Article

Neither Telling nor Describing
Reflective Passages and Perceived Reflectiveness 1700–1945

Benjamin Gittel [1] [iD]
Florian Barth [2, 5] [iD]
Tillmann Dönicke [2] [iD]
Luisa Gödeke [4]
Thorben Schomacker [3] [iD]
Hanna Varachkina [4] [iD]
Anna Mareike Weimer [4] [iD]
Anke Holler [4] [iD]
Caroline Sporleder [2] [iD]

1. Trier Center for Digital Humanities, Trier University [ROR], Trier, Germany.
2. Göttingen Centre for Digital Humanities, Göttingen University, Göttingen, Germany.
3. Computer Science Department, Hamburg University of Applied Sciences [ROR], Hamburg, Germany.
4. German Department, Göttingen University [ROR], Göttingen, Germany.
5. Göttingen State and University Library, Göttingen University [ROR], Göttingen, Germany.

Abstract. The paper analyses within-fiction reflections in 250 years of literary history. To this end, we formalise the concept of "reflective passage", demonstrate how our annotation categories are deduced from literary theory and derive three sub-phenomena – COMMENT, GENERALISATION, and NON-FICTIONAL SPEECH – that constitute literary reflection. A collaborative annotation serves (a) as the basis for the training of a neural classifier and (b) as the dataset for a reception experiment leading to the calculation of a "reflection score", a measurement for the perceived reflectiveness of a textual passage. The classifier is applied to a diachronic corpus of German-language literary fiction derived from the KOLIMO corpus through extensive metadata enrichment and filtering. The results suggest three boom periods of reflective passages (around 1755, 1835 and 1920) and show effects of text length, canonisation status and authors' sex.

1. Introduction

In 1795, Friedrich Schiller, in his famous poetological treatise *On Naïve and Sentimental Poetry*, claims that "ancient" and "modern" poetry differ in their degree of reflection. While the naïve poet moves us by imitating nature, "by sensuous truth, by living presence" (Schiller [1795] 1985, 194), [1]

> [t]he case is quite otherwise with the sentimental poet. He *reflects* upon the impression that objects make upon him, and only in that reflection is the emotion grounded which he self experiences and which he excites in us.

---

1. The German original reads: "durch sinnliche Wahrheit, durch lebendige Gegenwart" (Schiller [1795] 2004, 717).

(Schiller [1795] 1985, 196)[2]

This poetological distinction is linked in Schiller's treatise with a philosophy of history in such a way that naïve poetry is possible in the present but "latently anachronistic" (Prill 1994, 521): Under the conditions of modernity, in which a "correspondence between [...] feeling and thinking" is hardly possible any more,[3] poetry must increasingly become sentimental poetry, that is, a poetry that is moved "through ideas" (Schiller [1795] 1985, 194, 197).[4]

More than 220 years after Schiller formulated this influential thesis, which has found a diverse echo especially in discourses on the "reflexivity" of the modernist novel (see Beebe (1976), Orr (1981)), computational philological methods offer the possibility to study inner-literary reflections on a broad empirical basis. Using the example of German-language narrative fiction, the present paper will investigate whether literature indeed became more and more "sentimental" – as Schiller put it –, i.e. whether it exhibits an increasing degree of reflectiveness.

Of course, the concept of "literary reflectiveness" or – maybe more wide-spread – "literary reflexivity" is till today a very complex one and there is no direct route from Schiller's concept of sentimental (reflective) poetry, which is embedded in an entire anthropology and philosophy of history, to an annotation-based and narratologically underpinned approach like ours. The concept of "literary" or "narrative reflexivity" (Williams 1998) belongs to a whole semantic field of (often interchangeably used) 'big concepts' like "metatextuality", "metafiction", "self-reflexivity" on the on hand (see Julie Tanner (2022)), and rather text-passage oriented concepts like "authorial intrusions", "commentary" or "digression" on the other hand. This may be one of the reasons why there is little consens about the historical development of literary reflectiveness: While it is evident from a number of case studies that at least some early-modern works of literature exhibit significant traits of reflectiveness (see Zapf et al. (2005, 8), Henke (2005)), it is by no means clear how this phenomenon developed in the context of a rapidly growing book market in the 19th century and a mass market in the 20th century.

Our approach aims at measuring the degree of reflectiveness of a narrative by identifying so-called "reflective passages". In the next section, we will introduce our concept of a **reflective passage** and illustrate how we collaboratively annotated three different subtypes of reflective passages. Section 3 will present a questionnaire that was used to empirically assess the contribution of each of these subtypes (and their interplay) to readers' perception of a textual passage being a reflection. Based on the statistical analysis of the results of this questionnaire, we introduce the notion of **perceived reflectiveness** of a given text passage, which is measured by the **reflection score**. Section 4 will describe two neural classifiers: a multi-label and a binary classifier for identifying reflective passages. In section 5, we will present a diachronic analysis of reflective passages as well as perceived reflectiveness in German fiction based on these two classifiers, that allows for evaluating the hypothesis of a gradual increase of reflectiveness

---

2. The German original reads: "Ganz anders verhält es sich mit dem sentimentalischen Dichter. Dieser reflektiert über den Eindruck, den die Gegenstände auf ihn machen, und nur auf jene Reflexion ist die Rührung gegründet, in die er selbst versetzt wird und uns versetzt" (Schiller [1795] 2004, 720).
3. The German original reads: "Übereinstimmung zwischen [...] Empfinden und Denken" (Schiller [1795] 2004, 717).
4. The German original reads: "durch Ideen" (Schiller [1795] 2004, 717).

in the modern period. Finally, we will summarise our results and sketch prospects for future research.

2. Reflective Passages and Their Annotation

When speaking of reflective passages in the context of fictional literature, one may think of various things. Without a doubt, fictional narrative texts regularly stimulate reflections in readers. Authors of such texts also often engage in extensive reflection before or during writing. Reflective passages, in contrast, refer to those reflections that are present on the surface of the text in fictional narrative texts (Gittel 2022). The broad and complex field of the phenomenon of reflective passages becomes clear from the fact that they are referred to in research by many terms that are by no means synonyms, such as "authorial intrusion" (Dawson 2016), "commentary" (Chatman 1980, 226–252), "digression" (Esselborn 2007), "factual discourse" (Konrad 2017), "serious speech acts in fictional works" (Klauk 2015), "gnomic statement" (Mäkelä 2017), "narrator's comment" (Zeller 2007), or *Sentenz* ("aphorism", Reuvekamp 2007). Although reflective passages have been much discussed recently in connection with their specific manifestations in essayistic and encyclopaedic narrative (Ercolino 2014; Gittel 2015; Herweg et al. 2019), they are not a clearly delimited phenomenon either in narratology or in literary history. For a definition of **reflective passage**, however, one can draw on considerations of two more established terms in literary theory – 'comment'/'commentary' and 'non-fictional speech' – and one in linguistics, namely 'generalisation'. We consider a **reflective passage** as a textual passage that is either a comment, non-fictional speech, a generalisation or a combination of these three phenomena. Reflective passages greatly differ regarding their length, ranging from one clause to several sentences or whole paragraphs. The minimal length of a reflective passage being a clause, we will focus in our quantitative diachronic analysis (see section 5) on **reflective clauses** as the minimal unit of a reflective passage. Since the details of our annotation of these phenomena can be found elsewhere (cf. Barth et al. 2021, Gödeke et al. 2022, Weimer et al. 2022, Barth et al. 2022), we will introduce these phenomena through examples in the following and use the corresponding tags COMMENT, NON-FICTIONAL SPEECH, and GENERALISATION henceforth.

"Comment" is listed in narrative theory alongside "report", "description" and "speech" as a fourth so-called "narrative mode" (Bonheim (1975, 329), see also Bonheim (1982)). These four modes, which can overlap, are sufficient for a classification of all passages in a narrative text according to Bonheim. Comments express an evaluative attitude of the speaker towards diegetic state of affairs, illuminate his relationship to the diegesis, or the representation of the events. Thus, they can reveal the narrator's attitude towards characters or events, or his interpretations and explanations of them, as well as his relation to the concrete representation or to narration/fictionality in general. To illustrate what this main type of within-fiction reflections may look like, we use the beginning of Goethe's *Elective Affinities* as example (the corresponding passages are marked in square brackets here and in the following[5]):

(1) Eduard – [let that be the name we give to a wealthy baron in the best years of his life]COMMENT – Eduard had spent the loveliest hours of an April afternoon in his

---

5. For the original wording of all examples, please see Appendix A.

nursery grafting young trees with shoots newly arrived for him. (Goethe [1809] 2008, 3)

The account of Edward's April afternoon is interrupted here by a (metafictional) comment that identifies the speaker as an entity that exercises power of designation over the entities of the narrated world. Overall, however, "comment" is a relatively heterogeneous concept. In research, for example, comments on the story, which can have an interpretive, judgemental or generalising character, are distinguished from comments on the discourse (Chatman (1980, 226–252), see also the term "non-mimetic judgements" in Martinez-Bonati and Silver (1981, 32–33)). Because of this heterogeneity, two criteria are often involved in the identification of comments, one formal and one content-related: According to the formal criterion, comments are those passages of text that are neither speech, report nor description. Like descriptions, they belong to the static mode according to Stanzel and are accompanied by narrative pauses (Stanzel (1988, 66), Martínez and Scheffel (2007, 46)). One often speaks of "pure comment" in reference to such ex negativo identifiable passages (Bonheim 1975, 337). According to the criterion of content, these are passages that express an evaluative attitude of the speaker, his relationship to the event or the representation of the event. If this criterion is taken as a basis, comments can also occur within descriptions, character speech or narrator's report, so-called "integral comments" (Bonheim 1975, 337). The following dialogue in Theodor Fontane's *The Stechlin* can serve as an example, in which Woldemar, the son of the old Stechlin, expresses his astonishment:

(2) "Erratics?" "Yes, erratics," repeated Woldemar. "But if that word bothers you, you can call them monoliths too. [It´s really remarkable, Czako, how extremely discriminating you get about phrases when you´re not the one doing the talking at the moment]COMMENT..." (Fontane [1897] 2013, 10)

Please note that COMMENT is a relatively heterogeneous annotation category that comprises different sub-phenomena: ATTITUDE is annotated whenever the speaker comments on fictional events, characters, objects or itself. This type of comment can be easily interwoven with report, for example through certain adverbs (e.g. "Unfortunately, she had already left the house"). INTERPRETATION is annotated when explanations or interpretations are provided in a passage through which the diegesis can be understood anew. METACOMMENT is annotated whenever the narrator comments on the fictionality of the story or the process of writing or telling the story.

In addition to comment, there is a second phenomenon relatively well described in literary theory that can be used to formalise the concept of reflective passages: the phenomenon of non-fictional speech in fictional texts. According to many theorists, fictional texts consist not only of fictional speech, which – according to a common characterisation – serves to construct the fictional world but also of non-fictional speech (Searle 1975, Klauk 2015).[6] The typical case of non-fictional speech with an assertive character (in the speech act theoretical sense) is relevant to the question of reflections in literature. Characteristic of this phenomenon is that (1) an assertion/hypothesis about the real world is suggested in a clearly delimitable text passage and (2) the

---

6. Konrad also assumes the possibility of "fictional-factual text passages" (Konrad 2014, 447). Without being able to discuss this in detail here: Insofar as these fictional-factual passages have an assertive character, they also fall under the term "non-fictional speech" introduced in the following.

propositional content of the assertion/hypothesis can be read off from this text passage itself.[7] Corresponding (popular) examples are the following:

(3) [All happy families resemble one another, but each unhappy family is unhappy in its own way]_{NON-FICTIONAL SPEECH}. (Tolstoy [1878] 2017, 1)

(4) [Every country has its Samarkand and its Numancia]_{NON-FICTIONAL SPEECH}. That night, both places were here with us on the Morava. [Numancia, located in the Iberian highlands, had at one time been the last refuge from and bulwark against the Roman Empire, while Samarkand, whatever it may have represented in history, became and remains legendary, and will still be legendary when history is no more]_{NON-FICTIONAL SPEECH}. (Handke 2016, 3)

Example (4) – more precisely the third sentence of the Handke quote – demonstrates that NON-FICTIONAL SPEECH does not always have to take the form of GENERALISATION, even though this is the case most often discussed in research (e.g. Vesper 2014).

Third, the phenomenon of generalisation may be regarded as a subtype of reflective passages in its own right. Although generalisation is considered to be an indicator for 'non-fictional speech' and 'comment' (see Chatman (1980) and Vesper (2014)), its appearances in narrative fiction are much less explored than 'comment' and 'non-fictional speech' (see Gödeke et al. (2022) for a first attempt). As GENERALISATION, we annotate any statements not made about specific objects, individuals, time periods, or spaces, but about whole classes or groups of entities.

(5) Naphta responded, with disagreeable composure: "My good sir, [there is no such thing as pure knowledge]_{GENERALISATION}." (Mann [1924] 1969, 397)

As in this example, non-fictional speech often co-occurs with generalisation. However, generalisations can be about all sorts of entities (characters, spaces, events) in the fictional world as well. Generalisations and non-fictional speech (as comments) can also occur within characters' speech: Characters can make statements about whole classes or groups of entities, and characters can suggest in a clearly delimited text passage a hypothesis about the real world whose propositional content (e.g. "there is no pure knowledge") can be read off from this text passage itself.

Having examined the three reflection-constituting phenomena, we will give a brief overview of our annotation results. Our annotation corpus consists of 34 texts with more than 16,000 annotated sentences covering the time period from 1616 to 1942 (cf. https://gitlab.gwdg.de/mona/korpus-public/-/releases/v5.1, data publication (see section 7) and Table 1). In general, the first approximately 400 sentences of each text were annotated by two annotators with a background in German Philology. Two to three experts (authors of this paper) created gold standards for all texts and collaboratively adjudicated the initial annotations (i.e. review, accept, correct or delete).

We computed inter-annotator agreement on clause-level based on Fleiss' *Kappa* (κ, Fleiss 1971) and Mathet's *Gamma* (γ, Mathet et al. 2015, cf. Table 2). κ calculates agreement based on the differences for each clause, while γ respects the individual annotated

---

7. It should be noted that there is nothing attached to the term "non-fictional speech", which is particularly controversial among narratologists. One could also use another term, such as "passages with an assertive character", for the passages that fall under the above definition.

	n
annotated texts	34
tokens of annotated text parts	475,276
sentences of annotated text parts	16,481
clauses of annotated text parts	54,278
GENERALISATION clauses	6,932
NON-FICTIONAL SPEECH clauses	4,122
COMMENT clauses	10,095
GENERALISATION passages	3,696
NON-FICTIONAL SPEECH passages	1,060
COMMENT passages	2,730

Table 1: Descriptive statistics of annotation corpus.

	κ	(σ)	γ	(σ)
GENERALISATION	0.65	(0.19)	0.63	(0.16)
COMMENT	0.52	(0.25)	0.46	(0.21)
NON-FICTIONAL SPEECH	0.74	(0.21)	0.61	(0.17)

Table 2: Clause-level inter-annotator agreement for each phenomenon, averaged over all texts (standard deviations in parentheses).

passages as units in a continuum, and also partially overlapping passages are compared as units instead of disjointed clauses. We therefore consider that γ better represents the errors made by annotators for a category with rather long passages such as reflection.[8] Using Landis and Koch (1977)'s guideline for interpreting the results of κ, we achieve moderate values for COMMENT and substantial values for both, GENERALISATION and NON-FICTIONAL SPEECH for κ (see Table 2). In our perception, γ generally tends to yield more conservative values compared to κ.

So far, we have presented the theoretical background for and our operationalisation of 'reflective passages' and the associated phenomena of 'comment', 'non-fictional speech' and 'generalisation' as well as our annotation results. We stipulated that whenever at least one of these three phenomena is present, such a passage is a **reflective passage**. In the following section, we will introduce the second central term for the envisioned diachronic analysis: perceived reflectiveness, represented by the 'reflection score'.

3. Survey and Reflection Score

We tested the perception of reflectiveness in a reception experiment conducted via a survey. In particular, we were interested in the contribution of individual phenomena (GENERALISATION, COMMENT, NON-FICTIONAL SPEECH) to the overall reflectiveness of a text passage and whether the passages that were not annotated with any of the above-mentioned phenomena can be perceived as reflective. Our objective is to quantify the contribution of the three phenomena and their combinations to the perception of a textual passage as reflective.

The survey was designed as follows: First, we extracted passages from our corpus, more precisely, from texts after 1850 (because we assumed that our participants would more readily understand the language in these more modern texts than in many of the earlier

8. This assessment has already been presented in a similar form in Weimer et al. (2022).

Figure 1: Example question from the survey [In your opinion, is the following statement true: "In the highlighted text passage, something is reflected upon".].

texts). The extracted passages consisted of one sentence and were annotated with the tags GENERALISATION, COMMENT, NON-FICTIONAL SPEECH or their combinations.

Second, we manually chose ten sentences for each of the following groups:

- COMMENT only
- GENERALISATION only
- NON-FICTIONAL SPEECH only
- COMMENT & GENERALISATION & NON-FICTIONAL SPEECH
- COMMENT & NON-FICTIONAL SPEECH
- GENERALISATION & NON-FICTIONAL SPEECH
- COMMENT & GENERALISATION

Additionally, we extracted passages that do not carry any of these tags as negative examples. Altogether there were 100 passages in the survey.

For a better understanding of the passage, we provide the survey participants with the context of one sentence before and one sentence after the passage. The passage in question is highlighted (see Figure 1). Each passage is accompanied by the following question and the corresponding answer options on a scale of 1 to 5:

> In your opinion, is the following statement true: "In the highlighted text passage, something is reflected upon".[9]
>
> 1: false
> 2: somewhat false
> 3: neither true nor false
> 4: somewhat true
> 5: true

For our experiment, we used the web-based survey tool LimeSurvey (Limesurvey GmbH 2024). It allows us to give the participants 30 randomly selected passages. We chose 30 passages as a good trade-off between obtaining sufficient coverage for each passage in the survey while at the same time limiting the experimentation time for the participants. In total, we received 118 complete answers, in which the participants provided their assessments for all 30 passages.

---

9. The survey was conducted in German.

	corr. (p)	coef. (p)
COMMENT	0.61 (0.000)	1.29 (0.000)
GENERALISATION	0.35 (0.000)	0.72 (0.000)
NON-FICTIONAL SPEECH	0.29 (0.003)	0.34 (0.023)
GENERALISATION*COMMENT	–	-0.61 (0.039)
const.	–	-0.72 (0.000)

Table 3: Spearman's correlation coefficient (left) and logistic regression weights (right) for the three phenomena (main effects) and the only significant interaction effect. p-values are shown in parentheses.

For a statistical analysis, we averaged the ratings from all participants for each passage. When we speak of "reflection ratings", we refer to these averages. The left column in Table 3 shows that all three phenomena correlate with the reflection ratings, but to a varying degree. Using Dancey and Reidy 2004's naming convention, the correlation is weak for NON-FICTIONAL SPEECH and GENERALISATION, and moderate for COMMENT. This illustrates that none of our phenomena is perfectly congruent to (perceived) reflection.

In a next step, we created a logistic regression model to gain insights into the interplay between the phenomena. As features, we used the three phenomena as main effects as well as all combinations as interaction effects. We ran both forward selection and backward elimination to determine the best model in terms of the Akaike Information Criterion (AIC), both leading to the same result: a model using all main effects and the interaction effect GENERALISATION*COMMENT. The model's coefficients are shown in the right column of Table 3. Note that the regression coefficients of the main effects are sorted in the same way as their correlation coefficients.

Using the regression coefficients, we can calculate a reflection score r for any passage with known labels for GENERALISATION, COMMENT or NON-FICTIONAL SPEECH as follows:

$$r = \sigma \left([1.29 \cdot f_{\text{COMMENT}}] + [0.72 \cdot f_{\text{GEN.}}] + [0.34 \cdot f_{\text{NON-FICT. SP.}}] - [0.61 \cdot f_{\text{COMMENT}} \cdot f_{\text{GEN.}}] - 0.72 \right)$$

$\sigma(x)$ denotes the logistic sigmoid function $\frac{1}{1+e^{-x}}$. This means that, for example, a passage that is annotated as COMMENT but neither GENERALISATION nor NON-FICTIONAL SPEECH receives the following reflection score:

$$r = \sigma \left([1.29 \cdot 1] + [0.72 \cdot 0] + [0.34 \cdot 0] - [0.61 \cdot 1 \cdot 0] - 0.72 \right) = \sigma (1.29 - 0.72) = 0.64$$

The value of r lies between 0 and 1. Since $0.64 > 0.5$, the reflection score for COMMENT-only passages can be interpreted as "reflective". Table 4 shows that:

- passages that feature none of our phenomena or only non-fictional speech are not perceived as reflective;

- passages that feature only generalisation are equally often perceived as reflective or non-reflective;

- while passages that contain both non-fictional speech and generalisation as well as passages that contain comment are perceived as reflective.

Generally, the presence of each of our phenomena increases the reflection score, which runs between 0.33 and 0.73.

While further research would be necessary to understand why certain combinations

r	Phenomena
0.33	–
0.41	NON-FICTIONAL SPEECH
0.50	GENERALISATION
0.58	GENERALISATION & NON-FICTIONAL SPEECH
0.64	COMMENT
0.66	COMMENT & GENERALISATION
0.71	COMMENT & NON-FICTIONAL SPEECH
0.73	COMMENT & GENERALISATION & NON-FICTIONAL SPEECH

Table 4: Reflection scores for all label combinations.

tend to be perceived as reflective more often than others, another question is whether the perception of a reflective passage actually triggers reflection on the part of the reader. We have to leave such intriguing questions to (psychological) researchers, but we may emphasize two more general insights from our experiment: On the one hand, we can assume that our 'flexible' operationalisation of a "reflective passage" captures basic intuitions about what it is "to reflect upon something". On the other hand, this results in a hierarchisation of the sub-phenomena we examined, which have a varying degree of influence on whether a certain passage is perceived as reflective.

4. Neural Classifier for Reflection

So far, we developed a basic definition of "reflective passage" and a more complex reflection score in order to analyse literary reflection. Since both rely on the identification of the three reflective sub-phenomena (GENERALISATION, COMMENT and NON-FICTIONAL SPEECH), we trained two neural classifiers for the automatic tagging of these phenomena: one multi-tagger and, additionally, one binary tagger (reflective vs. non-reflective passage). To our knowledge, this has not been tried before. Each classifier takes a text span of three sentences as input, where one clause of the inner sentence is marked, and was trained to predict the categories of the marked clause.[10] We split our corpus text-wise into training, development and test set so that the distribution of GENERALISATION, COMMENT and NON-FICTIONAL SPEECH is similar in all sets. Wieland's *The History of Agathon* and Seghers' *The Seventh Cross* are held out for the evaluation of the models, and Fontane's *The Stechlin* and Mann's *The Magic Mountain* serve as development set, while the other texts are used for training.[11] The classifiers are available through the software package (Dönicke et al. 2022).[12]

We followed the approach of Schomacker et al. 2022. The multi-label classifier has three output neurons, where each neuron corresponds to one tag (GENERALISATION, COMMENT, NON-FICTIONAL SPEECH), and the binary classifier has one (REFLECTIVE). Both classifiers are based on a large BERT model, that was pre-trained on German data (Chan et al. 2020)[13] and were trained for 20 epochs with a batch size of 8. To increase the convergence speed,

10. The clauses are detected within our NLP-pipeline MONAPipe (cf. https://gitlab.gwdg.de/mona/pipy -public and software publication) using our algorithm for clause segmentation (Dönicke 2020).
11. We also excluded Kleist's *Michael Kohlhaas* from the training set because the annotated text part does not contain one of our phenomena (non-fictional speech).
12. See https://gitlab.gwdg.de/mona/pipy-public and software publication (see section 8).
13. See https://huggingface.co/deepset/gbert-large.

	GENERALISATION			COMMENT			NON-FICTIONAL SPEECH			Micro-avg.		
	P	R	F	P	R	F	P	R	F	P	R	F
NN-multi (all texts)	0.52	**0.74**	0.61	0.79	0.61	0.69	0.78	0.53	0.63	**0.68**	0.63	0.66
├─ Wieland	**0.53**	**0.74**	**0.62**	**0.80**	**0.68**	**0.74**	0.70	0.50	0.59	**0.68**	**0.66**	**0.67**
└─ Seghers	0.52	0.73	0.61	0.75	0.38	0.51	**1.00**	**0.59**	**0.74**	**0.68**	0.53	0.60
NN-binary (all texts)	–	–	–	–	–	–	–	–	–	0.77	0.62	0.69
├─ Wieland	–	–	–	–	–	–	–	–	–	0.77	**0.69**	**0.73**
└─ Seghers	–	–	–	–	–	–	–	–	–	**0.80**	0.42	0.55

Table 5: Clause-level Precision (P), Recall (R) and F-score (F) of our neural models for classifying clauses according to 'reflection' in the test texts.

we used the LAMB optimiser with a learning rate of 10^{-4} (You et al. 2020). Furthermore, we set the hidden dropout to 0.3 and the attention dropout to 0.0.

Table 5 shows the Precision, Recall and F-score of our classifiers on the test texts (cf. Sokolova and Lapalme 2009). For GENERALISATION, the multi-label reflection classifier performs with 61% F1 like the binary GENERALISATION-only classifier from Schomacker et al. 2022, which illustrates that the other two phenomena can be learned in addition without performance loss. The same classifier achieves with 69% F1 the best results for COMMENT, hereby outperforming the statistical COMMENT-only classifier from Weimer et al. (2022) by 10%. Overall, the multi-label reflection classifier achieves a micro-averaged F1 score of 66% and the binary reflection classifier adds 3% on top of that. While the multi-label classifier achieves similar performance on both test texts ($\pm$7%), the binary classifier shows a greater variation in F1 ($\pm$18%).

5. Diachronic Analysis

This section will first introduce our diachronic corpus "KOLIMO-selection" (1700–1945, see subsection 5.1). In a second step, we report the results of our diachronic corpus analysis (see subsection 5.2). In addition to the "reflection score", we analysed the presence of the three subtypes of reflective passages (COMMENT, GENERALISATION, NON-FICTIONAL SPEECH) that, according to our initial definition, constitute a "reflective passage". In a third step, we took into account potential covariates that may relate to the distribution of reflective passages in literary history, like text length, canonisation status and authors' sex (see subsection 5.3).

5.1 Corpus Building, Metadata Enrichment and Data Cleaning

For our analyses, we used a subset of the "German Corpus of Literary Modernism" (KOLIMO, Herrmann (2023)[14]), which comprehends more than 41,000 texts and spans the period mainly from 1500 to 1930. We filtered KOLIMO to obtain a subcorpus (KOLIMO-selection, see section 8) which fulfils the following criteria:

- only German fiction
- no translations into German
- only first editions

14. A current version of the corpus can be found in Herrmann and Göbel (2023).

- only works with known first publication year
- no duplicates
- balanced in the sense of single authors not being over-represented
- minimum text length of 10 sentences

Concretely, we proceeded as follows. For each step, either an annotation is performed or a filtering is applied (see Table 6):

1) Metadata enrichment: We identified texts with metadata on first publication year and enriched the corpus with data on canonisation status (see Brottrager et al. 2021) and data on the authors' sex (relying on publicly available data on German first names, Neumann 2018). We also relied on metadata concerning publication years from the Corpus d-Prose (Gius et al. 2021), a metadata-enriched subset of KOLIMO covering the period from 1870–1920.

2) Author annotation: We manually annotated at the author-metadata level "predominantly fiction-author" vs. "predominantly non-fiction-author". We filtered KOLIMO and excluded a) texts without author or title, b) duplicates, c) works from over- represented authors (> 500 texts) and d) works from predominantly non-fiction-authors such as Kant, Freud, or Hegel. The threshold of more than 500 texts is a qualitatively explored boundary set to exclude artefacts from highly productive authors that (apparently) were created by adding texts from text collections or chapters/paragraphs from books as separate texts from one author/editor to KOLIMO. This left us with 9,467 texts.

3) Neural classifier: We applied the neural classifier to the corpus, which tags reflective clauses. Some texts (196) could not be processed by the classifier due to artefacts in the text file, such as unexpected character encoding, etc. These texts were dropped.

4) Publication year annotation: We manually annotated the first publication year of texts without publication year relying on the following digitally available databases and multi-volume reference works: Arend et al. (2022), Arnold (2020), Kühlmann (2012), and only as last resort GoogleBooks. Annotators were also asked to mark non-German, non-narrative, non-fictional and translations into German. Based on this data, we filtered our corpus a second time, leaving us with 6,218 texts.

5) Fiction status annotation: Since we observed that our corpus still contains non-fictional narrative texts, we undertook a further annotation: We manually annotated the fictionality status (fiction/ non-fiction / unclear) of texts that contained more than 9.94% non-fictional speech at the clause level (the 75%-quantile) according to the results of our multi-label classifier, thereby using a disproportionately high share of non-fictional speech as a heuristic to identify the remaining non-fiction in our corpus. Subsequently, we removed texts that have been identified as non-fiction by our annotators from our corpus.

6) Data cleaning: In a last step, we removed outliers regarding the proportion of reflective clauses per text (interquartile range method), partly due to wrong or incomplete texts being part of the KOLIMO corpus (e.g. novel prefaces instead of the novel itself). The resulting subcorpus (KOLIMO-selection, 1700–1945, see section 8) contains 5,209 original German-language fiction texts with known first publication year.

Table 6 provides an overview of the filtering process and Figure 2 of the resulting

Step	Dropped	Remaining
1) Metadata enrichment	0	41,382
2) Author annotation		
Texts without author and without title	340	41,042
Texts without author-classification	23	41,019
Duplicates	924	40,095
Texts from non-fiction authors	15,740	24,355
Over-represented authors (> 500 texts)	12,789	11,566
Texts from non-German writing authors	2,099	9,467
3) Neural classifier		
Texts with exceptions during processing	196	9,271
4) Publication year annotation		
Texts without first publication*	2,633	6,639
Translations	44	6,595
Non-German language texts	0	6,595
Non-fictional texts	192	6,403
Non-narrative texts	4	6,399
Texts with less than 10 sentences	181	6,218
5) Fiction status annotation		
Non-fiction or texts with unclear fiction status	360	5,858
6) Data cleaning		
Texts before 1700	167	5,691
Texts after 1945	134	5,557
IQR-based outliers (> 61.68% reflective clauses)	348	5,209

Table 6: Overview of filtering the KOLIMO corpus; * at this step we additionally excluded 463 texts from one over-represented author with the same publication year.
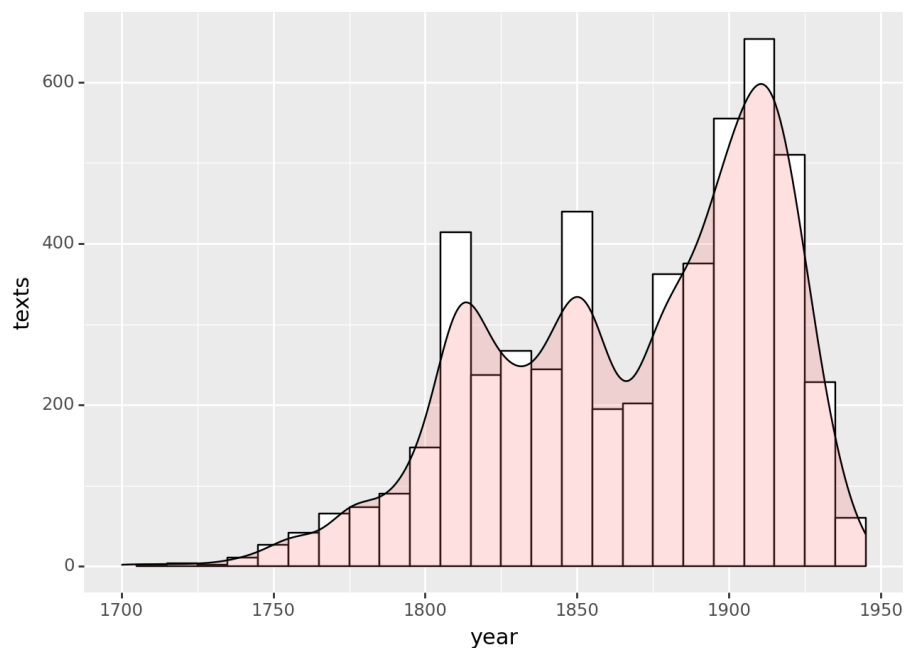
**Figure 2:** Distribution of texts in the corpus "KOLIMO-selection" over time.
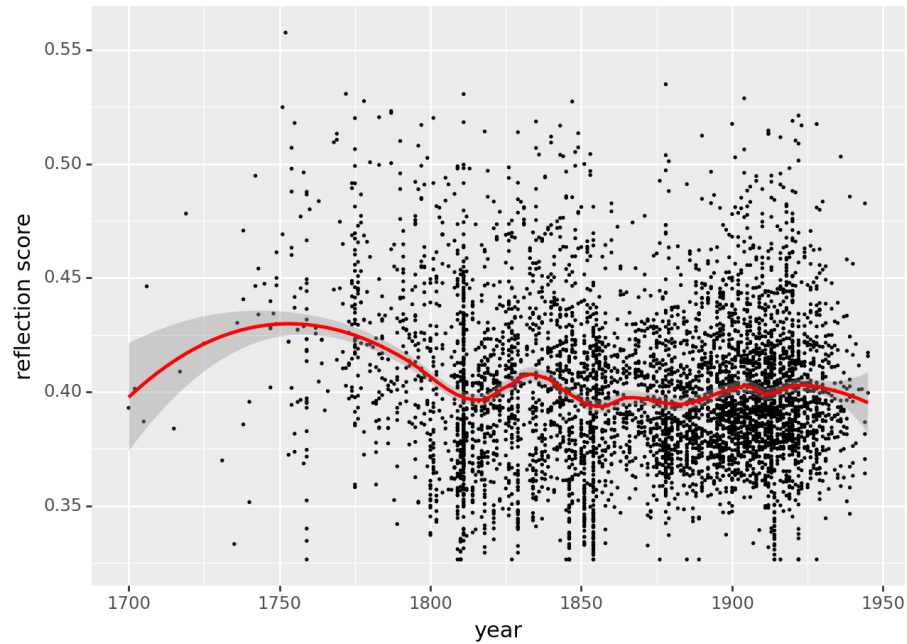
Figure 3: Perceived reflectiveness from 1700 to 1950.

KOLIMO-selection corpus.

5.2 Reflective Passages and Perceived Reflectiveness

Having set out our diachronic corpus and the assumptions built into it, in this section we will analyse the development of reflective passages over 250 years of literary history. In a first step, we take a look at the reflection score, which represents the perceived reflectiveness of a text, as explained above. Figure 3 shows the annual mean of the reflection score.

It can be observed that the average perceived reflectiveness is relatively stable (between 0.38 and 0.43) over time. Keeping in mind that the baseline reflection score, i.e. where none of our three phenomena is present, is 0.33 ($\sigma(-0.72)$) (cf. Table 4 above), this is very plausible: The average German fiction contains some reflections. A second interesting result is the three local maxima around 1755, 1830 and 1920. The first maximum may explain how Schiller, when he wrote *On Naïve and Sentimental Poetry* in 1795, arrived at his initially cited claim that literature was becoming more and more reflective: In fact, Schiller was looking back on a period in which fiction was more reflective than before. Although in his famous essay he cites mainly examples from antiquity – Homer as naïve and Horaz as a sentimental (reflective) poet – he does mention "the sentimental poets of the French, and the Germans, [...], of the period from 1750 to about 1780", who seemed long time more appealing to him than "the naïve Shakespeare" (Schiller [1795] 1985, 191). Figure 3 seems to confirm Schiller's subjective impression. The local peak around 1920 (which forms a saddle with the local peak shortly after 1900) dovetails nicely with the research thesis that there was a boom in essayism at the beginning of the 20th century that describes one aspect of the general trend toward the "dissolution of the boundaries of forms" (Kiesel 2004, 153): On the one hand, fictional essays emerged, and on the other, essayistic passages increasingly found
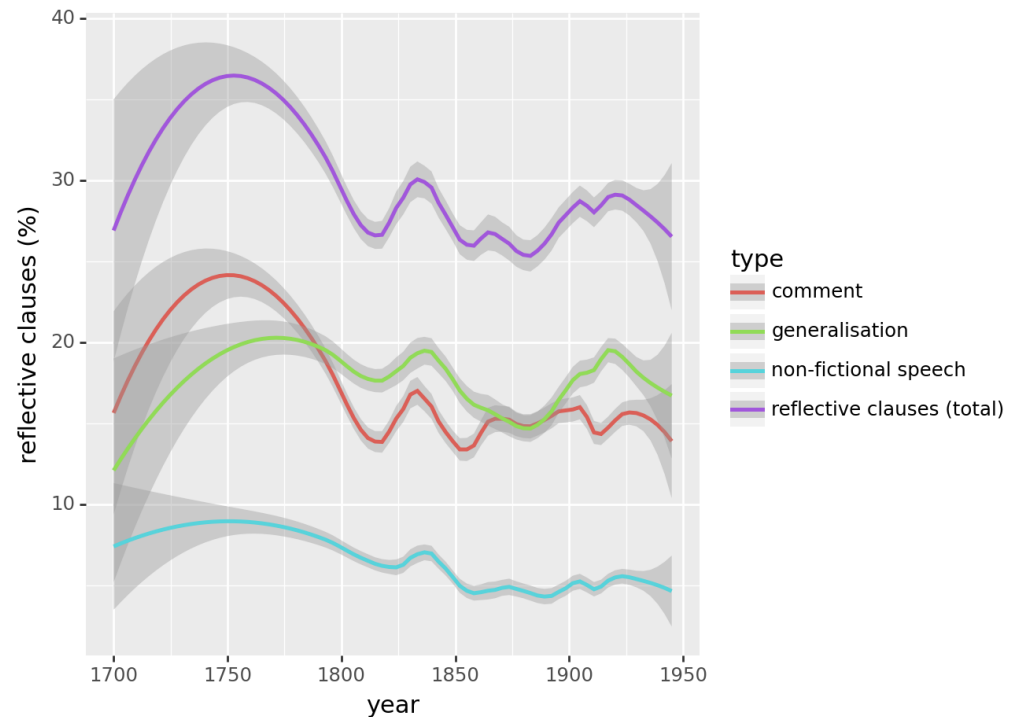
Figure 4: Reflective clauses and their subtypes over time.

their way into fiction, especially into the novel (see Ercolino (2014), Jander (2008), Just (1960), and Müller-Funk (1995)). However, the increase in perceived reflectiveness is less pronounced than one might have expected from the amount of research that exists on the phenomenon of essayism in that period. The peak around 1835 is an interesting finding, which may relate to a politicisation of literature during the *Vormärz* period. However, further research beyond the scope of this paper is needed to underscore such a hypothesis.

In a next step, we look at the frequency of reflective passages and their subtypes. Since reflective passages greatly differ regarding their length – ranging from one clause to several sentences or whole paragraphs – we carry out the following analyses at the clause level and speak of **reflective clauses**. Figure 4 represents the proportion of reflective clauses over time. Please note that we count a clause as reflective – according to our initial definition –, if at least one of our three phenomena (comment, generalisation, non-fictional speech) is present. The confidence intervals, here as in the following, are calculated with Python's ggplot2 implementation "plotnine" employing LOESS smoothing with a span parameter of 0.3.

One may observe four things: 1) The proportion of reflective passages (violet graph) is high over the 18th century (>30%), drops below 30% in 1800, reaches a local peak in 1830 and another in 1920. However, these local peaks in the 19th and 20th centuries never reach the level of the 18th century. The period of realism forms a tale, in which literary reflections are less widespread. 2) The shape of the graphs are very (or for comment relatively) similar one to another and to the reflection score graph in Figure 3. This indicates that the three phenomena do indeed co-evolve and represent different aspects of the overall phenomenon of reflection in fiction. 3) Only two graphs intersect:

GENERALISATION (green) and COMMENT (red). At the end of the 18ᵗʰ century, COMMENT loses its position as the most common subtype to GENERALISATION, which it more or less keeps until 1945. Only during the period of realism, GENERALISATION is less predominant, its 'pole position' being contested by COMMENT again. 4) As one might expect, NON-FICTIONAL SPEECH is the least frequent subtype. Interestingly, its development can be cut into two halves: Between 1700 and 1840, it has a significant share of between 7.5% and 10%, but after 1850 its proportion is more or less stable at around 5%.

5.3 Effects of Text Length, Canonisation Status and Sex

This section is dedicated to the analysis of three factors that may plausibly correlate with fictions' degree of reflectiveness: text length, canonisation status and authors' sex. For example, the fact that the phenomenon of within-fiction reflections has attracted attention primarily in novel research might indicate that reflective passages occur more often in novels than in shorter texts. To scrutinise this hypothesis, we calculated quantiles in the distance of 25% based on text length in word tokens separating our corpus into four parts: very short, short, long and very long texts. Very long texts have more than 58,000 tokens (i.e. > 4,800 sentences based on an estimate of 12 tokens per sentence). Since our diachronic corpus contains almost only prose fiction, this category can be interpreted as novels. Table 7 shows the proportion of reflective passages grouped by text length.

Longer texts tend to be more reflective than shorter texts, although differences are delicate, overall. There is almost no difference between long texts (e.g. novellas) on the one hand and very long texts (e.g. novels) on the other hand. A further analysis revealed that long and very long texts contain on average more COMMENT passages (almost 18%) than very short and short texts (12% and 14.6% respectively), while the values for the other subtypes are very similar.

Another plausible hypothesis is that canonical texts are more reflective than others because complexity is often seen as a text-related standard that may favour canonisation (see Winko 2002, 21–22). Therefore, we added information on the canonisation status (the so-called 'canonisation score', based inter alia on work-mentions in literary histories and anthologies, as proposed by Brottrager et al. 2021) of 357 texts that we were able to identify in our KOLIMO-selection. Table 8 compares these texts against all other (non-canonical) texts.

The group difference presented here is statistically significant as a t-test reveals that canonical texts contain on average 2.7% more reflective passages than non-canonical texts ($t(5207) = 4.22, p < .001, d = 0.23$). However, the relation between the degree of reflectiveness and canonisation is more complex as Figure 5 reveals. It represents the relation between the canonisation score (highest degree of canonisation, values from 0 to 1) and the proportion of reflective clauses of a text (taking only the 357 texts with canonisation score into account).

One observes that the relation is negative: The *less* reflective clauses a text contains, the more canonised the text is. Taking this result together with the previous one (that canonised texts contain on average more reflection), this seems to suggest that a *moderately* increased degree of reflectiveness favours canonisation. We intentionally formulate this

Text length	Mean	SD	SEM
Very short	26.63	14.70	0.41
Short	27.63	11.88	0.33
Long	29.26	9.88	0.27
Very long	29.22	9.46	0.26

Table 7: Proportion of reflective clauses (%) and text length.

Canonisation status	Mean	SD	SEM
Canonical	30.71	11.16	0.59
Non-canonical	28.00	11.74	0.17

Table 8: Proportion of reflective clauses (%) and canonisation.
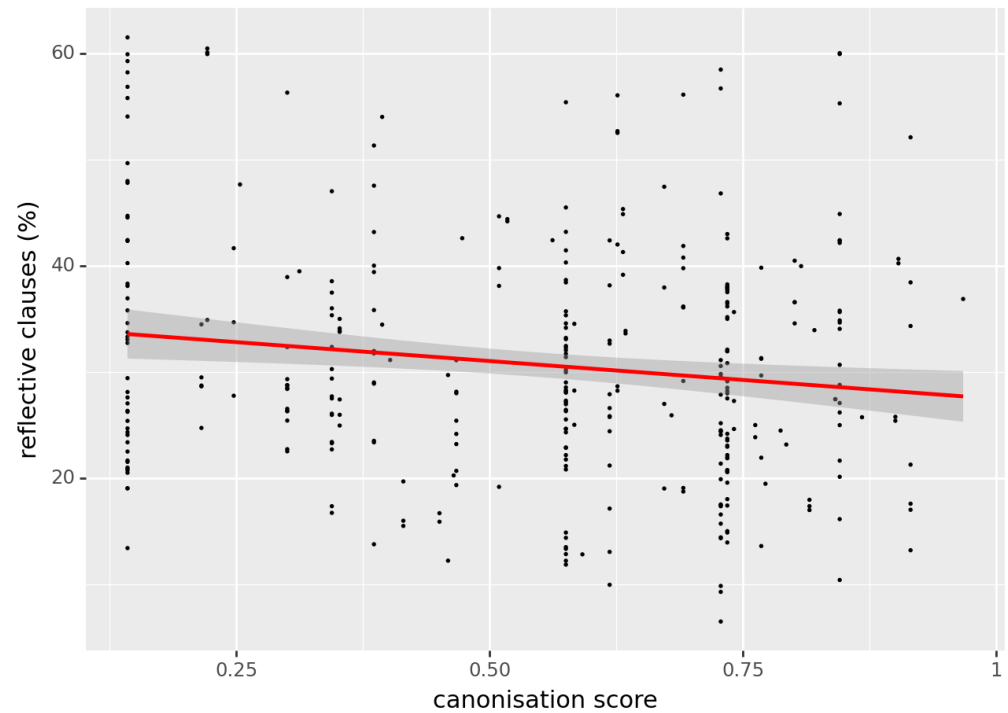
Figure 5: Proportion of reflective clauses in function of canonisation status, n = 357.
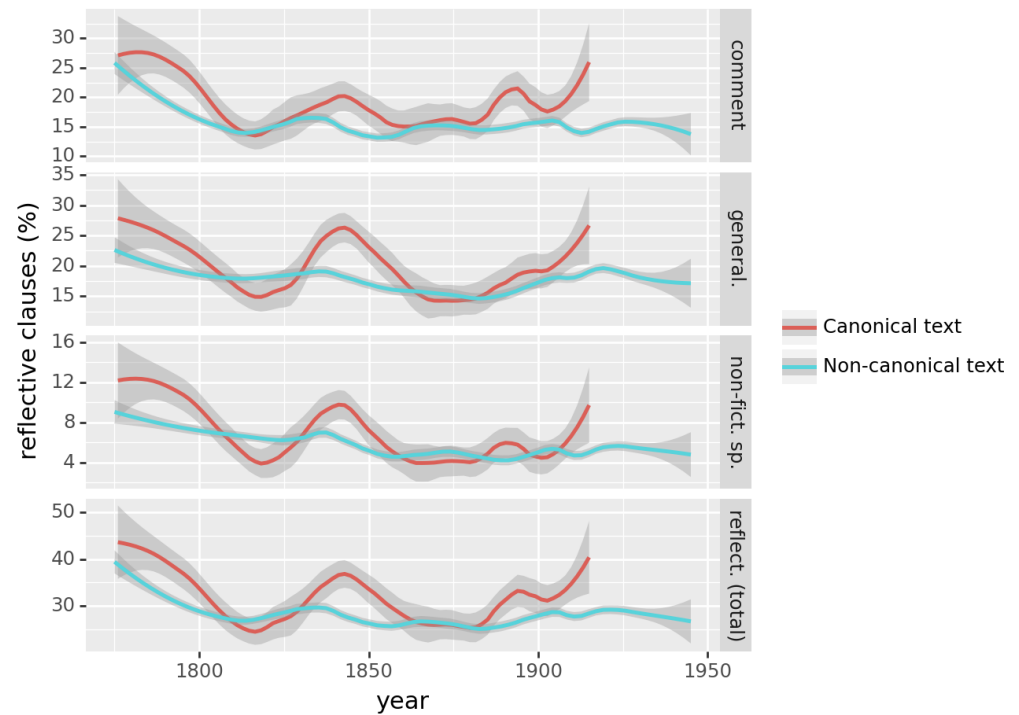
Figure 6: Proportion of reflective clauses and canonisation status over time.

| Authors' sex	Mean	SD	SEM
Female	26.28	12.25	0.33
Male	28.66	11.49	0.20

Table 9: Proportion of reflective clauses (%) and authors' sex.

hypothesis in cautious terms, because there are many other factors involved about which we have no information. However, there is one aspect of the complex relationship we can explore: the diachronic dimension (see Figure 6). The restricted temporal coverage is due to the fact that there are no canonical works before 1750 in our corpus.

Figure 6 reveals several things: 1) The observed mean difference for reflective clauses between canonical and non-canonical texts is due to relatively specific time periods, especially in the middle and in the end of the 19th century and in the beginning of the 20th century. 2) There is a remarkably steep increase for COMMENT and NON-FICTIONAL SPEECH for canonical texts in the beginning of the 20th century. For canonical texts, one may indeed witness the boom of reflection that one could have expected given the above-mentioned research. This underscores how much traditional research is driven by its attention to relatively few more or less canonical texts; the ratio between canonical texts and non-canonical texts in our KOLIMO-selection is 1 to 13.6 (357 to 4,852 texts).

As a third factor for analysis, we selected the authors' sex. From 5,200 texts more than 1,400 texts are from female authors. Table 9 shows that there is an association with the mean proportion of reflective clauses: Male authors tend to use reflective passages on average more often than female authors.

This finding is confirmed by a t-test ($t(4561) = 6.23, p < 0.001$), which reveals a small effect ($d = 0.20$). However, this is only a very general result in light of the highly varying
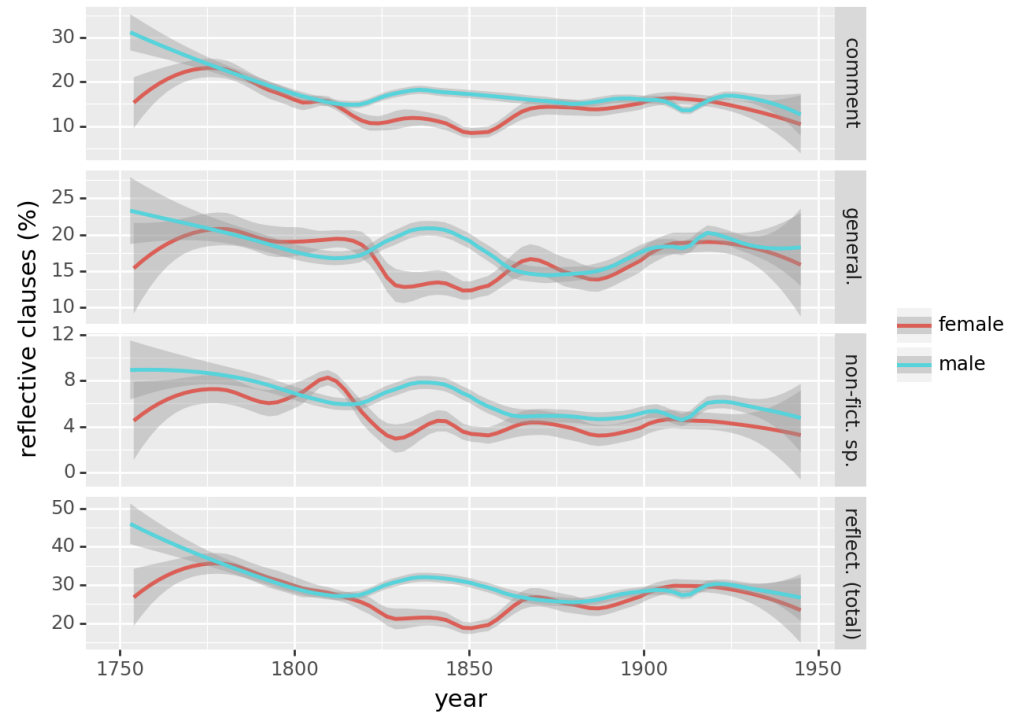
Figure 7: Proportion of reflective clauses and authors' sex over time.

presence of female authors in literary history. For this reason, Figure 7 enables the reader to take a closer look at the interrelations of reflective clauses and authors' sex over time.

From Figure 7 it becomes clear that the more frequent usage of reflective passages by male authors is mainly due to developments before 1875, where female authors – with one exception in the beginning 19th century – reflect less often on average in their fictions. From 1875 onwards, female authors use reflective passages as often as their male counterparts, on average. Only in the 1920s, a new discrepancy seems to loom, especially regarding NON-FICTIONAL SPEECH, which tends to be used less often by female authors.

6. Summary

A so far unfulfilled promise of Computational Literary Studies is to write a more empirically saturated history of literature. Our aim in this paper was to contribute to this new literary history through a diachronic analysis of the narratological phenomenon of reflective passages. Our approach illustrates how many different elements have to come together to get closer to this goal: After 1) a resource-intensive annotation of more than 16,000 sentences for the phenomenon of reflection, we were able 2) to build a multi-label and a binary classifier for reflective passages. 3) We studied how different types of reflective passages are perceived by actual readers and introduced the reflection score as a measure of the perceived reflectiveness of a textual passage. 4) Through a complex filtering process, we build a suitable diachronic corpus of 5,200 original German language fiction texts from the much larger KOLIMO corpus and 5) enriched their metadata regarding fictionality status, canonisation status and authors' sex. Finally, we were able

to analyse the frequency of reflective passages over 250 years of literary history. Our findings suggest three boom periods of reflective passages: around 1755, 1835 and 1920. GENERALISATION is the most common phenomenon (M=17.6% of all clauses), COMMENT the second common (M=15.6%), while NON-FICTIONAL SPEECH is rather rare (M=5.6%). In terms of perceived reflectiveness, all sub-phenomena contribute to a textual passage's reflectiveness, while COMMENT is the best indicator, GENERALISATION plus NON-FICTIONAL SPEECH also indicate reflectiveness. Important covariates of the proportion of reflective clauses are text length, canonisation status and authors' sex. On average, longer texts, canonised texts, and texts from male authors contain more reflective clauses than their respective counterparts. Since our diachronic corpus itself is only a (small) sample from the literary production in German language (Gittel 2021, 5), and – due to limited meta-data – does allow to control only a few potential covariates that steer literary production, our results should be regarded as motivation for further quantitative research in the future. Nevertheless, our research represents a step forward towards an empiricisation of literary studies. It demonstrates that quantitative research can underpin existing hypotheses in literary studies (like the one from a boom of essayism at the beginning of the 20th century) and set new questions on the agenda (e.g. about the nature of the boom of reflection in the *Vormärz* period). To answer such questions, Computational Literary Studies and hermeneutic research need to go hand in hand in our opinion. Quantitative research may in the future shed light on the thematic contents of the different subtypes of reflection and their combinations – a question deliberately put aside in the present paper – and hermeneutic research may formulate justified hypotheses about the functions of different types of reflective passages in specific contexts. In this way, literary studies may advance towards an empirically saturated functional literary history.

7. Data Availability

Data can be found here: `https://zenodo.org/records/10246193`, and here: `https://doi.org/10.5281/zenodo.11164190`.

8. Software Availability

Software can be found here: `https://doi.org/10.5281/zenodo.11163719`, and here: `https://doi.org/10.5281/zenodo.11164036`.

9. Acknowledgements

10. Author Contributions

Benjamin Gittel: Conceptualization, Funding acquisition, Formal analysis, Supervision, Visualization, Writing – original draft, Writing – review & editing

Florian Barth: Project administration, Data curation, Formal analysis, Resources, Software, Writing – original draft, Writing – review & editing

Tillmann Dönicke: Data curation, Formal analysis, Methodology, Resources, Software, Writing – original draft, Writing – review & editing

Luisa Gödeke: Conceptualization, Investigation, Writing – original draft, Writing – review & editing

Thorben Schomacker: Software, Writing – original draft, Writing – review & editing

Hanna Varachkina: Writing – review & editing

Anna Mareike Weimer: Conceptualization, Investigation, Writing – original draft, Writing – review & editing

Anke Holler: Funding acquisition, Supervision, Writing – review & editing

Caroline Sporleder: Funding acquisition, Methodology, Supervision, Writing – review & editing

References

Arend, Stefanie, Jahn Bernhard, Jörg Robert, Robert Seidel, Johann Anselm Steiger, Stefan Tilg, and Friedrich Vollhardt (2022). *Verfasserlexikon – Frühe Neuzeit in Deutschland 1620-1720*. De Gruyter. 10.1515/vdbo.

Arnold, Heinz Ludwig, ed. (2020). *Kindlers Literatur Lexikon* (*KLL*). J.B. Metzler.

Barth, Florian, Tillmann Dönicke, Benjamin Gittel, Luisa Gödeke, Anna Mareike Weimer, Anke Holler, Caroline Sporleder, and Hanna Varachkina (2021). *MONACO: Modes of Narration and Attribution Corpus*. https://gitlab.gwdg.de/mona/korpus-public (visited on 07/23/2024).

Barth, Florian, Hanna Varachkina, Tillmann Dönicke, and Luisa Gödeke (2022). "Levels of Non-Fictionality in Fictional Texts". In: *Proceedings of ISA-18 Workshop at LREC 2022*, 27–32.

Beebe, Maurice (1976). "Reflective and Reflexive Trends in Modern Fiction". In: *The Bucknell Review* 22 (2), 83–94.

Bonheim, Helmut (1975). "Theory of Narrative Modes". In: *Semiotica* 14 (4), 329–344.

— (1982). *The Narrative Modes: Techniques of the Short Story*. Brewer.

Brottrager, Judith, Annina Stahl, and Arda Arslan (2021). "Predicting Canonization: Comparing Canonization Scores Based on Text-extrinsic and -intrinsic Features". In: *CEUR Workshop Proceedings*, 195–205. https://ceur-ws.org/Vol-2989/short_paper21.pdf.

Chan, Branden, Stefan Schweter, and Timo Möller (2020). "German's Next Language Model". In: *Proceedings of the 28th International Conference on Computational Linguistics*, 6788–6796. 10.18653/v1/2020.coling-main.598.

Chatman, Seymour Benjamin (1980). *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press.

Dancey, Christine P. and John Reidy (2004). *Statistics Without Maths for Psychology: Using SPSS for Windows*. Pearson Education.

Dawson, Paul (2016). "From Digressions to Intrusions: Authorial Commentary in the Novel". In: *Studies in the Novel* 48 (2), 145–167.

Dönicke, Tillmann (2020). "Clause-level Tense, Mood, Voice and Modality Tagging for German". In: *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 1–17.

Dönicke, Tillmann, Florian Barth, Hanna Varachkina, and Caroline Sporleder (2022). "MONAPipe: Modes of Narration and Attribution Pipeline for German Computational Literary Studies and Language Analysis in SpaCy". In: *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)*, 8–15. https://aclanthology.org/2022.konvens-1.2 (visited on 07/23/2024).

Ercolino, Stefano (2014). *The Novel-Essay, 1884–1947*. Studies in European Culture and History. Palgrave Macmillan.

Esselborn, Hartmut (2007). "Digression". In: *Reallexikon der deutschen Literaturwissenschaft*. Ed. by G. Braungart, H. Fricke, K. Grubmüller, J.-D. Müller, F. Vollhardt, and K. Weimar. 1. De Gruyter, 363–364.

Fleiss, Joseph L (1971). "Measuring Nominal Scale Agreement Among Many Raters." In: *Psychological Bulletin* 76 (5), 378–382.

Fontane, Theodor [1897] (2013). *The Stechlin*. Trans. by William L. Zwiebel. Camden House.

— [1897] (2015). *Der Stechlin: Roman*. Ed. by Christine Hehle. 3rd ed. Aufbau.

Gittel, Benjamin (2015). "Essayismus als Fiktionalisierung von unsicheres Wissen prozessierender Reflexion". In: *Scientia Poetica* 19 (1), 136–171. 10.1515/scipo-2015-0106.

— (2021). "An Institutional Perspective on Genres: Generic Subtitles in German Literature from 1500–2020". In: *Journal of Cultural Analytics* 6 (1). 10.22148/001c.22086.

— (2022). "Reflexive Passagen in fiktionaler Literatur. Überlegungen zu ihrer Identifikation und Funktion am Beispiel von Wielands 'Geschichte des Agathon' und Goethes 'Wahlverwandtschaften'". In: *Euphorion* 116 (2), 175–191.

Gius, Evelyn, Svenja Guhr, and Benedikt Adelmann (2021). *d-Prose 1870-1920*. Version 2.0. 10.5281/zenodo.5015008.

Gödeke, Luisa, Florian Barth, Tillmann Dönicke, Anna Mareike Weimer, Hanna Varachkina, Benjamin Gittel, Anke Holler, and Caroline Sporleder (2022). "Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung". In: *Zeitschrift für digitale Geisteswissenschaften* 7. https://zfdg.de/2022_010.

Goethe, Johann Wolfgang [1809] (2008). *Elective Affinities: A Novel*. Trans. by David Constantine. Oxford University Press.

— [1809] (2021). *Die Wahlverwandtschaften. Ein Roman*. Reclam.

Handke, Peter (2008). *Die morawische Nacht: Erzählung*. 1. Aufl. Suhrkamp.

— (2016). *The Moravian Night: A Story*. Trans. by Krishna Winston. Farrar, Straus and Giroux.

Henke, Christoph (2005). "Self-reflexivity and Common Sense in A Tale of a Tub and Tristram Shandy: Eighteenth-Century Satire and the Novel". In: *Self-reflexivity in*

Literature. Ed. by Hubert Zapf, Werner Huber, and Martin Middeke. Königshausen & Neumann, 13–38.

Herrmann, Berenike (2023). *Digital Resources – jberenike.github.io*. https://jberenike.github.io/dig_res.html (visited on 07/22/2024).

Herrmann, Berenike and Mathias Göbel (2023). *KOLIMO*. 10.5281/ZENODO.10246193.

Herweg, Mathias, Johannes Klaus Kipf, and Dirk Werle (2019). *Enzyklopädisches Erzählen und vormoderne Romanpoetik (1400-1700)*. Wolfenbütteler Forschungen. Harrassowitz.

Jander, Simon (2008). *Die Poetisierung des Essays: Rudolf Kassner, Hugo von Hofmannsthal, Gottfried Benn*. Winter.

Julie Tanner (2022). "The Legacy of Literary Reflexivity; or, the Benefits of Doubt". In: *Textual Practice* 36 (10), 1712–1730. 10.1080/0950236X.2021.1972038.

Just, Klaus Günther (1960). "Die Geschichte des Essays in der europäischen Literatur". In: *Anstöße. Berichte aus der evangelischen Akademie Hofgeismar* 3, 83–94.

Kiesel, Helmuth (2004). *Geschichte der literarischen Moderne: Sprache, Ästhetik, Dichtung im zwanzigsten Jahrhundert*. C.H. Beck.

Klauk, Tobias (2015). "Serious Speech Acts in Fictional Works". In: *Author and Narrator*. Ed. by Dorothee Birke and Tilmann Köppe. De Gruyter, 187–222.

Konrad, Eva-Maria (2014). *Dimensionen der Fiktionalität: Analyse eines Grundbegriffs der Literaturwissenschaft. Dissertation, University of Regensburg, 2013*. Explicatio. Mentis.

— (2017). "Signposts of Factuality: On Genuine Assertions in Fictional Literature". In: *Art and Belief*. Ed. by Ema Sullivan-Bissett, Helen Bradley, and Paul Noordhof. Oxford University Press, 42–62.

Kühlmann, Wilhelm, ed. (2012). *Killy Literaturlexikon. Autoren und Werke des deutschsprachigen Kulturraums*. De Gruyter. 10.1515/9783110220292.

Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33 (1), 159–174.

Limesurvey GmbH (2024). *An Open Source Survey Tool*. http://www.limesurvey.org (visited on 07/22/2024).

Mäkelä, Maria (2017). "The Gnomic Space: Authorial Ethos Between Voices in Michael Cunningham's 'By Nightfall'". In: *Narrative* 25 (1), 113–137.

Mann, Thomas [1924] (1969). *The Magic Mountain / Der Zauberberg*. Trans. by H. T. Lowe-Porter. Vintage Books.

— [1924] (1991). *Der Zauberberg. Roman*. 24th ed. Fischer.

Martínez, Matías and Michael Scheffel (2007). *Einführung in die Erzähltheorie*. 7th ed. C.H. Beck.

Martinez-Bonati, Félix and Philip W. Silver (1981). *Fictive Discourse and the Structures of Literature: A Phenomenological Approach*. Cornell University Press.

Mathet, Yann, Antoine Widlöcher, and Jean-Philippe Métivier (2015). "The Unified and Holistic Method Gamma (γ) for Inter-annotator Agreement Measure and Alignment". In: *Computational Linguistics* 41 (3), 437–479.

Müller-Funk, Wolfgang (1995). *Erfahrung und Experiment: Studien zu Theorie und Geschichte des Essayismus*. Akademie.

Neumann, Felix (2018). *German Prenames as CSV Data*. https://github.com/fxnn/vornamen (visited on 07/22/2024).

Orr, Leonard (1981). "Vraisemblance and Alienation Techniques: The Basis for Reflexivity in Fiction". In: *The Journal of Narrative Technique* 11 (3), 199–215. http://www.jstor.org/stable/30225027 (visited on 07/22/2024).

Prill, Meinhard (1994). "Über naive und sentimentalische Dichtung". In: *Hauptwerke der deutschen Literatur*. Ed. by Rudolf Radler. Kindler, 520–521.

Reuvekamp, Silvia (2007). "Sentenz". In: *Reallexikon der deutschen Literaturwissenschaft: Neubearbeitung des Reallexikons der deutschen Literaturgeschichte.* Ed. by G. Braungart, H. Fricke, K. Grubmüller, J.-D. Müller, F. Vollhardt, and K. Weimar. De Gruyter, 425–427.

Schiller, Friedrich [1795] (1985). "On Naive and Sentimental Poetry". In: *Winckelmann, Lessing, Hamann, Herder, Schiller, Goethe*. Ed. by Hugh Barr Nisbet. German Aesthetic and Literary Criticism. Cambridge University Press, 180–232.

— [1795] (2004). "Über naive und sentimentalische Dichtung". In: *Erzählungen – Theoretische Schriften*. Ed. by Peter-André Alt et al. Friedrich Schiller – Sämtliche Werke 5. Hanser, 694–780.

Schomacker, Thorben, Tillmann Dönicke, and Marina Tropmann-Frick (2022). *Automatic Identification of Generalizing Passages in German Fictional Texts Using BERT with Monolingual and Multilingual Training Data*. 10.5281/zenodo.6979858.

Searle, John (1975). "The Logical Status of Fictional Discourse". In: *New Literary History* 6 (2), 319–332.

Sokolova, Marina and Guy Lapalme (2009). "A Systematic Analysis of Performance Measures for Classification Tasks". In: *Information Processing & Management* 45 (4), 427–437. 10.1016/j.ipm.2009.03.002.

Stanzel, Franz Karl (1988). *A Theory of Narrative*. Cambridge University Press.

Tolstoy, Leo [1878] (2017). *Anna Karenina*. Trans. by Louise Maude and Aylmer Maude. Macmillan Collector´s Library.

Толстой, Лев Николаевич [1878] (1998). Анна Каренина*. Романь. Типография Рис.

Vesper, Achim (2014). "Literatur und Aussagen über Allgemeines". In: *Wahrheit, Wissen und Erkenntnis in der Literatur*. Ed. by Christoph Demmerling and Ingrid Vendrell Ferran. Deutsche Zeitschrift für Philosophie / Sonderband. De Gruyter, 181–196.

Weimer, Anna Mareike, Florian Barth, Tillmann Dönicke, Luisa Gödeke, Hanna Varachkina, Anke Holler, Caroline Sporleder, and Benjamin Gittel (2022). "The (In-)Consistency of Literary Concepts. Operationalising, Annotating and Detecting Literary Comment". In: *Journal of Computational Literary Studies* 1 (1). 10.48694/jcls.90.

Williams, Jeffrey (1998). *Theory and the Novel: Narrative Reflexivity in the British Tradition*. Cambridge University Press.

Winko, Simone (2002). "Literatur-Kanon als 'invisible hand'-Phänomen". In: *Literarische Kanonbildung*. Ed. by Heinz Ludwig Arnold and Hermann Korte. edition text + kritik, 9–24.

You, Yang, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh (2020). "Large Batch Optimization for Deep Learning: Training BERT In 76 Minutes". In: https://openreview.net/forum?id=Syx4wnEtvH (visited on 07/22/2024).

Zapf, Hubert, Werner Huber, and Martin Middeke, eds. (2005). *Self-reflexivity in Literature*. Text & Theorie 6. Königshausen & Neumann.

Zeller, Rosmarie (2007). "Erzählerkommentar". In: *Reallexikon der deutschen Literaturwissenschaft: Neubearbeitung des Reallexikons der deutschen Literaturgeschichte.* Ed. by G. Braungart, H. Fricke, K. Grubmüller, J.-D. Müller, F. Vollhardt, and K. Weimar. De Gruyter, 505–506.

A. Examples in Original Wording

(1′) Eduard – [so nennen wir einen reichen Baron im besten Mannesalter]_{COMMENT} – Eduard hatte in seiner Baumschule die schönste Stunde eines Aprilnachmittags zugebracht, um frisch erhaltene Pfropfreiser auf junge Stämme zu bringen. (Goethe [1809] 2021, 7)

(2′) "Findlinge?" "Ja, Findlinge," wiederholte Woldemar. "Aber wenn Ihnen das Wort anstößig ist, so können Sie sie auch Monolithe nennen. [Es ist merkwürdig, Czako, wie hochgradig verwöhnt im Ausdruck Sie sind, wenn Sie nicht gerade selber das Wort haben]_{COMMENT}…" (Fontane [1897] 2015, 17)

(3′) [Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему]_{NON-FICTIONAL SPEECH}.(Толстой [1878] 1998, 7)

(4′) [Jedes Land hat sein Samarkand und sein Numancia]_{NON-FICTIONAL SPEECH}.
In jener Nacht lagen die beiden Stätten hier bei uns, hier an der Morava. [Numancia, im iberischen Hochland, war einst die letzte Flucht- und Trutzburg gegen das Römerreich gewesen; Samarkand, was auch immer der Ort in der Historie darstellte, wurde und ist sagenhaft; wird, jenseits der Geschichte, sagenhaft sein]_{NON-FICTIONAL SPEECH}. (Handke 2008, 7)

(5′) Naphta erwiderte mit unangenehmer Ruhe: "Guter Freund, [es gibt keine reine Erkenntnis]_{GENERALISATION}." (Mann [1924] 1991, 207)