









Modeling and Measuring Short Text Similarities On the Multi-Dimensional Differences between German Poetry of Realism and Modernism

Anton Ehrmanntraut¹ 
Thora Hagen¹ 
Fotis Jannidis¹ 
Leonard Konle¹ 
Merten Kröncke² 
Simone Winko² 

1. Institut für Deutsche Philologie, Julius-Maximilians-Universität Würzburg , Würzburg, Germany.
2. Seminar für Deutsche Philologie, Georg-August-Universität Göttingen , Göttingen, Germany.

Citation

Anton Ehrmanntraut, Thora Hagen, Fotis Jannidis, Leonard Konle, Merten Kröncke, and Simone Winko (2022). "Modeling and Measuring Short Text Similarities. On the Multi-Dimensional Differences between German Poetry of Realism and Modernism". In: *Journal of Computational Literary Studies* 1 (1). [10.48694/jcls.116](https://doi.org/10.48694/jcls.116)

Date published 2022-12-08

Date accepted 2022-03-30

Date received 2021-12-23


Keywords


short text, similarity, poetry, modernism, realism

License

CC BY 4.0 

Reviewers

Anne-Sophie Bories 

Thomas Weitin 

Note

This paper has passed through the conference track of JCLS. In addition to being peer reviewed, it was presented and discussed at the 1st Annual Conference of Computational Literary Studies 2022 at the Technical University of Darmstadt.

Abstract. This study contributes to the ongoing discussion on how to operationalize text similarity for the purposes of computational literary studies by defining, justifying theoretically and employing a multi-dimensional text model. Additionally, we evaluate a set of strategies to implement this model for very short texts like poetry using a range of methods from weighted sparse vectors up to very recent neural sentence embeddings based on annotations of emotions, genre and similarity. And finally, we show the relevance of using such a complex text model by applying the best method to a research question about the development of early modernism in German poetry. While we can confirm some important hypotheses from literary studies, we are also able to differentiate or relativize others. In particular, our findings do not support the widely held thesis that the change from realism to modernism was a revolutionary 'rupture'.

1. Introduction

This paper pursues two equally important goals: First, to find a suitable state-of-the-art method to model and analyze text similarity for poetry, and second, to contribute to the field of literary studies by studying the transition from realist to modernist poetry using the concept of similarity. The perception of similarity between texts is the basis for the construction of many literary terms like genre, author, or, as in our case, period. Grouping texts according to these terms usually presupposes that these texts have something in common and that these groups can be distinguished via these commonalities from other texts. Though the concept of similarity is ubiquitous in the practice of literary studies it has seldom been analyzed explicitly. While there are several studies that focus on the political aspects of comparison and (dis)similarity analysis (e. g. Felski and Friedman 2013), these contributions generally have little direct connection to our methodological goals. More applicable are some studies by scholars in Comparative Studies who reflected on similarity as part of their discipline defining practice (e.g. Corbineau-Hoffmann 2013). Similar attempts to model text similarity beyond the aspect of content have also been undertaken in computational linguistics (e.g. Bär et al. 2011). So one of the major contributions of this paper is our attempt to bring

these discussion threads together. But while it is possible to discuss these dimensions on a very abstract level, it is not possible to evaluate them on the same level. When we talk about structural aspects of a text, we look at very different elements depending on the genre we look at: speaker, stage directions, *dramatis personae*, etc. for drama, or stanza, verse, rhyme, etc. for poetry. Therefore, in order to discuss the phenomenon not only theoretically, but also to be able to apply it practically – and that means, above all, to include an evaluation method – it is more productive to limit the task to one genre – in our case, poetry.

The second goal of our research is to provide a broad foundation for a literary history of the beginnings of modernism. In the last years, we assembled a corpus of German poetry consisting of poems from realist and modernist anthologies. We are analyzing this corpus under the perspective of whether we can contribute to the discussion about the transition from realism to (early) modernism. We are using these period terms, as is the custom nowadays in literary studies, as useful constructions. That means: On the one hand it is understood that real breaks and disruptions are very rare and that history can be better understood as an evolutionary, gradual process with many small changes at each step. On the other hand, we assume that this process is not happening at the same speed all the time and that many of the changes in one time segment show some commonalities. Specifically, we will use the concept of similarity to describe the changes between the texts from the different corpora.

We structure our paper as follows: In a theoretical section, we first develop a four-dimensional model of textual similarity for poetry ([section 2](#)). We then describe our corpora; mainly the digitized anthologies of the poetry of realism and early modernism mentioned above ([section 3](#)). A selection of these poems was previously manually annotated with a hierarchical system of emotion labels. Within the context of our work, a subset of this selection was then additionally annotated using the dimensions of similarity described in the theoretical section. The following section discusses how each of these four dimensions can be measured in poetry ([section 4](#)). Poetry presents specific computational challenges even for semantics, a relatively traditional dimension of similarity. The main issue poses the shortness of the texts. Semantic similarity, which is usually modeled by using weighted terms to locate a document in vector space, does not work reliably on short texts. Additionally, working with poetry entails having to adapt to its specific language. This includes a high percentage of figurative speech, which makes the analysis of semantic similarity especially difficult, and also a high percentage of archaic words and expressions. For each of the four dimensions of similarity, we discuss and evaluate different methods to measure them using our poetry corpus in a first step. Among our methods used are traditional sparse document vectors, short dense feature vectors, and dense document embeddings, created either by computing them from token vectors or by using the recently proposed approach for sentence embeddings. In a second step, we adapt the best-performing models to each of the four dimensions ([subsection 4.4](#)). In the last section, we employ our final models from this two-step approach to assess the degrees of similarity and difference between realist and modernist poetry ([section 5](#)). In particular, we take up three specific research questions from literary studies and discuss our results with respect to the predominant hypotheses within the field. These questions are:

1. How does naturalist poetry relate to realism and modernism?
2. How homogeneous are realist and modernist poems?
3. How revolutionary is early modernism?

In summary, this study contributes to the ongoing discussion on how to operationalize text for computational literary studies by defining, theoretically justifying, and employing a multi-dimensional model of similarity. Additionally, we evaluate a set of strategies to implement this model for poetry using a range of methods from weighted sparse vectors up to the recent neural sentence embeddings based on extensive annotations of emotions, genre, and similarity. And finally, we show the relevance of using such a complex text-based model by employing the best method to provide new input for the continued research on the development of early modernism in German poetry.

2. Theoretical Considerations

As far as we can see, in literary studies, text similarity has been discussed mainly by Comparative Studies, where the concept of ‘comparison’ has been closely linked to ‘similarity’ (e.g. Zelle 2005). There seems to be a consensus that comparison is only possible on the basis of similarity in some specific aspects. Though principally many different aspects have been and can be used to compare literature, some have been established as especially useful for the study of literature. Corbineau-Hoffmann (2013), for example, groups them under three headings:

- I. Content (1. theme, 2. motifs, 3. settings, 4. characters, 5. concepts)
- II. Text-organization (1. narrative/description, 2. poetry/prose, 3. style levels, 4. instances of speech, 5. discourse)
- III. History (1. influences, 2. epochs, 3. other arts, 4. sciences, 5. genre).

While the first two groups are aspects of a text, the last group refers to typical contexts, often established again by analyzing groups of texts. To avoid the recursive loop hidden here, we focus on the two first aspects, ‘content’ and ‘text-organization’. It is important to note that these are open lists. There are other interesting aspects, but the ones mentioned are often used when people compare literature. The terms grouped under ‘content’ can be seen as parts of text semantics in general. A text has a theme, or there are specific motifs in a text, but usually, the meaning of text is more than each of these, it encompasses all of them. The terms grouped under ‘text-organization’ on the other hand cover quite heterogeneous aspects – even if you substitute the more common ‘form’ for it. In our experience, especially the term ‘style’ is hard to subsume under the same dimension as other text-organizational aspects.

Semiotics and linguistics support this position as they also distinguish between form and style (Nöth 2008; Sandig 2006), and the three aspects – content, structure, and style – are also distinguished in one of the very few attempts in computational linguistics to model text similarity (Bär et al. 2015). We propose to add one dimension which can only be subsumed with difficulties under one of the three headings and which is usually

highly important, especially for literature and especially for poetry, which has been defined as the prototypical medium to express subjective feelings: emotion.¹

Content, Form, Style, and Emotion are the four dimensions of similarity which we will use to describe the relations between texts. From the perspective of this study, it is more useful to explicate the dimensions via operationalizations and examples rather than ‘exact’ definitions. To this end, the annotation guidelines (see section 3) list specific components that make up the dimensions. Content consists of components such as theme, character, or setting; form is operationalized primarily through stanza structure, meter, and rhyme; style, in contrast, refers to components such as register or metaphor; and for emotion, we consider, among other things, the extent to which emotions are represented and their polarity. In further studies, these components could be analyzed individually and be integrated into an even more complex model of text similarity.

The heterogeneity of the four dimensions will have a direct influence on the inter-annotator agreement and the performance of any machine learning model trained to detect these aspects automatically. From a theoretical perspective, it is unclear how the dimensions relate to each other, or in the language of statistics, how much they correlate. Winko (2003), for example, assigns the aspect ‘linguistic shaping of emotions’ via the aspect ‘presentation of emotions’ to what is called ‘style’ in our model, while she assigns it to content via the aspect ‘thematization of emotions’. From this perspective, a relatively high correlation of emotion with content and style is to be expected.

3. Corpus and Annotation

The corpus is a collection of anthologies of contemporary poetry from the two epochs ‘realism’ and ‘modernism’.² The collections contain poems that the anthologists, i.e. contemporary experts in poetry, consider to be particularly typical, outstanding, or representative among other aspects. From the large amount of poetry anthologies in both epochs, the corpus was compiled³ according to the following criteria: The collections contain modern poetry, have a wide range of authors, and are all aimed at a general audience rather than a particular target group. The criteria minimize the risk that thematic constraints or specific addressee orientation could influence the poem selection as systematic factors. It is important to note that the corpus construction leads us to model ‘realism’ and ‘modernism’ from the perspective of the period of investigation. Ultimately, it is the anthologists who determine which texts are ‘realist’ and which are ‘modernist’, and their views do not always have to coincide with the perspective of today’s research. The corpus contains texts by both canonical and non-canonical authors. We call authors ‘canonical’ if they are frequently mentioned in recent literary histories. For early modernism, this applies to Stefan George, Hugo von Hofmannsthal, Arno Holz, Else Lasker-Schüler, and Rainer Maria Rilke.

1. Why emotion is a dimension of its own for the analysis of text is discussed in (Winko 2003).

2. Since this epoch is characterized by a multitude of literary trends, the more neutral label ‘turn of the century around 1900’ is preferred in literary studies. We choose the term ‘modernism’ because the anthologies we include claim to present modern poetry. In the following, ‘modernism’ always means ‘early modernism’, i.e. literature before expressionism.

3. For our corpus selection we used Günter Häntzschel’s comprehensive bibliography (Häntzschel 1991).

(1) Sub-corpus ‘Realism’: The first sub-corpus consists of 7 anthologies with German poems from the realist epoch: Avenarius 1882; Bern 1877; Kneschke 1865; Moltke 1882; Polko 1860; Prutz 1859; Willatzen 1875. The poems included in the anthologies cover the period under study, 1850 to 1880. Some of the anthologies, but especially Elise Polko’s widely distributed collection, also contain some poems written before the period of study; these have been excluded. This sub-corpus consists of 3,039 poems by a total of 484 different authors.

(2) Sub-corpus ‘Modernism’: Of the 941 anthologies of German-language poetry published in first edition between 1885 and 1912 (Häntzschel 1991, 587–589), twelve anthologies meet the selection criteria: Arent 1885; Benzmann 1904; Bethge 1905; Bierbaum 1893, 1894; Bonsels et al. 1905; Federmann 1908; P. Friedrich 1911; Gemmel 1898; Huch 1911; Jacobowski 1899; Renner 1899; Tille 1896. They all claim to contain ‘modern poetry’. This sub-corpus consists of 2,882 poems by a total of 361 authors. We annotated 1,278 poems from both sub-corpora for emotion and thematic genre. Thematic genres such as love poetry or nature poetry provide information about the content of the poems.⁴ The annotated emotions are not the readers’ emotions, but rather the emotions expressed in the text itself. The annotators used a list of 40 discrete emotions which we categorized into 6 groups, inspired by the emotion hierarchy in Shaver et al. 1987: love, joy, agitation/surprise, anger, sadness, and fear. First, emotions and genres were annotated independently by two annotators, then they merged annotations manually into a consensus annotation. Their agreement before creating the consensus annotation, measured with γ (Mathet et al. 2015), was 0.6445 for individual emotions, 0.7491 for the emotion groups, and 0.69 Krippendorff’s alpha (Krippendorff 2011) for the thematic genres.

Additionally, we annotated the similarity of the poems.⁵ The task was not to annotate absolute similarities (“These two poems are not at all/a little/very similar”), but relative similarities (“Poem A is more similar to poem B than poem C”), which is much easier. For each triple of poems, the annotators had to judge for each similarity dimension (content, form, style, and emotion) and for a comprehensive ‘overall’ category whether the focus poem was more similar to the one on the left, to the one on the right, or equally (dis)similar to both. The annotation guidelines specify for each dimension which components should be taken into consideration, e.g. stanza structure, rhyme, meter, and text length in case of the formal dimension, and which of these aspects are typically most important. Nevertheless, the annotators ultimately had to weigh the components on a case-by-case basis, which required considerable literary expertise.

We annotated 470 triples, consisting of a total of 866 poems. One constraint for the selection of the triples was that the poem length had to be quite short due to technical prerequisites.⁶ In addition, we selected triples for which we expected a strong similarity of the middle text with either the left or right text based on formal features such as text length or previous annotations of thematic genres and emotions. This second constraint

4. As the annotation is still ongoing to cover more poems, the entire corpus and a detailed report on the annotation guidelines for emotions and genre will be published at a later date.

5. The annotation guidelines can be found here: https://github.com/cophi-wue/jcls2022-poem-similarity/blob/main/annotation_guidelines_text_similarity.pdf.

6. The length of poems is bound to a maximum of 124 sentence-piece tokens used as input for paraphrase-*xlm-r-multilingual-v1*.

Content	1	0.24	0.68	0.42	0.82
Form	0.24	1	0.22	0.43	0.48
Style	0.68	0.22	1	0.4	0.76
Emotion	0.42	0.43	0.4	1	0.66
Overall	0.82	0.48	0.76	0.66	1
	Content	Form	Style	Emotion	Overall

Figure 1: Pearson correlation in annotated dimensions (majority vote).

ensured that the annotators could deal with reasonably clear cases. There were 400 triples covering both constraints available in our annotated poems. 70 triples were additionally annotated without similarity expectations. Each triple was annotated by at least two people.

The agreement, measured with Krippendorff's alpha, was 0.53 for content, 0.68 for form, 0.44 for style, 0.32 for emotion, and 0.48 for overall. Possible reasons for the differences in agreement are that the dimensions with lower agreement are more dependent on interpretation or that the weighting of the components is more ambiguous in their case. An experiment showed, however, that three annotators who created a consensus annotation after annotating 60 triples were able to increase their agreement when annotating another 30 triples from 0.49 to 0.63 on content, from 0.48 to 0.69 on emotion, from 0.32 to 0.41 on style, and from 0.45 to 0.68 on overall (only the agreement on form deteriorated from 0.77 to 0.71, but still remained high). Since the creation of consensus annotations seems to significantly improve the annotation quality, we plan to create consensus annotations for all triples in the future. Until then, for the triples without consensus annotations, we will only use annotations that the majority of annotators agree with. In the evaluation of the following section, we also omit all annotations that the majority of annotators found that 'the middle text is equally (dis)similar to both the left and the right text'.⁷ That leaves us with 346 usable annotations for content, 388 for form, 331 for style, 359 for emotion, and 381 for overall, with every annotation stating that the middle text is more similar to either the left text or the right text.

Some of the similarity dimensions correlate strongly with each other, according to the annotations. The 'overall' dimension correlates most strongly, especially with content and style. This is understandable since the annotation of the 'overall' dimension is usually based on annotations of the other similarity dimensions. Another relevant correlation exists between content and style. The most independent dimension is form, whose correlation with the other dimensions is the weakest.

7. More precisely, for each triple and similarity dimension, we calculate the mode of the annotation results. We use 'The middle text is more similar to the left text' (from now on: 'left') as the final annotation if the mode is 'left', but also if it is 'left' and at the same time 'The middle text is equally (dis)similar to both the left and the right text'. The same is true in reverse for annotations on the right. All other annotations are discarded.

4. Dimensions of the Similarity of Poems

Measuring the similarity of poems along the dimensions discussed above poses two challenges: First, the shortness of the texts makes it difficult to apply well-established approaches with high reliability. Research in natural language processing has proposed a set of methods for the measurement of short text similarity (Prakoso et al. 2021) usually complementing the texts with other sources which compensate for the lack of information in the text themselves. But research on text similarity, in general, focuses on the ‘content’ aspect. So the second challenge lies in finding methods that can be used to model the other dimensions.

Overviews of the research on short text similarity classify the methods in four groups: string-based, corpus-based, knowledge-based, and hybrid-based (Gomaa and Fahmy 2013; Prakoso et al. 2021, 1): 1) string-based methods use only the word or character tokens to create a representation of the text. We use f_{rVX} and $_X$. 2) Knowledge-based methods use an external knowledge base like WordNet. Our two models $X_{Sf}g_{dM}Z_{ad}_S^{\wedge}$ and $X_{Sf}g_{dM}Z_{af}a^{\wedge}_S$ can be seen as variants of this approach. 3) Corpus-based methods use an external corpus to create information-rich representations, nowadays usually word embeddings: $8S_{af}V_{f}$, $9^{\wedge}ahW94V_{f}$. Additionally, we experiment with document embeddings using different sentence embedding methods: $J >? \check{D}i_b^{\wedge} V_{f} ? [>? t U_{bae}Z_{W}Z_{M}Z_{h}T_{M}S$. The drawback of this approach is that we are limited to an input of 127 tokens, but it is reported to be the best representation for short texts. 4) Hybrid approaches, combining some of the strategies outlined above.⁸

The small number of poems we had annotated under the perspective of similarity made it inadvisable to use the typical finetuning approach. Instead, we opted for broader testing of how different text representations are able to mirror the results from our annotations, select the best representations, and then tweak the vector spaces with similarity learning based on our dimension annotations. So in sections 4.1 to 4.3, we introduce the different models we were able to use and their evaluation based on our similarity annotations. In subsection 4.4, we apply similarity learning to the best performing models.

4.1 Models

We evaluate the following embeddings, which can be roughly categorized into more simple baselines on the one hand, and some more complex embeddings derived from sophisticated deep-learning language models on the other. The baseline embeddings are defined as follows:

$F 8; 6 8 \check{Z}m \quad \dagger \quad \dagger \quad 0$, Poems are represented by a vector, where the dimensions correspond to the 1000 etc. most frequent terms in our corpus. Each individual vector component is the relative term frequency of that term in the poem, weighted by the inverse document frequency.

$? 8l \check{Z}m \quad \dagger \quad \dagger \quad \dagger \quad 0$, Defined like the embedding above, but the term frequencies are z-standardized for each term over all poems.

8. Bär et al. (2015) distinguish between compositional measures, which usually “compute pairwise word similarity between all words, and aggregate the resulting scores to an overall similarity score” (Bär et al. 2015, 5), and non-compositional measures, which project the texts into a shared space like the vector space model (Salton and McGill 1983). We concentrate here on the latter.

\mathbb{R}^4 . Poems are represented by a vector of the following four formal features: stanza count, verse count, word count, average stanza length in verses – each z-standardized over all poems.

\mathbb{R}^7 . Each poem is embedded with a vector of its verse-level relative frequency of shaver emotions (see section 3). These emotions derive either from annotations or, if no annotations are available, predicted by a machine learning model.⁹

The following deep-learning embeddings are derived from pre-trained static type-based embeddings:

\mathbb{R}^4 . For each term in the poem (minus stopwords), we obtain the embedding vector for that term with FastText (trained on the German OSCAR corpus with / # % as proposed by Ehrmanntraut et al. 2021, resp. a GloVe model with / %" provided by Deepset¹⁰). Finally, on that set of vectors, we compute the arithmetic mean (resp. median, resp. meannorm (Ehrmanntraut et al. 2021), resp. arithmetic mean weighted by smooth inverse frequency (Arora et al. 2017)) to obtain a single vector for a particular poem.

Similarly, the following embeddings are derived from the output of BERT, which generates vectors for each token, but also takes into consideration the textual context of the entire input sequence.

\mathbb{R}^{768} . For a particular poem, we plug in the tokenized poem into GBERT_{Base} (Chan et al. 2020), the currently best performing German BERT model. BERT then computes a contextualized output vector (i.e., the output of the last layer) for each token. We now aggregate all vectors by taking the arithmetic mean (resp. median, resp. meannorm), just like above. This results in a vector with 768 dimensions.

$\mathbb{R}^{12 \times 768}$. Defined just like above, except that we not only consider the final output vector, but the outputs of all layers. That is, for each token, we concatenate the input embedding with the 12 Transformer outputs to derive a vector with / #%) (*). Then, like above, we aggregate this sequence of token vectors into a single vector for a particular poem.

In contrast, the following embeddings result from pre-trained language models following a Sentence-BERT-architecture, as proposed by Reimers and Gurevych 2019, 2020.

\mathbb{R}^d . This model is a multilingual XLM-RoBERTa model that is fine-tuned to imitate Sentence-BERT-paraphrases, as described by Reimers and Gurevych 2020. We let paraphrase-XLM-R interpret our poems as sentences, which outputs a vector representation for each poem with /) (*). Note that in the case of paraphrase-XLM-R and all following Sentence-BERT models, fine-tuning was only performed for input sequences no longer than 126 SentencePiece tokens. Therefore, we also restrict our evaluation of these models to poems that are no longer than 126 SentencePiece tokens. (These are 29% of all poems in our corpus.)

9. The model achieves a performance of 0.73 (f1 score).
10. See: <https://www.deepset.ai/german-word-embeddings>.

Similarly, these are pre-trained Sentence-BERT models trained on a wide variety of sentence pair datasets and parallel multilingual data. Specifically, we use publicly available variants paraphrase-multilingual-mpnet-base-v2, paraphrase-multilingual-MiniLM-L12-v2 provided by Reimers and Gurevych 2019, and cross-en-de-roberta-sentence-transformer provided by T-Systems online. Again, we use these on poems of length ≤ 100 SentencePiece tokens to obtain vector representations with ℓ_2 norm (*).

4.2 Evaluation Setup

Evaluating the embeddings as described above requires us to formulate a task that probes each embedding space for its ability to represent certain dimensions of (dis-)similarity of poems via their distances in that particular embedding space, taking into consideration and comparing against the human annotations. As the embeddings define no particular distance function, we evaluate every embedding with each of the following three distance functions: Euclidean (L_2), Manhattan (L_1), and Cosine Distance.

We opted to replicate our prompts for the annotators by formulating a binary classification problem on a particular dimension of similarity, and checking if the model can replicate the majority vote. Note that these votes either take the value ‘annotated left’ or ‘annotated right’. Assume some embedding space and some distance function d fixed. For some annotated triple, let \mathcal{M} , S and Q denote the corresponding vectors in that embedding space. Now, we make the following prediction:

- Predict ‘annotated left’ if $d(\mathcal{M}, S) < d(\mathcal{M}, Q)$; i.e., \mathcal{M} is closer to S than Q .
- Otherwise, predict ‘annotated right’.

To compare the true majority annotations with the predicted ones, we use the F_1 score (arithmetic mean over the recall of both classes, cf. Grandini et al. 2020) as our metric. Note that the random ‘no skill’ classifier has a balanced accuracy score of 0.5.

We remark that variations on the above operationalization are possible as well, particularly if we do not omit cases where the majority of annotators chose ‘The middle text is equally (dis)similar to both the left and the right text’. However, while experimenting we observed that when including this third class ‘same’ in the operationalization, the balanced accuracy significantly drops. (We made the different balanced accuracies comparable by rescaling to the range $[0, 1]$ classes to 1, so that performance at random scoring is always at 0.) We suspect that this difference in performance is caused by the complexity of the triples that were labeled with ‘annotated same’: Human annotators agree on the features which make them classify a text as ‘more similar to the focus text’. But the label ‘same’ is given when neither of both comparison texts shows obvious similarities to the focus text, but that does not imply that the comparison texts have any features in common; they can be different to the focus text in very diverse ways.

In particular, we experimented with the following two variations of the original operationalization:

- (a) Probe whether the embedding can predict ‘annotated equally (dis)similar’ vs ‘an-

notated left or right' by evaluating $\cos(\mathbf{v}_i, \mathbf{v}_j) \geq \cos(\mathbf{v}_i, \mathbf{v}_k)$ against some optimal decision boundary θ .

(b) In a 3-class classification setup, probe whether the embedding space admits a classification using an optimal symmetric decision boundary θ . That is, predict 'annotated left' when $\cos(\mathbf{v}_i, \mathbf{v}_j) \geq \cos(\mathbf{v}_i, \mathbf{v}_k) + \theta$ (\mathbf{v}_i is closer to \mathbf{v}_j than \mathbf{v}_k by at least θ). Symmetric, when $\cos(\mathbf{v}_i, \mathbf{v}_j) \geq \cos(\mathbf{v}_i, \mathbf{v}_k) + \theta$, predict 'annotated right'. And otherwise, when $\cos(\mathbf{v}_i, \mathbf{v}_j) \geq \cos(\mathbf{v}_i, \mathbf{v}_k) - \theta$, predict 'annotated equally (dis)similar'.

As outlined above, variant (a) is solved with lower balanced accuracy than the original operationalization throughout all embeddings and variant (b) with even lower accuracy.

4.3 Results

The results show for all dimensions except 'form' a clear increase with the complexity of text representation: Word Embeddings are better than sparse representations – with dynamic embeddings based on BERT showing a better performance than static embeddings – and sentence embeddings are better than word embeddings. The best sentence embedding is showing an acceptable performance, especially if the cosine is used. As almost all the strategies of text representation, which we applied here, have been developed with the main focus on the semantic aspect, it is not too surprising that the best model is the best in all dimensions. The one big exception is form. Using only a very small set of features is enough to match the annotations. Discussions with the annotators revealed that they usually based their decision on a very small set of observations.

The best model is *bert-base-uncased*. To evaluate all German sentence embedding models,¹¹ which are available at this moment, we use the about 9,000 sentences of the Sick dataset (Marelli et al. 2014) which we had translated into German with DeepL. Our results show *bert-base-uncased* \geq *bert-base-multilingual-uncased* (correlation with human annotations: 0.82) slightly ahead of *bert-base-multilingual-uncased* (0.8165), which is why we include these two models and the best model based on static word embeddings (*glove-wiki-gigaword-6B*) in the next step.

4.4 Similarity Learning

To adapt the text representations to the specific textual dimensions (content, form, style, and emotion), we additionally apply similarity learning. The goal of this step is to learn a transformation of the vectors presented in the previous chapter that allows for better reproduction of the annotation. We use a siamese neural network (Bromley et al. 1993) for this purpose, which we modeled following the maaten network structure from (Szubert et al. 2019). The base model consists of three dense layers (500, 500 and 2,000 neurons) each followed by a normalization activation function (see Klambauer et al. 2017) and dropout. The input for the network consists of our annotated poem triples.

Regardless of the original size of its vector representation, each poem is transformed into a space with 128 dimensions. The loss, and hence the optimization objective of the

11. The multilingual models in Huggingface's sentence transformers; see <https://huggingface.co/sentence-transformers>.

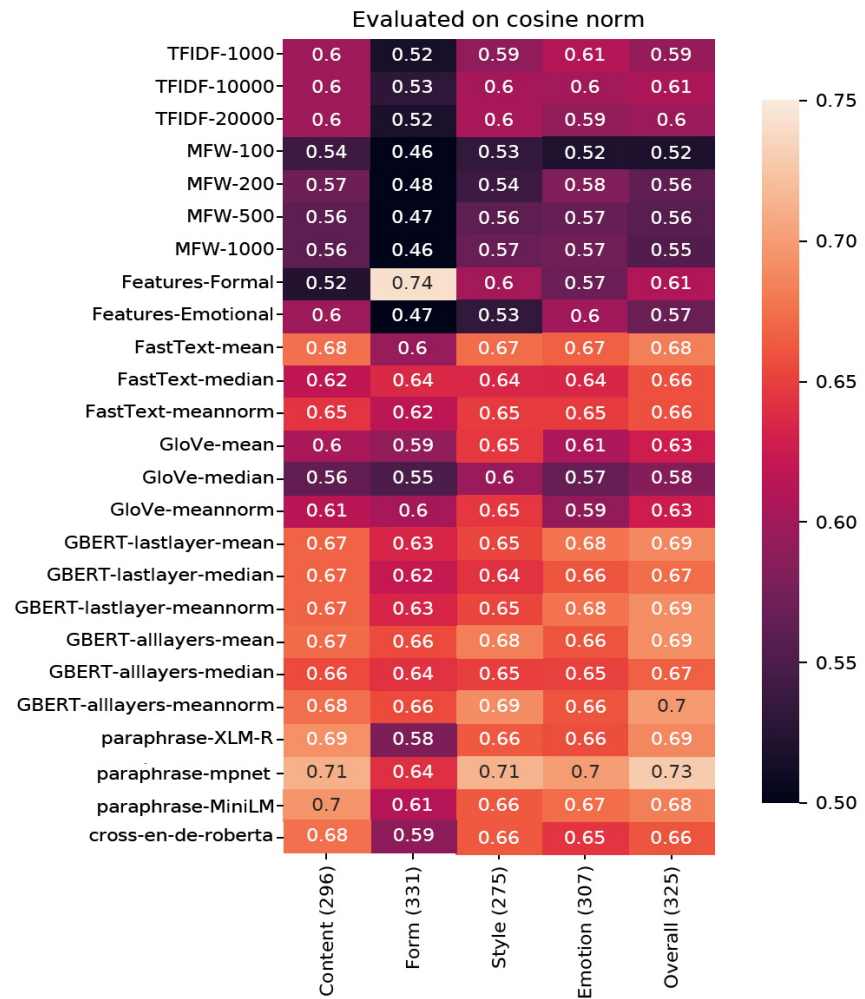


Figure 2: Balanced accuracy score for each model and dimension. Numbers on the x-axis indicate class support. For information on results with other distance metrics see [Figure 12](#) in [section 9](#).

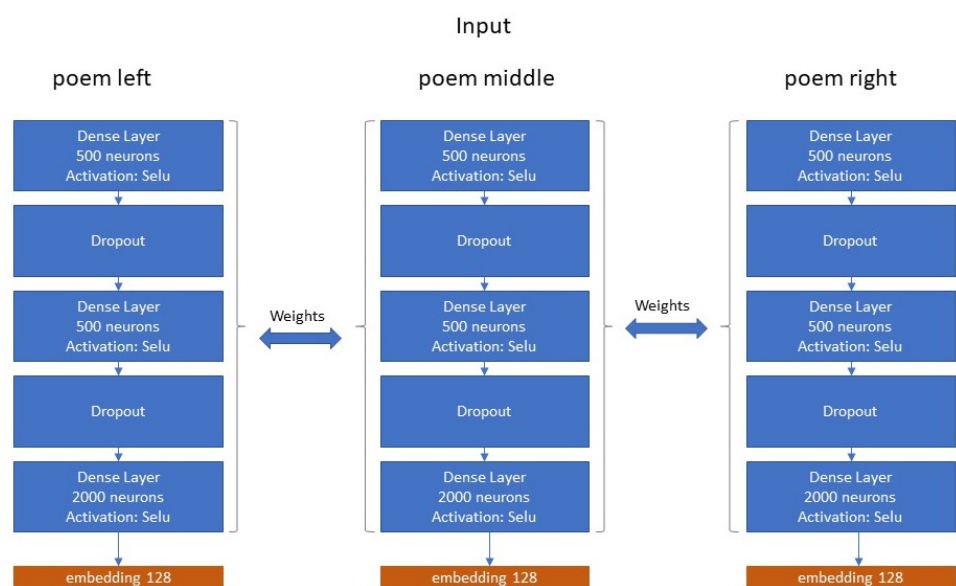


Figure 3: Architecture of the siamese neural network used for similarity learning.

Model	Content	Form	Style	Emotion	Overall
paraphrase-XLM-R	.69→.73	.58→.76	.66→.73	.66→.73	.69→.73
paraphrase-mpnet	.71→.75	.64→.68	.71→.71	.70→.74	.73→.74
FastText-mean	.66→.77	.59→.67	.65→.72	.66→.74	.66→.72
Formal-Features	-	.73→.73	-	-	-

Table 1: Similarity learning results (accuracy in 10-fold cross-validation). Format: best performance before similarity learning (see Figure 2) → performance afterwards.

network, is to maximize the difference between the focus text and the negative example while also minimizing the difference between the focus text and the positive example, i.e. the text which has been annotated as being more similar to the focus text. In short: in Euclidean distances ($\text{dist}(\text{anchor}, \text{negative}) - \text{dist}(\text{anchor}, \text{positive})$).¹² Learning rate decrease is bound to a *cosine* mechanism, which leads to strong performance gains compared to more common choices like constant or time-based decrease rates. The network’s performance is measured via the amount of correctly identified positive examples (accuracy, see Table 1).

4.5 Discussion

With our two-step approach, we are able to achieve good results for a complex task. It is probably open to discussion whether the restriction to 127 input tokens is acceptable compared to the small gain in performance. Future work will either improve on the input size or find a reliable way to compute representations for longer texts. Using one representation for three of the four aspects in the first step made us ask whether the representations after the second step are actually different. The correlations of distances (Figure 4) rate positive correlations between content and style, content and emotion, or style and emotion. In other words, the vector space was attuned to the specific dimension by similarity learning. The close relationship between ‘content’ and ‘overall’ was already noticed by the annotators.

It is unclear to us, why the different embeddings show significantly different improvements in the second step (mean values): 0.126 for *content* vs *overall*, 0.026 for *style* vs *overall*, 0.08 for *emotion* vs *overall*; on what factors does this capability for improvement depend? Which training data and training regime for the sentence embeddings enables the text representation to be adaptable to the text dimensions beyond content?

The results from Figure 2¹³ show that the best results are obtained using language models with transformer architecture and that they increase even more if those have previously been fine-tuned for sentence similarity. With the additional adaptation by similarity learning, we now perform a third tuning step of representations created this way. A next step would be instead of using the frozen output vectors of those networks, to include the network in the learning process and model the similarity learning as a fine-tuning step. Likewise, we should add another layer of pertaining before similarity learning and perform a domain adaptation (Gururangan et al. 2020) to our corpus.

12. Triplet margin loss.

13. Motivated by feedback during the conference, we additionally tested the combination of paraphrase-XLM-R and Formal-Features. This variant leads to an accuracy of 0.78 for the form aspect. It improves the result of paraphrase-XLM-R slightly, but remains below the value achieved by Formal-Features alone.

Content	1	0.6	0.62	0.58	0.79
Form	0.6	1	0.66	0.47	0.64
Style	0.62	0.66	1	0.55	0.71
Emotion	0.58	0.47	0.55	1	0.64
Overall	0.79	0.64	0.71	0.64	1
	Content	Form	Style	Emotion	Overall

Figure 4: Pearson correlation of distances in vector space after similarity learning.

5. (Dis)similarity between the Poetry of Realism and the Poetry of Modernism

5.1 Hypotheses from Literary Studies

In the following, we continue a discussion in German literary studies about the relationship of poems of realism to those of early modernism and the special position of naturalistic poetry in this development. We hope to contribute to this discussion by a mix of explorative methods and hypothesis testing. To enable the latter, we will condense positions in the debate into three hypotheses related to this transformation.

Hypothesis 1: The poetry of naturalism, as represented in the anthology μ

D ; , is predominantly traditional rather than modernist. The question of where exactly naturalism can be located between realism and modernism has been debated many times. In this context, the anthology μ [LZFWZ5ZScS] μ which is part of our corpus, is considered central to naturalist poetry. The anthology's introductions emphatically assert the novelty and revolutionary character of the texts (especially Conradi 1885, I–III). Research, on the other hand, is mainly of the opinion that these statements are exaggerated and that the poetry of the anthology is, on the whole, traditional (Vietta 1992, 194; Fähnders 1998, 36–37; Sprengel 1998, 621; Austermühl 2000, 350–51; Lamping 2000, 145–146; Andreotti 2014, 17). However, some scholars, even if they consider the anthology as a whole to be traditional, argue that it was at least innovative in terms of μ since new themes such as 'big cities' or 'social issues' were addressed (e.g. Fähnders 1998, 36–37).

Hypothesis 2: Modernist poetry is heterogeneous, that is, more heterogeneous than realist poetry.

While the poetry of realism, or at least the mass-produced poetry of this period, is considered by researchers to be relatively homogeneous (e.g. Stockinger 2010, 88), modernist poetry is highly diverse, according to many scholars, given the simultaneity of a wide variety of literary movements (Anz 2007, 330–331; Becker and Kiesel 2007, 30; Fähnders 1998, IX, 4). But the hypothesis of modernist heterogeneity

has its limitations. For example, some researchers support the view that modernism is homogeneous at least insofar as it responds to the same social-cultural problems (Vietta 1992, 30–31.; Fähnders 1998, 9–10; Becker and Kiesel 2007, 30; for further statements on the homogeneity of modernist poetry see H. Friedrich 1992, 140–142; Lamping 2008, 13). One researcher, therefore, argues that the period around 1900 was characterized by a “homogeneity of the heterogeneous” (Fähnders 1998, 11). Despite these limitations, most scholars would probably agree that modernist poetry is at least more heterogeneous than the poetry of realism.

Hypothesis 3: There is a fundamental ‘rupture’ between modernist poetry and earlier, more traditional poetry. This view was already held by contemporary authors, critics, and anthologists, who spoke of a ‘revolution’ in poetry (as an example from the corpus anthologies see Bethge 1905, 13–14; cf. on contemporary statements H. Friedrich 1992, 141; Anz 2007, 333; Lamping 2012; Wieland 2019, 17). Many researchers also emphasize major differences between modernism and previous literary periods, often using the metaphor of ‘rupture’ (H. Friedrich 1992, 20; Kiesel 2004, 141–142; Frick 2007, 97–98; Goltschnigg 2007, 169; Lamping 2012; Andreotti 2014, 5; without this metaphor: Klinger 2002, 160; Lamping 2000, 140; Lamping 2008, 11, 13). But the ‘rupture’-thesis is also partly qualified. For example, it is emphasized that modernism still refers to traditions (even though it uses them in new ways) (Kiesel 2004, 142–143; Frick 2007, 98–99; Goltschnigg 2007, 169). Others argue that many relevant authors were located somewhere *TM* *WW* realism and modernism or that they combined traditional as well as new elements, which implies a smoother transition between periods (see for C. F. Meyer Selbmann 1999, 149, 152; for Fontane (!) Selbmann 2007, 201; for Baudelaire, Rilke, Hofmannsthal, and Kafka Lamping 2012). Still, others relativize the novelty of modernism in general (Hiebel 2005, 27; Anz 2007, 333). Thus, hypothesis 3 is partly controversial in research.

It is possible to combine the aforementioned hypotheses in a visual model. The purpose of this model is threefold: it visually summarizes the research hypotheses, it relates the hypotheses to one another, and it demonstrates that all hypotheses about similarity and dissimilarity combined offer a fairly comprehensive interpretation of the transformation from realism to modernism, again underscoring the relevance of similarity as a category of analysis.

In the model, each point represents a poem. The greater the distances between the points, the more dissimilar the texts. The distances are not based on calculations but on a hermeneutic understanding of the research and are meant as rough approximations of general ideas. One can see that the distances within realism are smaller than the distances within modernism. It is also visible that there is a strong division between realism and modernism and that the naturalist poems tend to gravitate more towards realism than modernism.¹⁴

Admittedly, this model is not explicitly advocated in research. Only rarely does a single scholar state all the hypotheses that the model synthesizes. Like any model, it represents only a section of reality and neglects other aspects, such as the differentiation

14. Distances within naturalism should not be given any further significance; no research hypotheses were considered in this regard.

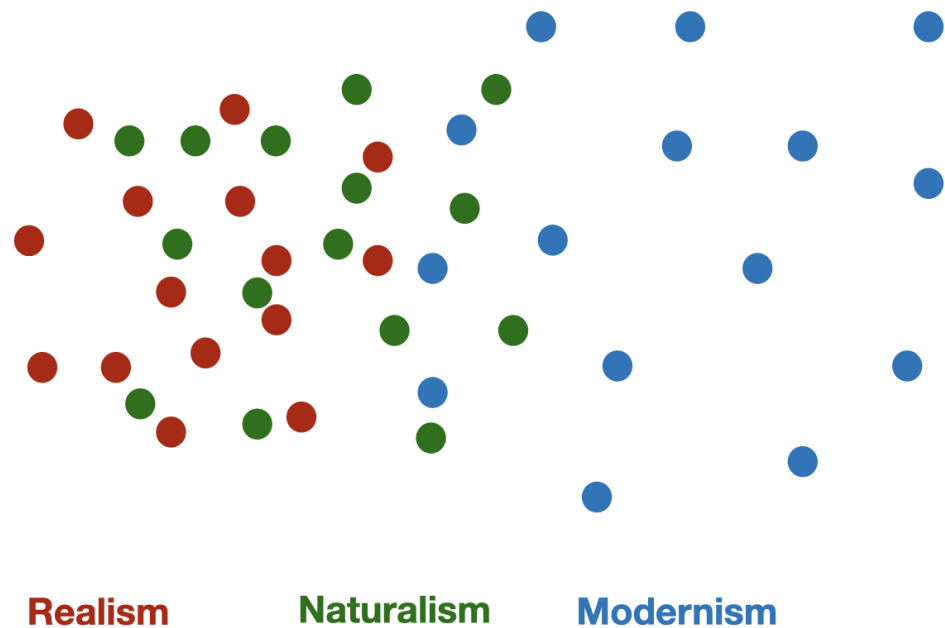


Figure 5: Model of the distances between poems of realism, naturalism and modernism according to research.

of individual dimensions of similarity, or synchronic and diachronic period-internal differentiations of, for example, individual authors, groups of authors, or literary movements. Some aspects of the model are, as explained, controversial in research, but it is all the more interesting to examine whether our results fit the model and the underlying hypotheses.

5.2 Results

For a first exploration of the (dis)similarities between realism and modernism, we project the poems into a two-dimensional space (Figure 6). Some similarities with the model derived from research (Figure 5) become apparent. In particular, a distinction between realism and modernism is evident, even though the separation is far from perfect since there are numerous overlaps between the two periods. Furthermore, it is consistent with the research model that the naturalist poems tend to stay within the realist spectrum and hardly enter ‘decidedly modernist’ areas. However, it is necessary to test the hypotheses from literary studies more precisely than just by explorative means.

Hypothesis 1: The poetry of naturalism, as represented in the anthology μ

D ; , is predominantly traditional rather than modernist To test the first hypothesis, we examined the similarity of the programmatically naturalistic anthology μ to the realism and modernism corpora. In addition, we measured the distances between the poems within the latter two corpora to be able to assess the comparative analyses more accurately.

The boxplots show that overall and for each individual dimension ‘content’, ‘form’, ‘style’, and ‘emotion’ the distances between naturalism and realism are smaller than the

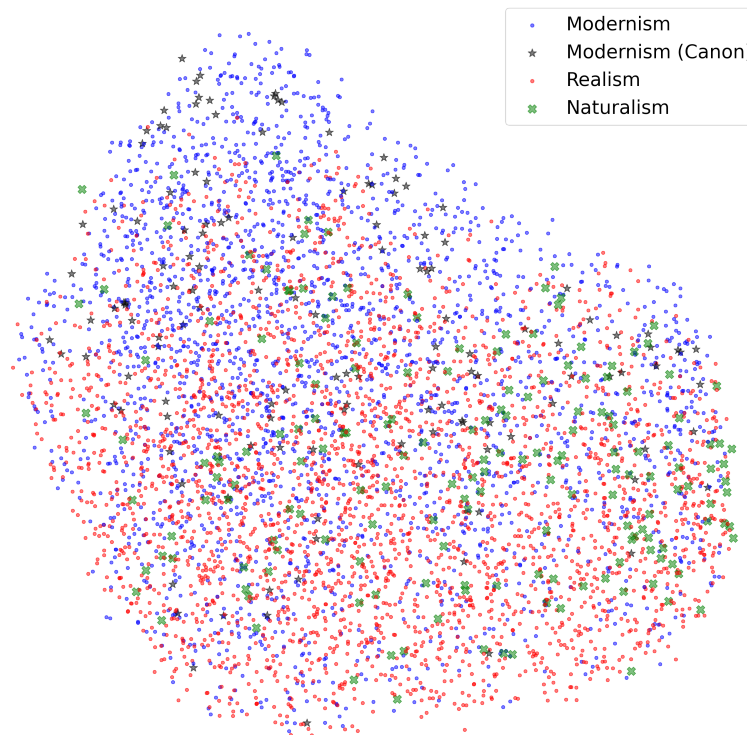


Figure 6: Poems embedded with both vanilla *GBERT-allayers-meannorm* (see Figure 2) and *FastText-meannorm* transformed to reflect the aspect 'content' (see Table 1) projected in 2-dimensional space using UMAP (McInnes et al. 2018).

distances¹⁵ between naturalism and modernism. At the same time, the distance between the naturalism and realism corpus is larger than the distance between the poems within the realism corpus. Surprisingly, in the dimension 'content' no higher proximity to the modernism corpus is seen.

A stronger similarity between naturalist and modernist poems would have been expected based on the literary-historical theses we have mentioned above. As expected, the analyses support the thesis that the naturalism corpus is more similar to the realism corpus than to the modernist corpus. However, a more detailed look shows that differences can be found in the individual dimensions. This could indicate that the naturalistic poems probably do not use the same means as realistic poems. What exactly these differences are should be investigated in a further study. However, equating naturalist with realist poetry falls short in any case since the internal distance in the realism corpus is smaller than that in the comparison between naturalism and realism. It should be emphasized that we have studied the effect only for the anthology *Walt Whitman's Poetry* and only using its short poems, as stated above. Further study would have to take into account that the modernism corpus also contains some naturalistic poems.

15. We tested for significance and all differences are highly significant. To make sure this is not solely an effect of the large sample size we randomly selected 100 texts, but the differences stay significant. New guidelines usually recommend complementing p-values with effect size. In our case this is not easy to apply, because the measure is not grounded in an intuitively comprehensible unit.

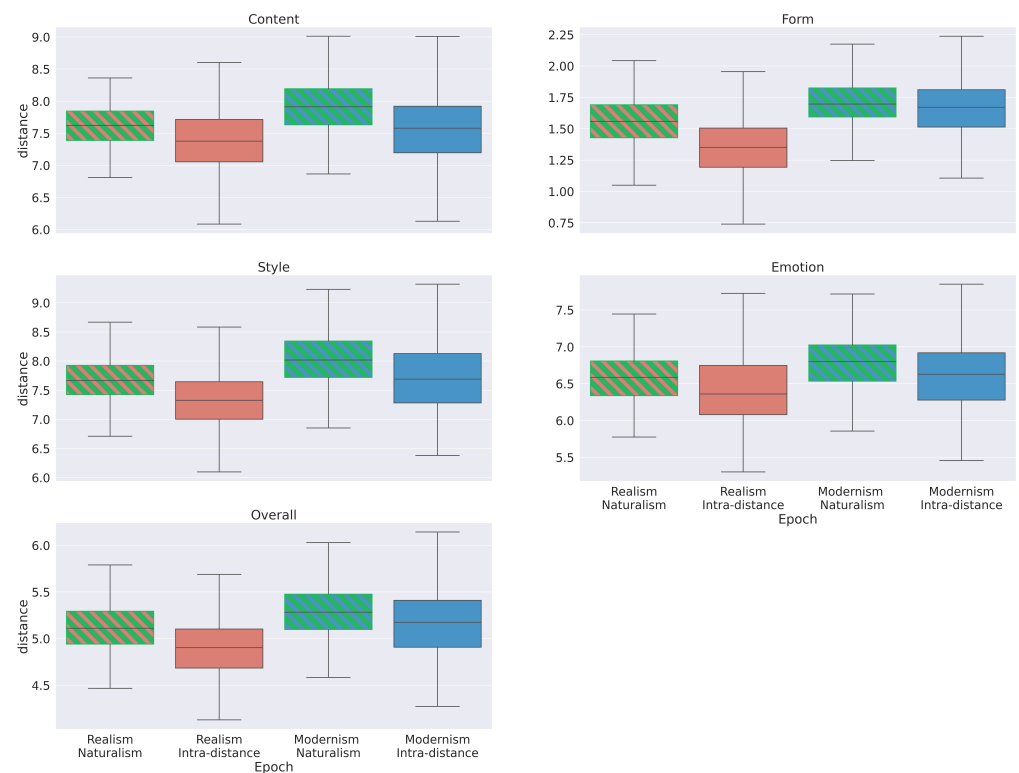


Figure 7: Distances between poems from Realism/Naturalism and Modernism/Naturalism and poems within Realism and Modernism. Distances in ‘content’, ‘style’, ‘emotion’ and ‘overall’ are measured in the space of *paraphrase-XLM-R* embeddings transformed via similarity learning (see section 4.4). Distances in ‘form’ are measured in the Feature-Form embedding space (see section 4.1). Each boxplot represents pairwise euclidean distances of 2,000 samples with a size of 20 poems.

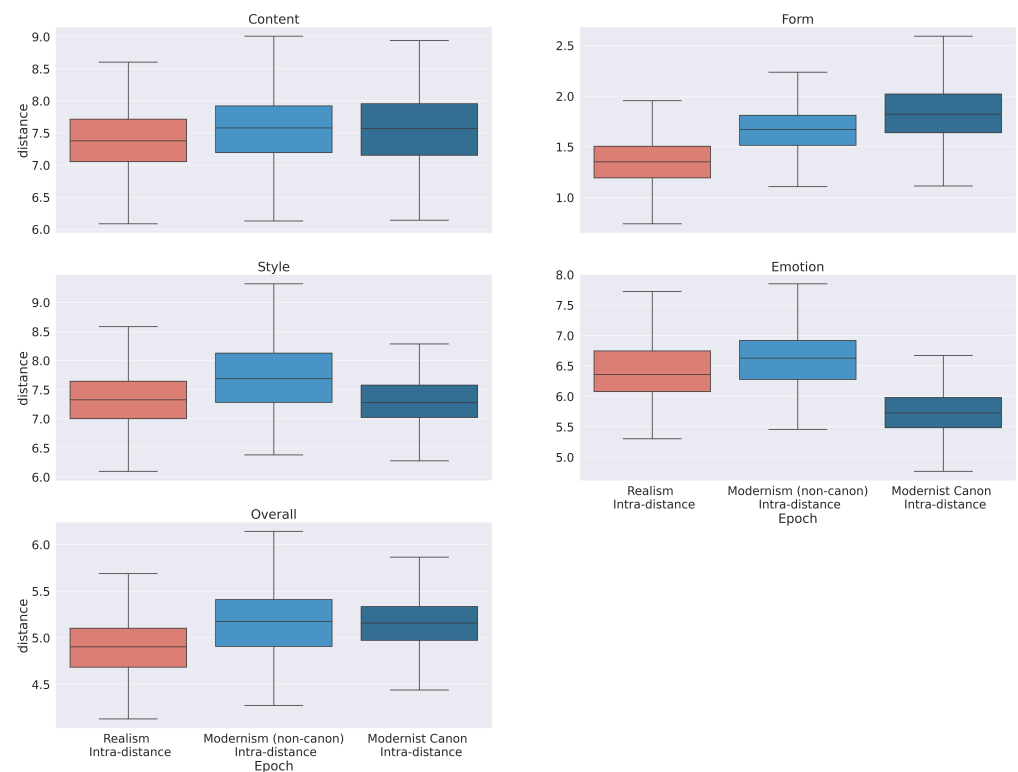


Figure 8: Distances within poems from Realism, Modernism and canonic Modernism. Distances in ‘content’, ‘style’, ‘emotion’ and ‘overall’ are measured in the space of *paraphrase-XLM-R* embeddings transformed via similarity learning (see section 4.4). Distances in ‘form’ are measured in the Feature-Form embedding space (see section 4.1). Each boxplot represents pairwise euclidean distances of 2,000 samples with a size of 20 poems.

Hypothesis 2: Modernist poetry is heterogeneous, that is, more heterogeneous than realist poetry. From now on, when we compare realism with modernism, we no longer include the naturalist poems in our calculations and visualizations, since we have seen that naturalism is located somewhere between realism and modernism. However, we now distinguish in modernist texts canonical and non-canonical authors in order to point out some peculiarities of the canonical poems.¹⁶

To test hypothesis 2, we compare the distances within realism with those within modernism (Figure 8). In all dimensions, the distances within modernism are greater than in realism, most clearly in the dimension 'form'. Thus, the hypothesis that modernist poetry is more heterogeneous than realist poetry can be confirmed by our data. However, the differences in heterogeneity are mostly small and should not be overemphasized. Modernist poems by canonical authors are slightly more heterogeneous than non-canonical poems regarding the dimension 'form'. Otherwise, the canonical poems are not characterized by greater distances among themselves than non-canonical modernist poems. On the contrary, the distances for the dimensions of style and especially emotion are much smaller within the canonical texts than within the non-canonical modernist poems. All in all, the canonical texts are no more heterogeneous than the non-canonical ones. This is surprising, since one might have expected a particularly high degree of individuality and thus heterogeneity in the canon. In any case, it must be kept in mind that the subcorpus of canonical modernist poems is very small (58 poems, 5 authors), which limits the validity of the results. Further research is needed here.

Hypothesis 3: There is a fundamental 'rupture' between modernist poetry and earlier, more traditional poetry. Some difficulties arise in testing hypothesis 3. It is not clear what exactly is meant by 'rupture' and how we should measure it. One possibility is to assume that the term 'rupture' in hypothesis 3 denotes a certain kind of literary change, namely a change that (a) is particularly large compared to other changes between periods (e.g. between romanticism and realism) and that (b) occurs abruptly, i.e. in a very short period of time. However, we do not have data on other changes between literary periods and we cannot analyse whether the shift between realism and modernism was a continuous, decades-long process or a matter of a few years.

So while we are not able to test hypothesis 3 directly, we can at least share some observations that are likely to be related. In particular, we will compare the distances between realism and modernism with distances within realism. If the distances between realism and modernism are greater than within realism, it can at least be said that modernism is different from realism.

In all dimensions, the distances between realism and modernism are larger than the distances within realism. However, the differences are not enormous. Moreover, the two-dimensional plot above (Figure 6) shows that modernist poems appear not only outside realism, but often within the realist spectrum as well. While these observations cannot falsify hypothesis 3 directly, they certainly do not confirm the hypothesis either. If anything, our results suggest that the notion of a fundamental 'rupture' between realism and modernism might be exaggerated.

16. In our study, in accordance with German literary history, Stefan George (6 poems), Hugo von Hofmannsthal (6 poems), Arno Holz (19 poems), Else Lasker-Schüler (3 poems), and Rainer Maria Rilke (24 poems) represent canonical modernism.

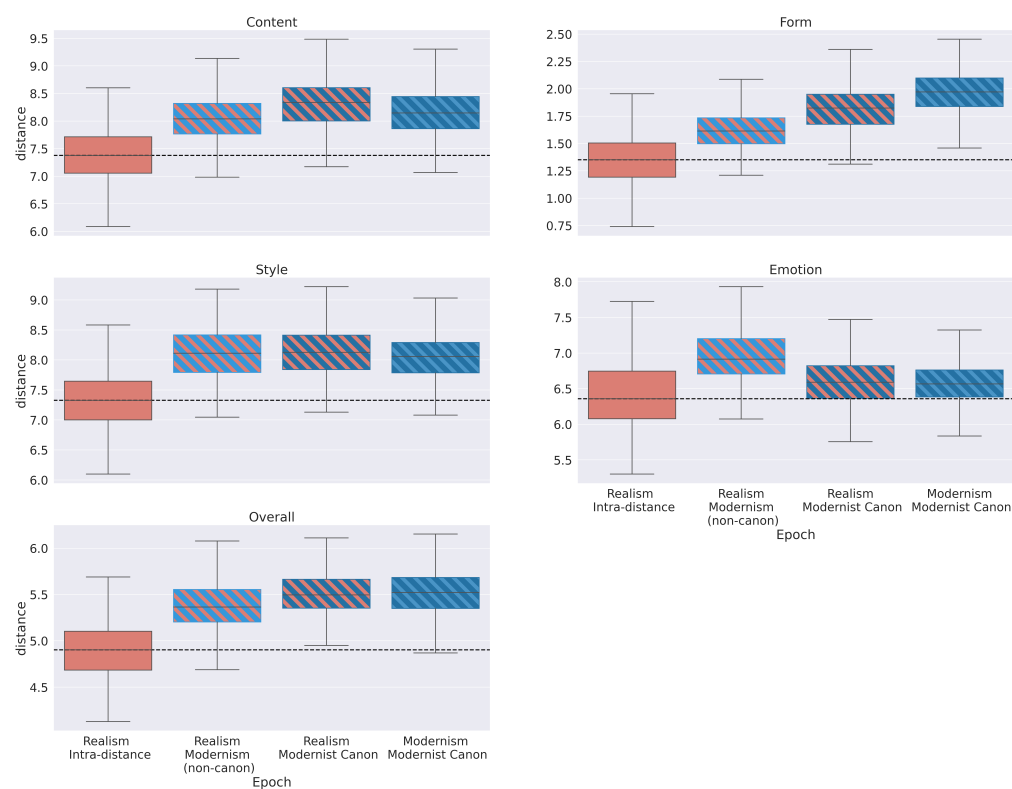


Figure 9: Distances within poems from Realism and between Realism/non-canon Modernism, Realism/canonic Modernism and non-canon Modernism/canonic Modernism. Distances in 'content', 'style', 'emotion' and 'overall' are measured in the space of *paraphrase-XLM-R* embeddings transformed via similarity learning (see section 4.4). Distances in 'form' are measured in the Feature-Form embedding space (see section 4.1). Each boxplot represents pairwise euclidean distances of 2,000 samples with a size of 20 poems.

böÖ" | " e|W që šö-" ö3"œ ÑD| šë" %D3š"3š W3• ÑD|- •ö-"3œ öD3œ •W
œö-ö W|öšö"œ DÑ ,r qD"-œ x œW-q "• ÑD| "W%ë Šýæ"W| šö-"œ qW3x •
Û" %šD pœ/D W dãnW ~E% D3š P3š ñ W3• šë" ÑD|-W Ñ" Wš-" UÖ" %š D3œ ĩÑD|
œ" %š öÑD|š W W|Ö" | Û" |œ öD3 DÑ šëöœ Ö" |" Ö

0OF DPVME BTTVNF UIBU SFTFBS DIFST VTF UIF NFUBQIPS PG
PUIFS OBNFMZ DBOPOJDBM UFYUT 5IFEJTUBODF CFUXFFO S
JOEFFE MBSHFS UIBO UIF EJTUBODF CFUXFFO SFBMJTN BOE
JOH UIF GPSN BOE BU MFBTUB UJOZ CJU MBSHFS GPS UIF EJI
JOUFSNT PG TUZMF DBOPOJDBM NPEFSOJTN JT OP GVSUIF
NPEFSOJTN BOE JO SFHBSET UP FNPUJPO UIF EJTUBODF
NPEFSOJTN JT FWFO TNBMMFS UIBO CFUXFFO SFBMJTN BOE
PVS SFTVMUT EP OPU TIPX UIBU UIF EJTUBODFT CFUXFFO DI
BSF TZTUFNBUJDBMMZ MBSHFS UIBO CFUXFFO OPO DBOPO
JEFB UIBU UIF DBOPOJDBM UFYUT TFU BUSFOE UIBU UIF OP
EFDJTWFMZ DBOOPUCFDPOaSNFE

0OF NJHIU FYQFDU UIF DBOPOJDBM NPEFSOJTU QPFNT UP CFB
NPEFSOJTU UFYUT UIBO UP UIF SFBMJTU QPFNT CVU UIJT
EBUB 5IFEJTUBODFT GSPN DBOPOJDBM NPEFSOJTU QPFN
BOE UP OPO DBOPOJDBM NPEFSOJTU UFYUT PO UIF PUIFS I
UIF DBTF PG UIF EJNFOTJPO qGPSNr UIF DBOPOJDBM NPEFS
SFBMJTU POFT UIBO UP UIF OPO DBOPOJDBM NPEFSOJTU PO

5IF SFTVMUT GPS UIF DBOPO BSF DPVOUFS JOUVJUJWF BOE
PCTFSWBUJPOT NBZ IBWF TPNFUIJOH UP EP XJUI UIF GBDU U
UFYUT JT WFSZ TNBMM BOE UIBU XF POMZ BOBMZ[F TIPSU QP

5P GVSUIFS FYQMPSF UIF EJêFSFODFT CFUXFFO NPEFSOJT
TQBDF XFDPOTUSVDUFE BUJNFMJOF GSPN BHSBQI OFUXPS
BMMQBJSXJTFEJTUBODFT PSTJNJMBSJUJFT NPSFQSFDJTF
BMM EJNFOTJPOT FYDFQU qGPSNr UIF EJTUBODFT SFQI SFTFC BT.F3
BGUFS UIF BEBQUBUJPO XJUITJNJMBSJUJZ MFBSSOJOH 'PS q
TJNJMBSJUJFT XFSF VTFE "MM EJTUBODFT XFSF TUBOEBS
BOE EVF UP UIF EJêFSFOU NFUSJDT VTFE UP EFUFSNJOF U

&BDIOPEF JO UIF HSBQIT SFQSFTFOUT BTQBOPG ZFBST J
5IF FEHF CFUXFFO UXP ZFBS TMJDFT JTEFQJDUFE CZ UIF NI
QPFNT v JG MFTT UIBO QPFNT XFSF BWBJMBCMF QPFNT XI
BMQIB PG POF FEHF CFUXFFO UXP ZFBST WJTVBMJ[FT UIF EF

böÖ" | "æ|Vqë šö-" ö3"Œ ÑD|šë" "-DšöD3 W3• Œšæ " •ö-"3ŒöD3Œ•W
Œö-ö W|öšö"Œ DÑ ,r qD"-Œx ŒW-q "•ÑD| "W%ë Šýæ"W|šö-"ŒqW3x•
Ū"%šDpŒVŒW dānW-Œ)šöŒP| "özŒ "%šöŒD|šW W|Ö" | Ū" |ŒöD3 DÑ šëöŒ

UIF DIPTFO QPFNT 8F POMZ VTFF QPFNT XIFSF UIF DPSSFTQ
DIFDLFE BOE DPSSFDFUFE CZ VT JG OFDFTTBSZ 5IJT BNPVO
BOE TQFDJaDBMMZ

'SPN UIJT WJTVBMJ[BUJPO XIJDI JT CBTFE OPU PO UIF BTTJH
CZ UIF FEJUPST PG UIF BOUIPMPHJFT CVU PO UIF QVCMJD
NBLF TPNF PCTFSWBUJPOT *O UFSNT PG GPSN XFDBO TVSN
QPFNT BSF NPSF TJNJMBS UP FBDI PUIFS BOE UIVT NPSF IPNPH
BSF DPJODJEJOH XJUI PVS aOEJOHT GSPN IZQPUIFTJT
OPEFT BSF BXBZ GSPN SFBMJTN UIF XFBLFS UIF TJNJMBSJ
NPEFSOJTU QPFNT CFDPNF FWFO NPSF FTUSBOHFE GSPN U
OFUXPSLT GPS DPOUFOU BOE TUZMF TFFN TJNJMBS CPUIT
FQPDIT IJOUJOH BU UIF QPTTJCJMJUZ UIBU NPEFSOJTU BO
UIBO FYUSB FQPDIBM TJNJMBSJUZ DPJODJEJOH XJUI PVS
UJNFMJOFT DPVME QPUFOUJBMMZ OPU POMZ IFMQ XJUI JEF
UIF FQPDIT FYJTUT PS OPU CVU BMTP XIFO FYBDUMZ TVDIB SV
JUT TQMJU BSPVOE UIF TQMJU GPS qDPOUFOU r BQQFE
UIBU UIF DIBOHF GSPN SFBMJTN UP NPEFSOJTU aSTU CFDBM
DPOUFOU 'PS qFNPUJPO r XFDBOOPU EJTDPWFS BOZ LJOE PG
UIBU FNPUJPOT UIFNBUJ[FE PS FYQSFTTFF JO UIF QPFNT N
CFUXFFO UIF UXP FQPDIT

*O TVNNBSZ XF XFSF BCMF UP DPOaSN TPNF JNQPSUBOU IZQ
XIJMF EJêFSFOUJBUJOH PS SFMBUJWJ[JOH PUIFST 0VS EBU
QPFUSZ JT DMPTFS UP SFBMJTN UIBO UP NPEFSOJTU IPXFW
SFBMJTU QPFUSZ XPVME OPU CF BQQSPQSJBUF 8F TIPXFE
NPSF IFUFSPHFOPVT UIBO SFBMJTU QPFUSZ FWFO UIPVHI U
PVS aOEJOHT TVHHFTU UIBU UIF DIBOHF GSPN SFBMJTN UP
USBOTJUJPO SBUIFS UIBO B SFWPMVUJPOBSZ EJTSVQUJJP
BUUFOUJPO JO MJUFSBSZ IJTUPSZ UP QSPDFTTFT PG HSBEVB
POMZ JO UFSNT PG FJUIFS TUBTJT PS SVQUVSF

5IF BTTVNQUJPOT NBEF JO UIJT TFDUJPO BSF TUJMM POMZ C
BOE DPNQBSBUJWFMZ MJUUMF EBUB 4VCTFRVFOU SFTFB
ZFBS BOOPUBUFE QPFNT NPTU JNQPSUBOUMZ JODMVEJOHM

XIJMF GVSUIFS SFTFBSDI RVFTUJ POT DPVME JOWFTUJHBUF
SVQUVSF CFUXFFO UIF FQPDIT DPVME IBWF IBQQFOFE BU TMJ
EJêFSFOU EJNFOTJ POT PS XIFUIFS qGPSNr SFBMMZ JT UIF NP
IPNPHFOFJUZ BOE IFUFSPHFOFJUZ XJUIJO SFBMJTN BOE NP

*O B SFDFOU BSUJDMF 6OE FESJTDVET BOE F4 RVFTUJPO PG XIFU
QJOH PG DVMUVSBM BSUJGBDUT UP TPNF TQBUJBM SFQSFTF
DVMUVSBM SFMBUJ POTIJQT PCFZ B TQBUJBM MPHJD BU BM
JG XF IBWF TPNF TFFNJOHMZ DPOWJODJOH BSHVNFOUT BH
BDDVNVMBUF NPSF BOE NPSF FNQJSJDBM FWJEFODF UIBU J
XFMM 0VS QBQFS BEET UP UIJT FWJEFODF 5FYUUBM SFQS
TQBD F TFFN XFMM TVJUFE UP FYQSFTT FWFO DPNQMFY UF
XPSL NBZ FYQPTF JUT TIPSUDPNJOHT JO UIF GVUVSF *O UIFN
DBO CF VTFE UP SFFWBMVBUF PVS VOEFSTUBOEJOH PG UIF
OPU POMZ JO \$PNQVUBUJPOBM -JUFSBSZ 4UVEJFT

Ž Ö W š W Û W ö W • ö ö š æ

%BUB DBO CF GPWÖE IFS SFVJWD EQO EQRJK YWG LENU RQGO UK

'Öó D Ò š Ý W | " Û W ö W • ö ö š æ

4PGUXBSF DBO CF GPWÖE IFS SFVJWD EQO EQRJK YWG LENU RQGO
KV[

- Ö % 3 D Ý " • Ö " - " 3 š Ö

5IJT XPSL XBT GVOEFE CZ UIF %FVUTDIF 'PSTDIVOHTHFNJOT
\$PNQVUBUJPOBM -JUFSBSZ 4UVEJFTUOOWOHQ\$PKNFDELFSO QPF
MJUFSBSZ IJTUPSZ XJUI UFYU TJNJMBSJUJFT

š Ö ¯ š ë D | , D 3 š | ö • ¯ š ö D 3 Ö

"OUPO & ISNBO4FUSBXBUSF 8SJUJOH v PSJHJOBM ESBGU
5IPSB)BH7F0VBMJ[BUJPO 8SJUJOH v PSJHJOBM ESBGU
'PUJT +BOS\$PJE D FQUVBMJ[BUJPO 4VQFSWJTJPO 8SJUJOH v P
TJUJPO

-FPOBSE , PPSNFBM "OBMZTJT 4PGUXBSF 8SJUJOH v PSJHJO
.FSUFO , SÓÖBUB DVSBUJPO .FUIPEPMPHZ 8SJUJOH v PSJH
4JNPOF 8JÖLFB \$VSBUJPO \$PODFQUVBMJ[BUJPO 4VQFS
ESBGU 'VOEJOH BDRVJTJUJPO

é"Ñ"| "3% "Œ

"OESFPUUJ .BS%BF 4USVLUVSEFS NPEFSOFO -JUFSBUVS /F
EFT 4DISFJCFOT &S[ÁIMQSEPT)BVQEB 7FSMLBH
"O[5IPNBT s5IFTFO [VS FYQSFTTJPO-JTUJSTDSIFD.IIFE P&E
#FHSJê VOE 11&EDONZF4BCJOB #FDLFS BOE)FMNVUI ,JFTFM %
"SFOU 8JMIFMN FEEFSOF %JDIUFS,BOE)BLUFSF
"SPSB 4BOKFFW :JOHZV -JBOH BOE 5FOHZV .B s" 4J
#BTFMJOF GPS 4FOUFODFU&NOBES,OBHTPO&M \$POGFSFODF
TFOUBUJPOT *\$-3 5PVMPO 'SBODF "QSJM
JVVRU QRGPTGXKGY PGV RFH!KW J]JUKF EZPO
"VTUFSNØIM &MLF s-ZSJL EFS +BBSIMSOBEMJSTUNXFTO EJO E"K
&YQSFTTJPOJTNV&TE CZv:PSL (PUIBSU .JY)BOTFST 4P[JB
EFVUTDIFO -JUFSBUVS WPN +BISIVOEFSU CJT [VS (FHFO
"WFOBSJVT 'FSEJOB&EVUETDIF -ZSJL EFS (FHFOXBSU TFJU
NJUCJPHSBQIJTDIFO VOE CJCMJPHSBQ&JMFISNOB&OJ[FO
#ÁS %BOJFM 5PSTUFO ;FTDI BOE *SZOB (VSFWZDI :
4JNJMBSJUSZDF&EJOHT PG 3FDFOU "EWBODFT JO /BUVSBM
JVVRU CENCPVJQNQI[QTI WJTJUFE PO
w \$PNQPTJOH .FBTVSFT GPS \$PNQVUBUOH 5RYU 4VJN JMB SW
FCTOUVCFV FG WJTJUFE PO
#FDLFS 4BCJOBBOE)FMNVUI ,JFTFM s-JUFSBSJTDIF .P
*O-JUFSBSJTDIF .PEFSOF #F&HSJGZ/DECJIA&B#N&D&FS BOE)
,JFTFM %F (SVZUFS v
#FO[NBOO)BOT F.PEFSOF EFVUTDIF&DFM&ZSJL
#FSO .BYJNJMJBO%F&EVUETDIF -ZSJL TFJUF&DFM&NFT 5PEF
#FUIHF)BOT F&EVUETDIF -ZSJL TF&DFM&JFODSPO
#JFSCBVN 0UUP +VMJ&EF&S&FS .VTFOBMNBOBDI"MVCF&BT +B
w FE .PEFSOFS .VTFOBMNBOBDI"MVCF&BT +BIS
#POTFMT 8BMEFNBS)BOT #SBOEFOCVSH #FSOE *TFNBOO
%JF &#POTFMT
#SPNMFZ +BNFT *TBCFMMF (VZPO :BOO -F\$VO &EVBSE 4
s4JHOBUVSF 7FSJaDBUJPO 6TJOH B q4JBNFTFr 5JNF
1SPDFFEJOHT PG UIF UI *OUFSOBUJPOBM \$POGFSFODF
UFNT v
\$IBO #SBOEFO 4UFGBO 4DIXFUF S BOE 5JNP .ÓMMFS
.PEFMt B\$9JW QSFQSJ&UZXKX
\$POSBEJ)FSNBOO s6Q.PEF\$S&F&P&D&I&O&F&S&S&I&S&B&U&M&F&S&M&N
"SFOU 8JMIFMN 'SJFESJDI *v*7
\$PSCJOFBV)P&NBOO "O&J&F&M&G&O&B&S&V&O&H&J&O&E&J&F&S&F&E&N&F&C&E&B&S&S&U&D&T&U
4DINJEU
&ISNBOOUSBVU "OUPO 5IPSB)BHFO -FPOBSE ,POMF BOE
BOE 5PLFO CBTFE 8PSE &NCFEEJOHT JO UI&F&S&B&D&H&F&J&E&B&M&H&V&M
UIF \$POGFSFODF PO \$PNQVUBUJPOBM)VNBOU&J&F&E&G&W&F&T&F&B&S
U QTI 8QN NQPIARCRGWJ&R&J&U&FE PO
'ÁIOEFST 8BMU'F&S&BOU&H&B&S&E&F&V&O&E .PEF&S&O&F&C&V&D&I (FSNBOJTUJL
.FU[MFS

'FEFSNBOO)FSUB %FES 4DIBU[4EIBJDUJDS F -FINLVIM
'FMLTJ 3JUB BOE 4VTBO 4UBOGPS\$P\$NQBES#BOMQJENTPSF +P
)PQLJOT 6OJWFSTJUJZ 1SFTT
'SJDL 8FSOFS s"WBOUHBSEF VOE MPOHVF EVSÆF 'CFSS
CSBVDI EFS LMBTTJTDIF OUP\$B\$S ODFDIF OPEFSOF #FHSJê
&E CZ 4BCJOF #FDLFS BOE)FMNVUI ,JFTFM %F (SVZUFS
'SJFESJDI)VH%JF 4USVLUVS EFS NPEFSOFO -ZSJL 7PO EFS
[VS .JUUF EFT [XBO[JHTUFO P\$P\$B\$S IVOEFSUT
'SJFESJDI 1BVM/FVEMBOE &JO #VDIKØOH#RUSEFVNSÁO FSS -ZS
(FNNFM -VEXJH F%JF 1FSMFOTDIOVS &JOF "OU4DIMP THWFSN F
-PFîFS
(PMUTDIOJHH %JFUNBS s5SBEJUJ POT[VTBNNFOIÁOH
FSOF BN #FJTQJFM EFS)FJOF VOE #OJUSFSS3FFDIF OUP E
#FHSJê VOE 11&E ONZF4BCJOF #FDLFS BOE)FMNVUI ,JFTFM

(PNBB 8BFM BOE "MZ 'BINZ s" 4VSWFZ PG 5FYU 4JNJM
*OUFSOBUJPOBM +PVSOBM PG \$PNQVUFS "QQMJDBUJ POT
(SBOEJOJ .BSHIFSJUB &OSJDP #BHMJ BOE (JP SHJP 7JTBOJ
\$MBTTJaDBUJPO BOB\$V\$E\$W\$U\$F\$X\$Q\$S\$J\$O\$T\$K\$X
(VSVSBOHBO 4VDIJO "OB .BSBTPWJÅ 4XBCIB 4XBZBNEJQU
%PXOFZ BOE /PBI " 4NJUI s%POrU 4UPQ 1SFUSBJ
.PEFMT UP %PNBJOT B\$E\$B\$T\$Q\$S\$F\$Q\$S\$J\$O\$T\$K\$X
)ÁOU[TDIFM (ØO#F\$M\$J\$P\$H\$S\$B\$Q\$I\$J\$F\$E\$F\$S\$E\$F\$V\$U\$T\$D\$I\$T\$Q\$S\$B\$D\$I\$J\$H\$F\$O\$-
, (4BVS
)JFCFM)BOT)%BT 4QFLUSVNEFS NPEFSOFO 1,ØFOTJHT 5FVIMFO
/FVNBOO
)VDI .BSHBSFUIF F\$B\$V\$F\$O\$M\$Z\$S\$J\$L\$E\$F\$S\$Q\$E\$B\$S\$Ø\$B\$S\$U
+BDPCPXTLJ -VEXJH#FVE -JFEFS EFS CFTUFO OFV F\$F\$N\$Ø\$%\$Ø\$D\$U
,JFTFM)FMNVUFTDIJDIUF EFS MJUFSBSJTDIFO .PEFSOF 4Q
JN [XBO[JHTUFO -\$B)S#V O EFSU
,MBNCBVFS (ØOUFS 5IPNBT 6OUFSUIJOFS "OESFBT .BZS
s4FMG /PSNBMJ[JOH /FVSB \$19/FW XPSFLQTS J O O: KX

,MJOHFS \$PSOFMJB s.PEFSO .PEF STOJH F.Ø.EF\$SIOJ(SNVØT
HSJêF #E CZ ,BSMIFJO[#BSDL .BSUJO 'POUJVT %JFUFS 4
8PMG[FUUFM + # .FU[MFS v
,OFTDILF &NJM FØUIPMPHJF EfvUTDIFSP\$B\$Ø\$LLFS TFJU
,SJQQFOEPSê ,M\$B\$F\$N\$Q\$V\$U\$J\$O\$H\$,SJQQFOEPSê#TVRMQIBG\$B\$M\$K\$B\$C
QT[WRGPP GFW CUEARWRTJUFE PO
-BNQJOH %JFU%\$T MZSJTDIF (FEJDIU %FaOJUJPOFO [V 5I
(BUUVØE FE 7BOEFOIPFDL 3VQSFDIU
w .PEFSOF -Z\$Ø\$E\$FO\$IP\$FD\$L 3VQSFDIU
w ,MBTTJLSEFS .PEFSOF 'CFSEJF ,BORJØVJUF\$S\$W\$K\$H\$N\$F
GTCVWMTKVKM FG WJTTJUFE PO
.BSFMMJ .BSDP 4UFGBOP .FOJOJ .BSDP #BSPOJ -VJTB #
BOE 3PCFSUP ;BNQBSFMMJ s" 4*\$, DVSF GPS UIF FW
EJTUSJCVUJPOBM TFNB\$Ø\$P\$D\$F\$E\$P\$E\$H\$M\$T\$R\$G\$*Ø\$IF /JOUI *OUFSO

- a` >S` YgSYWVWagDUW S` V 7hSgSf[a` />D75... fi European Language Resources Association (ELRA), 216–223. http://www.lrec-conf.org/proceedings/lrec2014/pdf/363_Paper.pdf (visited on 04/22/2022).
- Mathet, Yann, Antoine Widlöcher, and Jean-Philippe Métivier (2015). “The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment”. In: *5a_bgfSf[a` S^>[` Yg[ef[Ub3* (41), 437–479. [10.1162/COLI_a_00227](https://doi.org/10.1162/COLI_a_00227).
- McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger (2018). “UMAP: Uniform Manifold Approximation and Projection”. In: *FZWagD` S^aXA bW EagDUW Eaxi` S`M6g* (3). [10.21105/joss.00861](https://doi.org/10.21105/joss.00861).
- Moltke, Max, ed. (1882). *@VjWV WjfeLZW Bsd` SÖE[`TWIT`[U` VSGe` Wtd`>kd]` g` eWtd`FSYW` Rühle*.
- Nöth, Winfried (2008). “Stil als Zeichen”. In: *DZVfad[US` VEFk`[ef[Ub3: S` VTaa]eAX>[` Yg[ež` f[UbS` V5a`_g` [Lsf[a` EUWUM* Ed. by Ulla Fix, Andreas Gardt, and Joachim Knappe. De Gruyter, 1178–1196.
- Polko, Elise, ed. (1860). *6 [LZFWWd`ÖW`@VjWV WjfeLZW>kd]`*. Amelang.
- Prakoso, Dimas Wibisono, Asad Abdi, and Chintan Amrit (Mar. 2021). “Short text similarity measurement methods: a review”. In: *Eax` 5a_bgf[` Y6* (25), 4699–4723. [10.1007/s00500-020-05479-2](https://doi.org/10.1007/s00500-020-05479-2).
- Prutz, Robert, ed. (1859). *6 WjfeLZW6 [LZFWWd`9Wwi` Sdfz7[` `kd[eLZW6`3` Tg`_`*. Kober & Markgraf.
- Reimers, Nils and Iryna Gurevych (2019). “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”. In: *BcbUWW[` YeaxfZW` 5a` XWVUWA` 7`_b[d[US^?` VIZave[` @SfgdS^>S` YgSYWBcbUWte[` YS` VfZWfZ;` fWtd` Sf[a` S^a[` f5a` XWVUWA` @SfgdS^>S` YgSYW` BcbUWte[` Y/7?` @>Bž` <5@>Bfi`* Association for Computational Linguistics, 3980–3990. [10.18653/v1/D19-1410](https://doi.org/10.18653/v1/D19-1410).
- (2020). “Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation”. In: *BcbUWW[` YeaxfZW` 5a` XWVUWA` 7`_b[d[US^?` VIZave[` @SfgdS^>S` YgSYWBcbUWte[` Y/7?` @>Bfi`* Association for Computational Linguistics, 4512–4525. [10.18653/v1/2020.emnlp-main.365](https://doi.org/10.18653/v1/2020.emnlp-main.365).
- Renner, August, ed. (1899). *6 Se` kd[eLZW` [Wž7[` W_` aWtd` W>Wtd`* Georg Szelinski.
- Salton, Gerard and Michael J. McGill (1983). *;` fcbVgUf[a` fa`?` aWtd` ;` Xad` Sf[a` DVtd[MS^`* McGraw-Hill.
- Sandig, Barbara (2006). *FVWef[`[ef]` Wtd`6` WjfeLZW`*. De Gruyter.
- Selbmann, Rolf (1999). *6 [Me[`_g`[WfW` [d[`[LZ]`VfzL`gd>kd]` Wtd`DVSt`e`_ge`* Aisthesis.
- (2007). “Die Lyrik des Realismus”. In: *DVSt`[e`_gež`7baLZW%3`gfadW`%b` Wtd` Wtd`* Ed. by Christian Begemann. Wissenschaftliche Buchgesellschaft, 189–206.
- Shaver, Phillip, Judith Schwartz, Donald Kirson, and Cary O’Connor (1987). “Emotion Knowledge. Further Exploration of a Prototype Approach”. In: *<agd` S^aXBVtda` S`[fk` S` VEa[US^BekLZa`aYk` 6* (52), 1061–1086. [10.1037/0022-3514.52.6.1061](https://doi.org/10.1037/0022-3514.52.6.1061).
- Sprengel, Peter (1998). *9` Wtd`[LZFWWd` WjfeLZebdSU[YW`>[fWtd`Sfgd` ž` žHa` Wtd`DVtd`LZēž` Ycb` Vg` Y` T[e]gd` <SZZg` Wtd`fi` Wtd`* Geschichte der deutschen Literatur von den Anfängen bis zur Gegenwart 9.1. C.H. Beck.
- Stockinger, Claudia (2010). *6 Se` ž`<SZZg` Wtd`fzL` Vfs`fWtd` Wtd`DVSt`[e`_ge`* Akademie Verlag.
- Szubert, Benjamin, Jennifer E. Cole, Claudia Monaco, and Ignat Drozdov (2019). “Structure-preserving visualisation of high dimensional single-cell datasets”. In: *EUWf[r`UDVž` badfe1* (9). [10.1038/s41598-019-45301-0](https://doi.org/10.1038/s41598-019-45301-0).
- Tille, Alexander, ed. (1896). *6 WjfeLZW>kd]` ha` :` Wjfwg` V?` adW`*. Neumann.

- Underwood, Ted and Richard Jean So (2021). "Can We Map Culture?" In: *agda S^aX 5g fgdS^3` Skf[Le3* (6), 32–51. [10.22148/001c.24911](https://doi.org/10.22148/001c.24911).
- Vietta, Silvio (1992). *6 [Wf[MSd[eLW? aVWf V7[VbcbT\W YWdZ[LZf/[LZW6 ScfWg` Y Wd WjfeLZebcSLZ[YW >[fMSfgdha` : i VWf[T[eFZa_Se 4Wf ZSdV*. J.B. Metzler.
- Wieland, Klaus (2019). "Die deutschsprachige Lyrik der Frühen Moderne (1890-1930)". In: *DVZVdLZV#9Wd_S [cgV#14*, 5–27. [10.4000/rg.976](https://doi.org/10.4000/rg.976).
- Willatzen, Peter Johann, ed. (1875). *4`bfZWI i VjYWWjfeLZVd>kd[] `SLZ 9aVZV7[VB` Ž fZa`ay[WKühtmann*.
- Winko, Simone (2003). *=aV[MfV9VbZVZL g Vj Vd BaVf[] Wd 7_ af[a` W [`kd[eLW g` V baVb`ay[eLZW FjWfW g_* . Erich Schmidt.
- Zelle, Carsten (2005). "Komparatistik und 'comparatio' – der Vergleich in der Vergleichenden Literaturwissenschaft: Skizze einer Bestandsaufnahme". In: *=a_bScSf[ež f[]ž-SZdTgLE Wd 6 WjfeLZW 9VWVdLZSf Xgd 3`^YW Vj Vg` VHVdYVjLZWVWb[fVbSfgd [ež eWdLZSf*, 13–33.

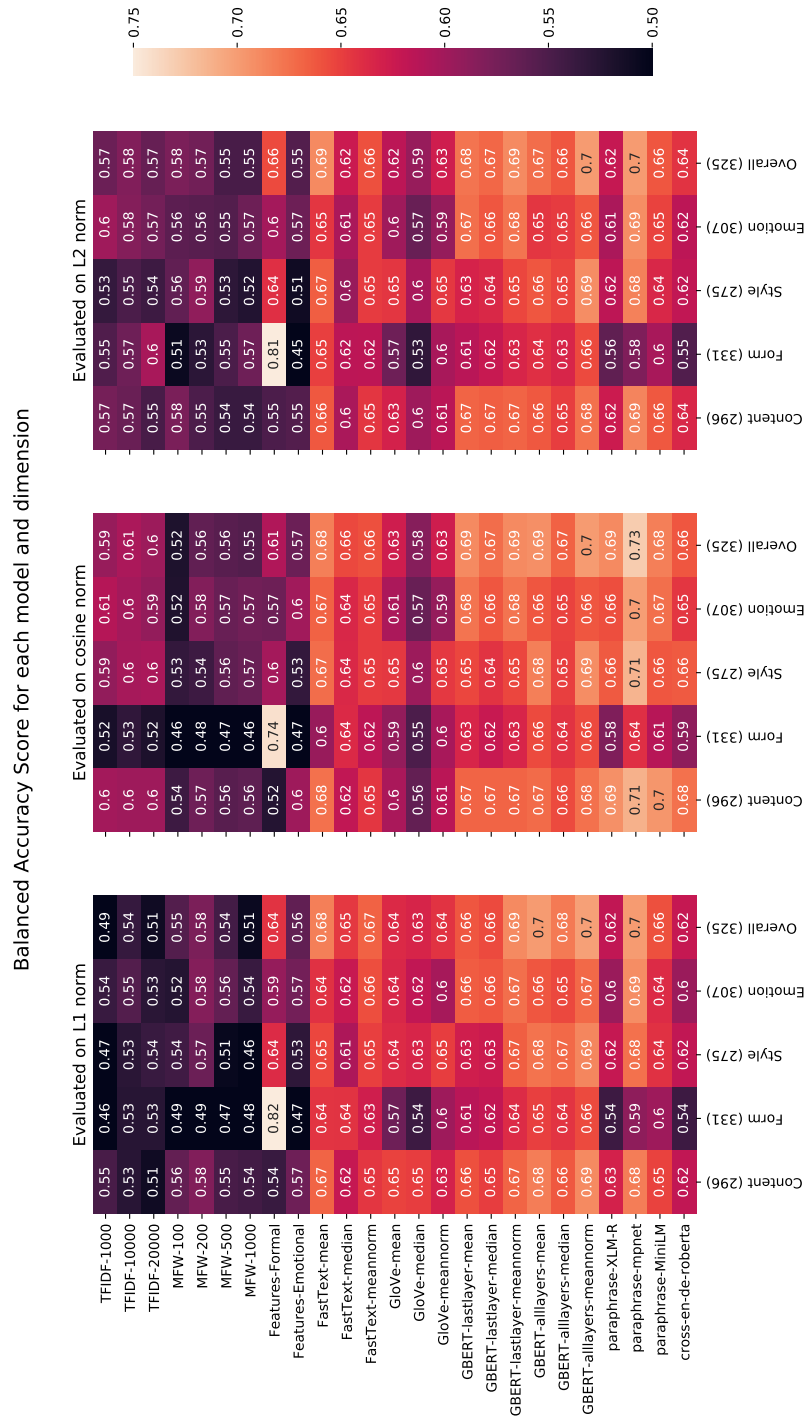


Figure 12: Balanced Accuracy Score for each model and dimension and three distance metrics: L1, Cosine, and L2. Number on the x-axis indicate class support.

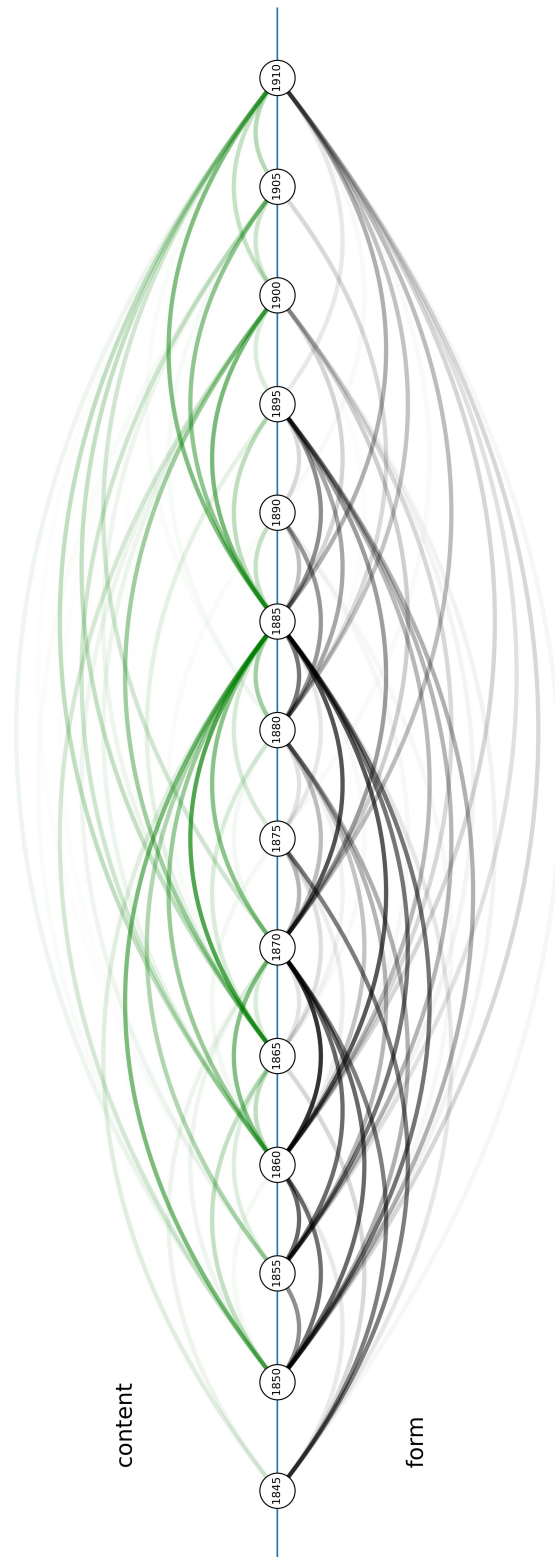


Figure 13: Graph timelines for the ‘content’ and ‘form’ dimensions based on the mean pairwise similarities of 30 poems, sampled for each 5-year time span, based on the similarity-adapted vectors of *paraphrase-XLM-R* (content) and the formal feature vectors (form).

Figure 14: Graph timelines for the 'emotion' and 'style' dimensions based on the mean pairwise similarities of 30 poems, sampled for each 5-year time span, based on the similarity-adapted vectors of *paraphrase-XLM-R*.