“Joy” and “Fear” in Thomas Bernhard’s autobiographies: Aspects of a Computational Sentiment Analysis

Manfred B. Sellner
University of Salzburg
Manfred.Sellner@sbg.ac.at

Abstract

This pilot-study of a computational analysis of literary texts presents the results of aspects of a “sentiment analysis”. The data of analysis are the autobiographies of the Austrian novelist Thomas Bernhard. The primary object of attention are the sentiments “joy” and “fear”. We elaborate on and demonstrate the impact of several preprocessing procedures, describe the characteristics of the dictionary and the annotations of its entries conceived and used for analysis. We specify the general methodology and the steps involved for quantifying of its result by the use of the functions of the R-package “Quanteda”. The descriptive output of the procedures is examined with several statistical measures to compare the counts of “joy” vs “fear” that were found in the texts individually, contrastively and in combination as a corpus. We conclude that there is a proportional and relative difference between the frequencies of the sentiments of the individual texts, but that this observation is insignificant if interpreted on the basis of the non-parametric Wilcoxon rank-sum test. A “goodness of fit” test, on the other hand, shows that the two sentiments show a homogeneous distribution across the corpus.

Keywords: computational analysis, literary sentiment analysis, sentiment dictionary

Sentiment Analysis is a computational method for analyzing and interpreting aspects of textual characteristics. While this N(atural L(anguage) A(pproach) has its roots and main application in the detection of positive and negative “sentiments” in the world of commerce, it has also experienced a recent surge in popularity in the field of literary studies, covering many diverse topics of novels and plays.[1] Following up on previous work that concentrates on the detection and interpretation of emotional categories, we shall explore the impact of the lexical expressions of the two basic emotions “joy” and “fear/anxiety”, as expressed and lexicalized in the original German version of the five autobiographical novels by the Austrian novelist and playwright Thomas Bernhard as “Angst” and “Freude”. The novels, in temporal order of publication, are “Die Ursache. Eine Andeutung” (1975), “Der Keller. Eine Entziehung” (1976), “Der Atem. Eine Entscheidung” (1978), “Die Kälte. Eine Isolation” (1981), and finally, “Ein Kind” (1982).[2] The computational method employed makes use of the list of near-synonyms of a widely used German synonym lexicon in tandem with the computing environment and functions of the open-source programming language “R”, the R-package “Quanteda”, and the plotting

[1] For surveys of sentiment analysis see Pang & Lee (2008), Liu & Zhang (2012), and Mohammad (2015), (2016), (2007), Lehmann, Mittelbach & Schmeier (2017). For an analysis on a commercial topic featuring a German business environment, see Wolfgruber (2015). Some examples of work on various aspects of literary texts are the following works listed in the bibliography: Jacobs (2019), Kim & Klinger (2019), Klinger et al (2016), Mohammad (2011), Nalisnick & Baird (2013), Schmidt et al (2018), Schmidt & Burghardt (2018a), Schmidt et al (2018), & Winko (2003). A synopsis thereof is far beyond the scope of this paper.

[2] The titles of the English translations of Bernhard’s series of autobiographies in the compilation “Gathering Evidence: a memoir” (Translated from the German by David McLintock), are as follows: “An Indication of the Cause”, “The Cellar: An Escape”, “Breath: A Decision”, “In the Cold”, and finally “A Child”. For a competent general account of Bernhard’s writings, and his autobiographic novels, in particular from the viewpoint of literary criticism, see the “Bernhard Handbuch” edited by Huber & Mittermayer (2018).

program “ggplot2”. We see the ultimate aim of such analyses in the computational modeling of the sentiment detection ability and interpretation of literary texts by conscious readers of literary works of art. The analysis presented here can be seen as a rudimentary pilot study towards this distant goal.

Computational analyses require first of all a digitized text that is uploaded to the R-computing environment of a PC. We use the function implemented in the R-package, “readtext”, for this job. We immediately verify the correct import of the texts by calling up a list of the imported text-files and the first 10 elements of the texts. A copy of the results of this methodological step is given in Table 1:

```
readtext object consisting of 5 documents and 0 docvars.
#Description: df [,2]
doc_id                      text
<chr>                       <chr>
1 1975_Die_Ursache.txt    “\“Die Ursach\”…”
2 1976_Der_Keller.txt      “\“Der Keller\”…”
3 1978_Der_Atem.txt       “\“Der Atem, \”…”
4 1981_Die_Kälte.txt        “\“Eine Isola\”…”
5 1982_Ein Kind.txt          “\“Ein Kind. \”…”
```

Table 1

The table shows that we have loaded 5 “objects” into the environment of R and “Quanteda” and that we have created a data frame (df) object with a document identification label (doc_id) for the texts. In addition, we see the names of the files that have been imported (i.e., “the novels”), as well as the first 10 characters of the beginnings of the individual texts. Preprocessing for the sentiment analysis proper is then continued by removing the extension “txt” and by adding the numbers 01-05 for easier identification of the texts in the graphical illustration to be presented in the following sections. We also convert the mixture of the “small and big letter orthography” of German to conform to a general small letter convention of spelling. Finally, we tokenize the texts to create a “Quanteda-Corpus-Object” that gives us an overview of the data for analysis and a first look at some descriptive statistics of the “objects” of the corpus. This information is presented in Table 2 in detail.

Text	Types	Tokens	Sentences	Novel#
75-Die_Ursache	5193	29313	479	01
76-Der_Keller	5258	30152	1016	02
78-Der_Atem	4455	27827	949	03
81-Die_Kälte	5233	27729	1121	04
82-Ein_Kind	6178	31670	1909	05
mean	5263.40	29338.17	1084.80	
sd	611.53	1481.45	517.27	

Table 2: Frequencies of types, tokens, and sentences, mean, and standard deviation (sd)

The table shows the frequency of the types, the tokens, the sentences, the means of the columns, as well as the standard deviations of the texts, in order to facilitate the use of these figures as reference in follow-up studies. For later visualization of the individual novels, we have also added the column “Novels#” that associates each novel with a corresponding number from (01) to (05).

Inspecting the data of the corpus to flesh out some quantitative aspects of a descriptive nature, we find that the individual autobiographies do not have the same number of sentences, tokens and types, so that there is apparent variation between the texts in these respects. Variation is not unexpected in the contexts of literary works that are not confined to length in principle, though, intuitively speaking, the standard deviation of the tokens and the types of the corpus seems rather large. In any case, there is a minimum of 5193 types and a maximum of 6178, which makes a difference of 16%, a minimum of 27729 and a maximum of 31670 tokens, resulting in a difference of 12.5% between minimum and maximum. While standard deviation and mean are difficult to interpret when judged non-comparatively, we do see that the frequencies of the tokens and the types of our corpus of novels do neither show a consequent rise nor fall. They vary in either direction when compared over the time span of their publication, which is always given as the first figure in the identification label of the respective novel. The line graph of Figure 1 demonstrates this characteristic, showing that the line points do not follow a straight horizontal line. We can thus recognize that later publications, with the exception of "Ein Kind", are not of increased length, although we can imagine that increased expertise and fame on the part of a novelist can result in greater freedom not only in selecting the topics, but also in the length of a manuscript submitted to the publisher under contract.[3] In contrast, there is a continuing rise of the frequencies of the types found in the novels after the publication of "Der Atem", which itself marks the lowest level of types of all the novels.

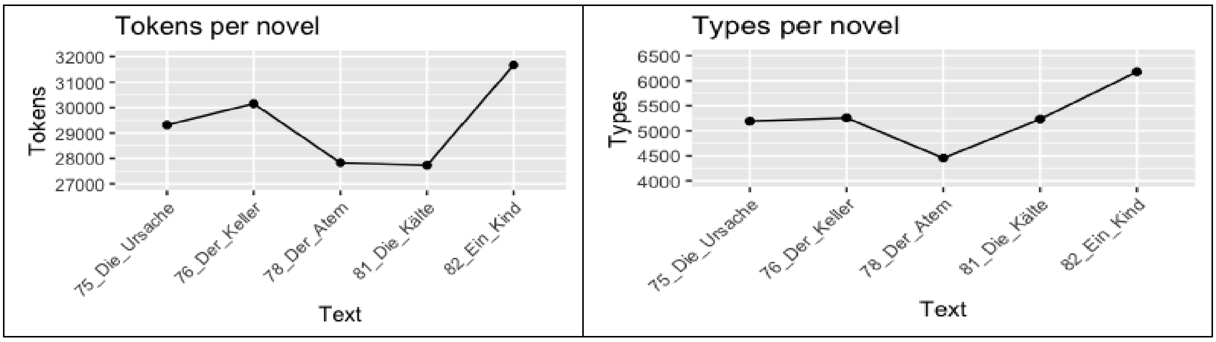

Figure 1: Types and Tokens per novel (non-standardized raw data)

While the inspection of the graph and the corresponding figures of Table 2 with respect to the frequency of types and tokens of the individual novels seem to indicate differences of a quantitative nature, we want to know if the distribution that spreads from 4455 tokens ("Der Atem") to 6178 tokens ("Ein Kind") is founded and supported in a significant way statistically speaking. For this purpose, we apply a "goodness-of-fit-test" to the frequency data and postulate the conventional null hypothesis that "there will be no frequency difference between the elements of the corpus".[4] As indicated by the figures below in detail, the results of the chi-square test of the base program of R disconfirm the null hypothesis, enabling us to safely assume that the differences between the variables, i.e., sentences, tokens, and types, are indeed highly significant (sentences: x-squared=977.61, df=4, p=<2.2e-16, expected in even distribution=1094.8; tokens: x-squared=374.03, df=4, p=2.2e-16, expected in even distribution=29338.2; types: x-squared=403.03, df=4, p=<2.3e-16, expected in even distribution=5253.2). Thus, basing our conclusions on intuitive interpretation and statistical calculation, we conclude confidently that the novels have not been "constructed" to become quantitative equivalents, as all the frequencies of the data elements tested do differ significantly. It is therefore assumed that the subtle differences between the variables that we realized so far substantiate the requirement for a standardization of the frequencies of the texts if we compare the elements of the texts quantitatively. This will be done in due course.

Turning now in greater detail to the characteristics of the "types" of the corpus, we want to entertain and initially support the idea that the general choice of different words is a possible criterion

[3] See the list of novels and plays that Thomas Bernhard wrote between 1975 and 1982 in Mittermayer (2015) account of "Thomas Bernhard" listed in the bibliography.
[4] See, among others, Hatzinger, et. al. (2014[2]) & Meindl (2011).

measuring the increasing linguistic skill of an author with respect to (near)-synonymic vocabulary variation. This possibility leads us to the calculation of the "Type-Token-Relationships" (TTR) of the five autobiographies, as TTR is regarded as a basic measure of the diversity of the vocabulary of written texts. This relationship between types and tokens is determined on the basis of the proportion between the total number of tokens and the total number of diverse tokens ("types"). In a TTR-calculation we obtain a result that lies between "0" and "1", where "1" is the (unlikely) result of a computation in case every "token" is also a "type". This situation might be interpreted as an indicator of "perfect style", because the vocabulary is very diverse and not many tokens of the content category have been used many times.[5] A TTR-value close to zero then means that many tokens have been used very often. The numerical result of the TTR-calculation is presented in Table 3 in the column "TTR including stopwords". It shows that the TTR in the individual texts is rather low and not very diverse, as the values of the TTR all lie below 0.20.

document	TTR (including stopwords)	TTR (excluding stopwords)	Tokens (excluding stopwords)	Types (excluding stopwords)	Novel#
75_Die_Ursache	0.1771288	0.3904702	12823	5007	1
76_Der_Keller	0.1743831	0.4671207	12774	5967	2
78_Der_Atem	0.1600963	0.3574418	11946	4270	3
81_Die_Kälte	0.1887194	0.4187864	12030	5038	4
82_Ein Kind	0.1950742	0.4122348	14516	5948	5
mean	0.1790804	0.4092108	12817	5253	
sd	0.0135527	0.04026671	1032	727	

Table 3

As the TTR values seem very low, we decided to entertain the possibility that this result is due to the inclusion of so-called "stopwords" in the calculation of the TTR. Thus, we delete the stopwords from the data for calculating the TTR.[6] The result is given in the column "TTR excluding stopwords". The calculation shows that the result is quite different from the calculation that included the stopwords. In eliminating them from the texts, a practice quite frequent in corpus linguistics, we now find that the TTR does rise to reasonable values in all the novels and to a maximum of 0.46 in one novel ("Der Keller"). Interestingly, though, the TTR-hierarchy has changed, as well. This means that we have found a TTR-hierarchy of 5<4<1<2<3 for absolute tokens and types and a hierarchy of 2<4<5<1<3 when stopwords are removed. Thus, the presence of stopwords in the data has indeed influenced the vocabulary-related results to become more realistic for a literary writer of fame. The difference between the two ways of calculation of the TTR is also clearly visible in the figures 2 and 3 where we compare the frequency of the TTRs to the means of their respective distributions. In Figure 1 we find that the TTR of novel 1 is very close to the means of the distribution, that novels 4 and 5 are above the means, and that novels 2 and 3 are below the means. In comparison to the distribution that includes the stopwords. But as the data differed only with respect to the inclusion or exclusion of the set of stopwords, we can safely conclude that this variable is responsible for the Type/Token-differences though we will show over the course of this analysis that the issue of "stopwords" is not detrimental to the sentiment analysis proper, because we have chosen an adequate corpus size for data normalization, so that we can compare the frequencies of the corpus texts without restraint. Concluding briefly the characteristics of the text, we can state that

[5] This assumption clearly is in need of further qualification and investigation.
[6] Quanteda has a built-in list of German stopwords. This is a list of "unwanted closed-class words" that are deleted from the original, as they are seen to be irrelevant for analysis. (See, for ex., Desagulier 2017: 110)

the individual texts are rather short, and that there is remarkable variation both with respect to token and type, and their relationship.

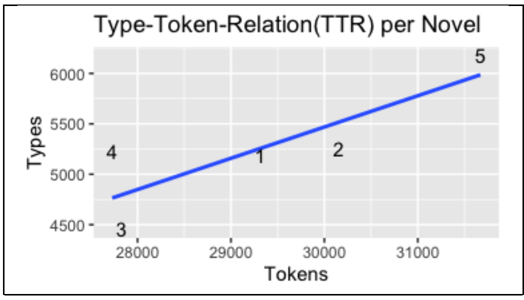

Figure 2: TTR-Analysis including stopwords (mean: ≈0.18)

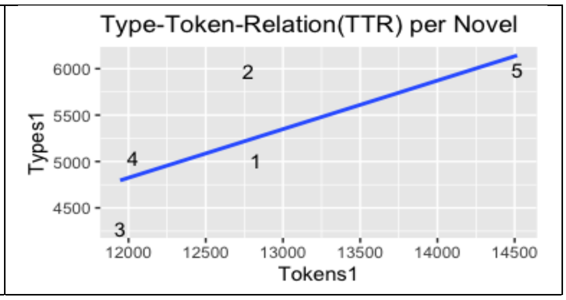

Figure 3: TTR-Analysis excluding stopwords (mean: ≈ 0.41)

The aim of the sentiment analysis here is to specify the frequencies, the distribution and the share of the emotions “Freude”/“joy” and “Angst”/“fear” in the five autobiographic novels by Thomas Bernhard.[7] The method for achieving this information from the surface elements of the texts is rather straightforward. In essence, it is a process that performs an automatic matching procedure between a dictionary and a text. It begins with an adaptation of the list of elements of a dictionary, so that the algorithm of the program will perform the matching job. We thus apply the “dictionary construction function” that is implemented in the R-package “Quanteda” to prepare the elements of the dictionary for the “equivalence matching”-procedure to the elements of the texts of the corpus. This means that a general annotation of the text elements is not necessary, but that we devote our attention – first of all – to the preparation of the list of elements to make them maximally general for the “look-up”-function. In doing this, we devise and adopt a system of annotation that will avoid the spelling-out of most of the possible variations of a morphological form, and at the same time it will accommodate possible lexical innovations and extensions that might be encountered in the texts of analyses.

The number of NLP-Lexicons for German that are of potential use for our task is manageable, though judged not to be ideal for the research at hand. One of the reasons for this is that most of them are not designed for the general calculation of sentiment shares but for the “positive” and “negative” interpretation of the specific contents of texts, including their “granularities/valence”.[8] Consequently, we resolved to devise a basic dictionary of our own, and to start off towards this interim goal using the basic entries of the (German) “Duden Dictionary of Synonyms”. This dictionary is “linguistically oriented” in the sense that it was assembled with a theoretical orientation towards aspects of general semantic prototype principles. This means that the lexemes in the dictionary entries are regarded as “near-synonyms” of an abstract headword that is also realized as a concrete member of the category. The members of the prototype can be judged to be situated in the center of the prototype but also in or towards its periphery. This means that the actual, possible use of an individual lexeme is thought to depend on the appropriate pragmatic and linguistic context that conforms to the linguistic intuition of the user. If searching for “an appropriate word” in a specific context, the popular Duden dictionary of

[7] The emotions “fear” and “joy” are taken from Ekman’s (1999) set of “The Big Six”, which are also referred to as “basic emotions”. They include in addition to the two emotions “fear” and “joy/happiness” also “sadness”, “surprise”, “anger”, and “disgust”. For details and a general overview see for example Piórkowska & Wróbel (2017).

[8] See Puschmann & Haim (2010) in this respect: “The aim of sentiment analysis is to determine the polarity of a text (i.e., whether the emotions expressed in it are rather positive or negative).” See exemplary references to dictionaries, such as Rauh (2018), Remus (2010), Sidarenka & Stede (2016), Schmidt & Burghardt (2018b), and Werlen et al (2019) and the examples of dictionaries in Pushman (2019). Mohamad’s (2020:6) opinion is noteworthy in this context when he summarizes the issue of annotation of emotions as follows, “The degree of agreement between annotators is significantly lower when assigning valence or emotions to instances, as compared to tasks such as identifying part of speech and detecting named entities.” For this reason, we opt for an approach that circumvents the problem of the annotation of possibly relevant text sequences.

synonymous lexemes offers its users the choice of approximately 2000 headwords with associated lexemes and phrases for use. These are furthermore grouped and categorized as nominals (e.g., "Angstgefühl"/"feeling of anxiety"), stative ("Angst haben"/ "be frightened"), transitive ("Angst einjagen"/"to frighten"), and as qualitative-adverbial ("ängstlich"/"anxious"). Furthermore, there is an important reference (p. 79) to the effect that the German words "Angst" ("anxiety") and "Furcht" ("fear") are used interchangeably in everyday language. On the basis of this observation, we have conflated the entries under the head word "Furcht" ("fear"), and take the list of (near)-synonyms, and regard the entries as the basic matches for the sentiment analysis of Bernhard's autobiographies.

While we assumed from the outset that Quanteda would work well with single tokens ("unigrams"), it seemed advisable to create a small mock-test to test the functions of the program beforehand and to check if the package would also work with a dictionary that contains multiple elements ("multigrams"). Such a test can also give us the opportunity to test the annotation scheme that was planned for the entries of the dictionary list proper. The dummy elements used for annotation, i.e. the asterisk, is necessary to cover the generative characteristics and possibilities of creative language use and to keep the number of dictionary entries as small as possible and at the same time "open and general". This will speed up the computational process if the texts are of a greater length dimension. Towards these aims, we create two short texts and call them "doc1" and "doc2" to mimic the texts of a fictitious corpus.[9] The workflow presented in Workflow 1 is identical to the one used for the later sentiment analysis of the corpus of Bernhard's autobiographic novels. They illustrate the functions as such, and especially their interactive and interdependent workflow. The text of the mock-test is given in Text 1. The list of dictionary entries that makes up the actual input of the program are in the column "Dictionary entry" of Table 4. The column to its right lists the presumed matches of the dictionary entries when applied to the texts of the mock-test:

```
text1<-c(doc1="Er ist ein Angsthase. Ihm ist Angst und Bange. Er hat mir Angst gemacht.
Ich werde dir noch Angst machen.", doc2= "Wir sind sehr ängstlich. Ich bin nicht
verängstigt. Mach ihm Angst!")
```

Text 1: Mock-test

Examples	Dictionary entry	Predicted match & pick up in docs1 and docs2
1	Angst*	Angsthase
2	angst und bange	Angst und Bange
3	Angst *mach*	Angst gemacht
4	Angst mach*	Angst mach
5	*ängst*	ängstlich, verängstigt
6	*mach* Angst	Mach ihm Angst

Table 4

We assume that the algorithms of the programs are successful, when the dictionary entries of the keyword "ANGST" correctly replace the lexical items listed in the column "Predicted Match" of Table 4 with the keyword. The functions of the workflow that are anticipated to accomplish this match and the replacement are marked in italics and shown below (Workflow 1)[10]:

[9] A rough and ready translation of the text is as follows: doc1. "He is a scaredy-cat. He is very worried. He scared me. I'll frighten you". Doc2. "We are very anxious (about sth.). We are very scared. Scare the hell out of him!"
[10] The workflow uses the so-called pipe convention (%>%) to shorten the command sequence.

```
a.  convert the entries of the keyword ANGST to a quanteda-dictionary:
> dict1<-dictionary(list(ANGST=c(“Angst*”, “Angst und Bange”, “Angst
*mach*”, “Angst mach*”, “*ängst*”,  “*mach * Angst”)
Dictionary object with 1 key entry.
- [Angst]:
  - angst*, angst und bange, *ängst*, angst mach*, *mach ihm angst, angst
*mach*
b.  tokenize the text, remove the punctuation of the text, and convert the
tokens to lower case:
> text1.toks<-tokens(text1, remove_punct=T)%>%
+ tokens_tolower()%>%
c.  call for the match between text and dictionary:
+ tokens_lookup(dictionary=dict1, exclusive = F)
> text1.toks
d.  Show the results of text1 and text2:
> head(text1.toks[[1]], 30)
 [1] "er"    "ist"   "ein"   "ANGST" "ihm"   "ist"
 [7] "ANGST" "er"    "hat"   "mir"   "ANGST" "ich"
[13] "werde" "dir"   "noch"  "ANGST"
> head(text1.toks[[2]], 30)
[1] "wir"   "sind"  "sehr"  "ANGST" "ich"   "bin"
[7] "nicht" "ANGST" "ANGST"
```

Workflow 1

Inspecting the matches between the text and the dictionary entries, we find that the quanteda algorithm picked up and replaced the annotated unigrams, as well as multigrams. This means that we can start from the assumption that such markups, and the coding scheme as such, will also work for the analyzed novels. What we now have to test is whether the quanteda functions will sum up and output the correct sentiment frequencies. For this part of the mock test, we use the following workflow (Workflow 2):

```
a. Tokenize the text, convert entries to lower case:
> text2<-tokens(text1, remove_punct=T)%>%
+ tokens_tolower()%>%
b. use function “tokens_lookup”:
+ tokens_lookup(dictionary=dict1)%>%
c. use function “dfm” (=document feature matrix)
+ dfm()
d. call up the text:
> text2
Document-feature matrix of: 2 documents, 1 feature (0.0% sparse).
     features
docs   angst
 doc1     4
 doc2     3
```

Workflow 2

As the manual and the computational count of the sentiments of the mock test give us the same result of seven sentiments (i.e., 4 matches in doc1 and 3 matches in doc2), we can confidently use the same functions and procedures for the analysis of the sentiments “joy” and “fear” in the autobiographies selected for sentiment analysis of the respective head words. The methodology of the computation is now regarded as comprehensible. In addition, it seems adequate for the goal set in this pilot study.

The automatic computation procedure for the sentiments of the text and the following summarization of the counts of the two sentiments “joy” and “fear” begins with the listing and adaptation of the senses of the two headwords “Angst” (“fear”) and “Freude” (“joy”). As outlined above, they were taken from the “Duden Dictionary of Synonyms”. This is done irrespective of their putative membership in the center or the periphery of the prototype that is represented by the two headwords. We ignore any indication of the “granularity” of the lexemes, such as a possible numeric or qualitative scale of “very fearful” to “little fearful”, or 1 to 5 and limit the specification of the dictionary entries to their lexical values and assume that the entries represent – at least to a great extent – the intuition of readers who interpret the content of the relevant passages of the texts as “fearful”, or “joyful” so that the computational analysis can be regarded an explication of the intuitions of the readers of the novels. To achieve this goal computationally, we thus need to preprocess the entries of the dictionary list and annotate them with the dummy element “asterisk” for multiple element entries, and the obligatory morphological additions to the stems of lexemes, or for complex word building possibilities. Before turning to the process and result of this “dictionary making process” that is in part exemplified in Table 5, we must admit that we should not overlook the fact that the dictionaries differ quantitatively, as “Joy” has 181 items and “Fear” has 138 items. This obvious numerical difference is in our mind not judged as problematic for the forthcoming analysis, as it merely reflects the linguistic fact that some semantic fields are more pronounced and detailed than others. Also, some entries have to be “spelled out” to avoid false overgeneration, such as “Freude”/“joy” vs “freudlos”/“without joy”, if listed as “freud*”, while others will cover many lexemes coded with a dummy asterisk for lookup in the text, and thereby reducing the number of entries. Furthermore, the addition of case, gender, or derivational specifications can also lead to differences in the overall size of the dictionary, so that the number of the entries of the dictionary can only be a limited indicator of its appropriateness for the following sentiment analysis.

Dictionary object with 2 key entries. - [Angst]: - ahn*, geahnt, alpdruck, angespannt*, angst*, ängst*, ängsten, angst und bange, angst und schrecken, aftersausen*, apathisch*, apprehensiv*, aufgewühlt*, aufschrecken*, aufgeschreckt*, bammel*, bangnis*, bang*, gebangt, bangend* [... and 118 more] - [Freude]: - alberei*, albern*, amüsant*, amüsement*, amüsier*, angenehm*, aus dem häuschen, aus hetz, aufgedreht*, aufgekratzt*, ausgelassen*, auskunftsfreudig*, aus vergnügen, begeistert*, begeisterung*, beglück*, begrüßenswert*, behagen*, belustig* [... and 161 more]

Table 5: Excerpt of Dictionary for the Sentiments “Fear” (“Angst”) and “Joy” (“Freude”) and entries after loading into R

Table 5 illustrates the initial part of the dictionary after its importation into the computing platforms of “R” and “Quanteda”, and after the application of the constructor function “dictionary” to the preprocessed corpus data. Table 5 is a copy of the first 20 entries of the respective keywords “Angst” and “Freude” and shows the tokens of the dictionary object and their annotations with the asterisk notation. As already indicated above, the dummy elements are intended to keep the dictionary as small as possible, while, at the same time, allowing for the pickup of morphological additions of lexemes and of multiple lexical entries with irrelevant parts in between them. The entry also shows the counts of the number of elements found in the dictionary. This is indicated as 118 entries for “Angst” and 181 entries for “Freude”.

The sentiment analysis proper was performed with the original, preprocessed “raw data” of the five novels. We opted against the use of the “non-stopword” version of the data, because their elimination from the texts can result in token objects with tokens that are in the dictionary but not in the texts, so that no match and consequently no count would have been possible. In addition, we face the general problem of homonymity of automated corpus analysis, as the algorithm would have to distinguish between the German particle "an" in a phrase like "ich rufe dich morgen *an*" ("I'll call you tomorrow") and a local adverb, such as in "*an* der Uni" ("at the university"). A solution to this distinction is far beyond the NLP-framework used in this analysis. We follow the workflow introduced in the "mock analysis" of "Text 1" and analyze the tokens of the individual novels with an alike workflow. The result of the output of the analysis of the sentiment counts detected by the automatic matching and consequent counting of the hits is recorded in columns three (3) and four (4) of Table 6:

		Raw Data Col.3	Raw Data Col.4	Normalized Data (50000) Col.5	Normalized Data (50000) Col.6	Proportion/ Difference in % (Normalized Data)
Novels	Novel#	Freq. of "Fear-Counts"	Freq. of "Joy-Counts"	Freq. of "Fear-Counts"	Freq. of "Joy-Counts"	
75_Ursache	01	117	36	199.57	61.40	3.25:1(69.23%)
76_Keller	02	61	65	101.15	107.87	1:1.07(06.23%)
78_Atem	03	66	47	118.59	84.45	1.40:1(28.79%)
81_Kälte	04	56	36	101.00	64.91	1.56:1(35.74%)
82_Kind	05	59	84	93.14	132.62	1:1.42(29.77%)
Sum		359	268	613.45	451.25	1.36:1(26.44%)
Mean		71.8	53.6	122.69	90.25	
Median		61	47	101.15	84.45	
sd		25.53	20.72	43.97	30.06	
%		57.25	42.75	57.62	42.38	1:1.35(24.63%)

Table 6: Overview of results of Sentiment Analysis

The figures of columns three (3) and four (4) of Table 6 show that the sentiment "fear" has been detected more often in novel 1 (117:36), in novel 3 (66:47), and in novel 4 (56:36), while "joy" is more frequently found in novel 2 (65:61), and novel 5 (84:59). Taking the whole corpus as the basis of the result of the counts of the sentiment analysis, we calculate a relative frequency of 57.25% against 42.25% in favor of the "fear" counts. But in order to make a statistically sound comparison across the five novels, we decided go beyond the descriptive figures of the raw data and normalize the frequencies (number of tokens) of the program output, because the novels do not have the same number of elements. Though there are statements in the relevant literature suggesting that the standard of normalization commonly is 1,000,000 tokens, we decided against this high level of standardization, as it would greatly inflate, and thus possibly misrepresent, the results of a sentiment analysis that has been performed on texts with relatively few tokens. In brief, a standardization of up to 1,000,000 tokens might make the facts quite irreal.[11] Thus, we decided on a normalization scale of 50,000 tokens, because this level of standardization was hypothesized to still resemble the original numeric values and their distribution. The result of these standardized sentiment scores is given in columns (5) and (6) of Table 6. We see that the normalization scale was quite sound, as the resulting figures have not changed greatly after normalization of the individual texts. The plots of Figures 4a and 4b visualize the information coded in the relevant columns and show the distribution and the counts of the sentiments as a line graph. In other

[11] See Perkuhn et al (2012) pp. 78-80.

words, the normalization of the tokens up to an equal level of 50,000 tokens still represents the character of the original distribution of the sentiments that were found in the original raw data, even though now in a descriptively correct and statistically viable way for a possible in-between comparison.

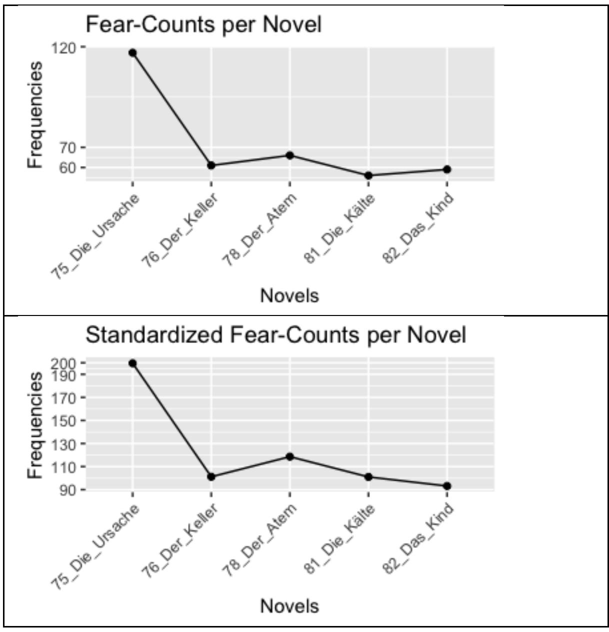

Figure 4a

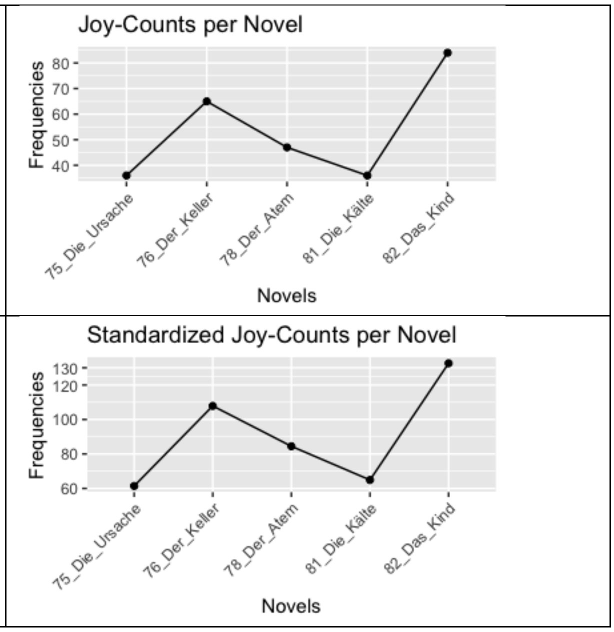

Figure 4b

Summarizing the result of these calculations, we conclude that that the proportion between "fear" and "joy" shows the biggest difference in "Ein Kind" ("A Child"), as the frequency of the "'fear' sentiment" is more than three times higher than the "'joy' sentiment", which has a calculated proportion of 3.25:1, which indicates a difference of 69.23%. The difference between the sentiments of novels 3 and 4 is 28.79% and 35.74%, respectively. In contrast, the sentiment "joy" wins out over "fear" in novel 2 (06.23%) and in novel 5 (29.77%). This means that the sentiment "fear" is without a doubt more frequently expressed in the autobiographies that we have analyzed, when we base our conclusions on both the raw data and the normalized data. The general fear/joy proportion of the normalized data thus amounts to approximately 1:1.35, or a difference in percentages of roughly 25%, in favor of the sentiment "fear". It is an empirical question that cannot be answered within the boundaries of this paper whether unbiased readers will also end up with the impression that the novels analyzed here with very simple computational methods have a greater proportion of "fear" than of "joy", although we are quite confident that they will do so.

The next question that seems of interest in this context is whether the frequencies of the sentiment groups "fear" and "joy" can still be regarded as a homogenous group of counts. In order to find a substantiated answer to this question on the basis of a statistical test, we applied a chi-squared test for "goodness of fit" and obtained the following highly significant results: ("fear": x-squared=63.045, df=4, p-value=6.643-13, expected:122,69; "joy": x-squared=40.042, df=4, p-value=4.243e-08, expected:90.25). These results indicate unambiguously that the frequency differences between the novels are statistically real and that the frequencies of sentiments do not occur in alike numerical values, and that they differ between the novels. But when we contrast the sentiments "joy" vs "fear" in our corpus of novels "across the board" with the non-parametric "Wilcoxon Rank Sum Exact Test", we detect that the test does not output a statistically significant result: (w=7, p-value=0.3095) for the "normalized data", as well as the "raw data" (w=7, p-value=0.2948). In addition, a calculation of the effect strength on the same data with Cliff's delta supports this statistic, as it produces only a "medium result" for the differences between the sentiments in the novels (delta estimate: 0.44 (medium); 95

percent confidence interval). This result thus indicates that we must distinguish between a general impression of the sentiments as a corpus and a particular impression of the sentiments of concern when contrasting them between the novels.

While the foregoing calculations of some of the statistics that can be applied to the results of the computationally performed sentiment analysis allowed for generalizations from an empirical viewpoint, we also want to find out and present some specifics of the actual lexis that has been detected by our dictionary of sentiments. Such an illustration is provided in Table 6 and 7, where we list a top-to-bottom hierarchy of the 20 topmost tokens from the sentiment classes "joy" and "fear". Table 6 is an example of the lexical expressions of "joy". It shows that the lexeme "gut" ("good") was picked up 44 times, followed by "glücklich" ("happy") which was detected 31 times in the corpus. As for the "relevance" of the set of lexemes of Table 6 with respect to the emotion "joy", we do note that there is indeed a clear semantic affinity to sentiments that can intuitively be regarded as belonging to their center, or to near the center. Therefore, we can confidently assume that the dictionary as such, and the linguistically oriented theoretical foundations that it is based on, are worthy of extension in follow-up textual analyses.

Tokens/frequencies of the 20 topmost "Joy-Tokens" in the corpus
gut(44)<glücklich(31)<gefallen(29)<glück(27)<freude(15)<lust(14)<guter(10) <positiv(10)<schönheit(9)<schöne(9)<vergnügen(9)<gute(9)<schönsten(7) <gutes(7)<schön(7)<glückliche(7)<glücklichen(6)<gefiel(6)<froh(5)<guten(5)

Table 6

Directing our attention now to the textual expression of the sentiment "fear" in Table 7, we see that the items of the dictionary are relevant to the semantic dimension of "fear" when the lexemes are seen as members of a prototypical chain of differing distances from the center to the periphery of the prototype postulated. The top hit "Angst" ("fear") occurs 79 times in the corpus of the five autobiographies. As we also see, some morphological variation of the stem "Furcht" ("fear") is variously used throughout the sample of the hierarchy of the 20 topmost tokens, and thus much more often than "Freude" ("joy"), which occurs just 15 times. Thus, while the analysis may not be a perfect indicator of the sentiments of analysis at all times, this list does give us a basic idea of the concrete lexical results of sentiment analysis in general and the tone of the five autobiographic novels, which was judged to be one of moderate "fear". While we do not get an impression of the environment of Thomas Bernhard's upbringing by his kinsfolk, his experiences in school and work, his health problems and the societal pressures of the time on the basis of the lexemes used in the form of a detailed text, we do nevertheless obtain a general indication of the "tone" of the novels as a mixture of "joyful" and "fearful" descriptions of his childhood life with an overweight of "fear" in comparison to "joy", even if the contrast is not statistically significant when the novels are examined one by one.

Tokens/Frequencies of the 20 topmost "Fear-Tokens in the corpus
angst(79)<fürchterlichkeit(21)<fürchterliche(17)<ahnung(16)<fürchtete(12) <fürchterlichsten(11)<entsetzen(9)<fürchterlich(9)<fürchterlichen(8) <gefürchtet(7)<grauen(7)<entsetzt(6)<ahnte(6)<fürchterlichste(5) <grausamkeit(5)<ahnungslosen(5)<fürchterlicher(4)<schreckliche(4) <vermutete(4)<angstzustand(4)

Table 7

At the end of this pilot study, we want to point out that there is still room for improvement in future when it comes to sentiment analysis in general and the analysis of literature in particular. This includes, for instance, the lexicalized surface expression of "sentiments" in texts. We relied on the explicit coding of sentiments with lexical means and thus disregarded the possible indirect transmission of emotions, as well as the influence of the context on its detection and interpretation. Such a strategy might only be partially responsible for the impact and the overall interpretation of sentiments of the texts, however. We also performed a general sentiment analysis of the novels ignoring the possible (un)importance of the characterization of the individual protagonists with respect to the two sentiments expressed. In addition, we did not include the notion of modality in the context of sentiments, and so might have missed an aspect of textual expression that could have influenced the result, as well. As we relied on the literal meaning of expressions, we could not ascertain the influence of non-literal meaning, such as metaphorically used lexis. Finally, we are aware of the fact that we did not investigate the occurrence of negative elements (e.g., "kein" ("no"), "nicht" ("not"), "ohne" ("without", etc.). They could indeed be responsible for sentiment frequencies that are different, possibly even lower, from the ones we have obtained.[12] In addition, as we relied on a "yes/no" approach, adverbs of degree were also of no concern to this sentiment analysis.

To conclude, this paper provided a pilot study of the sentiments "fear" and "joy" of the autobiographical novels of the Austrian playwright Thomas Bernhard. We presented first-hand accounts of the data including descriptive and analytical statistics of several aspects of the novels for possible, and necessary comparison to other works in similar domain and methodological background. We exemplified the computational process involved in the analysis and concluded – not surprisingly – that Thomas Bernhard's early life as presented in the lexical corpus of texts was rather one of "fear" than of "joy". In the end, we concluded this computational analysis of pilot character with an enumeration of textual aspects of the novels that call for inclusion and detailed exploration within the bounds of a computational sentiment analysis. Finally, we have thus to admit that this quantitatively oriented work might at present only supplement qualitatively based research pursued under the heading of traditional literary by asking new questions whose answers may be obtained by computational means.

Bibliography

Barnes, Jeremy, Roman Klinger & Sabine Schulte im Walde. 2017. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Copenhagen, Denmark, pp. 2-12.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. *Quanteda: An R package for the quantitative analysis of textual data. R package version: 2.12.*

Desagulier, Guillaume. 2017. *Corpus linguistics and statististics with R: Introduction to Quantitative Methods in Linguistics*. Berlin: Sprinter International.

Duden. 2004. *Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter (Bd. 8)* Bibliografisches Institut: Mannheim.

Ekman, Paul. 1999. Basic emotions. In T. Dalgleish & M. Power (eds) *Handbook of Cognition and Emotion*. Sussex, U. K.: John Wiley & Sons. Pp. 45-60.

Gries, Stefan. Th. 2017^2. *Quantitative corpus linguistics with R: A practical introduction*. New York: Routledge.

Hatzinger, Reinhold, Kurt Hornik, Herbert Nagel & Marco J. Maier. 2014^2. R: Einführung durch angewandte Statistik. Halbergmoos: Pearson

Huber, Martin & Manfred Mittermayer (Hg.) 2018. *Bernhard-Handbuch: Leben-Werk-Wirkung. Section F: Autobiographie*. Metzer: Stuttgart. Pp. 166-192.

Jacobs, Arthur M. 2019. *Sentiment analysis for words and fiction characters from the perspective of computational (neuro)poetics*. https://www.frontiersin.org/articles/10.3389/frobt.2019.00053/full.

Kim, Evgeny & Roman, Klinger. 2019. A survey on sentiment and emotion analysis for computational literary studies. *Zeitschrift für Digitale Geisteswissenschaften*, 4.

[12] A cursory look at the occurrence revealed that the dictionary entry "ahn" which matched the noun "Ahnung" ("hunch", "suspicion", "have an idea of sth.") did occur in combination with a negation element several times. This means that that the data are still in need of supplementary analysis to obtain a final word on Thomas Bernhard's interpretation of his childhood as either "joyful" or "fearful".

Kleiber, Georges. 1993. *Prototypensemantik: Eine Einführung*. Gunter Narr Verlag: Tübingen.
Klinger, Roman, Surayya Samat Suliya & Nils Reiter. 2016. Automatic Emotion Detection for Quantitative Literary Studies: A case study on Franz Kafka's "Das Schloss" and "America". *Digital Humanities 2016: Conference Abstracts*, Kraków, Poland.
Lehmann, Jörg, Moritz Mittelbach & Sven Schmeier. 2017. Quantifizierung von Emotionswörtern in Texten. *DARIAH-DE working papers*. 24. http://resolver.sub.uni-goettingen.de/purl/?dariah-2017-4.
Liu, B & Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In Charu C. Aggarwal & ChengXiang Zhai (eds.) *Mining text data*. New York: Springer, pp. 415-463.
McLintock, David. 2003. *Thomas Bernhard. Gathering Evidence: A Memoir*. Vintage: London. Translations of "Die Ursache. Eine Andeutung" (An indication of the cause), "Der Keller. Eine Entziehung" (The cellar: An escape), "Der Atem. Eine Entscheidung" (Breath: A Decision), "Die Kälte: Eine Isolation" (In the cold), and "Ein Kind" (A child).
Meindl, Claudia. 2011. Methodik für Linguisten: Eine Einführung in Statistik und Versuchsplanung. Tübingen: Narr Francke Attempto Verlag.
Mittermayer, Manfred. 2015. Thomas Bernhard: Eine Biografie. Wien-Salzburg: Residenz Verlag.
Mohammad, Saif M. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural H, Social Science, and Humanities.* Portland, OR. Pp 105-114.
Mohammad, Saif M. 2015. Sentiment analysis: Detecting valence, Emotions, and other affectual states from text. In H. Meiselman (ed.) *Emotion measurement*. Amsterdam, The Netherlands: Elsevier.
Mohammad, Saif M. 2017. Challenges in sentiment analysis. In E. Cambria et.al (eds) 2017. *A practical guide to sentiment analysis*. Springer International Publ. pp. 61-84.
Mohammad, Saif M. 2020. *Sentiment Analysis: Detecting valence, emotions, and other affectual states from text*. (https://arxiv.org/abs/2005.1182v1).
Nalisnick, Eric T. & Henry S. Baird. 2013. Character-to-character sentiment analysis in Shakespeare's plays. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, pp. 479-483.
Ortony, Andrew. 1990. What's basic about basic emotions. *Psychological Review* 97:3, 315-331.
Pang, Bo, Lillian Lee & Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, pp.
Pang, Bo & Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1-2), 1-135.
Perkuhn, Rainer, Holger Keibel & Marc Kupietz. 2012. *Korpuslinguistik*. Paderborn: Wilhelm Fink.
Piórkowska, Magda & Monika Wróbel. Basic emotions. Virgil Zeigler-Hill & Todd K. Shakelford (eds.) *Encyclopedia of Personality and Individual Differences*. New York: Springer International Publishing AG.
Pushmann, Cornelius and Mario Haim. 2020. *Automated content analysis with R: Sentiment analysis* (https://content-analysis-with-r.com/3-sentiment.html)
Pushmann, Cornelius. 2019. *Automatisierte Inhaltsanalyse mit R*. (https://inhaltsanalyse-mit-r.de/)
R Core Team. 2020. R: *A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. (https://www.R-project.org)
Rauh, Christian. 2018. Validating a sentiment dictionary for German political language – a workbench note. *Journal of Information Technology & Politics*. 15:4:319-343.
Remus, Robert, Uwe Quasthoff & Gerhard Heyer. 2010. SentiWS – A publicly available German-language resource for sentiment analysis. In *Proceedings of the 7th International Language Resources and Evaluation (LREC10)*, pp. 1168-1171.
Rudkowsky, Elena, Martin Haselmayer, Matthias Wastian, Marcello Jenny, Štefan Emrich & Michael Sedlmair. 2018. More than bags of words: Sentiment analysis with word embeddings. Conference: *2018 IEEE 7th International Conference on Adaptive Science &Technology (ICAST)*
Sidarenka, Uladzimir & Manfred Stede. 2016. Generating sentiment lexicons for German Twitter. *Proceedings of the Workshop on Computational Modeling of Peopl's Opinions, Personality, and Emotions in Social Media*. Osaka, Japan, pp.80-90.
Schmidt, Thomas & Manuel Burghardt. 2018a. Towards a tool for sentiment analysis for German historic plays. In M. Piotrowski, (ed.), *COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018* Lausanne, Switzerland: Laboratoire laussannois d'informatique et statistique textuelle, pp. 46-48.
Schmidt, Thomas & Manuel Burghardt. 2018b. An evaluation of lexicon-based sentiment analysis techniques for the plays of Gotthold Ephraim Lessing. *Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Santa Fe, New Mexico, pp.139-149.

Schmidt, Thomas, Manuel Burghardt & Katrin Dennerlein. 2018. Kann man denn auch nicht lachend sehr ernsthaft sein? – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In *Book of Abstracts, DHd 2018*. Pp.244-249.

Schmidt, Thomas, Manuel Burghardt & Christian Wolff. 2018. Herausforderungen für Sentiment Analysis bei literarischen Texten. In Manuel Burghardt & C. Müller-Birn (Hrsg.) *INF-DH 2018-Workshopband* (https://doi.org/10.18420/infdh2018-16)

Schwarz, Monika & Jeanette Chur. 2007[5]. *Semantik*. Tübingen: Gunter Narr Verlag.

Werlen, Egon, Christof Imhof, Fernando Benites & Per Bergamin. 2019. The reader's feeling and text-based emotions: The relationship between subjective self-reports, lexical ratings, and sentiment analysis. In *SwissText 2019: 4th Swiss Text Analytics Conference*

Winko, Simone. 2003. Über Regeln emotionaler Bedeutung in und von literarischen Texten. In Fotis Jannidis, Gerhard Lauer, Matias Martinez & Simone Winko (Hrsg.) *Regeln der Bedeutung*. Berlin: de Gruyter, S. 329-348.

Wolfgruber, Marlene. 2015. *Sentiment Analyse mit lokalen Grammatiken: Wissensbasierter Ansatz zur Extraktion von Sentiments in Hotelbewertungen*. München: Dissertationen der LMU, vol. 3.