Aspect-based Sentiment Analysis for German: Analyzing "Talk of Literature" Surrounding Literary Prizes on Social Media

Lore De Greve*
Pranaydeep Singh**
Cynthia Van Hee**
Els Lefever**
Gunther Martens*

** Department of Literary Studies, Faculty of Arts and Philosophy, Ghent University, Belgium*

*** LT³ Language and Translation Technology Team, Department of Translation, Interpreting and Multilingual Communication, Faculty of Arts and Philosophy, Ghent University, Belgium*

Abstract

Since the rise of social media, the authority of traditional professional literary critics has been supplemented – or undermined, depending on the point of view – by technological developments and the emergence of community-driven online layperson literary criticism. So far, relatively little research (Allington 2016, Kellermann et al. 2016, Kellermann and Mehling 2017, Bogaert 2017, Pianzola et al. 2020) has examined this layperson user-generated evaluative "talk of literature" instead of addressing traditional forms of consecration. In this paper,[1] we examine the professional and layperson literary criticism pertaining to a prominent German-language literary award: the Ingeborg-Bachmann-Preis, awarded during the Tage der deutschsprachigen Literatur (TDDL).

We propose an aspect-based sentiment analysis (ABSA) approach to discern the evaluative criteria used to differentiate between 'good' and 'bad' literature. To this end, we collected a corpus of German social media reviews retrieved from Twitter, Instagram and Goodreads and enriched it with manual ABSA annotations: aspects and aspect categories (E.g., TEXT_Motifs_Themes, JURY_Discussion_Valuation), sentiment expressions and named entities. In a next step, the manual annotations are used as training data for our ABSA pipeline including 1) aspect term extraction, 2) aspect term category prediction and 3) aspect term polarity classification. Each pipeline component is developed using state-of-the-art pre-trained BERT models.

Two sets of experiments were conducted for the aspect polarity detection: one where only the aspect embeddings were used and another where an additional context window of five adjoining words in either direction of the aspect was considered. We present the classification results for the aspect category and aspect sentiment prediction subtasks for the Twitter corpus as well as the next steps to tackle aspect term extraction. These preliminary experimental results show a good performance and accuracy for the aspect category classification, with an F1-score of 0.81, and for the aspect sentiment subtask, which uses an additional context window, with an F1-score of 0.72.

1. Introduction

In recent times, the knowledge of a limited number of professional 'pundits' is being challenged by technological developments and the 'wisdom of the crowds'. Ample research has been devoted to the ways in which technology causes shifts in the consecration of literary texts, affecting gatekeepers such as literary prizes (English 2009, Sapiro 2016), or professional critics' position of authority (Dorleijn

1. which ties in with the FWO-funded research project entitled "Evaluation of literature by professional and layperson critics. A digital and literary sociological analysis of evaluative talk of literature through the prism of literary prizes (2007-2017)" (2019-2023).

et al. 2009, Löffler 2017, Thomalla 2018, Schneider 2018, Kempke et al. 2019, Chong 2020). Nevertheless, comparatively little research (Allington 2016, Kellermann et al. 2016, Kellermann and Mehling 2017, Bogaert 2017, Pianzola et al. 2020, e.g.) has actually attempted to directly ingest and mine the content of user-generated online literary criticism. Text mining could help to examine the role of peer-to-peer recommendation systems and layperson critics as new literary gatekeepers and cultural transmitters. It is an important case in point to study the differences between professional critics and this 'wisdom of the crowd', especially since traditional gatekeepers of the literary field (e.g., publishers, reviewers) are increasingly trying to tap the potential of online reading communities.

We aim to present the language technologies used in the context of the FWO-funded research project entitled "Evaluation of literature by professional and layperson critics. A digital and literary sociological analysis of evaluative talk of literature through the prism of literary prizes (2007-2017)" (2019-2023)[2], which compares, analyses and mines the evaluative 'talk of literature' of both professional and layperson critics surrounding six prominent literary prizes in three different languages. In this paper, we will present our annotated corpus and suggest a sentiment analysis-based methodology to examine the professional and layperson literary criticism pertaining to the German-language *Ingeborg-Bachmann-Preis and the Tage der deutschsprachigen Literatur (TDDL)*[3]. During this event, all nominated contenders read an unpublished narrative text in front of a jury and a live (television) audience. This literary prize is unique because the contributions are discussed by the professional jury in the presence of the author and the live audience, but also, and even increasingly so, by an online audience. A devoted following of ca. 1000 Twitter followers react, by using the #tddl-hashtag, both to the literary text and its discussion by the official jury.

Our ultimate goal is to gain insights into the evaluative criteria used by both professional and layperson critics to tell 'good' from 'bad' literature, as well as to engage with the differences in evaluation practices across platforms and media. In order to do this, we aim at performing fine-grained aspect-based sentiment analysis (ABSA) on an annotated corpus consisting comments and reader reviews on social media platforms, such as Twitter, Instagram and Goodreads. In future, this system should make it possible to detect which sentiment is being expressed about a certain aspect or *feature expression* (e.g. contender, nominated book, jury, etc.) and named entities mentioned in such comments, and by whom. Consequently, we search to construct literary value through evaluative diction by using ABSA. In this paper, we mainly focus on both advantages technical challenges raised by the nature of the corpus and the annotation system. Furthermore, we aim to describe the preliminary conditions for arriving at a model that will allow to perform fine-grained sentiment analysis on our corpus, which will be the main focus of machine learning experiments at a later stage and which are beyond the scope of the current contribution.

2. Related Research

While ABSA is a common task with regard to domains such as restaurants, consumer technology and, to a lesser extent, movies, there have been few attempts to apply ABSA to domains that express sentiment in less lexicalized and/or straightforward ways. Jurafsky (2016) has done similar work on "linguistic markers of status in food culture" (Jurafsky et al. 2016). In an article on Australian Book Reviews, Stinson argued that"[c]omputational sentiment analysis—at least the kind enabled by off-the-shelf software tools—does not yet present an adequate means for determining polarity of book reviews" (Stinson 2016). Stinson also argues for the necessity of going beyond sentence-level sentiment mining and doing ABSA in view of a recurring trait of corpora containing literary criticism, namely their tendency to voice criticism by means of the "compliment sandwich"(2016), in

2. for more information: https://research.flw.ugent.be/en/projects/evaluation-literature-professional-and-layperson-critics-digital-and-literary-sociological or https://www.talklitmining.ugent.be/.
3. Translation: "Days of German-Language Literature".

other words, by making use of elements of the epideictic discourse strategy of praising and blaming for purposes of nuance and comparison.

3. Corpus Construction and Annotation

While the overarching project's entire corpus comprises comments on literary prizes in three different languages (i.e. English, Dutch and German), we focus on the specific challenges raised by performing sentiment mining on our German-language subsection of the corpus, namely on the Tweets about the Ingeborg-Bachmann-Prize. The time frame for the collected data ranges from 2007, when Twitter was founded and the very first tweets about the Bachmann Prize were created, up until 2019.The Bachmann-Preis has had its own official Twitter account, @tddlit, and encourages the online audience to use #tddl as the hashtag when tweeting about the TDDL and the Bachmann-Preis (Diener 2020) .[4] In addition to the official #tddl-hashtag, we scraped similar relevant hashtags, for example by adding the year to the official hashtag, such as #tddl16 and #tddl2016, or by looking for other terms that might refer to the prize, e.g., #bachmannpreis or #bachmannwettbewerb. This led to a definitive list of 46 hashtags used between 2007 and 2019 (see Figure 1). We then collected all tweets containing these hashtags and removed those created outside of the examined time frame.

All scraped TDDL-Hashtags on Twitter (2007-2019)	
• bachmannbewerb	• tddl08
• bachmannpreis	• tddl09
• bachmannpreis2010	• tddl10
• bachmannpreis2013	• tddl11
• bachmannpreis2014	• tddl12
• bachmannpreis2015	• tddl13
• bachmannpreis2016	• tddl14
• bachmannpreis2017	• tddl15
• bachmannpreis2018	• tddl16
• bachmannpreis2019	• tddl17
• bachmannpreisträger	• tddl18
• bachmannpreisträgerin	• tddl19
• bachmannpreisträgerinnen	• tddl2009
• bachmannpreiswettbewerb	• tddl2011
• bachmannwettbewerb	• tddl2012
• bachmannwettbewerb2018	• tddl2013
• ingeborgbachmannpreis	• tddl2014
• ingeborgbachmannpreis2018	• tddl2015
• ingeborgbachmannpreisträgerin	• tddl2016
• tagederdeutschsprachigenliteratur	• tddl2017
• tagederdeutschsprachigenliteratur2018	• tddl2018
• tddl	• tddl2019
• tddl07	• tddlkanon

Figure 1: Overview of the scraped TDDL-related hashtags.

So far, we have manually annotated the 2019 run of the literary prize's online back-channel. In our paper, we present the annotation procedure. In addition, we present the following steps towards automatising (using a semi-supervised learning system) annotation of this corpus. We will employ a tripartite polarity, using the labels 'positive', 'neutral' and 'negative' to label the sentiment expressions or *Polarity Triggers* in our corpus. Eventually, these will then be linked to the aspects or

4. To safeguard the personal and privacy rights, tweets will be cited by mentioning only the tweet-ID, name of the website, date and last access, e.g. 867326032038199297. *Twitter*, 24 May 2017. Accessed 14 September 2020.

feature expressions as well as the named entities occurring in the comments. In order to categorise the aspects that are mentioned and evaluated in the tweets, we will identify the relevant target words and label these, using a layered labelling system consisting of 7 main categories that are relevant in the context of literary prizes, namely “Text”, “Reading”, “Contender”, “Jury”, “Onsite Audience”, “Meta” and ”Allo-References”. These categories group all aspects referring respectively to the nominated and competing texts, the live author-readings of said texts, the competing authors, the official jury of the prize, the audience present in the Bachmann-Preis studio and the meta-aspects of the prize. Each main category is then divided into smaller and more specific subcategories, as illustrated by Figure 2. For the Text category, for example, there are specific subcategories for those feature expressions that concern the characters, the general content or plot, the language of style etc. One of our future aims will be to discover how fine-grained the automatised identification and labelling of such aspects or feature expressions may be.

Feature Expressions: Categories	
Main Category	**Subcategories**
Text	• Characters • Flow/ Rhythm/ Punctuation • Form • General • General Content/ Plot • Language/ Style • Motifs/ Themes • Point of View /Narration • Quote • Title
Reading	• Flow/ Rhythm/ Punctuation • General • Pronunciation/ Intonation/ Understandability
Contender	• Age • Appearance/ Clothing • Gender • General • Quote • Voice/ Language Use
Jury	• Age • Appearance/ Clothing • Behaviour • Discussion/ Valuation • General • Quote • Voice/ Language Use
Onsite Audience	• Age • Appearance/ Clothing • Behaviour • General
Meta	• Literature/ Literary Prizes • Location • Longlist • Main Event • Montage • Music • Online Assessment • Opening/ Opening Speech • Shortlist • Side Event • Technology/ Social Media • Videoportrait • Voting • Weather • Winner/ Award ceremony

Allo-References	• General
	• Music
	  ◦ Musician
	  ◦ Music
	• Other Person
	• Screen
	  ◦ Director/Actor
	  ◦ Film/Tv
	• Text
	  ◦ Other Author
	  ◦ Other Text

Figure 2: Overview of all aspect or feature expression (FE) categories

As the annotations for this corpus are not yet completely finished, the finalised corpus statistics regarding the polarity triggers, feature expressions, etc. are not yet available at this point of the research. However, these will follow as soon as the annotations are finalised.

4. Experiments

Mining for sentiment is feasible because of the somewhat ritualised and formulaic nature of the communication involved. However, there is a fair deal of ambiguity in the actual rhetoric of praise and blame. In a small-scale experiment, we performed sentence-level sentiment analysis by means of the standard bert-base-uncased model from the Transformers repository by HuggingFace [5]. The models in question are trained for the objective of optimising MLM (Masked Language Modelling). The MLM objective gives the models an unprecedented understanding of the language and vocabulary while also adding contextualisation to the generated embeddings. The advanced language learning makes these transformers very efficient at being trained for downstream tasks like part-of-speech tagging, named-entity recognition, or in this case, sentiment analysis.

```
0] sentiment_model30('Like the characters, the dialogue can be stilted and unconvincing, all too obviously serving the novel’s themes')

   [{'label': 'NEGATIVE', 'score': 0.9996254444122314}]

1] sentiment_model30('I hate this movie')

   [{'label': 'NEGATIVE', 'score': 0.9996687769889832}]

3] sentiment_model30('I finished The Snow Kimono with a queasy sense of discomfort, and not, I sense, of the sort intended.')

   [{'label': 'NEGATIVE', 'score': 0.9994954466819763}]

4] sentiment_model30('Wilson’s vision of Launceston town is hellish, but this is not to detract from the novel’s vitality or its perfectly rendered dialogue.')

   [{'label': 'POSITIVE', 'score': 0.9917587637901306}]

]
```

Figure 3: Use of BERT to predict the sentiment of English-language literary criticism

The examples used in Figure 3 stem from the previously mentioned study of English-language literary criticism (Stinson 2016). From a structural point of view, the sentences are highly similar to the sentences in our corpus: They are variations of epideictic discourse (rhetoric of praise and blame); they contain positive and negative evaluations within one sentence and make use of indirectness such as litotes (understatements). All of these elements contribute to a very nuanced assessment that is probably typical of art criticism in general. The sentences rarely contain superlatives, exclamation marks, predicative structures (e.g., “The new iPhone is really great”) or straightforward evaluations, discourse elements typical of social media discourse. Notwithstanding this relative degree of complexity, the Transformers language model proves to be very efficient in establishing the overall sentiment of these statements: despite the verbosity and indirectness of the nuanced assessments of quality, the overall sentiment is guessed correctly. This means that also negation and attribution through adjectives and adverbs is taken into account automatically.

Although English sentiment and opinion mining can be dealt with using BERT-based models out of the box, this methodology has, however, proven less reliable with regard to our German corpus. We did a similar experiment with sample sentences from the German-language Bachmann Prize corpus. The German language model is obviously less comprehensive, although it does manage to deal with some intricate cases of negation, as can be seen in Figure 4:

But in general, the German model errs on the safe side of things by labelling utterances that are not predicative as neutral. This holds valid even for quite damning and polemical assessments that are untypically blunt in their (d)evaluation of literary artefacts. So what you gain on the swings of the large Transformers architecture, you lose on the roundabouts of the actual training that has gone into the actual model. Moreover, the models did not provide insights into the typical aspects or feature expressions talked about in the corpus. In order to have a better understanding of how evaluations are made in our corpus, and to which specific aspects they refer; we aim to apply ABSA and look into domain adaptation techniques. Although there is some ambiguity attached to the actual rhetoric of praise and blame, the actual evaluation is somewhat similar to the discussion of food by (Jurafsky et al. 2016) and (Vásquez and Chik 2015), with specific standard phrases. One example of this

5. https://github.com/huggingface/transformers

```
[7] from germansentiment import SentimentModel

    model = SentimentModel()

    texts = [
        "Das ist Kitsch. #AdaDorian #tddl16",
        "Ja aber Der Roman eines Schicksallosen ist doch gerade kein kitsch, gerade stilistisch brilliant #tddl #tddl19",
        "Katharina Schultens ist eine sehr gute Lyrikerin - und eigentlich müsste sich der Stil gut in Prosa übertragen lassen. Aber das hier ist nichts, stilistisch. #t
        "Das kommt dabei raus, wenn du schreiben willst, aber nichts zu sagen hast. Oder nicht schreibst, was ist. #tddl",
        "Ich war noch nie so froh darüber, dass ein Text endlich aus ist. #tddl",
        "Ich habe auch Bock, nonchalantes Gelaber runterzuschreiben. Welchen Genozid hätten‘s denn gerne, liebe Leser*innen? Ich setze nur schnell den Kaffee auf. #tddl"
        "Wäre ein schöner Text, wenn das denn gelungen wäre - hier jedenfall nicht. Könnte man wunderbar schreiben, Propaganda und Stiegenhausplatitüden, selektive Beoba
        "Federer und Heitzler auf der Shortlist und Birkhan fehlt? Unverständlich. Das muss ein Fall von Quotenmännern sein. #tddl",
        "Und alles, was am Ende vom Bachmannpreis bleibt, sind Literaturbetriebler, die abends im Strandbad den Handtuchhüfttanz aufführen, um völlig eleganzbefreit nass
        "@Bov Jetzt ist's mir klar, der Text ist nur eine reine SEO-Maßnahme ;-) #tddl #kurznasenseefledermaus",
        "Eindrucksvoller poetischer Essay. Aber Thema verfehlt. #tddl",
        "nicht schlecht genug, um so richtig loszulästern, aber heute in textgesellschaft weitaus gelungenerer beiträge #kränzler #tddl",
        "Schon sehr viel Feier des Handwerks. Dachte immer, das sei die Mindestvoraussetzung. #tddl",
        "handwerk, genau. autorInnen sind handwerkerInnen. #tddl"]
    result = model.predict_sentiment(texts)
    print(result)

    ['negative', 'neutral', 'neutral', 'negative', 'neutral', 'negative', 'negative', 'neutral', 'neutral', 'negative', 'neutral', 'neutral', 'neutral', 'neutral']
```

Figure 4: Use of BERT to predict the sentiment of German-language literary criticism

is the phrase “well-made” (or ”handwerklich gut gemacht” in German), which somewhat counter-intuitively is a negative marker, connoting effort and hard work rather than spontaneous creativity, as illustrated by this Tweet implicitly discussing the jury discussion and TDDL’s literary criteria ”Falls ich mal in Klagenfurt lese, ich will nicht für Handwerk gelobt werden. #tddl”[6].

To be able to tackle the challenges posed by our initial experiments, we aim to conduct a new set of machine learning experiments to automatise a three-step analysis of our corpus: 1) extract all aspects or feature expressions, 2) categorise the aspects into their relevant categories and 3) determine whether the sentiment expressed about the aspects is positive, negative or neutral.

5. Future Work and Experiments

Specifically, there are two directions to improve sentiment-based learning of the model. Firstly, we hypothesize that German BERT[7] which is trained on the German Wikipedia and other similar corpora with formal language, may lack an inherent understanding of social media language and it’s nuances. To this end, we propose to re-train German BERT with additional Twitter data. Secondly, as mentioned above, sentiment can be a multi-layered construct in the context of art reviewing. We therefore also hypothesise that simply a fraction of our corpus may not be sufficient to instil an understanding of the complications with polarities and aspect detection. To solve this issue, we propose to pre-train our version of German BERT with related English sentiment datasets, by translating them to German. We believe these two additional learning signals would be sufficient to out-perform previous approaches to ABSA for literary reviews.

References

Allington, Daniel (2016), ”’Power to the Reader’ or ’Degradation of Literary Taste’? Professional Critics and Amazon Customers as Reviewers of ’The Inheritance of Loss’ ”, *Language and Literature* **25** (3), pp. 254–278. https://doi.org/10.1177/0963947016652789.

Bogaert, Xiana (2017), *‘ICH WÜRDE AM LIEBSTEN MIT DER JURY DISKUTIEREN! #TDDL’ - Der Ingeborg-Bachmann-Preis: Ein Vergleich zwischen der Jury- und Laienkritik auf Twitter.*, Master’s thesis, Ghent University.

6. Translation:”In case I ever read in Klagenfurt [read: as a competitor], I do not wish to be praised for handwork #tddl” 882909365191180288. *Twitter*, 6 July 2017. Accessed 14 September 2020.
7. https://huggingface.co/bert-base-german-cased

Chong, Phillipa K. (2020), *Inside the Critics' Circle: Book Reviewing in Uncertain Times*, Princeton University Press. http://www.jstor.org/stable/j.ctvkwnph1.

Diener, Andrea (2020), "Bachmannwettbewerb 2.0. Fast alles wie immer", *Frankfurter Allgemeine Zeitung*. www.faz.net/aktuell/feuilleton/buecher/bachmannwettbewerb-statt-in-klagenfurt-im-internet-16767473.html.

Dorleijn, Gillis J., Dirk De Geest, and Koen Rymenants (2009), *Kritiek in crisistijd: Literaire kritiek in Nederland en Vlaanderen tijdens de jaren dertig*, Nijmegen: Vantilt.

English, James F. (2009), *The Economy of Prestige: Prizes, Awards, and the Circulation of Cultural Value*, Harvard University Press. https://books.google.be/books?id=vY3UOFDA2sAC.

Jurafsky, Dan, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith (2016), "Linguistic Markers of Status in Food Culture: Bourdieu's Distinction in a Menu Corpus", *Cultural Analytics*.

Kellermann, Holger and Gabriele Mehling (2017), *"Laienrezensionen auf amazon.de im Spannungsfeld zwischen Alltagskommunikation und professioneller Literaturkritik." Die Rezension: Aktuelle Tendenzen der Literaturkritik*, Würzburg: Königshausen und Neumann, pp. 173–202.

Kellermann, Holger, Gabriele Mehling, and Martin Rehfeldt (2016), *"Wie bewerten Laienrezensenten? Ausgewählte Ergebnisse einer inhaltsanalytischen Studie." Was wir lesen sollen: Kanon und literarische Wertung am Beginn des 21. Jahrhunderts*, Würzburg: Königshausen und Neumann, pp. 229–238.

Kempke, Kevin, Lena Vöcklinghuis, and Miriam Zeh (2019), *Institutsprosa: Literaturwissenschaftliche Perspektiven auf akademischen Schreibschulen*, Leipzig : Spector Books.

Löffler, Sigrid (2017), "Danke, kein Bedarf? Wie die totgesagte Literaturkritik ihr Ableben überleben könnte", *Stimmen der Zeit <Freiburg, Breisgau>* **235** (12), pp. 850–814.

Pianzola, Federico, Simone Rebora, and Gerhard Lauer (2020), "Wattpad as a Resource for Literary Studies. Quantitative and Qualitative Examples of the Importance of Digital Social Reading and Readers' Comments in the Margins", *PLOS ONE* **15** (1), pp. 1–46, Public Library of Science. https://doi.org/10.1371/journal.pone.0226708.

Sapiro, Gisèle (2016), "The Metamorphosis of Modes of Consecration in the Literary Field: Academies, Literary Prizes, Festivals", *Poetics* **59**, pp. 5–19. https://www.sciencedirect.com/science/article/pii/S0304422X16000103.

Schneider, Ute (2018), *"Bücher zeigen und Leseatmospären inszenieren. Vom Habitus enthusiastischer Leserinnen und Leser." Gelesene Literatur: Populäre Lektüre im Zeichen des Medienwandels*, München: Edition Text+Kritik, pp. 113–123.

Stinson, Emmett (2016), "How Nice is too Nice? Australian Book Reviews and the 'Compliment Sandwich"', *Australian Humanities Review* **60**, pp. 108–126.

Thomalla, Erika (2018), *"Bücheremphase: Populäre Literaturkritik und Social Reading im Netz." Gelesene Literatur: Populäre Lektüre im Zeichen des Medienwandels*, München: Edition Text+Kritik, pp. 124–136.

Vásquez, Camilla and Alice Chik (2015), "'I Am Not a Foodie...': Culinary Capital in Online Reviews of Michelin Restaurants", *Australian Humanities Review* **23** (4), pp. 231–250.