

Machine Learning as a Measure of the Conceptual Looseness of Disordered Genres

Studies on German *Novellen*

Julian Schröter 

Abstract This essay discusses how procedures of computational and literary genre stylistics can be built and implemented in order to reconstruct the ways in which disordered, heterogeneous, and historically discontinuous genres undergo historical change. The discussion consists of three parts. In the first part, I will introduce a distinction between classificatory and aesthetic interest and show that genre stylistics should in general be based on aesthetic interest. This is an argument against the common claim that genre concepts should be defined prior to historiographical work. The second part outlines the specific historical situation of the German *Novelle* as well as the basis of an aesthetic historiography of this disordered genre. The third part gives an operationalization of one fundamental task in computational and literary genre stylistics—namely, the psychological question of the extent to which historical readers were able to gain an understanding of different genres, based on texts that I grouped according to genre labels. This issue is processed here as a supervised learning task. With this method, low accuracy scores in classification tasks are usually interpreted as methodic flaws, and optimal accuracy scores are regarded as the criterion for adequate modeling. This essay argues, however, that if the question concerns the aesthetic history of disordered genres, then any adequate classification task based on supervised machine learning is expected to yield low accuracy scores. In the next step, the criterion of adequateness must therefore be revised so that low accuracy scores can be attributed to a specific genre itself and not to inappropriate selection of features, problematic parametrization, and other aspects of modeling. Finally, accuracy scores are not to be optimized, but they can be interpreted, in the case of disordered genres. Machine learning tasks can be integrated into a psychological framework so that accuracy scores can be interpreted as a measure of the semantic looseness of the concept of a specific genre within historical literary communities.

Keywords supervised machine learning, German *Novelle*, computational and literary genre stylistics

1. Introduction

In terms of conceptual extension, genre can be thought of as a group or cluster of texts that can be separated from other clusters. Clustering texts by means of their features is one major task of computational stylistics (Schöch 2017). Herrmann et. al. (2015, 46) emphasize that “style can be associated with categories such as genre, epoch, author, and many more.” For quantitative text analysis, it is not the only option but, I believe, essential that style be investigated as *style of one or several categories* such as author, period, or genre (Jockers 2013; Evert et. al. 2017). The practice of investigating the style of a certain genre in the realm of computational and literary genre stylistics (CLGS) can be characterized by a common premise: we presume that genres behave like logical classes insofar as we presume that texts of the same genre share a set of characteristic and specific features. This set yields the style of that genre and can be described “in terms of a quantitative profiling of formal features” (Herrmann et al. 2015, 46). However, for the German *Novelle*, recent research indicates quite a different situation: over the last decades, studies have increasingly questioned an understanding of the *Novelle* as a genre in literary studies that had largely escaped critical scrutiny.¹ Earlier studies of the nineteenth and early twentieth century (Borchardt 1926; Klein 1960; Himmel 1963; Kunz 1970) as well as recently published introductions to the genre (Freund 2009; Rath 2008; Füllmann 2010; Meier and Vrckovski 2014) claim that the German *Novelle* is a very stringent and clearly delineated type of text that must be distinguished from the genre of *Erzählungen*.² In contradiction to this position, several studies have suggested that, from a historical rather than a normative perspective, the label ‘Novelle’

1 As ‘Novelle’ and ‘novella’ do not have the same meaning, I will use, or rather refer to, the German term. Single quotation marks are used in this paper according to conventions in analytic philosophy. If a term or a sentence is addressed as the object of a statement, it is covered by single quotation marks, for example in statements such as: “The term ‘novella’ is used by person *X* in situation *S*.” Both the whole sentence and the term ‘novella’ within this sentence are addressed as different observations. In contrast, if a person *X* claims, “Though included to Heyses’s *Deutscher Novellenschatz*, Goethe’s *Die neue Melusine* is not a novella but a tale,” *X* uses the terms ‘novella’ and ‘tale’ and thus makes use of his concepts of novella and tale. Within the quoted claim, these terms are not covered by single quotation marks.

2 See Baker (1999, 34): “The *Novelle* is regarded as one of the most stringent forms of prose as it deals with a closed, narrow segment of reality, where one particular event is important, and where there is one central conflict around which the whole is organized. Few characters are involved in the *Novelle* so the effect is heightened, and the action is generally of short duration.”

does not signify any clear-cut literary genre (Polheim 1965; Meyer 1987; 1998; Lukas 1998). This paper asks who is right: the proponents of a strict genre of *Novellen* or their more recent and skeptical opponents? For CLGS, and more precisely, for analyses based on supervised machine learning, this endeavor presents a methodological challenge. If the skeptical position is right, which means that the *Novelle* does not behave like a category, we would expect that classification tasks will yield bad results. Usually, bad results require the learning or clustering algorithm to be optimized. This paper must thus elaborate a strategy to prove that bad results are due to the semantics of the genre concepts and not to methodic flaws. This essay accordingly develops such a strategy, outlining a psychological framework that will enable us to investigate the idiosyncratic structure of genres by means of classification and clustering algorithms.³ This will include a clear notion of a baseline that makes sense of the distinction between bad, good, and expectedly weaker or stronger accuracy scores. In order to make clear why the looseness of genre concepts is worth being studied at all, this essay starts with taking one step back from common presuppositions to clarify our general interest in genres.

2. Why Genres Matter—Classificatory Versus Aesthetic Interest

In the case of loose genre concepts, scholarly work usually begins by offering a definition in order to get rid of any vague use of language that might threaten conceptual clarity (for the *Novelle*, see Lukas 1998, 252; Rath 2008). This first section of my essay will reject this default strategy and offer a better alternative. This requires distinguishing between two common interests: classificatory interest and aesthetic interest. Classificatory interest aims at a systematic concept of the metalanguage that is used to describe a specific object language.⁴ Thus, the metalanguage is also called description language. In communities that stand in the tradition of the Vienna Circle or the analytic philosophy of science, the terms of our description language should be unambiguous,

3 Underwood (2016 and 2019) proposed the idea of perspectival modeling to analyze the historical change or the “life spans” of historically discontinuous genres. This paper owes much to Underwood’s idea but suggests a more psychological interpretation of the results of classification tasks.

4 The distinction between metalanguage and object language relates to Tarski (1935). The notion of metalanguage shall roughly refer to a more or less strictly organized field of a regulated semantics, which is operatively used to conduct research. The notion of object language refers here to the semantics of historical use of genre concepts.

which can most suitably be achieved by definitions based on necessary and sufficient conditions. Thus, a definition of any term in a description language should meet the following requirement:

(R1) A definition of a term ‘*c*’ within a language *L* has to enable the users of *L* to determine whether any item *T* of any kind (in case of literary studies, mostly texts) is an instance of the concept which is expressed by the term ‘*c*.’

I call this the criterion of *decidability*. Fricke (2010b, 19) regards this requirement as essential to classificatory genre concepts. It is worth noting that he abandons another requirement postulating taxonomic simplicity so that any item *A* would be an instance of only one category on each categorical level (for example, level 1: epic, drama, poetry; level 2: novel, novella, short story, tale, tragedy, comedy, sonnet, elegy, etc.; level 3: adventure novel, education novel, historical novel, etc.). One reason why Fricke dispenses with the requirement of systematic simplicity is that another requirement obtains priority. This is the requirement of continuing ordinary language use:

(R2) One important task of establishing an academic description language is to clarify vague ordinary language use. Carnap (1950) developed the procedure of clarifying vague concepts as explication.

This requirement (R2) is not a logical one but a formalization of a pragmatic aim of scholarly work. Physics and chemistry, for example, should help people better understand what water, carbon combustion, lightning, etc. actually are. By analogy, literary studies should help people come to a better understanding of what sonnets, elegies, novels, novellas, comedies, tragedies, and so on are. Explications clarify what the ordinary language words ‘water’ and ‘novella’ mean if described within scholarly description language. Most theorists of genre postulate that the concepts of concrete literary genres must necessarily be established as terms that serve the classificatory interest (Fricke 1981, Fricke 2010a, Fricke 2010b, Zymner 2003). At the same time, these theorists advocate for the second requirement. In short, most theories of genre claim that genre concepts must be *explicated as classificatory terms*.

This essay will question the claim this classificatory interest makes to such a general scope. The shortcomings of the classificatory stance can be inferred from arguments made by its prominent opponents. From the 1960s on, an approach oriented toward natural language and based on the notion of family resemblance as elaborated in Wittgenstein’s *Philosophical Investigations* (1953) became more and more influential in theories of literary genre. Among the first supporters of this approach was Fishelov (1991). Hempfer and Strube refined this Wittgensteinian approach (Hempfer 2010b; 2014; Strube 1986). Their reasoning is built upon the presumed structure of ordinary

language. Several genre concepts (such as elegy, *Novelle*, comedy) do not rest on the structure of classificatory concepts but on the structure of family resemblance, or on the structure of prototype theory, which is also based on Wittgenstein's thoughts (Rosch 1978). Although their Wittgensteinian line of reasoning is moving in the right direction, Hempfer and Strube run into a variant of the *is-ought-problem*.⁵ As classificatory terms fulfill the function of more specifically describing and designating objects, classificatory definitions can be accepted as a pragmatically solid instrument. Thus, the empirical finding that everyday or even scholarly readers are using a concrete genre concept in a rather vague way does not entail that the concept of that genre must be used in the same vague manner in every possible analytical situation. Strube and Hempfer make the mistake of starting dogmatically from Wittgenstein's results when they assert that all concepts and, a fortiori, all genre concepts are vague. Instead, they should have started with Wittgenstein's way of asking questions. The proper question would be: which function is the concrete genre concept supposed to fulfill in a concrete analytical practice?

My claim is that for the task of interpreting texts and of reconstructing a concrete genre history, the concept of the respective genre need not fulfill any classificatory function but rather an aesthetic one. The aesthetic function is delineated by Fishelov, too: "In order to understand and to evaluate the writer's work, we are expected to take into account the generic background against which he operates" (1991, 135). The best reason why "we are expected to take into account the generic background against which" the author operates is provided by Walton (1970). At the core of Walton's argument is the "thesis that what aesthetic properties a work seems to have [...] often depends (in part) on which of its features are standard, which are variable, and which are contra-standard for us" (ibid., 343). Which features are standard, variable, or contra-standard is regulated by categorical expectations. One of the best examples of a category of art might be the sonata-allegro form. Walton shows that judgments on aesthetic properties, in particular on novelty, originality, and perfection, require that we know which features would be expected to be realized in that respective work. In Walton's sense, a work is perfectly shaped relative to a form, and innovative relative to the techniques and procedures that were operative at the time the work was produced. Walton is consequently able to formulate four requirements that must be satisfied if a work *W* is to be ascribed to a certain category of art *C*:

5 I refer here to Hume's classic notion of the *is-ought-problem* in his *Treatise of Human Nature* (1739), which addresses the fallacy of unreasonably and often implicitly changing from descriptive statements (such as 'scholars do use vague concepts') to 'ought'-claims with the same propositional content (such as 'scholars ought to use vague concepts') without further reasoning (Hume, *A Treatise of Human Nature* [1739] 469–70, URL: https://en.wikisource.org/wiki/Treatise_of_Human_Nature/Book_3:_Of_morals/Part_1/Section_1).

- i) The work *W* fits in with the category *C* inasmuch as “it has a minimum of contra-standard features” of *C* (ibid., 357).
- ii) “*W* is better, or more interesting or pleasing aesthetically, or more worth experiencing when perceived in *C* than it is when perceived in alternative ways” (ibid.).
- iii) “[T]he artist who produced *W* intended or expected it to be perceived in *C*, or thought of it as a *C*” (ibid).
- iv) “*C* is well established in and recognized by the society in which *W* was produced” (ibid.).

Point (i) is a quasi-classificatory requirement, (ii) expresses the rule of maximizing aesthetic value, and (iii) and (iv) are historicizing requirements. In a previous study, I elaborated in more detail on how overemphasizing the maxim (ii) is very common to literary criticism because (ii) regulates the degree of originality and relevance of an interpretation but leads to ahistorical or anachronistic constructions of genre concepts (Schröter 2019, 232).

It might seem that referring to an aesthetic category requires a classificatory definition of that category. The distinction between *use* and *mention*, which is fundamental to analytic philosophy, comes into effect here. Genre theorists such as Fricke and Zymner claim that in scholarly work and interpretive practice we necessarily make use of the concepts of concrete genres whenever we designate and describe a concrete text as a novel, comedy, sonnet, etc. Hempfer’s claim (2010a, 16) that whenever we *use* a concept, we use it within our description language, is an analytic statement. I agree with Fricke (2010a, 7) that the concepts of a description language should at least in part be defined as classificatory terms. Fricke’s line of reasoning suggests that interpreting a text would be an instance of ascribing a work to a category. However, I consider this latter assumption to be a serious error.⁶ Scholarly interpreters who are reading texts as *Novellen* seek to find out in what way these texts refer to certain historical understandings of ‘Novelle.’ Hence we have to take into account different kinds of reference: a text can be an instance of a genre, a provocation, or a challenge to a genre concept. Moreover, each text can be a typical or a very innovative instance of the respective genre and, finally, the reference to the genre can be merely ironic. This latter aspect has

6 The same seems to apply when we are dealing with the history of a concrete genre because talking about the historical changes of a certain object presupposes that the object is determined. This affects this essay itself: if this text is about the history of German *Novellen*, the word *Novelle* seems to be used to designate the object of research. Therefore it appears that this essay should define the concept of *Novelle* in its description language. I do agree that scholarly work has to clarify the way it refers to the object which is to be investigated. However, I maintain that such a clarification does not have to be composed of a classificatory definition. Nor does it have to be composed of an explication in Carnap’s sense. I offered a solution in another essay (Schröter 2019).

already been emphasized, or rather over-emphasized, by Derrida (1980, 59) and, for the *Novellen*, by Kiefer (2010, 14).⁷ According to Walton's requirement (i), interpreting a text with reference to its genre seems at first glance to be an act of *using* a genre concept. However, whenever we really try to interpret a text in a historical way according to the author intentional requirement (iii) or the reader response requirement (iv), we do not actually *make use* of the genre concept. Instead, we are establishing one of several distinguishable connections between a work and a historical category. Establishing a connection between an artifact and a historical category can be regarded as a somewhat odd kind of *mentioning* a historical genre concept. We refer to the historical use of the genre concept and not to our systematic use.⁸ However, this kind of reference usually entails a description of the historical use within our systematic language because we need to reconstruct the historical semantics of the concept.

In several cases of clear-cut genres such as the sonnet, the ode in the Asclepiadean style, or—perhaps—the joke, the classificatory interest matches the aesthetic interest by and large. Mentioning and using a genre concept amount to the same thing if the respective intensions of both the genre understanding and the classificatory term are consistent. Clear-cut genres might even be a regular case because poetological definitions mostly had the conceptual structure of classificatory concepts until the nineteenth century. This essay is concerned with the case of problematic or disordered genres where the aesthetic and the classificatory interest do not align. This leads me to the following two conclusions:

1. Reading literary works as an aesthetic endeavor that seeks to avoid crude anachronisms requires reference to historical understandings of genre. This implies that Walton's requirements (iii) and (iv) are of higher priority than (i) and (ii). In terms of cultural studies, this insight was already stated by Ryan: "The significance of generic categories thus resides in their cognitive and cultural value, and the purpose of genre theory is to lay out the implicit knowledge of the users of genres" (1981, 112). However, it is not only a matter of prioritization but also a logical issue. Whereas the academic practice of reconstructing historical concepts uses the same techniques as the practice of explication in Carnap's sense, its aim is quite different from the practices of explication and definition (Pawlowski 1980, 18–28). Reconstructing conceptual history tries to elucidate historical language

7 Derrida might have overstated the possibility of ironic references because such references are very seldom. Neither the period of Biedermeier, nor that of poetic Realism rely on aesthetic deviance.

8 In the German-speaking world, a distinction has been proposed between text type (*Textsorte*) and genre; the first designates the metalinguistic classificatory transhistorical concept, the latter refers to the respective historical understanding (Fricke 1981; 2010a). As I have demonstrated, this distinction is neither necessary nor useful to historiographical methodology (see Schröter 2019).

use, whereas explication and definition aim at regulating ordinary language use. My intention here is to make clear that neither scholarly nor lay readers need a classificatory concept of the respective genre defined in their description language when they interpret a literary text. In case the classificatory interest threatens to distort the reconstruction of a historical genre concept from an aesthetic point of view, the classificatory interest should be abandoned.

2. One of the main tasks of a literary history that supports reading as an aesthetic endeavor is to supply readers with all the various understandings of genre that are historically relevant. Of course, each historical understanding of genre itself has a conceptual structure, but this structure is not necessarily classificatory. It could be the structure of a concept of prototype or of loose family resemblance. In contrast to Strube (1986), I do not introduce the theory of conceptual structure on the level of our academic description language but on the level of historical semantics. The main constituent of each historical understanding of a concrete genre is the respective set of standard, variable, and contra-standard features that historical readers and authors linked with the genre labels. The following section elaborates the general requirements for investigating the genre of the *Novelle* with historical interest, beginning with the situation of the research specific to that genre.

3. The Situation of the *Novelle* and the Basic Outline of Its Aesthetic Historiography

The controversial situation of the German *Novelle* as outlined in the introduction, namely that of the *Novelle* as a clear-cut genre in contrast to skeptical opponents, raises several problems. First, this situation demands some clarification of both positions because this kind of contradiction raises suspicion that both parties are dealing with quite different objects. I assume that the older position is right according to its own terms, if applied exclusively to historical genre poetics. Indeed, historical poetics postulates, in a prescriptive manner, that the genre of the *Novelle* *should* be established as a highly valuable literary genre. However, the older position misses two central facts. First, that the actual text production did not meet the prescriptive claims. And second, that historical poetics of the *Novelle* mostly included a dialectical view of the situation and recognized that text production was widely heterogeneous and unstructured. A historiography of the German *Novellen* should therefore test five hypotheses as a first step:

1. German literary communication developed two labels, 'Novelle' and 'Erzählung' for mid-length prose fiction.

2. Whereas there was no poetics of the 'Erzählung,' the concept of 'Novelle' was subjected to intense theorizing and was included within cultural political semantics. Thus, the latter concept was invested with claims to high literary value.
3. The labels 'Novelle' and 'Erzählung' do not mark any substantial generic differences on the level of textual features (Polheim 1981; Lukas 1998).
4. The historical concepts of the 'Novelle' differ synchronically between literary groups and change diachronically from generation to generation.
5. Generic differences have to be traced back to medial constraints of the almanacs, newspapers, and journals in which these texts were published, rather than to poetics (Meyer 1987, 1998).

The primary task of a history of German *Novellen* is to support these hypotheses with empirical evidence. Testing these hypotheses requires that we separate the history of the poetological genre concept from the generic and aesthetic expectations that readers based on their reading of texts they perceived as 'Novellen' or 'Erzählungen.' As I am interested in aesthetic expectations that were communicated by genre concepts in historical cultures, the apparent conflict between the second and the third hypothesis requires me to take into account that poetological reflection, as well as reading practice, might have served as two different and conflicting sources for generalizing aesthetic expectations. As the poetological concept and reading experience do not match in the case of *Novellen*, I regard the relationship between poetics and reading practice as dialectical. By dialectical, I mean that the strong aesthetic expectations historical readers gained from their poetological knowledge of the concept of *Novelle* were undermined by their reading practice, so that these readers either had to ignore, decide, or mediate this conflict.

4. Inductive Procedure

CLGS comes into play on the level of inferring categorical expectations from the recurrent features of text groups. It is reasonable to ask why textual analysis should be preceded by computational analysis and not by close reading alone. The text group that has generally been taken as the basis for constructing the aesthetic category of *Novellen* in literary studies is rather small indeed. It consists of ten to twenty texts that were not explicitly published as *Novellen* in the first place.⁹ Inferring categorical

9 This holds for most canonized texts such as Johann Wolfgang von Goethe's tale *Die neue Melusine*, or the novellas by Adalbert von Chamisso, E. T. A. Hoffmann, Heinrich von Kleist, Franz Grillparzer, Annette von Droste-Hülshoff, Gottfried Keller, and Franz Kafka. By contrast,

expectations based on extensive close reading of these texts would lead to the same concept of the *Novelle* that is common to literary studies. The fourth hypothesis mentioned above claims that the concepts of *Novelle* differ between groups and change from generation to generation. If this thesis holds, the inferences based on a small canon of high literature will not elucidate the historical understanding of genre. Thus, an operationalization has to be chosen that makes it possible to test this hypothesis. The first step is to define historical groups and generations. Theories about system change (Titzmann 1991) and media change (Meyer 1987) suggest the following periodization: 1790–1820, 1820–1850, and 1850–1880–1910 (Schröter 2019, 233–35). There is also strong evidence that from 1850 on the social and medial system was subject to complex differentiation (Mellmann and Reiling 2016). Thus, historical situations S_i , defined by temporal as well as social group parameters, have to be generated heuristically and based on prior knowledge. As the formation of different situations is meant to identify stable language use within each situation, but change between two situations, it might turn out that two provisionally distinguished situations should be either merged or further differentiated. In the next step, we must determine which genre labels were used within each situation S_i for the set of texts to be considered. Though ugly in regards to readability, terms such as ‘Novelle_ S_i ’ (in contrast to ‘Erzählung_ S_i ’) must be used to designate the extension of relevant text groups. We do not know the intension or meaning of these terms in advance, but only the extension (‘Novelle_ S_i ’ designates all texts that were regarded as *Novellen* in the historical situation S_i). This technique comes close to what is called “inductive procedure” in genre theory, as it is opposed to the deductive procedure (Müller 2010). In order to rid the inductive method of its methodological deficiencies, I reformulated this method in terms of quantity theory in such a way that the procedure fulfills its essential requirements (Schröter 2019, 240–47).¹⁰

the equally canonical novellas by Adalbert Stifter, Theodor Storm, C. F. Meyer, Thomas Mann, Arthur Schnitzler, and Robert Musil were published as ‘Novellen’ in the first print. Works by these authors are contained in all collections; see Wiese (1963); Swales (1977); Freund (1998); Reclam (2011).

- 10 The problem with the inductive procedure is that, in the case of inconsistent historical use of names for genres, it is not suitable for inferring generic characteristics from text groups. My set-theoretical reformulation of the inductive procedure makes it possible to reconstruct the generic expectations historically associated with the generic label by making use of classificatory predicates. However, it is not the concept of genre that this procedure aims to determine prior to historiographic work. Rather, the aim is to relate the set of texts that were communicated as ‘Novelle’ or ‘Erzählung’ in the relevant historical situation S_i to classificatory sets by means of intersections. Each of these intersections—for example, the intersection of the texts referred to as ‘Novelle’ and of fictional journal prose—can be described with regard to the relevant text characteristics of the defined set (for example, fictional journal prose). It is important not to summarily define such intersections as a systematic concept of novella, as is often done. Such a definition would result in a classificatory concept of the respective genre, which in turn would

The set-theoretical reformulation of the inductive procedure enables us to reconstruct the generic expectations in the cases where genre labels have determined the perception of connected text groups.¹¹ It is, of course, another empirical and social-psychological assumption that the genre labels of the original publication caused the perception of text groups and led readers to produce genre concepts through a process of abstraction. In contrast to the way literary studies and CLGS have until now often modeled literary genres, ascribing a genre to a text is neither an ontological nor an epistemic matter in the case of disordered genres. It should not be regarded as simply true or false with regard to textual features and a defined genre concept that a certain text is a *Novelle*. Instead, it is our task, first, to collect original genre labels in heuristically defined historical situations and, second, to work out the rules for the use of these labels within these situations.

5. A Machine Learning Task

The following section deals with the problem of how a concrete machine learning task can be modeled in order to determine the aesthetic understanding of genre in a certain period. As indicated at the end of the last section, the fundamental question is twofold: did the members of S_i base their historical genre concept on inferences that consist of the following steps: (1) texts were grouped by the genre labels ‘Novelle,’ ‘Erzählung,’ ‘Märchen’ (tale), and ‘Roman’ (novel); and (2) recurrent textual patterns and text types were inferred from this type of grouping? This twofold question is also suited for elaborating in more detail the third of the hypotheses mentioned above—that there was no difference between *Erzählungen* and *Novellen* on a textual basis. According to the psychological frame I am suggesting, this thesis means that people within the historical situation S_i were not able to perceive any differences on the basis of grouping texts according to their original labels.

The corpus includes 509 texts from the long nineteenth century. What is important and new to this corpus is that it consists up to 30 percent (120 texts) of journal prose fiction texts that are currently being digitized in the course of a postdoctoral project and have not yet been discussed in literary studies.¹² The corpus design followed

no longer be suitable for comprehending the historical use of the generic label and hence the semantics of a genre in historical literary communication.

11 This strategy is also pursued by Underwood (2019, 37), who defines a genre as a “group of books recognized by some specific, historically situated groups of readers.”

12 The project (*Habilitation*) “Ästhetische und soziale Funktionen der Erzählungen und Novellen im 19. Jahrhundert” (“Toward a functional history of nineteenth-century German novellas”) started in 2018 and has been funded by the research fund of the Philosophical Faculty of the

Table 1 Corpus grouped by periods and genre labels

Situation/period	S ₁ : 1790–1820	S ₂ : 1820–1850	S ₃ : 1850–1880	S ₄ : 1880–1910	S ₅ : 1880–1910	S ₀ : total
Novelle_S _i	4	86	83	17	7	197
Other-novellas_S _i	34	31	35	15	5	120
Erzählung_S _i	30	45	13	3	4	95
Märchen_S _i (tales)	28	9	0	1	0	38
Romane_S _i (novels)	13	6	9	10	3	41
Non-fictional-report_S _i	5	4	0	0	-	9
Dorfgeschichte_S _i	0	2	3	0	-	5
Kriminalerzählung	0	0	2	0	-	2
Total	114	183	145	46	19	509

two main criteria: the share of genre labels should be representative of the historical situation in the media market, and canonized texts should not be overrepresented. These maxims are close to being satisfied for the periods from 1790 to 1850 but not yet for the periods from 1850 to 1940. The following Table 1 shows the proportions of the corpus grouped by periods and genre labels (see Table 1).

In the following, the accuracy scores of a supervised machine learning algorithm trained on texts grouped by genre labels shall be interpreted as a proxy for the cognitive ability of historical readers and authors to perceive these groups as genres. Thus, two borders have to be defined: first, that of the best possible classification; and second, a baseline that indicates the indistinguishability of two or more genres. I tested bootstrapped accuracy scores for three classification algorithms (logistic regression, support vector machine (SVM), and $k (=5)$ nearest neighbors (KNN)), for pairs of two genres, and for different feature sets. Table 2 shows the three best predicted genre pairs for all algorithms and feature sets. As feature sets, I chose the absolute counts and the normalized frequency of 2,000 most frequent words with and without stop words:¹³ (a) 2000_mfw_abs_with_stop_words, (b) 1615_mfw_abs_without_stopwords, (c) 2000_mfw_norm_with_stop_words, (d) 2000_mfw_norm_without_stop_words.¹⁴

University of Wuerzburg. Over the course of the project, more than 2000 texts will be digitized. These 509 texts were the intermediate result of the digitization process.

- 13 As normalization according to what is known as the Manhattan style (called 'l1' in *Python*'s sklearn) yielded results below the baseline (see below), I used the default quadratic normalization ('l2' in sklearn).
- 14 Two thousand most frequent words without stop words and without proper names resulted in 1,615 words. For the justification of the use of logistic regression see Underwood (2016) and Underwood (2019).

Table 2 Bootstrapped accuracy scores for different feature sets, learning algorithms, and text groups

Features	Roman (novel) versus Märchen (tale)			Novelle versus Märchen (tale)			Novelle versus Roman (novel)		
	Log Reg	SVM	KNN	Log Reg	SVM	KNN	Log Reg	SVM	KNN
a) abs	0.75	0.84	0.97	0.78	0.84	0.82	0.65	0.72	0.92
b) abs, no stop words	0.83	0.95	0.93	0.83	0.89	0.80	0.68	0.73	0.87
c) norm	0.72	0.74	0.70	0.76	0.79	0.74	0.56	0.48	0.54
d) norm, no stop words	0.49	0.47	0.52	0.48	0.45	0.54	0.45	0.46	0.52

We see accuracy scores above 0.9 for absolute counts of most frequent words in particular for SVM and KNN.¹⁵ This effect is due to the average length of text groups. Tales are significantly shorter than *Novellen* and *Novellen* are shorter than novels.¹⁶ If we compare the stronger accuracy scores for feature sets (a) and (b) with the weaker scores for (c) and (d), it appears that all algorithms are trained basically on text length in case of (a) and (b). Thus, accuracy scores above 0.9 are possible in principle. However, I am interested here in the difference between text groups based on compositional strategy instead of text length.

In the next step, the baseline has to be established. This baseline also operationalizes the third hypothesis, that the labels ‘Novelle’ and ‘Erzählung’ do not mark any substantial generic differences on the level of textual features (see section 2). This hypothesis serves as the null hypothesis. In a pairwise classification task, the accuracy scores are expected to be 0.5 if the null hypothesis is true. In order to test the algorithm, I randomized the attribution of the genre labels in the sample before training and in each iteration of the bootstrapping. As expected, the empirical mean accuracy score with randomized genre labels is for all feature sets, algorithms, and for all pairs of text groups 0.5.

Surprisingly, the feature set (d) (normalized most frequent words with stop words removed) is on the level of that baseline (see Table 2). Table 3 shows the distribution of 1,000 iterations of bootstrapped sampling, training, and validating pairs of text groups with feature set (c) (2,000 most frequent words, with quadratic normalization and with stop words) and with logistic regression algorithm, which is preferred by Underwood (2019) because of its interpretability.¹⁷

15 Improvement for SVM and logistic regression, when stop words are removed (feature set b), is not significant.

16 An ANOVA test (analysis of variance) showed that the difference between average length is significant also with regard to the extremely large variance of text length in each group.

17 Each sample consists of a pair of two groups that are grouped by their original genre label. Both groups are of equal size in the sample. The size of the validation set was set to 0.2 of the respective sample.

Table 3 Distribution of accuracy scores for pairwise training and validation for S_0 : 1790–1940 (bootstrapping with 1,000 iterations, logistic regression)

Pair	Sample size	Mean of bootstrapped accuracy scores	Standard deviation	0.95 Empirical confidence interval for the distribution of bootstrapped accuracy scores
Novellen_ S_0 versus Erzählungen_ S_0	190	0.64	0.08	0.47–0.81
Novellen_ S_0 versus other novellas_ S_0	242	0.56	0.07	0.43–0.69
Novellen_ S_0 versus Romane_ S_0 (novels)	88	0.55	0.13	0.28–0.78
Novellen_ S_0 versus Märchen_ S_0 (tales)	76	0.76	0.11	0.50–0.93
Novellen_ S_1+S_2 (before 1850) versus Novellen_ S_3-S_5 (after 1850)	180	0.67	0.07	0.50–0.80
Novellen_ S_1+S_2 versus Erzählungen_ S_1+S_2 (before 1850)	180	0.64	0.09	0.43–0.80

All accuracy scores for the pairs listed in Table 3 clearly miss the level of 0.9, which was achieved with features set (a) for *Märchen* ('tales') versus *Novellen* or novels. In fact, the results are much closer to the baseline, so that we have to ask whether the null hypothesis can be rejected. According to the empirical 0.95 confidence interval, the difference between *Novellen* and *Romane*, *Erzählungen* and other novellas without a genre label is clearly insignificant.¹⁸ Surprisingly the difference between *Novellen* and tales is as significant as that between *Novellen* before 1850 and *Novellen* after 1850. This indicates that the semantic change of the concept of *Novelle* in the nineteenth century is stronger than the semantic difference between *Erzählung*, *Novelle*, and *Roman* with regard to textual features (and compositional strategy) beyond text length. Hence the null hypothesis cannot be rejected for the pairs of *Novellen*, *Erzählungen*, other novellas and novels.¹⁹ There is, however, one more interesting detail: the difference between *Novellen* and *Erzählungen*, though not strongly significant, might in effect be larger than the difference between *Novellen* and *Romane* (novels) if the effect size is calculated as the difference between two means. This result corresponds to a common assumption

18 This can be directly drawn from Table 4, as for the mentioned pairs, in more than 2.5 percent of all iterations the accuracy scores are below the baseline of 0.5.

19 In contrast to genre categories, media types (such as almanacs, journals, or anthologies) prove more stable within limited periods. This might support the fifth thesis (see section 3 above), which was put forward by Meyer. However, this essay is about genre labels and not about the role of media.

that in the first half of the nineteenth century, the labels *Novelle* and *Roman* were used synonymously (Meyer 1998). However, due to small sample sizes for the periods after 1850 (S_3 and S_4 , see Table 1), it is not yet possible to compare text groups relative to shorter periods except for the largest group of the *Novellen* for all periods and *Erzählungen* for the time before 1850 (see last two rows of Table 3).

Although greater effort could be put into validating and refining the technical procedures, which I cannot provide here, I maintain that we should be able to interpret the weak results represented here from a psychological and aesthetic perspective. From a psychological point of view, the results suggest that readers in the nineteenth century were able to learn what they should expect from tales compared to what they should expect from novels. Likewise, they were able to identify *Novellen* compared to tales. In contrast, readers might have had difficulty understanding the differences between *Novellen* and novels, even more so than between *Erzählungen* and *Novellen*. Thus, historical genre semantics that were induced from reading texts with certain labels (see section 3) do not yield blurred or clear-cut genres per se. Instead, genre semantics based on label use rely upon relating two or more text groups. In the nineteenth century, the *Novelle* might have appeared as a clear-cut and a blurred genre at the same time: historical readers saw it as clear-cut when they compared *Novellen* with tales, and they saw it as blurred when they compared *Novellen* with novels or *Erzählungen*.

In order to improve methods, optimization by means of starting from well-formed classes should not meet the requirements of the classificatory interest as presented in the first section of this essay. Rather, it should be oriented toward historical language use. Thus, I construed the group of *novellas_merged*, which is a sample of n texts from the set of *Novellen*, *Erzählungen*, and no-label-novellas. This sample is intended to map a historical practice that was common to nineteenth-century reading culture. Realist mid-length narrative prose was merged and regarded as ‘*Novellen*’ or ‘*Erzählungen*’ irrespective of the original genre labels. Although ‘*novella*’ is introduced here as classificatory terms of my description language,²⁰ it is intended to represent the historical practice of merging all fictional prose texts in journals, which was mostly labeled as ‘*Novelle*’ or ‘*Erzählung*.’ Based on this grouping, we can test whether the *novellas_merged* group helped historical readers to distinguish novellas not only from tales, but also from novels. Thus, in Table 4, the *novellas_merged* group is compared to novels and tales.

This result does not give a definitive answer. At first glance, it seems that the *novellas_merged* group is not significantly different from novels. In order to see things

20 The term ‘*novella*’ serves a classificatory interest as elaborated in the first section of this paper and is defined as realist narrative mid-length prose fiction, published in journals and anthologies or as a separate book publication. The term ‘no-label-novella’ designates novellas that had no genre label in their first publication.

Table 4 Novellas_merged compared to novels and tales

Pair	Sample size	Mean of bootstrapped accuracy scores	Standard deviation	95% Empirical confidence interval for the distribution of bootstrapped accuracy scores
Novellas_merged_S ₀ versus tales_S ₀	76	0.73	0.12	0.50–0.93
Novellas_merged_S ₀ versus novels_S ₀	88	0.57	0.12	0.33–0.78

more clearly, the following scatter plot of a principle component analysis based on the absolute counts of most frequent words (feature set a) gives a visual sense of the obtained classification results for the original groups according to actual genre labels.

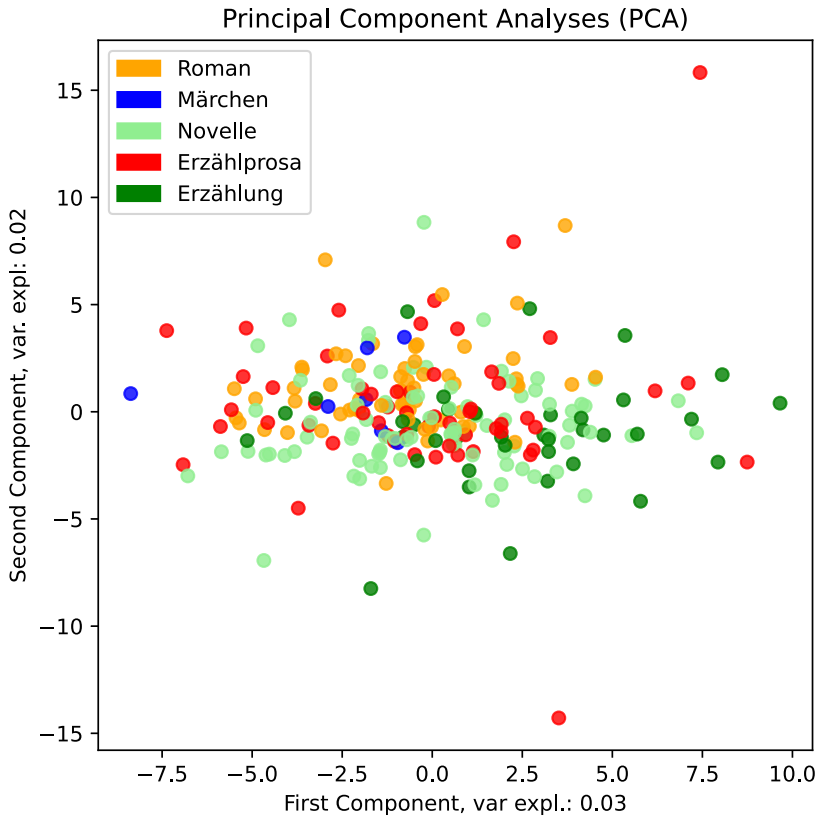


Fig. 1 Principal component analysis and genre labels (Schröter, CC BY).

Figure 1 provides all texts with their original genre labels. This visualization suggests that there is no genre structure at all in the data. If all mid-length prose fiction is regarded as one group (novellas merged) (see Figure 2), the plot looks more uncluttered.

Now we can understand why novellas_merged are not clearly different from novels. Many novellas spread into the group of novels. Nevertheless, we get the picture that, by and large, novellas as the text type of mid length prose fiction are slightly different from fairy tales (*Märchen*) and novels (*Romane*) and that the internal differentiation of novellas into *Novellen* and *Erzählungen*, which is so characteristic of German culture, does not correspond to different text types.

Finally, a considerable bias within the corpus can be seen in Figure 2. Fairy tales, which are represented only in a very small sample of 7 instances here, seem to be closer to each other compared to the remainder of the genre groups. This suggests that fairy

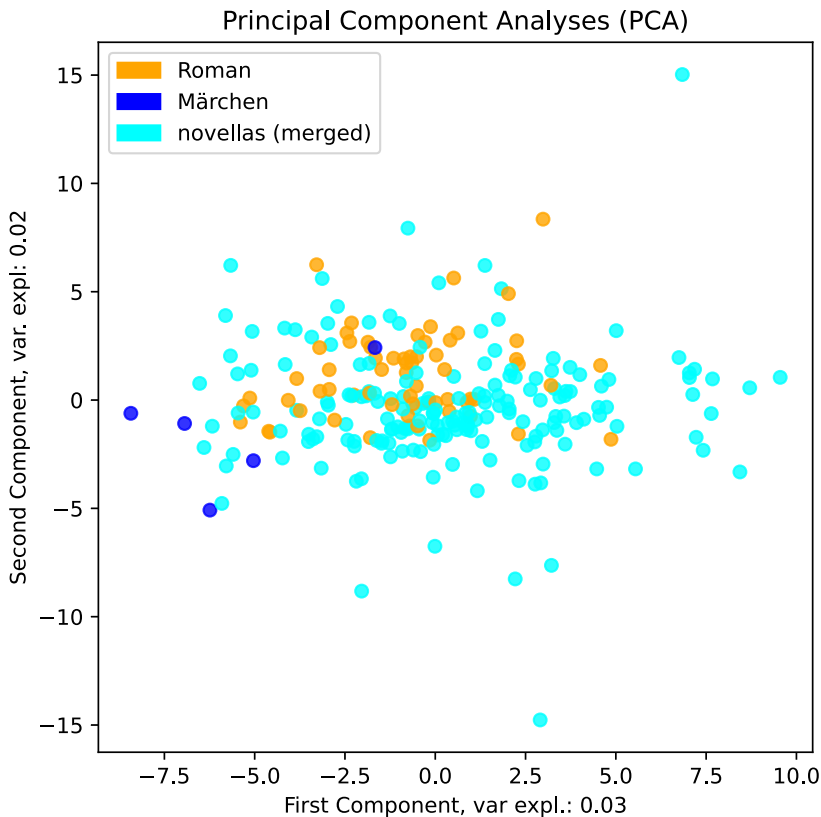


Fig. 2 Principal component analysis for merged genres (Schröter, CC BY).

tales might be slightly more homogeneous in style than novels. As the 38 fairy tales in the whole corpus have been written by only ten different authors (17 by the Brothers Grimm; by contrast the 41 novels are written by 37 different authors), this result should be traced back to the style of the author rather than that of the genre. This bias is already controlled here by sampling only one text written by each author. From the perspective of literary history, things are intricate because this kind of bias might be essential to nineteenth-century knowledge: most of the well-known fairy tales of this period were written or transmitted by the Brothers Grimm. Thus, the composition of fairy tales in the corpus might indeed be regarded as fairly representative of nineteenth-century culture. The well-known fairy tales are rather homogeneous in style also, but not only, because they were written by a small number of authors. However, the main insight here is that, from the perspective of CLGS, and featuring a bag of words model, PCA and classification tasks indicate unanimously that it is more reasonable to assume only one text type of the novella, which can be defined as mid-length prose fiction, rather than assuming different novella sub-types.

6. Conclusion: Toward a Psychological Frame and Why Aesthetic Interest Matters

At first glance the results shown above seem to suggest that a wide classificatory concept of novella encompassing all narrative mid-length prose fiction should be favored. This would weaken my claim that the aesthetic interest should be prioritized over the classificatory interest. On closer observation, it becomes clear that these results actually support my claim. Initially, it was necessary to use a definition of the term ‘novella’ in the description language because I had to identify the group of no-label-novellas. Several scholars have chosen ‘Novelle’ as a term to designate all mid-length prose fiction in journals (see Lukas 1998). The extension of this definition is captured by the `novellas_merged` group, which covers all texts labeled as ‘Novellen,’ ‘Erzählungen,’ and prose fiction in journals without label or with different labels. CLGS could be tempted to do the same, because this definition would yield much better clustering and slightly better classification results. As definitions are arbitrary, there is nothing wrong with this procedure from the perspective of genre theory. From the perspective of genre history however, Lukas’s decision is fatal because it blocks the access to the historical use of the term ‘Novelle’ once and for all. It is possible for ‘novella’ to serve as a term, but ‘Novelle’ should be referred to as a historical genre label that can be used to construct samples of texts grouped by original genre labels. If the intention is to reconstruct historical genre understanding from historical language use, the group of texts that were labeled

as *Novellen* in a certain historical situation S_i must not be defined by textual features but by the historical use of the genre label. I maintain that this inductive procedure should be applied to all disordered genres.

The fourth section of this essay elaborated a technique to answer the question of whether authors, readers, and other actors in a historical situation S_i might have been able to infer categorical genre expectations based on the procedure of grouping texts according to their original genre labels. I maintain that this question should be the first step to reconstruct historical genre understanding. However, this question actually pertains to the *psychological* issue of whether the members of S_i were able to perceive categorical differences between novels and *Novellen*. Walton already emphasized this psychological dimension (see 1970, 338). However, up to this point, my answer has consisted of bootstrapped accuracy scores for classification tasks and of a scatter plot based on principal component analysis. Therefore, in the next step I will have to clarify how statistical significance correlates with psychological significance as outlined in Figure 3.

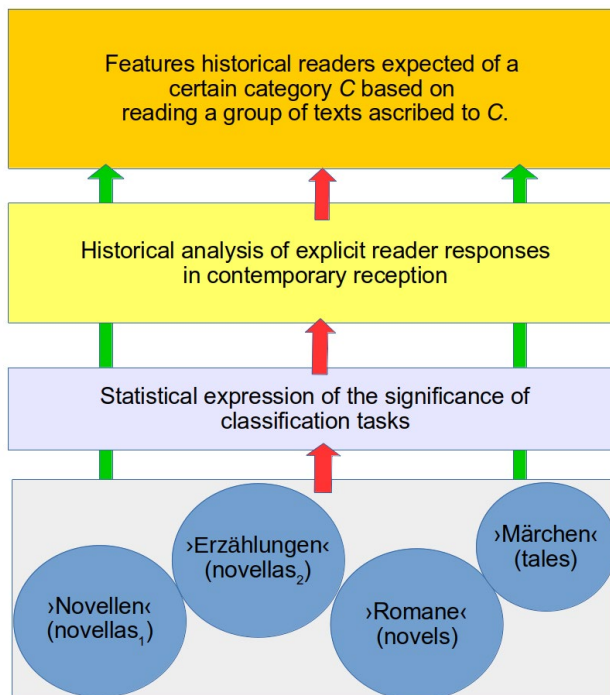


Fig. 3 Outline of correlating statistical and psychological significance (Schröter, CC BY).

The vertical axis of Figure 3 reflects two different types of processes. The green arrows in the background indicate a latent historical process that is intended to be revealed between the lower and the upper box. The lower box shows the input: the texts that were read and grouped by genre labels. As a result, readers come to their genre concepts based (at least in part) on reading and grouping texts. From an aesthetic and historiographic perspective, genre research should determine the semantics of the category *C* (in the upper box). The levels between the upper and the lower box are latent and need to be revealed by historical research because these levels are suitable to explain the development and change of historical semantics. This essay has taken the first step by giving a statistical expression of the distinctiveness of text groups (second box from the bottom). The third box from the bottom refers to the psychological task of formulating statistical results as psychological hypotheses. I would like to propose that two procedures can be combined here: empirical studies of literary response might help to translate the statistical results of computational text analysis to psychological pattern recognition. There is occasional evidence that authors and readers in the nineteenth century, such as Gottfried Keller, doubted that there was any actual difference between *Novellen* and *Erzählungen* and between *Novellen* and novels. More systematic evidence from explicit reader responses could help to validate statistical results. This procedure is shown in the yellow box. If combined with methods of CLGS as elaborated in the fourth section of this paper and as illustrated in the second box from the bottom in Figure 2, classification tasks can be interpreted as a metric that measures the degree of looseness of historical genre concepts. This would lead to a new way of interpreting supervised learning, that comes close to Underwood's (2019) proposal. This method of interpretation would use supervised machine learning to make hermeneutical sense of conceptual looseness within and—if further refined—between historical reading cultures, rather than using it to determine any presumed textual differences between texts in order to reproduce distinctions that have already been established. I argue that research into aesthetic genres should focus on this kind of historical semantics even in cases where the use of genre concepts seems to be arbitrary.

Data repository

Julian Schröter. 2024. julianschroeter/19CproseCorpus: 19th Century Prose Corpus (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.10801777>

Code repository

Julian Schröter. 2023. julianschroeter/PyNovellaHistory: Python Code for the Project on the history of the German 19th-century novella (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.7945348>

ORCID®

Julian Schröter  <https://orcid.org/0000-0003-0168-2608>

References

- Baker, Christine L. 1999. *Landscapes in Theodor Storm's Novellen*. Leicester: University of Leicester.
- Borchardt, Hans Heinrich. 1926. *Geschichte des Romans und der Novelle in Deutschland*. Leipzig: Weber.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Derrida, Jacques. 1980. "The Law of Genre." *Critical Inquiry* 7 (1): 55–81.
- Evert Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. "Understanding and Explaining Delta Measures for Authorship Attribution." *Digital Scholarship in the Humanities* 32 (2): ii4–ii16.
- Fishelov, David. 1991. "Genre Theory and Family Resemblance – Revisited." *Poetics* 20 (2): 123–138.
- Freund, Winfried 1998. *Deutsche Novellen*. München: Fink.
- Freund, Winfried. 2009. *Novelle*. Erweiterte und bibliographisch ergänzte Ausgabe. Stuttgart: Reclam.
- Fricke, Harald. 1981. *Norm und Abweichung. Eine Philosophie der Literatur*. München: Beck.
- Fricke, Harald. 2010a. "(A) Aspekte der literaturwissenschaftlichen Gattungsbestimmung. 1 Methodische Aspekte. 1.1 Definitionen und Begriffsformen." In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 7–10. Stuttgart, Weimar: Metzler.

- Fricke, Harald. 2010b. "(A) Aspekte der literaturwissenschaftlichen Gattungsbestimmung. 1 Methodische Aspekte. 1.5 Invarianz und Variabilität von Gattungen." In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 19–21. Stuttgart, Weimar: Metzler.
- Füllmann, Rolf. 2010. *Einführung in die Novelle*. Darmstadt: WBG.
- Hempfer, Klaus W. 2010a. "Generische Allgemeinheitsgrade." In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 15–19. Stuttgart, Weimar: Metzler.
- Hempfer, Klaus W. 2010b. "Zum begrifflichen Status der Gattungsbegriffe: Von 'Klassen' zu 'Familienähnlichkeiten' und 'Prototypen.'" *Zeitschrift für französische Sprache und Literatur* 120 (1): 14–32.
- Hempfer, Klaus W. 2014. "Some Aspects of a Theory of Genre." In *Linguistics and Literary Studies/ Linguistik und Literaturwissenschaft, Interfaces, Encounters, Transfers/ Begegnungen, Interferenzen und Kooperationen*, edited by Monika Fludernik and Jacob Daniel, 405–22. Berlin, Boston: De Gruyter.
- Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. 2015. "Revisiting Style, a Key Concept in Literary Studies." *Journal of Literary Theory* 9 (1): 25–52.
- Himmel, Hellmuth. 1963. *Geschichte der deutschen Novelle*. Bern, München: Francke.
- Jockers, Matthew Lee. 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Kiefer, Sascha. 2010. *Die deutsche Novelle im 20. Jahrhundert*. Köln, Weimar, Wien: Böhlau.
- Klein, Johannes. 1960. *Geschichte der deutschen Novelle von Goethe bis zur Gegenwart*. 4th ed. Wiesbaden: Steiner.
- Kunz, Josef. 1970. *Die deutsche Novelle im 19. Jahrhundert*. Berlin: Schmidt.
- Lukas, Wolfgang. 1998. "Novellistik." In *Zwischen Restauration und Revolution 1815–1848. Hansers Sozialgeschichte*, vol. 5, edited by Gerd Sautermeister et. al., 251–80. München: Hanser.
- Meier, Albert, and Simone Vrckovski. 2014. *Novelle: eine Einführung*. Berlin: Schmidt.
- Mellmann, Katja, and Jesko Reiling, eds. 2016. *Vergessene Konstellationen literarischer Öffentlichkeit zwischen 1840 und 1885. Studien und Texte zur Sozialgeschichte der Literatur*. Berlin, Boston: De Gruyter.
- Meyer, Reinhart. 1987. *Novelle und Journal, I: Titel und Normen: Untersuchungen zur Terminologie der Journalprosa, zu ihren Tendenzen, Verhältnissen und Bedingungen*. Stuttgart: Steiner.
- Meyer, Reinhart. 1998. "Novelle und Journal." In *Zwischen Restauration und Revolution 1815–1848. Hansers Sozialgeschichte*, vol. 5, edited by Gerd Sautermeister et. al., 234–50. München: Hanser.
- Müller, Ralph. 2010. "Korpusbildung." In *Handbuch Gattungstheorie*, edited by Rüdiger Zymner, 23–25. Stuttgart, Weimar: Metzler.
- Pawłowski, Tadeusz. 1980. *Begriffsbildung und Definition*. Berlin, New York: De Gruyter.
- Polheim, Karl Konrad. 1965. *Novellentheorie und Novellenforschung: Ein Forschungsbericht*. Stuttgart: Metzler.
- Polheim, Karl Konrad. 1981. "Gattungsproblematik." In *Handbuch der deutschen Erzählung*, edited by Karl Konrad Polheim, 9–16. Düsseldorf: Bagel.
- Rath, Wolfgang. 2008. *Die Novelle: Konzept und Geschichte*. 2nd ed. Göttingen: Vandenhoeck & Ruprecht.

- Reclam. 2011. *Erzählungen und Novellen des 19. Jahrhunderts* Vol. 1. Stuttgart: Reclam.
- Rosch, Eleanorl. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by Eleanor Rosch and Barbara B. Lloyd, 27–48. Hillsdale, NJ: Lawrence Erlbaum.
- Ryan, Marie L. 1981. "Introduction: On the Why, What, and How of Generic Taxonomy." *Poetics* 20 (2–3): 109–126.
- Schöch, Christof. 2017. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly* 11 (2). <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>.
- Schröter, Julian. 2019. "Gattungsgeschichte und ihr Gattungsbegriff am Beispiel der Novellen." *Journal of Literary Theory* 13 (2): 227–57.
- Strube, Werner. 1986. "Sprachanalytisch-philosophische Typologie literaturwissenschaftlicher Begriffe." In *Zur Terminologie der Literaturwissenschaft*, edited by Christian Wagenknecht, 35–49. Stuttgart: Metzler.
- Swales, Martin. 1977. *The German Novelle*. New Jersey, Princeton: Princeton University Press.
- Tarski, Alfred. 1935. "Der Wahrheitsbegriff in den formalisierten Sprachen." *Studia Philosophica* 1: 261–405.
- Titzmann, Michael. 1991. *Modelle des literarischen Strukturwandels. Studien und Texte zur Sozialgeschichte der Literatur*. Tübingen: Niemeyer.
- Underwood, Ted. 2016. "The Life Cycles of Genres." *Journal of Cultural Analytics*. 2 (2) <https://doi.org/10.22148/16.005>.
- Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago, London: The University of Chicago Press.
- Walton, Kendall L. 1970. "Categories of Art." *Philosophical Review* 79 (3): 334–67.
- Wiese, Benno von. 1963. *Die deutsche Novelle von Goethe bis Kafka*. Düsseldorf: Bagel.
- Zymner, Rüdiger. 2003. *Gattungstheorie. Probleme und Positionen der Literaturwissenschaft*. Paderborn: mentis.

