Woolf	3,578	1925	prose
Chicago corpus	9,000	1880-2000	prose
Imagist	1,195	1915	poetry
Modern Love	1,126	1896-1939	poetry

Table 3: Data

For Experiment III, we conduct a literary case study aimed at distinguishing *imagist poetry* from more topic-based, modern love-themed poetry using embedding-based metrics and dictionary-derived scores. The choice of Imagist poetry for testing the distinguishing power of dictionary- and embedding-based features was also made because of previous works supporting that Imagist poetry stands out in these dimensions (Gleason 2007, 2009; J. T. Kao and Jurafsky 2015). To this end, we utilize two distinct poetic datasets:

- *Some Imagist Poets* (1915), an anthology compiled by Ezra Pound, which includes 37 poems authored by six poets. Imagist poetry is characterized by its emphasis on precise and concrete imagery, minimalism, and a rejection of abstraction, making it an ideal test case for computational measures of imageability.⁸
- A selection of 74 *modern love-themed poems* by 22 authors, collected from *The Poetry Foundation’s* curated category ‘Love’.⁹

By juxtaposing these two corpora, we aim to assess whether computationally derived imageability and concreteness scores can effectively differentiate poetic traditions that

6. It should be noted that the dataset is predominantly English-language fiction, potentially limiting its generalizability to other linguistic traditions.

7. Specifically, we employed the SpaCy `en_core_web_sm` model, which provides robust sentence segmentation for literary texts, ensuring consistent parsing across different styles.

8. We use the 1915 Pound anthology primarily as a test case to explore computational representations of imageability. We do not claim that this selection fully captures the complexity or historical breadth of Imagism as a movement—nor that it represents the full range of poetic strategies associated with it. While alternative anthologies, such as those edited by Amy Lowell, might offer broader coverage, our aim is not to provide a literary-theoretical account of Imagism, but to use this corpus as a controlled experimental setting.

9. To maintain clear genre distinctions, we excluded poets in *Some Imagist Poets* from the love-themed poetry dataset. For a complete list of included poems and authors, see subsection A.2. The dataset is available at: <https://huggingface.co/datasets/merve/poetry>.

prioritize sensory evocation (Imagism) from those that may contain a broader range of abstract or figurative language (general love poetry). This comparison provides insight into the applicability of our methods for distinguishing stylistic and thematic variations in literary texts.

		imageability	concreteness	visuality
Hemingway		4359.2 ± 3241.6	38.1 ± 28.4	33.7 ± 26.0
	<i>normalized</i>	352.6 ± 38.0	2.8 ± 0.3	2.4 ± 0.3
Woolf		5156.8 ± 6167.0	45.5 ± 55.0	42.3 ± 51.5
	<i>normalized</i>	350.6 ± 47.7	2.7 ± 0.4	2.5 ± 0.4
Chicago		5173.8 ± 5292.2	47.3 ± 46.3	43.1 ± 43.1
	<i>normalized</i>	345.7 ± 37.1	2.7 ± 0.3	2.5 ± 0.3
Imagist		1870.6 ± 1027.3	17.1 ± 8.7	14.9 ± 7.9
	<i>normalized</i>	376.2 ± 64.5	3.0 ± 0.5	2.6 ± 0.6
Love		2112.53 ± 846.7	18.3 ± 6.6	16.2 ± 6.1
	<i>normalized</i>	363.7 ± 53.1	2.8 ± 0.5	2.5 ± 0.5

Table 4: Sentence-level average (and SD) imageability, concreteness, and visuality of datasets.

Note that in gauging the relationship between the dictionary-based features and the embeddings, we sum the imageability, concreteness, and visuality scores assigned via the dictionaries across sentences without normalizing for sentence length. This approach allows us to capture the total intensity of these features in the sentence, rather than averaging the intensity of individual words. We do this because we are interested in the overall presence or weight of these features in a given sentence. Literary texts may rely on the cumulative effect of imagery across sentences, and this approach allows us to reflect that broader, contextual presence, which we also expect the embeddings to capture as well. In Experiment III, where we compare Imagist and Love poetry lines, we do normalize for line length to replicate the methodology used in J. T. Kao and Jurafsky (2015), which assigns scores by dividing the summed imageability score of words with the number of words (extant in the dictionary).

It is important to note that whether features are assigned based on sums or length, normalized sum has significance when differentiating between groups or authors – which is what we do in Experiment III. For example, in the data summary (Table 4), we see that Love poetry, on average, has higher imageability than Imagist poetry. However, when normalizing the scores for line length, this trend is reversed, with Imagist poetry averaging 376.2 and Love poetry 363.7.

The decision to normalize feature scores when using dictionaries in literary experiments is an question of operationalization that we want to underline. For example: is a long sentence with many low- and high-imageable words *more or less imageable* than a shorter sentence with high-imageable words, if their summed scores are the same? In other words, do factors like density or brevity affect the perceived imageability of sentences? This question can only be answered by comparing human imageability judgments against both methods of score assignment. We leave this to future work and focus here only on the relationship between different systems (embeddings vs. dictionaries), not their relation to human judgment.

4.3 Embeddings 336

When it comes to embeddings, we employ the CLIP model (Radford et al. 2021), a multimodal vision language model trained on large-scale image-text pairs. CLIP is designed to align textual and visual representations, making it particularly suitable for capturing visual and concrete dimensions of language, which are directly relevant to our study of imageability.

Given its training objective, the semantic space of CLIP’s embeddings is expected to encode visual salience and concreteness more effectively than purely text-based models. This suggests that CLIP-based embeddings may provide a more explicit representation of imageability compared to traditional word embeddings derived from text-only corpora.

However, the extent to which multimodal representations differ from text-based embeddings in encoding sensory and imagistic properties remains an open question. To address this, we compare CLIP-based embeddings against those generated by a text-only model, specifically BERT. BERT embeddings provide a useful contrast, as they are derived solely from linguistic contexts without access to visual grounding.¹⁰ This comparison allows us to evaluate whether imageability-related features emerge naturally in textual embeddings or whether multimodal supervision enhances their representation.

It’s important to underline that psycholinguistic lexica used in Experiments I–III enter our pipeline only for validation. They provide a widely accepted yard-stick against which to gauge whether a candidate metric is even plausible. Crucially, the mapping from embedding shape to imageability is not fitted on dictionary scores—indeed no fitting is required, because norm and entropy are closed-form functions of the raw vectors. In downstream applications (e.g. analysing an unedited novel draft, or surfacing vivid quotations for digital exhibits) the dictionaries can be dropped entirely.

5. Methods 360

We analyze the *shape* of both word and sentence embedding representations to determine how imageability manifests in textual and multimodal embeddings. Specifically, we examine whether embeddings of highly imageable and concrete words exhibit distinct distributional properties compared to those of abstract or less imageable words.

Our initial hypothesis (H_1) is that words with higher imageability and concreteness might have more *localized values* in the embedding space, meaning that they might cluster more tightly within specific regions of the vector space. In contrast, embeddings of more abstract words, that might lend themselves to a larger array of contexts, may be more dispersed, leading to higher entropy in their distribution and lower norm (i.e., strength). To illustrate this, consider the word *dog*: especially in a multimodal model like CLIP, its embedding is likely to concentrate most of its information on specific dimensions, while a more abstract term like *beautiful* is less directly tied to a specific visual referent and may exhibit a broader, more diffuse representation across the semantic space.

10. We selected BERT due to its widespread use in word embedding research and literary analysis. In particular, the `bert-base-cased` model is frequently applied for semantic and stylistic investigations in computational literary studies Grisot et al. 2022; Paragini and Kestemont 2022; Silva et al. 2023. <https://huggingface.co/google-bert/bert-base-cased>

The difference in embedding structure between these two cases can be quantified by analyzing the distribution of activation values across all dimensions.¹¹

The opposite hypothesis is also a possibility (H_2). Under this view, concrete words such as *dog* or *tree* may activate a broader range of dimensions due to their rich sensory associations across multiple modalities. In contrast, abstract words like *justice* or *hope* may activate fewer, more specific dimensions, resulting in a sharp activation profile with, for example, higher norm but lower entropy — akin to a *spike* in certain representational axes. This could occur if embeddings for abstract concepts rely on a small number of high-level semantic features (e.g., valence, affect, discourse function), while embeddings for concrete words require a more distributed, multimodal representation that increases their variance and entropy.

To formally test this, we compute various vector shape metrics on the sentence embeddings. These include:

- **Norm** ($\|x\|_2$). The Euclidean norm measures the overall magnitude of the embedding vector. It provides a sense of how “large” the values in the vector are. This does not necessarily tell us about distribution across dimensions.

It is defined as:

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

A sparse embedding (where most values are near zero and only a few are large) might have a lower norm, while more uniform activation leads to higher norm.

- **Entropy** ($H(e)$). Entropy is defined as:

$$H(\mathbf{e}) = - \sum_{i=1}^n p_i \log(p_i + \epsilon)$$

where \mathbf{e} is the embedding vector, and p_i are probabilities obtained by applying a normalization to the elements of \mathbf{e} . We primarily construct these probabilities by taking the absolute value of each dimension and normalizing to sum to 1, thus ensuring non-negative values interpretable as a pseudo-probability distribution over dimensions. This approach was chosen over the softmax transformation to avoid amplifying the largest embedding values exponentially, which can distort the distribution and reduce sensitivity to variations across smaller dimensions.¹² n is the number of components in the vector and ϵ is a small constant to avoid $\log(0)$. The entropy reflects how evenly the embedding’s values are distributed across dimensions.

11. This is also the hypothesis of Hessel et al. (2018, p. 2194) in quantifying visual concreteness, who write: “Intuitively, a visually concrete concept is one associated with more locally similar sets of images; for example, images associated with “dog” will likely contain dogs, whereas images associated with “beautiful” may contain flowers, sunsets, weddings, or an abundance of other possibilities.”

12. To verify the robustness of this approach, we also computed probabilities using a softmax transformation, which produced near-identical entropy values across our experiments. This confirms that our simpler absolute-value normalization provides a consistent and interpretable proxy for measuring the spread or concentration of embedding activations. While this method introduces nonlinearity, embeddings with primarily positive values (e.g., CLIP) may yield systematically different entropy scores. Nonetheless, this approach balances interpretability and computational simplicity.

- **Variance** (σ^2): It measures the spread of values in the embedding vector, indicating how much they deviate from the mean: 404
405

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

where x_i is the value of the embedding at dimension i , n is the total number of dimensions, and \bar{x} is the mean of these values. Higher variance suggests a more dispersed representation, potentially reflecting greater contextual flexibility, while lower variance may indicate a more compact, feature-specific encoding. This is crucial in evaluating whether highly imageable embeddings are tightly clustered or broadly distributed in a semantic space. 406
407
408
409
410
411

- **Sparsity** ratio. Sparsity ratio can be defined as: 412

$$\text{sparsity_ratio} = \frac{\|\mathbf{e}\|_1}{\sqrt{n} \cdot \|\mathbf{e}\|_2}$$

where \mathbf{e} is the embedding vector, n is the number of components in the vector, $\|\mathbf{e}\|_1 = \sum_{i=1}^n |e_i|$ is the Manhattan norm (the sum of the absolute values of the elements of \mathbf{e}), and $\|\mathbf{e}\|_2 = \sqrt{\sum_{i=1}^n e_i^2}$ is the Euclidean norm (the square root of the sum of the squared elements of \mathbf{e}). The sparsity ratio gives us an idea of how densely populated the embedding is, with lower values indicating higher sparsity. 413
414
415
416
417

Note: For norm and entropy, which are less intuitive measures, we show the distribution of values over embedding dimensions for the embeddings with the highest/lowest entropy and norm of the Chicago corpus sentence samples (see [Appendix A](#), Figures 4-5). 418
419
420
421

To evaluate the validity of these embedding-based metrics as indicators of imageability, we correlate them with dictionary-based imageability, concreteness, and visuality scores. Specifically, we compare values derived from the lexical resources of the MRC Psycholinguistic Imageability Lexicon (Coltheart 1981), the Concreteness Lexicon (Brysbaert et al. 2014), and the Lancaster Sensorimotor Norms (Lynott et al. 2020) against our computed embedding shape properties (see Section 3). By computing Spearman correlations, we assess the degree to which embedding metrics reflect known psycholinguistic properties of the lexicon. 422
423
424
425
426
427
428
429

Since dictionary-based scores are primarily word-level ratings, we then extend our analysis to the sentence level by aggregating word-level values across sentences. Specifically, for each sentence, we compute the sum imageability, concreteness, and visuality of the sentence based on the words that are present in the dictionaries (see the end of Section 4.2 again for more details). This allows us to compare *sentence* embeddings with dictionary-based metrics computed over entire sentences. 430
431
432
433
434
435

It is important to note that the transition from word-level representations to sentence-level embeddings is non-trivial. The imageability of a sentence is not simply the sum of its individual words' scores; rather, it depends on syntactic structure, compositionality, and context-dependent meaning shifts which cannot be captured by sums of dictionary scores. For example, a sentence like "The sky darkened before the storm" contains words 436
437
438
439
440

with varying individual imageability scores, but their combined effect creates a vivid, 441
 scene-setting description. Conversely, a sentence with highly imageable words may still 442
 lack clear imagery if its structure is abstract or ambiguous. 443

At the same time, sentence embeddings are not simple averages of word representa- 444
 tions: they incorporate contextual interactions, modifying the contribution of each 445
 word depending on its grammatical role and semantic dependencies. This means that 446
 embedding-based metrics may diverge significantly from dictionary-based scores when 447
 applied at the sentence level. 448

6. Results 449

6.1 Experiment I: Correlations at the Word Level 450

Figure 1a presents the correlation matrix for dictionary-based and embedding-derived 451
 metrics, computed over the MRC dictionary lemmas (i.e., the subset of words appearing 452
 across all dictionaries used in this study). The figure illustrates the relationships both 453
 within dictionary-based scores and within embedding-based metrics, as well as their 454
 inter correlations. 455

Internal Correlations in Dictionary-Based Scores. We observe a strong mutual corre- 456
 lation among dictionary-based metrics. Notably, **Cortese Imageability** (*imag R*) exhibits 457
 a correlation with **MRC Imageability** (*imag*) comparable in magnitude to its correlation 458
 with **Concreteness**. This suggests that, in practice, imageability and concreteness are 459
 not sharply distinguished in these resources — at least at the word level. While both 460
 measures are conceptually distinct, their empirical overlap aligns with prior research 461
 indicating a strong connection between how vividly a word evokes imagery and how 462
 concrete its referent is. 463

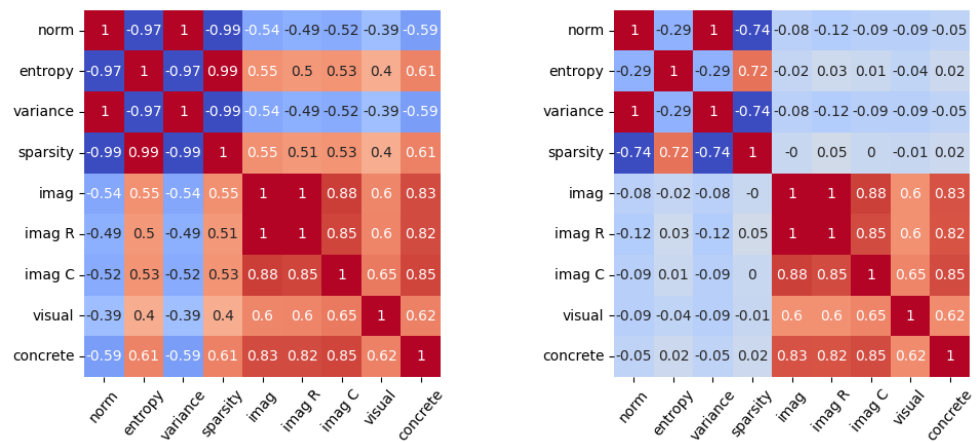
Interestingly, **Visuality** shows a weaker correlation with both **Imageability** and **Con-** 464
creteness, suggesting that the dictionary-based concept of imageability is more strongly 465
 associated with tactile or sensorimotor properties than with purely visual modalities. 466
 This reinforces the idea that imageability, as defined in psycholinguistic lexica, captures 467
 a broader range of sensory experiences beyond just the visual salience. 468

Internal Correlations in Embedding-Based Metrics. Turning to the embedding-derived 469
 metrics, we find an even stronger internal correlation structure. For instance, **norm** 470
 and **entropy** exhibit an inverse relationship, indicating that embeddings with higher 471
 activation magnitudes (higher norm) tend to have more localized values, while those 472
 with lower norm tend to have more evenly spread, high-entropy distributions. Because 473
 our entropy calculation involves absolute-value normalization, direct comparison of 474
 entropy values across embedding types (e.g., CLIP vs. BERT) should be interpreted 475

cautiously, though relative trends within each model remain informative.¹³¹⁴ 476

This is consistent with our hypothesis that abstract words may have sharp activation 477
spikes in fewer dimensions, whereas concrete words may activate a broader set of 478
features across the embedding space. 479

Correlations Between Dictionary and Embedding Metrics. Across all embedding- 480
based metrics, we find a significant correlation with dictionary-derived scores, particu- 481
larly with Imageability and Concreteness ($\rho = .55 - .61$). This suggests that embedding 482
norms, entropy, and related properties encode information that aligns with human- 483
annotated word imageability and concreteness ratings. As we transition to the sentence 484
level in subsequent experiments, we investigate whether these correlations persist when 485
compositional effects come into play. 486



(a) Using the multimodal model.

(b) Using BERT.

Figure 1: Comparison of embedding metrics and dictionary scores using the multimodal model and BERT. Numbers refer to the Spearman coefficient. Note that Imag C and Imag R refers to the two expansion dictionaries of imageability, see subsection 4.1, while Imag refers to the general MRC imageability dictionary.

Moreover, when comparing our results with a text-only model, we find that the pre- 487
viously observed correlations do not fully hold. While there are slight correlations 488
between Imageability (MRC_Reilly) and embedding-derived metrics such as norm, 489
entropy, and variance, these are notably weaker than those found using the multimodal 490
CLIP model (Fig. 1b). 491

Additionally, the internal correlations between embedding-based metrics are less pro- 492
nounced in BERT. Specifically, the relation between variance and entropy drops, as 493

13. See section 5 for details on entropy computation and normalization procedures. Note that: The relation between embeddings' norms and entropy is not necessary, but a by-product of CLIP's training objectives. The strong correlation between norm and entropy when using the multimodal CLIP model likely stems from the way the model processes text. Since CLIP relies on softmax at various stages to encode textual inputs, its embeddings inherently carry a probabilistic structure. When computing entropy, which also transforms embeddings into a probability distribution, this process can introduce an automatic dependency on the norm. Specifically, embeddings with higher norms tend to distribute probability mass differently, leading to a systematic correlation between norm and entropy. To avoid enforcing this relation, we ensured that we did not use softmax in the process of computing entropy, although the softmax approach was also tested. See section 5.

14. As with the relation between norm and entropy, we see the relation between variance and entropy strongly in the CLIP model, but not in the BERT model. Again, this may be related to the normalization process in generation CLIP embeddings, where a normalization will also fix the variance dependent to the entropy – both then relying on the scaling of the data.

well as the relation between norm and entropy (suggesting that the strong correlations between these in the CLIP model do relate to the model’s reliance on probability conversion, see footnotes 9 & 10).

Given the strong internal correlations of the embedding metrics in the CLIP model embeddings, we retain only norm and entropy for the next two experiments, while maintaining imageability, visuality, and concreteness as dictionary-based features.¹⁵

6.2 Experiment II: Sentence-level Analysis in Literary Texts

	data	imageability	visuality	concreteness
Norm	<i>Hemingway</i>	-0.61	-0.63	-0.62
	<i>Woolf</i>	-0.44	-0.46	-0.46
	<i>Chicago</i>	-0.31	-0.37	-0.36
Entropy	<i>Hemingway</i>	0.60	0.63	0.62
	<i>Woolf</i>	0.42	0.44	0.44
	<i>Chicago</i>	0.31	0.37	0.37

Table 5: Spearman correlations between dictionary scores of sentences and the norms and entropies of sentence embeddings across our literary data.

For correlations of metrics across sentences, we find very similar correlations as in Experiment I, presented in table 5. That is, across all of our 3 literary datasets, we find that Imageability has a negative relationship to embedding norm, and a positive relation to embedding entropy. The direction and strength of these correlations is the similar for Concreteness and Visuality. We find the strongest correlations within the Hemingway sentences (min. $\rho \pm 0.61$) and the weakest for the Chicago sentences (min. $\rho \pm 0.31$) (table 5). This suggests that embedding norm and entropy maintain their correlation direction with the dictionary-based features also when aggregating scores at the sentence level, but that the strength of the correlation might differ according to the type of literature. Among Chicago corpus sentences, where correlations between embedding- and dictionary-based metrics is the lowest, we might expect the diversity – across both genres, styles, and decade – to have an effect.

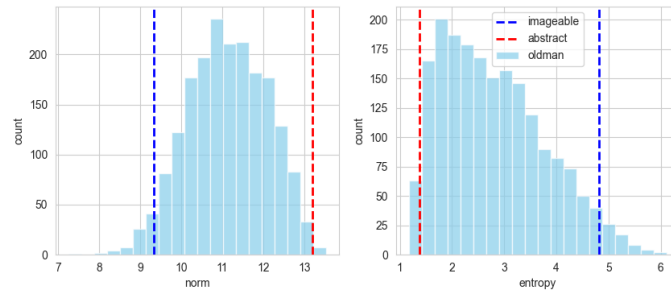
Still, within each group of literary data (Hemingway, Woolf, Chicago), we find more or less imageable sentences (according to the dictionary) have a relation to the level of norm and entropy of their embedding. To examine how norm and embedding distribute within the groups, and to illustrate this effect, we selected two example sentences as reference points for high and low imageability:

- Highly Imageable: The thin white surgical gloves he wore as he pumped the gas looked like pale skin.
- Non-Imageable: Wishful thinking as the saying goes.

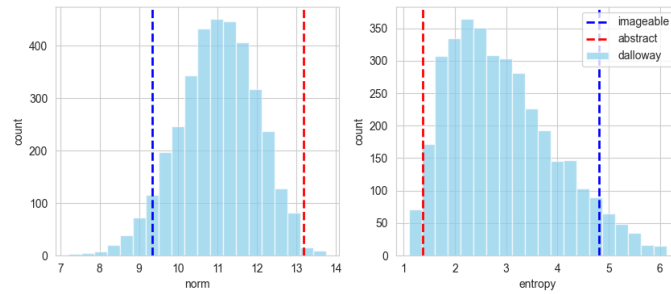
These two sentences were among the top and bottom 10 in terms of dictionary-based imageability scores among all sentences in our data.

The contrast between these examples highlights the degree to which descriptive, sensory-

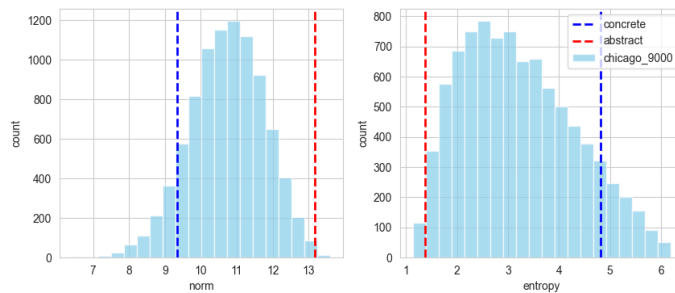
15. Alternative imageability dictionaries were excluded due to the smaller size of their lexica and due to seeing their high correlation with the MRC imageability dictionary, making them redundant for our analysis.



(a) Hemingway



(b) Woolf



(c) Chicago

Figure 2: Histogram using the norms and entropies of example sentences among the most and least imageable sentences from the 9,000 sampled sentence of the Chicago Corpus as references.

rich language correlates with embedding structure, where highly imageable sentences appear to cluster in regions with lower entropy and higher norm. 524 525

As shown in Figure 2, our analysis indicates that all our 3 groups of literary data tend to be skewed toward lower entropy, meaning that we predominantly observe a tail distribution at the entropy levels of highly imageable sentences. 526 527 528

6.3 Experiment III: Comparing Poems 529

In our final experiment, we compare Imagist poems to an assorted set of Modern Love poems not constrained by any specific literary movement. Previous research by J. T. Kao and Jurafsky (2015) found that, compared to 19th-century poetry, Imagist poetry exhibits higher levels of object mentions, abstraction, imageability, and concreteness – particularly when measured using the MRC Imageability Dictionary Coltheart 1981 and the Brysbaert Concreteness Lexicon Brysbaert et al. 2014 – both of which are also used in this study. To ensure consistency with previous methodologies, we compute 530 531 532 533 534 535 536

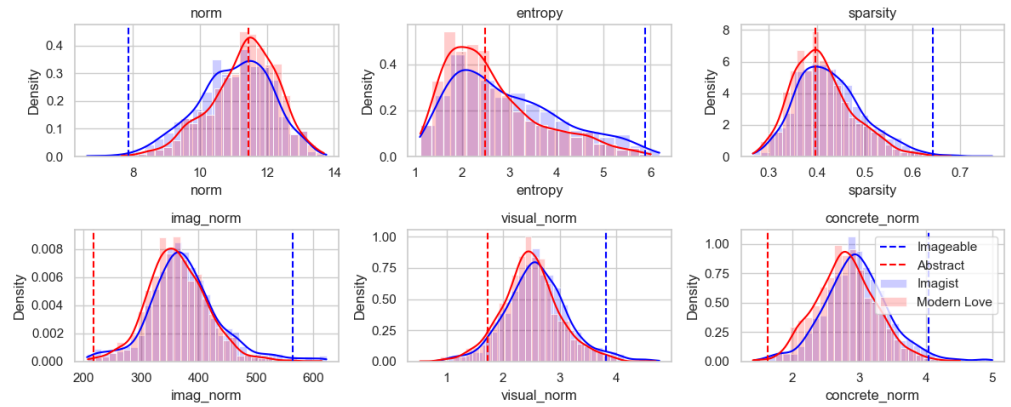


Figure 3: Imagist vs. love poems in terms of dictionary features and embedding-based metrics. We use the feature values of the example poem lines among the most and least imageable lines as reference points in each case.

imageability, visuality, and concreteness using the same approach as J. T. Kao and Jurafsky (2015).¹⁶

Figure 3 presents the juxtaposition of the two poetic traditions in terms of their embedding- and dictionary-based features. The lines indicate two poem lines which were found among the top 10 most imageable and bottom 10 least imageable lines of poetry in the full set (Imagist & Love poems). These were:

- Highly imageable: Homespun, dyed butternuts dark gold color.
- Non-imageable: Of insidious intent

Here we are showing where these two sentences are positioned in terms of each measure.

This comparison allows us to assess that embedding-based metrics do distinguish Imagist poetry, and that Imagist poetry does exhibit higher imageability and concreteness at the sentence level, aligning with prior findings on its heightened emphasis on sensory detail and concrete imagery. These findings are supported by conducting a t-test between the groups (Table 6).

	T-test	Mann-Whitney U
embedding norm	-6.69	567900.00
embedding entropy	<u>7.19</u>	<u>783363.00</u>
embedding sparsity	<u>6.44</u>	<u>773000.00</u>
imageability	5.07	734321.50
visuality	4.98	749496.00
concreteness	7.93	796327.00

Table 6: T-test and Mann-Whitney U results (for comparison) for dictionary-based features and embedding-based metrics between Imagist & Love poems groups. The largest statistic for each variable is in bold, and the second-largest is underlined. All tests were significant with $p < 0.01$.

16. Unlike previous studies, however, we conduct our analysis at the poem-line level rather than at the poem level. That is, we calculate the average imageability score across the words in each line that appear in the dictionary, rather than aggregating at the level of entire poems. This allows us to maintain higher granularity and a larger number of data points, providing a more detailed view of how Imagist poetic discourse — not just entire poems — manifests imageability.

7. Discussion 551

Our findings suggest that embedding structure meaningfully reflects psycholinguistic properties of words, particularly in relation to imageability, concreteness, and visuality. Across our experiments, we observe that embedding norm and entropy serve as reliable indicators of a text's sensory specificity, but in a manner contrary to our initial hypothesis (H_1). Instead, our results provide stronger evidence for H_2 , indicating that concrete and imageable words and sentences exhibit higher entropy and lower norm, while abstract words show lower entropy and higher norm. This suggests that highly concrete words are encoded in a more diffuse and broadly distributed manner, engaging multiple representational dimensions, whereas abstract words activate fewer dimensions more intensely, leading to sharper activation peaks (high norm) but a more compressed distribution (low entropy). We see this pattern when visualizing the embeddings with the highest and lowest norms, as well as the highest and lowest entropy, where low-entropy/high-norm embeddings appear to exhibit longer tails and more dimensions with zero values; while high-entropy/low-norm embeddings show values more evenly centered around zero (see [subsection A.1](#), Figures 4-5).

This pattern challenges the assumption that concrete words, due to their contextual constraints, would be more compactly represented in embedding space. Instead, it appears that concreteness leads to a more dispersed activation profile. This might reflect semantic affordances — that is, concrete words can be associated with a rich variety of semantic features, leading to a broader spread across dimensions. In contrast, abstract words tend to be semantically constrained to fewer, high-level conceptual dimensions, which results in embeddings with spiky, high-norm activations concentrated in a limited set of representational axes. This raises new questions about how different embedding models distribute meaning across dimensions — particularly whether multimodal training systematically encourages broader activation patterns for sensory-rich words compared to text-only embeddings (e.g., BERT).

At the sentence level, we observe a similar effect: literary texts exhibit a strong skew toward lower norm and higher entropy, with particularly imageable sentences spreading their representational load across more dimensions, while abstract sentences cluster in sharper, more concentrated regions of the embedding space.

Our analysis of Imagist vs. Modern Love poetry provides further confirmation that the shape of semantic embeddings encapsulates imageability-related psycholinguistic features. Consistent with prior research, we find that Imagist poetry exhibits higher overall imageability and concreteness, with embedding structures reflecting a more diffuse, multimodal distribution (low-norm/high-entropy). This finding reinforces the idea that imageability is not merely a product of genre conventions but is actively shaped by individual sentence composition. In contrast, Modern Love poetry — while still employing rich figurative language — tends to contain more conceptual abstraction and affective expression, which aligns with a more sharply clustered, locally spiky, embedding representation — reflected in the generally high-norm/low-entropy shape of their embeddings.

Taken together, these findings suggest that norm/entropy may act as a dictionary-free proxy for readers' experience of vividness. The Imagist–Love case study demonstrates

genre-level separability even under lexical control, indicating that the signal is not 595
 reducible to word-level concreteness counts. At the same time, benchmarking against 596
 legacy dictionaries offers an interpretable bridge to prior literature. 597

8. Conclusion 598

Our study suggests that the computational representation of sensory experience in 599
 embeddings follows distinct structural patterns for concrete vs. abstract language. 600
 Specifically, we find that highly concrete and imageable words exhibit greater entropy 601
 and lower norm, reflecting a more distributed, multimodal representation, whereas 602
 abstract words show lower entropy and higher norm, indicative of sharper, more lo- 603
 calized activation patterns. These findings challenge our original assumptions about 604
 the compactness of concrete word representations and the way linguistic meaning is 605
 distributed across high-dimensional embedding spaces. Moreover, our results reinforce 606
 the role of multimodal models like CLIP in capturing sensorimotor properties, while 607
 text-only models like BERT appear to encode imageability less systematically. 608

Finally, if this approach is valid, it can constitute a method for dictionary-free inference 609
 on text imageability. Once the mapping from embedding shape (norm, entropy) to 610
 an imageability score is learned, no external lexicon is required at inference time. Any 611
 sentence—whether it contains out-of-vocabulary words, creative neologisms, or code- 612
 switched phrases—can be scored in a single forward pass through a pre-trained model.¹⁷ 613

A key next step is to directly evaluate embedding-based metrics – alongside dictionary- 614
 derived features – against human judgments of sentence imageability. While our study 615
 establishes that embedding norms and entropies exhibit trends similar to dictionary- 616
 based imageability and concreteness, it remains unclear how well these computational 617
 features actually predict human-perceived sensory vividness. 618

This issue is particularly pressing because dictionaries, though widely used in psycholin- 619
 guistics and NLP, are inherently limited – especially when extended to sentence-level 620
 interpretation, where contextual and compositional effects play an important role to 621
 their human interpretation. We have demonstrated the relationship between these met- 622
 rics, but it remains an open question whether embeddings might actually outperform 623
 lexicon-based methods in capturing human imageability judgments – or whether they 624
 introduce biases or artifacts not present in traditional resources. 625

Further work should also explore a broader range of multimodal architectures, including 626
 models with more fine-grained visual-text alignment (e.g., DALL·E’s prior networks, 627
 BLIP, or fine-tuned vision-language transformers). 628

If multimodal embeddings systematically encode sensory experience, they could offer 629
 a scalable alternative to the hand-annotated psycholinguistic resources that are costly 630
 and relatively limited in scope. This is particularly relevant for literary studies, where 631
 large-scale human annotation of imageability, concreteness, or perceptual vividness 632

17. Scalability therefore stems from the billions of image–text pairs used to fit CLIP, not from the 5 k–40 k entries of psycholinguistic dictionaries. In this sense our approach is data-driven in deployment, and we use the legacy lexica only as a hold-out benchmark during evaluation. The distinction mirrors practice in automatic speech-recognition research, where acoustic models are trained on broadcast audio but validated against a much smaller, human-transcribed test set.

remains impractical outside of standard use of modern English and few other languages. 633

Further validation is needed before applying these embedding-based metrics to broader 634
literary studies – but this may perhaps also be said of applying imageability dictionaries 635
at the sentence level to broader literary studies. If proven reliable, these methods 636
could enable large-scale investigations into the evolution of prose styles, genre-specific 637
imageability trends, and historical shifts in literary sensory encoding. Additionally, it 638
would be valuable to compare literary texts to non-literary domains, such as journalistic 639
writing, political rhetoric, or scientific discourse, to better understand how imageability 640
and perceptual concreteness vary across communicative registers. 641

A. Data Availability 642

Data can be found here: [https://github.com/centre-for-humanities-computing/i](https://github.com/centre-for-humanities-computing/imageability_jcls) 643
[mageability_jcls](https://github.com/centre-for-humanities-computing/imageability_jcls) 644

B. Software Availability 645

Software can be found here: [https://github.com/centre-for-humanities-computi](https://github.com/centre-for-humanities-computing/imageability_jcls) 646
[ng/imageability_jcls](https://github.com/centre-for-humanities-computing/imageability_jcls) 647

C. Author Contributions 648

Yuri Bizzoni: Conceptualization, Methodology, Formal analysis, Validation, Resources, 649
Writing 650

Pascale Feldkamp: Conceptualization, Methodology, Formal analysis, Resources, Visu- 651
alization, Writing 652

Kristoffer L. Nielbo: Methodology, Formal analysis, Validation, Funding aquisition, 653
Writing 654

References 655

Auracher, Jan and Hildegard Bosch (Dec. 2016). “Showing with words: The influence 656
of language concreteness on suspense”. In: *Scientific Study of Literature* 6.2, 208–242. 657
ISSN: 2210-4372, 2210-4380. [10.1075/ssol.6.2.03aur](https://doi.org/10.1075/ssol.6.2.03aur). [http://www.jbe-platform.c](http://www.jbe-platform.com/content/journals/10.1075/ssol.6.2.03aur) 658
[om/content/journals/10.1075/ssol.6.2.03aur](http://www.jbe-platform.com/content/journals/10.1075/ssol.6.2.03aur) (visited on 04/10/2024). 659

Bizzoni, Yuri, Pascale Feldkamp Moreira, Ida Marie S. Lassen, Mads Rosendahl Thom- 660
sen, and Kristoffer Nielbo (May 2024). “A Matter of Perspective: Building a Multi- 661
Perspective Annotated Dataset for the Study of Literary Quality”. In: *Proceedings* 662
of the 2024 Joint International Conference on Computational Linguistics, Language Re- 663
sources and Evaluation (LREC-COLING 2024). Ed. by Nicoletta Calzolari, Min-Yen 664
Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue. Torino, 665
Italia: ELRA and ICCL, 789–800. <https://aclanthology.org/2024.lrec-main.71> 666
(visited on 09/28/2024). 667

- Blackwell, Simon E. (2020). "Emotional Mental Imagery". In: *The Cambridge Handbook of the Imagination*. Ed. by Anna Abraham. Cambridge Handbooks in Psychology. Cambridge: Cambridge University Press, 241–257. ISBN: 978-1-108-42924-5. [10.1017/9781108580298.016](https://doi.org/10.1017/9781108580298.016). <https://www.cambridge.org/core/books/cambridge-handbook-of-the-imagination/emotional-mental-imagery/E619467FC94941DF59F5ED37434335DD> (visited on 02/04/2025).
- Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (Sept. 2014). "Concreteness ratings for 40 thousand generally known English word lemmas". In: *Behavior Research Methods* 46.3, 904–911. ISSN: 1554-3528. [10.3758/s13428-013-0403-5](https://doi.org/10.3758/s13428-013-0403-5). <https://doi.org/10.3758/s13428-013-0403-5> (visited on 01/21/2025).
- Burroway, Janet (1987). *Writing Fiction: A Guide to Narrative Craft*. Little, Brown. ISBN: 978-0-316-11770-8.
- Chen, Yufeng, Guanghui Yue, Weide Liu, Chenlei Lv, Ruomei Wang, Fan Zhou, and Baoquan Zhao (2025). "Predicting Plain Text Imageability for Faithful Prompt-Conditional Image Generation". In: *PRICAI 2024: Trends in Artificial Intelligence*. Ed. by Rafik Hadfi, Patricia Anthony, Alok Sharma, Takayuki Ito, and Quan Bai. Singapore: Springer Nature, 89–95. ISBN: 978-981-9601-22-6. [10.1007/978-981-96-0122-6_9](https://doi.org/10.1007/978-981-96-0122-6_9).
- Coltheart, Max (1981). "The MRC psycholinguistic database". In: *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 33A.4. Place: United Kingdom Publisher: Taylor & Francis, 497–505. ISSN: 1464-0740. [10.1080/14640748108400805](https://doi.org/10.1080/14640748108400805).
- Cortese, Michael J. and April Fugett (Aug. 2004). "Imageability ratings for 3,000 monosyllabic words". In: *Behavior Research Methods, Instruments, & Computers* 36.3, 384–387. ISSN: 1532-5970. [10.3758/BF03195585](https://doi.org/10.3758/BF03195585). <https://doi.org/10.3758/BF03195585> (visited on 01/21/2025).
- Daoshan, MA and Zhang Shuo (2014). "A Discourse Study of the Iceberg Principle in *A Farewell to Arms*". In: *Studies in Literature and Language* 8.1, 80–84.
- Dellantonio, Sara, Claudio Mulatti, Luigi Pastore, and Remo Job (July 2014). "Measuring inconsistencies can lead you forward: Imageability and the x-ception theory". In: *Frontiers in Psychology* 5, 708. ISSN: 1664-1078. [10.3389/fpsyg.2014.00708](https://doi.org/10.3389/fpsyg.2014.00708). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4097956/> (visited on 02/05/2025).
- Ellen J. Esrock (1994). *The reader's eye*. Johns Hopkins University Press. ISBN: 978-0-8018-4669-4. <http://archive.org/details/readerseyevisual00esro> (visited on 02/06/2025).
- Ellis, Nick (Jan. 1991). "Chapter 21 Word meaning and the links between the verbal system and modalities of perception and imagery or In verbal memory the eyes see vividly, but ears only faintly hear, fingers barely feel and the nose doesn't know". In: *Advances in Psychology*. Ed. by Robert H. Logie and Michel Denis. Vol. 80. Mental Images in Human Cognition. North-Holland, 313–329. [10.1016/S0166-4115\(08\)60521-X](https://doi.org/10.1016/S0166-4115(08)60521-X). <https://www.sciencedirect.com/science/article/pii/S016641150860521X> (visited on 01/21/2025).
- Feldkamp, Pascale, Ea Lindhardt Overgaard, Kristoffer Laigaard Nielbo, and Yuri Bizzoni (2024). "Sentiment Below the Surface: Omissive and Evocative Strategies in Literature and Beyond". In: *Computaitonal Humanities Research 2024*. Forthcoming. CEUR Workshop Proceedings.
- Fludernik, Monika (1996). "Towards a 'Natural' Narratology". In: *JLSE* 25.2, 97–141. ISSN: 0341-7638, 1613-3838. [10.1515/jlse.1996.25.2.97](https://doi.org/10.1515/jlse.1996.25.2.97). (Visited on 07/07/2024).

- Gleason, Daniel W. (2007). "Seeing Imagism: A Poetics of Literary Visualization". Ph.D. Dissertation. Evanston, Ill: Northwestern University. 715
- (Sept. 2009). "The Visual Experience of Image Metaphor: Cognitive Insights into Imagist Figures". In: *Poetics Today* 30.3, 423–470. ISSN: 0333-5372, 1527-5507. 10.1215/03335372-2009-002. <https://read.dukeupress.edu/poetics-today/article/30/3/423/20994/The-Visual-Experience-of-Image-Metaphor-Cognitive> (visited on 02/05/2025). 716
- Goetz, Ernest T., Mark Sadoski, Michael L. Stowe, Thomas G. Fetsco, and Susan G. Kemp (Sept. 1993). "Imagery and emotional response in reading literary text: Quantitative and qualitative analyses". In: *Poetics* 22.1-2, 35–49. ISSN: 0304422X. 10.1016/0304-422X(93)90019-D. <https://linkinghub.elsevier.com/retrieve/pii/0304422X9390019D> (visited on 02/04/2025). 717
- Grisot, Giulia, Federico Pennino, and J. Berenike Herrmann (2022). "Predicting sentiments and space in Swiss literature using BERT and Prodigy". In: <https://pub.uni-bielefeld.de/record/2969114> (visited on 02/05/2025). 718
- Hessel, Jack, David Mimno, and Lillian Lee (June 2018). "Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Ed. by Marilyn Walker, Heng Ji, and Amanda Stent. New Orleans, Louisiana: Association for Computational Linguistics, 2194–2205. 10.18653/v1/N18-1199. <https://aclanthology.org/N18-1199/> (visited on 01/19/2025). 719
- Kao, Justine and Dan Jurafsky (June 2012). "A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry". In: *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Ed. by David Elson, Anna Kazantseva, Rada Mihalcea, and Stan Szpakowicz. Montréal, Canada: Association for Computational Linguistics, 8–17. <https://aclanthology.org/W12-2502/> (visited on 01/20/2025). 720
- Kao, Justine T. and Dan Jurafsky (Oct. 2015). "A computational analysis of poetic style: Imagism and its influence on modern professional and amateur poetry". In: *Linguistic Issues in Language Technology* 12.3. <https://aclanthology.org/2015.lilt-12.3/>. 721
- Kuzmičová, Anežka (2014). "Literary Narrative and Mental Imagery: A View from Embodied Cognition". In: *Style* 48.3. Publisher: Penn State University Press, 275–293. ISSN: 0039-4238. <https://www.jstor.org/stable/10.5325/style.48.3.275> (visited on 02/06/2025). 722
- Lacey, Simon and Rebecca Lawson (2013). "Introduction". In: *Multisensory Imagery*. Ed. by Simon Lacey and Rebecca Lawson. New York, NY: Springer, 1–8. ISBN: 978-1-4614-5879-1. 10.1007/978-1-4614-5879-1_1. https://doi.org/10.1007/978-1-4614-5879-1_1 (visited on 02/04/2025). 723
- Liu, Xingyu "Bruce", Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, and Ruofei Du (Apr. 2023). "Visual Captions: Augmenting Verbal Communication with On-the-fly Visuals". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, 1–20. ISBN: 978-1-4503-9421-5. 10.1145/3544548.3581566. <https://dl.acm.org/doi/10.1145/3544548.3581566> (visited on 01/21/2025). 724
- Lynott, Dermot, Louise Connell, Marc Brysbaert, James Brand, and James Carney (June 2020). "The Lancaster Sensorimotor Norms: multidimensional measures of percep-

- tual and action strength for 40,000 English words". In: *Behavior Research Methods* 762
52.3, 1271–1291. ISSN: 1554-3528. [10.3758/s13428-019-01316-z](https://doi.org/10.3758/s13428-019-01316-z). 763
- Magyari, Lilla, Anne Mangen, Anežka Kuzmičová, Arthur M. Jacobs, and Jana Lüdtké 764
(2020). "Eye movements and mental imagery during reading of literary texts with dif- 765
ferent narrative styles". In: *Journal of Eye Movement Research* 13.3, 10.16910/jemr.13.3.3. 766
ISSN: 1995-8692. [10.16910/jemr.13.3.3](https://doi.org/10.16910/jemr.13.3.3). [https://www.ncbi.nlm.nih.gov/pmc/art- 767
icles/PMC7886417/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7886417/) (visited on 02/06/2025). 768
- Martínez, María-Angeles (Jan. 2024). "Imagining emotions in storyworlds: physiological 769
narrated perception and emotional mental imagery". In: *Frontiers in Human Neu- 770
roscience* 18, 1336286. ISSN: 1662-5161. [10.3389/fnhum.2024.1336286](https://doi.org/10.3389/fnhum.2024.1336286). [https://w 771
ww.frontiersin.org/articles/10.3389/fnhum.2024.1336286/full](https://www.frontiersin.org/articles/10.3389/fnhum.2024.1336286/full) (visited on 772
02/04/2025). 773
- Paivio, Allan, John C. Yuille, and Stephen A. Madigan (1968). "Concreteness, imagery, 774
and meaningfulness values for 925 nouns". In: *Journal of Experimental Psychology* 775
76.1. Place: US Publisher: American Psychological Association, 1–25. ISSN: 0022-1015. 776
[10.1037/h0025327](https://doi.org/10.1037/h0025327). 777
- Paragini, Margherita and Mike Kestemont (2022). "The roots of doubt : fine-tuning a 778
BERT model to explore a stylistic phenomenon". In: *Proceedings of the Computational 779
Humanities Research Conference 2022 (CHR 2022), 12-14 December, 2022, Antwerp, 780
Belgium*. CEUR workshop proceedings ; 3290. CEUR-WS.org, p. 72–91. [https://ane 781
t.be/record/opacirua/c:irua:192413](https://ane.t.be/record/opacirua/c:irua:192413). 782
- Pound, Ezra (1913). "A Few Don'ts by an Imagiste". In: *Poetry* 1.6, 200–206. ISSN: 0032- 783
2032. <https://www.jstor.org/stable/20569730> (visited on 07/07/2024). 784
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sand- 785
hini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen 786
Krueger, and Ilya Sutskever (July 2021). "Learning Transferable Visual Models From 787
Natural Language Supervision". In: *Proceedings of the 38th International Conference on 788
Machine Learning*. ISSN: 2640-3498. PMLR, 8748–8763. [https://proceedings.mlr.p 789
ress/v139/radford21a.html](https://proceedings.mlr.press/v139/radford21a.html) (visited on 01/21/2025). 790
- Reilly, Jamie and Jacob Kean (2007). "Formal Distinctiveness of High- and Low- Im- 791
ageability Nouns: Analyses and Theoretical Implications". In: *Cognitive Science* 31.1. 792
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1080/03640210709336988>, 157– 793
168. ISSN: 1551-6709. [10.1080/03640210709336988](https://doi.org/10.1080/03640210709336988). [https://onlinelibrary.wiley 794
.com/doi/abs/10.1080/03640210709336988](https://onlinelibrary.wiley.com/doi/abs/10.1080/03640210709336988) (visited on 01/20/2025). 795
- Sharma, Pratyusha, Tamar Rott Shaham, Manel Baradad, Stephanie Fu, Adrian Rodriguez- 796
Munoz, Shivam Duggal, Phillip Isola, and Antonio Torralba (Jan. 2024). *A Vision 797
Check-up for Language Models*. arXiv:2401.01862 [cs]. [10.48550/arXiv.2401.01862](https://doi.org/10.48550/arXiv.2401.01862). 798
<http://arxiv.org/abs/2401.01862> (visited on 02/05/2025). 799
- Sharma Paudyal, Homa Nath (July 2023). "The Use of Imagery and Its Significance in 800
Literary Studies". In: *The Outlook: Journal of English Studies* 14, 114–127. ISSN: 2773- 801
8124, 2565-4748. [10.3126/ojes.v14i1.56664](https://doi.org/10.3126/ojes.v14i1.56664). [https://www.nepjol.info/index.ph 802
p/ojes/article/view/56664](https://www.nepjol.info/index.php/ojes/article/view/56664) (visited on 02/04/2025). 803
- Silva, Kanishka, Burcu Can, Frédéric Blain, Raheem Sarwar, Laura Ugolini, and Ruslan 804
Mitkov (July 2023). "Authorship Attribution of Late 19th Century Novels using GAN- 805
BERT". In: *Proceedings of the 61st Annual Meeting of the Association for Computational 806
Linguistics (Volume 4: Student Research Workshop)*. Ed. by Vishakh Padmakumar, Gisela 807
Vallejo, and Yao Fu. Toronto, Canada: Association for Computational Linguistics, 310– 808

320. [10.18653/v1/2023.acl-srw.44](https://aclanthology.org/2023.acl-srw.44). <https://aclanthology.org/2023.acl-srw.44> 809
4/ (visited on 02/05/2025). 810
- Strychacz, Thomas (2002). ““The sort of thing you should not admit”: Ernest Hemingway’s Aesthetic of Emotional Restraint”. In: *Boys Don’t Cry? Rethinking Narratives of Masculinity and Emotion in the U.S.* Ed. by Milette Shamir and Jennifer Travis. Columbia University Press, 141–166. [10.7312/sham12034-009](https://doi.org/10.7312/sham12034-009). 811
812
813
814
- Verma, Gaurav, Ryan Rossi, Christopher Tensmeyer, Jiuxiang Gu, and Ani Nenkova (Dec. 2023). “Learning the Visualness of Text Using Large Vision-Language Models”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, 2394–2408. [10.18653/v1/2023.emnlp-main.147](https://doi.org/10.18653/v1/2023.emnlp-main.147). <https://aclanthology.org/2023.emnlp-main.147/> (visited on 01/19/2025). 815
816
817
818
819
820
- Wu, Si and David Smith (July 2023). “Composition and Deformance: Measuring Imageability with a Text-to-Image Model”. In: *Proceedings of the 5th Workshop on Narrative Understanding*. Ed. by Nader Akoury, Elizabeth Clark, Mohit Iyyer, Snigdha Chaturvedi, Faeze Brahman, and Khyathi Chandu. Toronto, Canada: Association for Computational Linguistics, 106–117. [10.18653/v1/2023.wnu-1.16](https://doi.org/10.18653/v1/2023.wnu-1.16). <https://aclanthology.org/2023.wnu-1.16> (visited on 05/31/2024). 821
822
823
824
825
826
- Wu, Yaru, Yuri Bizzoni, Pascale Moreira, and Kristoffer Nielbo (Mar. 2024). “Perplexing Canon: A study on GPT-based perplexity of canonical and non-canonical literary works”. In: *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*. Ed. by Yuri Bizzoni, Stefania Degaetano-Ortlieb, Anna Kazantseva, and Stan Szpakowicz. St. Julians, Malta: Association for Computational Linguistics, 172–184. <https://aclanthology.org/2024.latechclfl-1.16> (visited on 09/28/2024). 827
828
829
830
831
832
833
- Zhao, Baoquan, Songhua Xu, Shujin Lin, Ruomei Wang, and Xiaonan Luo (July 2019). “A New Visual Interface for Searching and Navigating Slide-Based Lecture Videos”. In: *IEEE Computer Society*, 928–933. ISBN: 978-1-5386-9552-4. [10.1109/ICME.2019.00164](https://doi.org/10.1109/ICME.2019.00164). <https://www.computer.org/csdl/proceedings-article/icme/2019/955200a928/1cd0J1kUety> (visited on 01/21/2025). 834
835
836
837
838

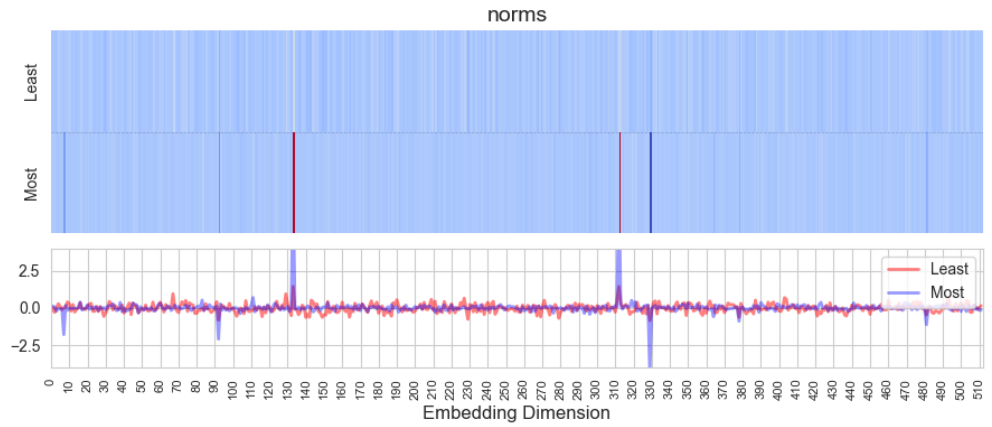
A. Appendix

839

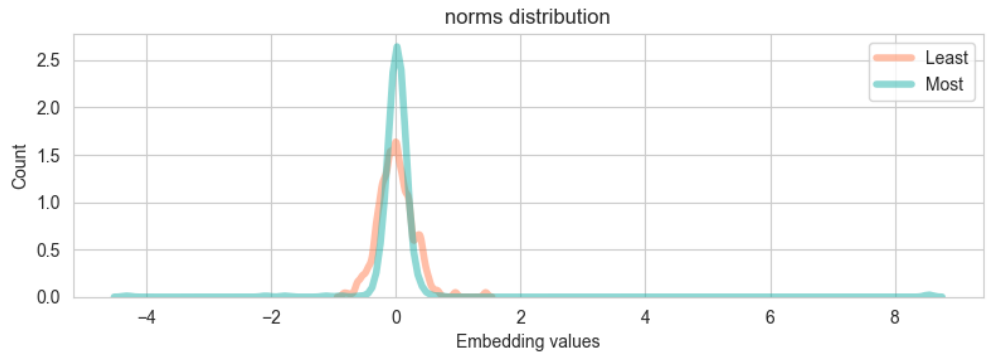
A.1 Norm & entropy of embeddings

840

We visualize the full embedding for extreme cases of norm and entropy values to give an idea of what these measures imply.



(a) Heatmap of the embedding values over dimensions, giving a sense of the strength and density of dimensions.



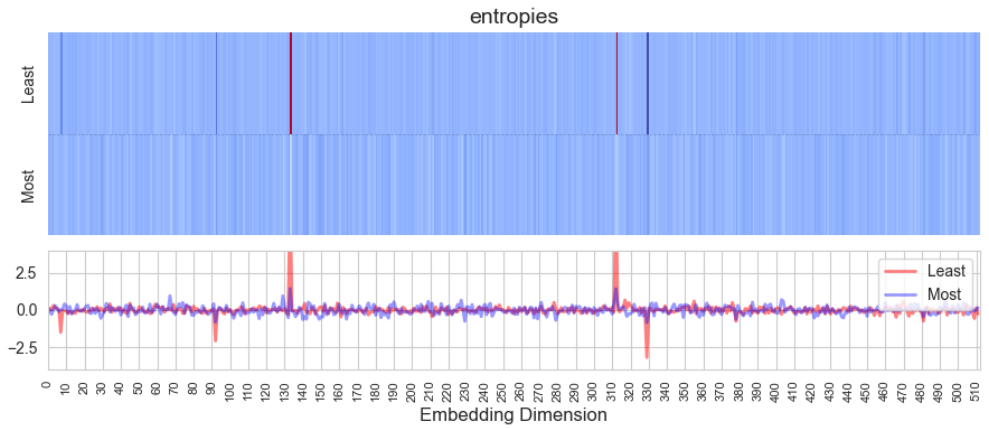
(b) Distribution (kde plot) of embedding values, giving a sense of the range.

Figure 4: The two sentence embeddings with the highest and lowest norm of the 9,000 Chicago corpus sentences. Note how the embedding with *most norm* appears to contain stronger values (i.e., is bluer)(a) and more extreme values (in either direction)(b).

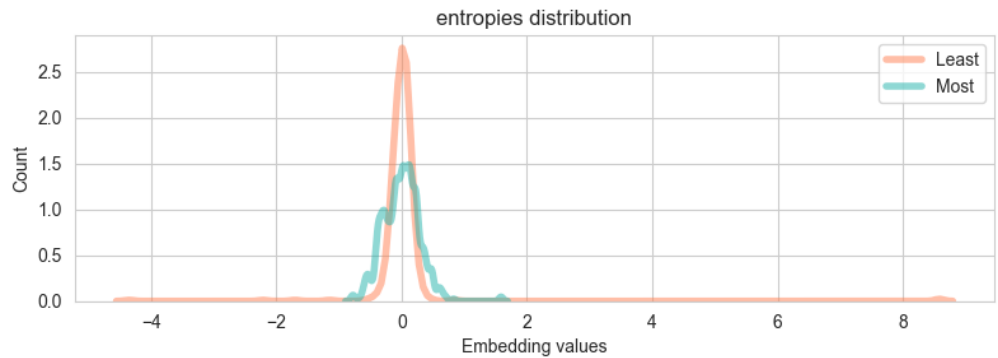
A.2 Experiment III Data

843

conference version



(a) Heatmap of the embedding values over dimensions, giving a sense of the strength and density of dimensions.



(b) Distribution (kde plot) of embedding values, giving a sense of the range.

Figure 5: The two sentence embeddings with **the highest and lowest entropy** of the 9,000 Chicago corpus sentences. Note how the embedding with *least entropy* appears to contain stronger values (i.e., is bluer)(a) and more extreme values (in either direction)(b), mirroring the shape of the embedding with most norm above.

Author	Poem
Michael Anania	Motet
Louise Bogan	To a Dead Lover Leave-Taking Juans Song Epitaph for a Romantic Woman Knowledge Song for the Last Act A Tale
Asil Bunting	from Odes: 30. The Orotava Road
Hart Crane	Voyages from The Bridge: Southern Cross
E. E. Cummings	as freedom is a breakfastfood i carry your heart with me(i carry it in) love is more thicker than forget
Paul Laurence Dunbar	The Old Front Gate A Negro Love Song Invitation to Love Night of Love Thou Art My Lute Song (Wintah, summah, snow er shine)
T. S. Eliot	Portrait of a Lady The Love Song of J. Alfred Prufrock
Kenneth Fearing	Aphrodite Metropolis (2) X Minus X
Ivor Gurney	Photographs
Stephen Spender	Song
James Joyce	Tutto Sciolto
D. H. Lawrence	Last Words to Miriam Gloire de Dijon Cruelty and Love Tortoise Gallantry The Bride Song (Love has crept...) Tortoise Shout
Edgar Lee Masters	Lydia Puckett Lucinda Matlock Mrs. Meyers Sarah Brown
Marjorie Pickthall	Adam and Eve
Carl Sandburg	Bilbea At a Window How Much? New Magic
Kenneth Slessor	The house was just twinkling in the moon light
Gertrude Stein	Idem the Same: A Valentine to Sherwood Anderson
Wallace Stevens	Hymn from a Watermelon Pavilion Peter Quince at the Clavier
Sara Teasdale	Union Square Spring in War-Time The Old Maid Since There Is No Escape The Look Over the Roofs Faults Eight O'Clock Old Love and New Debt
Louis Untermeyer	Infidelity Feuerzauber
Elinor Wylie	Wild Peaches Valentine
William Butler Yeats	When You Are Old Politics The Circus Animals Desertion He wishes his Beloved were Dead Never give all the Heart To an Isle in the Water Reconciliation The Cap and Bells Down By the Salley Gardens The Song of Wandering Aengus Adam's Curse No Second Troy A Drinking Song

Table 7: Poems and Authors of the 'Modern Love' poems

Author	Poem
Richard Aldington	Childhood The Poplar Round-Pond Daisy Epigrams The Faun sees Snow for the First Time
H. D.	Lemures The Pool The Garden Sea Lily Sea Iris Sea Rose Oread Orion Dead
John Gould Fletcher	The Blue Symphony London Excursion
F. S. Flint	Trees Lunch Malady Accident Fragment Houses Eau-Forte
D. H. Lawrence	Ballad of Another Ophelia Illicit Fireflies in the Corn A Woman and Her Dead Husband The Mowers Scent of Irises Green
Amy Lowell	Venus Transiens The Travelling Bear The Letter Grotesque Bullion Solitaire The Bombardment

Table 8: Poems and Authors of the 'Imagist' poems