Sentiment Analysis of German Texts in Finance: Improving and Testing the BPW Dictionary

Matthias Pöferlein[1]
University of Bayreuth, Germany
matthias.poeferlein@uni-bayreuth.de
https://orcid.org/0000-0002-8967-6087

Received: 4 November 2021 / Revised: 13 December 2021 / Accepted: 15 December 2021 / Published online: 20 December 2021

ABSTRACT

Using the dictionary-based approach to measure the sentiment of finance-related texts is primarily focused on English-speaking content. This is due to the need for domain-specific dictionaries and the primary availability of those in English. Through the contribution of Bannier et al. (2019b), the first finance-related dictionary is available for the German language. Because of the novelty of this dictionary, this paper proposes several reforms and extensions of the original word lists. Additionally, I tested multiple measurements of sentiment. I show that using the edited and extended dictionary to calculate a relative measurement of sentiment, central assumptions regarding textual analysis can be fulfilled and more significant relations between the sentiment of a speech by a CEO at the Annual General Meeting and subsequent abnormal stock returns can be calculated.

JEL Classification: G12; G14

Keywords: textual analysis, textual sentiment, sentiment analysis, content analysis, annual general meeting, CEO speeches.

1. INTRODUCTION

In recent years, textual analysis has become an important part of accounting and finance research. This is due to the fact that the availability and quantity of digitally available texts are constantly increasing. Additionally, the information encoded in those texts in the form of sentiment can be obtained in an easier and more targeted way through recent developments in the field of textual analysis (Bannier et al., 2019b, pp. 82f.; Gentzkow et al., 2019, p. 535; Loughran & McDonald, 2015, p. 1).

Algaba et al. (2020, p. 2) define sentiment "[…] as the disposition of an entity toward an entity, expressed via a certain medium. […] This disposition can be conveyed numerically but is primarily expressed qualitatively through text, audio, and visual media." The two most common methods for transforming qualitative sentiment data into quantitative sentiment variables are the dictionary-based approach (also referred to as bag-of-words) and machine learning (Kearney

[1] Matthias Pöferlein – University of Bayreuth, Germany, matthias.poeferlein@uni-bayreuth.de, University of Bayreuth, Chair of Finance and Banking, Universitätsstr. 30, 95447 Bayreuth, Germany, Tel.: +49 (0)921/55-6258, Fax: +49 (0)921/55-6272.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

& Liu, 2014, pp. 174f.). The dictionary-based approach is a rule-based approach that uses an algorithm to classify a text's words or phrases into different categories based on predefined rules or categories like dictionaries[2] (Li, 2010, p. 146). More specifically, the dictionary assigns words into different categories like positive or negative. Using the total count of positive, negative, and all words, several measurements of sentiment can be calculated (Loughran & McDonald, 2015, p. 1). The machine learning or statistical approach relies on statistical techniques to classify the content of documents (Kearney & Liu, 2014, p. 175; Li, 2010, p. 146).

When using the dictionary-based approach, the chosen dictionary has a specific importance (Bannier et al., 2019b, p. 80; Loughran & McDonald, 2015, p. 1). As described in the following section, the newly developed word list provided by Bannier et al. (2019b) (BPW Dictionary) gives researchers the possibility to analyze German-speaking texts in finance in a more targeted way.

Due to the novelty of this BPW Dictionary, I propose several reforms and extensions with the objective of improving its performance. Therefore, the main hypothesis of this paper is that the edited version of the BPW (BPW_N) can improve results compared to its original (BPW_O). So far, the BPW Dictionary has been used primarily to analyze the market reaction to the sentiment of CEO speeches held at the Annual General Meeting (AGM) of German stock companies (Bannier et al., 2017, 2019a). Therefore, this paper also uses comparable speeches for testing the possible improvements.

As stated in the following course of this paper, there are several different possibilities to measure the sentiment of textual documents in a dictionary-based approach. Given the fact that this is the first German domain-specific dictionary for the field of finance, the additional research question is which sentiment measure is the most appropriate for measuring the tone of textual documents in the field of finance using a German domain-specific dictionary. This topic is especially relevant, given the previous use of exclusively four different measurements of sentiment using the BPW Dictionary (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

The contribution of this paper to the literature on textual analysis of German texts is the extension and reform of the only existing German finance-related dictionary and testing the performance of the original against the new dictionary. Additionally, the suitability of the primarily used measures of sentiment in a business context is analyzed. This should make it possible for researchers to measure the sentiment of German texts in finance more accurately and more thoroughly.

The paper proceeds as follows. In the following section, I will give a short review of the relevant literature regarding textual analysis with a particular focus on analyzing financial texts. The data and the parsing procedure applied to it, as well as the used dictionaries form the third section. The used measurements of sentiment and the empirical approach to obtain the results given in section five are presented in the fourth section. Section six concludes.

2. LITERATURE REVIEW

The extensive field of textual analysis in finance is ideally pictured in the surveys of Kearney and Liu (2014) and the online appendix of Bannier et al. (2019b). Other important surveys giving additional information and areas of caution regarding textual analysis in finance are Algaba et al. (2020) and Loughran and McDonald (2016).

One of the first steps in measuring the tone of a text is selecting a dictionary or word list (Loughran & McDonald, 2015, p. 1). According to Loughran and McDonald (2016, p. 1200), four different word lists have been primarily used by researchers classifying English finance-related

[2] As stated in Loughran and McDonald (2015, p. 10), the terms dictionary and word list are used interchangeably.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

texts. These are the two general dictionaries – General Inquirer (Stone et al., 1966) and DICTION (Hart, 2000) – and the two word lists generated for finance-related texts: Henry (Henry, 2006, 2008) and Loughran and McDonald (Loughran & McDonald, 2011).

In the contributions of Henry (2006, 2008) and Loughran and McDonald (2011), the usage of general word lists for different forms of textual content like news, earnings press releases or annual reports was widely criticized in favor of domain-specific word lists, because of the high possibility of misclassification (Algaba et al., 2020, pp. 13–15; Lewis & Young, 2019, pp. 598f.; Mengelkamp et al., 2016, p. 7; Price et al., 2012, p. 1006). Loughran and McDonald (2011, p. 49) analyzed that 73.8% of negative words in the general dictionary General Inquirer do not have a negative meaning in a business context.

Despite the fact that the Henry word lists have been used for different purposes like conference calls (Davis et al., 2015, pp. 641, 647; Price et al., 2012, pp. 996f.) or news (Jandl et al., 2014, pp. 4, 7), the lists provided by Loughran and McDonald have become predominant (Kearney & Liu, 2014, p. 175) in the field of finance. They have been used in the classification of many different kinds of written financial content like news (Garcia, 2013, pp. 1272, 1274; Gurun & Butler, 2012, pp. 562, 566), conference calls (Mayew & Venkatachalam, 2012, pp. 2, 20) and annual reports (Ahmed & Elshandidy, 2016, p. 179; Jegadeesh & Wu, 2013, pp. 713, 715).

Due to the absence of a German domain-specific dictionary for the field of finance, research was limited to different versions of general dictionaries like LIWC (Meier et al., 2018; Wolf et al., 2008) or SentiWS (Remus et al., 2010), resulting in little research (Ammann & Schaub, 2016; Dorfleitner et al., 2016; Fritz & Tows, 2018). The first public available business-related dictionary for the German language was introduced by Bannier et al. (2019b). The introduced word lists are based on the predominant lists by Loughran and McDonald (Bannier et al., 2019b, p. 79) and have already been successfully used (Bannier et al., 2017, 2019a; Röder & Walter, 2019; Tillmann & Walter, 2018, 2019).

As stated in Bannier et al. (2019a, p. 2), the contributions of Bannier et al. (2017, 2019a) are the primary studies analyzing the information content of CEO speeches delivered at the Annual General Meeting. Thus, this paper is also an essential complementary contribution to the information content of CEO speeches.

3. DATA

3.1. Data Source

I collected the transcripts of the CEO speeches from the companies' homepages, since there is no database for German CEO speeches delivered at the AGM. I screened the web pages of all companies listed in the DAX, MDAX, SDAX or TECDAX between 2008 and 2019 for transcripts of CEO speeches delivered at the AGM. Since not all companies publish transcripts on their homepage, I could find 976 speeches of 139 companies for the initial sample. I had to remove 53 speeches that were not delivered by the CEO. All available additional information, such as annotations, audio and video material provided by the company or other providers, was evaluated to confirm that the speeches were initially delivered in German. Therefore I had to exclude another 50 speeches. Additionally, 49 transcripts contained speeches of several speakers and required filtering of the relevant parts. Due to a delisting, I had to delete one additional speech. The final sample consists of 872 speeches from 125 companies. Comparing the contributions of Bannier et al. (2017, p. 10) (338 speeches) and Bannier et al. (2019a, p. 7) (457 speeches), this is the most comprehensive collection of German CEO speeches so far. An overview of the sample creation is given in Table 1. I obtained all other variables from Thomson Reuters Datastream.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 1
Sample creation

Source/Filter	Sample Size	Removed Observations
CEO speeches found on the companies' homepages	976	
Speeches not held by the CEO	923	53
Speeches held initially in English	873	50
Speeches where no CAR or CAV could be calculated	872	1
Final Sample	872	

Source: Author's calculation.

3.2. Used Dictionaries

The mutated vowels "ä", "ö" and "ü" in the German language can alternatively be written as "ae", "oe" and "ue". To get the updated form of the BPW_O (BPW_N), the first step is to add the alternative spelling of words with mutated vowels because the BPW_O does not include those. As a part of the parsing procedure, I deleted hyphens. Therefore, stop words written with hyphens had to be included without hyphens. Overall, I deleted 21 words that also appear on the positive and negative list of the BPW_O from the stop word list. In total, 144 stop words occurred twice and had to be deleted, because 110 surnames match company or given names (e.g. "kummer"). After extending for mutated vowels and hyphens, another 34 words occurred twice. Finally, I added 244 additional stop words through a translation of the generic list provided by Loughran and McDonald (2020) (LMD stop words). A summary of the conducted steps and the resulting alteration of the number of words on the different lists is given in Table 2.

Table 2
Updating of the BPW

	Positive	Negative	Stop words
BPW_O total words	2,223	10,147	3,682
Adding mutated vowels	+ 626	+ 2,514	+ 218
Including words without hyphens			+ 153
Delete doubles (positive/negative)			- 21
Delete doubles			- 144
Adding additional LMD stop words			+ 244
BPW_N total words	2,849	12,661	4,132

Source: Author's calculation.

Due to the update of the BPW_O, this paper examines the suitability of two different dictionaries.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

3.3. Parsing

Given expressed criticism regarding unspecified parsing rules and the related difficulty to replicate existing studies (Loughran & McDonald, 2015, p. 2), I give a detailed overview of performed text manipulation.

In the first step, the collected PDF files were transferred into TXT files using UTF-8 encoding (Bannier et al., 2017, p. 10, 2019a, p. 9; Meier et al., 2018, p. 29). In order to automatically process the speeches, they need to be parsed. Due to the unique and unsystematic character of the collected texts, manual corrections need to be conducted before using an automated parser. Those include the removal of headlines, disclaimers, legal notices, and additional information (e.g. the positioning of slides).

The subsequent automated parser was programmed using python. First of all, I replaced typographic ligatures (Bannier et al., 2017, p. 10, 2019a, p. 9) and hyphens (Loughran & McDonald, 2011, internet appendix) and converted all words to lowercase (Fritz & Tows, 2018, p. 61; Picault & Renault, 2017, p. 139). Additionally, I removed special characters (Allee & Deangelis, 2015, p. 247; Mengelkamp et al., 2016, p. 4), numbers (Boudt & Thewissen, 2019, p. 84; Schmeling & Wagner, 2016, p. 8), punctuation (Gentzkow et al., 2019, p. 538; Loughran et al., 2009, p. 41), and multiple whitespaces (González et al., 2019, p. 7; Schmeling & Wagner, 2016, p. 8). Finally, I removed words with fewer than three characters (Bannier et al., 2017, p. 10, 2019a, pp. 9f.; Loughran et al., 2009, p. 42). Depending on the used dictionary (BPW_O or BPW_N), I deleted the predefined individual stop words. Stop words are very common words but have relatively little meaning or rarely contribute information on their own, despite being essential to the grammatical structure of a sentence (Bannier et al., 2017, p. 10; Gentzkow et al., 2019, p. 538).

Furthermore, I included an important automated alteration[3] of the words "betrug" and "sorgen" prior to the automated parser. When written in lowercase, the words were changed to "betrugnoneg" and "sorgennoneg." This is because of the very frequent occurrence of those words in the analyzed texts (betrug: 812, sorgen: 344) and the characteristics of the German language. When written with a first capital letter, both words are nouns, where the word "Betrug" means "fraud" and the word "Sorgen" means "sorrow," which are both negative words in a business context and due to that are justifiably on the list of negative words. But when written entirely in lowercase, both words are verbs. In this case, the word "betrug" means "amounted" and "sorgen" means "care," which does not have a negative connotation. Without this automated alteration, the exclusive use of lowercase words would lead to a wrong and exaggerated number of negative words.

4. METHDOLOGY

4.1. Measurement of Sentiment

Using python, I counted the occurrence of positive (p) and negative (n) words from each of the two dictionaries as well as the total number of words (w) for each document. By using those three numbers, a variety of measurements of sentiment can be calculated. Even though the notations differ in several contributions, this paper focuses on the most widely used measurements to evaluate which sentiment measure is the most appropriate for the tone of textual documents in finance.

[3] Note that this automated alteration was only implemented when using the updated form of the dictionary provided by Bannier et al. (2019b) (BPW_N).

DOI: 10.7172/2353-6845.jbfe.2021.2.1

First of all, I calculated a simple share of negative and positive words as in Loughran and McDonald (2011, p. 46), Ferguson et al. (2015, p. 7) and Ammann and Schaub (2016, p. 2):

$$N = \frac{n}{w} \tag{1}$$

$$P = \frac{p}{w} \tag{2}$$

Other studies, as stated below, use the relation of positive and negative words rather than their individual fractions. However, there are different approaches to measure this relation. In this paper, I used the three most prominent relative measurements of sentiment.

Following the approach of Davis et al. (2015, p. 646), Loughran and McDonald (2015, p. 4), and Picault and Renault (2017, p. 141), I measured the sentiment of a text as the number of positive words minus the number of negative words divided by the total number of words:

$$Tone = \frac{p - n}{w} \tag{3}$$

Other contributions switch the numerator while retaining the notation "*Tone*" (Franke, 2018, p. 9; Kim & Meschke, 2014, p. 33). To prevent misinterpretations, this paper uses the term *ITone* for inverted tone.

$$ITone = \frac{n - p}{w} \tag{4}$$

In contrast to *Tone* and *ITone*, the variable *NTone* used by Henry (2008, p. 386), Price et al. (2012, p. 998), and Henry and Leone (2016, p. 159) only focuses on the number of positive and negative words and is not altered by the length of the analyzed text. It therefore gives the NetTone:

$$NTone = \frac{p - n}{p + n} \tag{5}$$

Also, a fourth relative variable *NToneSQ* as in Henry (2008, p. 393) is estimated, by squaring the variable *NTone*.

Given this variety of six different measurements of sentiment, this paper adds the two measurements *InvTone* and *NToneSQ* to the four already tested calculations, when using the BPW_O (Bannier et al., 2017, p. 11, 2019a, p. 10; Röder & Walter, 2019, p. 396; Tillmann & Walter, 2018, pp. 9, 21, 2019, pp. 69f.).

In this paper, following Apel and Blix Grimaldi (2012, p. 9), Davis et al. (2015, p. 653), and Bannier et al. (2017, p. 15), all words found are weighted equally. This approach makes it possible for other researchers to replicate and further develop the results of this contribution, due to the independence of the weighting scheme from the used texts. This approach and the superiority of equal weighting is also supported by Henry and Leone (2016, p. 166).

4.2. Empirical Approach

By using linear regressions, I conduct one of the most common approaches for analyzing the impact of sentiment on stock prices (Kearney & Liu, 2014, p. 177). Therefore, I performed several linear regressions for ten different dependent variables in the following form:

$$Dep_j = \alpha_0 + \alpha_1 Sentiment_j + \sum_{k=1}^{K} \propto_k Control_{kj} + \varepsilon_j \tag{6}$$

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Dep represents two different forms of variables to measure the effect of speech sentiment on stock prices and trading.

To obtain the effect on stock prices, I calculated cumulative abnormal returns (*CAR*). The abnormal returns are calculated by the market adjusted model using the value weighted market index CDAX. Following Henry (2006, p. 5, 2008, p. 385), Loughran and McDonald (2011, p. 41), Henry and Leone (2016, p. 159), and Bannier et al. (2017, p. 12, 2019a, p. 8), the CARs are calculated through cumulating the abnormal returns (*AR*) over a predefined event period (event window) with length *T*. I obtained the individual *AR*s by subtracting the returns (*R*) of the analyzed stock (*j*) from the return of the CDAX for a given day (*t*):

$$AR_{j,t} = R_{j,t} - R_{CDAX.t} \tag{7}$$

$$CAR_{j,T} = \sum_{t=0}^{T} AR_{j,t} \tag{8}$$

Based on Loughran and McDonald (2011, p. 41), Boudt and Thewissen (2019, p. 95) and Bannier et al. (2019a, p. 9), this paper solely uses event windows beginning on the day of the AGM ($t = 0$), to only measure the effect of the CEO speeches. Therefore, the five different trading day event windows [0,1], [0,3], [0,5], [0,15], and [0,30] were used following contributions examining similar texts like CEO letters or CEO conference calls (Bannier et al., 2019a, p. 9; Boudt & Thewissen, 2019, p. 95; Doran et al., 2012, p. 412; Loughran & McDonald, 2011, p. 41; Mayew & Venkatachalam, 2012, p. 20).

Additionally, I performed all regressions with cumulative abnormal trading volumes (*CAV*) for the five different event windows. I calculated the different CAVs according to Bannier et al. (2017, p. 47, 2019a, p. 38) and Price et al. (2012, p. 1000) as:

$$AV_{j,t} = \frac{VOL_{j,t}}{\overline{VOL_{j,t}}} - 1 \tag{9}$$

$$CAV_{j,T} = \sum_{t=0}^{T} AV_{j,t} \tag{10}$$

Here $VOL_{j,t}$ is the trading volume for firm *j* at day *t*, and $\overline{VOL_{j,t}}$ is the mean volume for firm *j* from trading day $t = -252$ to $t = -2$. Due to different estimation windows in the primary studies of Bannier et al. (2017, p. 47, 2019a, p. 38), I selected a combined period of time in accordance with Price et al. (2012, p. 1000).

I used the six above mentioned measurements of sentiment separately for each of the ten different dependent variables *Dep*.

The comprehensive set of control variables *Control* consist of eleven different variables (*K*), which include the firm size (*SIZE*), the market to book value (*M2B*), leverage (*LEV*), volatility (*VOLA*), volume (*VOL*), number of words (*COUNT*), individual words (*IND*), return on assets (*ROA*), the earnings surprise (*EPS_SP*), and the dividend surprise (*DIV_SPP* and *DIV_SPN*) (Bannier et al., 2017, p. 47, 2019a, pp. 38f.; Doran et al., 2012, p. 426; Loughran & McDonald, 2011, p. 63). The calculation of the individual control variables can be found in the appendix.

I used the variables *SIZE*, *VOL*, and *COUNT* in a logarithmic form. When using *CAV*, the variable *VOL* is excluded from the regression. Additionally, I used year fixed effects.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

5. RESULTS

5.1. Summary Statistics

I report summary statistics for the analyzed sample of 872 CEO speeches in the following three tables.

Table 3 provides descriptive statistics for all calculated CARs and CAVs. While I could calculate CARs for all different event windows, the calculation of CAVs is only partially possible based on the availability of data. As stated in Bannier et al. (2017, p. 16), the means of all CARs are economically small, indicating no market reaction due to the AGM. In comparison, CAVs are in the mean higher than 1, indicating an abnormal trading volume caused by the AGM.

Table 3
Descriptive statistics for CARs and CAVs

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
CAR01	872	0.001	0.027	-0.184	0.104	-0.013	0.015
CAR03	872	-0.0002	0.031	-0.285	0.116	-0.017	0.018
CAR05	872	-0.002	0.037	-0.171	0.138	-0.021	0.018
CAR015	872	-0.004	0.059	-0.271	0.229	-0.035	0.033
CAR030	872	-0.005	0.087	-0.459	0.321	-0.057	0.046
CAV01	849	2.790	2.192	0.041	32.141	1.654	3.195
CAV03	841	4.825	3.076	0.054	37.987	3.130	5.645
CAV05	839	6.787	3.705	0.087	41.084	4.604	7.927
CAV015	827	16.498	7.859	0.595	82.829	12.060	19.007
CAV030	817	30.614	12.434	0.931	124.574	23.843	35.132

Source: Author’s calculation based on data from Thomson Reuters Datastream.

Because of the extension of the stop word list, the mean words counted are 22.7% lower for BPW_N, as given in Table 4. In addition to the change of sentiment measures, the reduction of words also improves calculation times of algorithms for measuring textual sentiment. The deletion of positive words from the stop words list leads to an increase in the number of positive words. In contrast, the mean number of negative words decreases due to the treatment of the words “betrug” and “sorgen.” The combination of those changes leads to an increase in all six sentiment measures on average. The mean number of positive and negative words combined with positive means for the measurements *Tone*, *NTone*, and *NToneSQ* show that the speeches delivered by the CEOs are on average positive. This positivity of speeches is slightly higher for the BPW_N dictionary. As stated in Doran et al. (2012, p. 414) for earnings conference calls using the Henry word list, it is not surprising that the general sentiment is positive, reflecting the effort of CEOs to present their information as positive as possible. This positive wording is also reflected in the characteristics of values of *NTone*, which by construction is bounded between -1 and 1. While the minimum value is -0.455 and thus relatively far from the highest possible minimum, the maximum value of 0.941 for BPW_O and 0.943 for BPW_N shows that in the most positive speeches hardly any negative words were used. This finding is additionally confirmed by the positivity of the 25% quartile and by the minimum number of one negative and eleven positive words.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 4
Descriptive statistics for sentiment variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
COUNT_BPW_O	872	2,411.709	834.021	759	5,625	1,817.5	2,909
IND_NUM_BPW_O	872	1,153.603	334.053	433	2,402	920.8	1,331.5
IND_BPW_O	872	0.490	0.046	0.368	0.642	0.457	0.519
P_NUM_BPW_O	872	90.142	32.124	11	206	65	112
N_NUM_BPW_O	872	38.556	25.082	1	152	21	49
N_BPW_O	872	0.015	0.007	0.001	0.046	0.010	0.019
P_BPW_O	872	0.038	0.009	0.010	0.068	0.032	0.044
Tone_BPW_O	872	0.023	0.013	-0.029	0.062	0.014	0.032
NTone_BPW_O	872	0.428	0.241	-0.455	0.941	0.283	0.606
ITone_BPW_O	872	-0.023	0.013	-0.062	0.029	-0.032	-0.014
NToneSQ_BPW_O	872	0.241	0.188	0.000	0.886	0.083	0.367
COUNT_BPW_N	872	1,864.443	646.324	589	4,431	1,405	2,247.2
IND_NUM_BPW_N	872	1,098.989	326.592	399	2,323	873	1,277
IND_BPW_N	872	0.602	0.052	0.456	0.777	0.566	0.634
P_NUM_BPW_N	872	92.905	32.992	11	212	68	116
N_NUM_BPW_N	872	37.361	24.830	1	149	20	48
N_BPW_N	872	0.019	0.010	0.001	0.062	0.012	0.024
P_BPW_N	872	0.051	0.011	0.015	0.095	0.043	0.058
Tone_BPW_N	872	0.031	0.017	-0.039	0.090	0.020	0.043
NTone_BPW_N	872	0.454	0.238	-0.455	0.943	0.304	0.630
ITone_BPW_N	872	-0.031	0.017	-0.090	0.039	-0.043	-0.020
NToneSQ_BPW_N	872	0.263	0.195	0.000	0.889	0.095	0.396

Source: Author's calculation.

I conducted a dependent-samples t-test to compare the alteration of positive and negative words found. There was a significant difference in the number of positive words found concerning the use of the BPW_O (M = 90.142, SD = 32.124) and BPW_N (M = 92.905, SD = 32.992), $t(871) = -22.939$, $p < .001$. This also applies to the number of negative words found when using the BPW_O (M = 38.556, SD = 25.082) and the BPW_N (M = 37.361, SD = 24.830), $t(871) = 18.471$, $p < .001$.

Table 5 gives the descriptive statistics for the additional control variables used in the regression. In accordance with Bannier et al. (2017, p. 17), the number of observations in which the dividend per share is unchanged compared to the previous year is 31.1%. In 51.4% the dividend per share increased, and in 17.5% decreased.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 5
Descriptive statistics for control variables

Statistic	N	Mean	St. Dev.	Min	Max	Pctl(25)	Pctl(75)
SIZE	870	9,883.827	16,996.830	30.200	104,226.900	845.245	10,287.470
M2B	869	2.208	2.267	-17.640	19.070	1.160	2.930
LEV	865	0.637	0.209	0.094	1.811	0.519	0.753
VOLA	872	0.020	0.010	0.002	0.130	0.014	0.024
VOL	852	2,108.435	4,949.786	0.100	47,270.600	67.925	1,518.850
ROA	865	0.037	0.065	-0.483	0.679	0.007	0.063
EPS_SP	848	1.685	16.275	-140.625	196.193	-1.607	2.625
DIV_SPP	872	0.514	0.500	0.000	1.000	0.000	1.000
DIV_SPN	872	0.175	0.381	0.000	1.000	0.000	0.000

Note: The definitions of all variables are given in the appendix.

Source: Author's calculation based on data from Thomson Reuters Datastream.

Overall, editing stop words leads to a word reduction of 22.7% (477,216 words), as stated in Table 6. Deleting the 21 words from the stop word list that are also on the positive and negative list leads to 3.1% (2,409) more positive words found, with only eight more individual words. Although there are three more individual negative words, the number of negative words found decreases by 3.1% (1,042). This is because of the correction for "betrug" and "sorgen" described in the parsing process.

Table 6
Total number of words

	BPW_O	BPW_N
	All words	
Number of words	2,103,010	1,625,794
Individual words	100,151	99,970
	Positive words	
Number of words	78,604	81,013
Individual words	1,123	1,131
	Negative words	
Number of words	33,621	32,579
Individual words	2,180	2,183

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 7 displays the number and cumulative fraction of the ten most frequent positive words in all speeches after correcting for stop words. The only difference is the deletion of the word "große" from the stop word list of the dictionary BPW_N.

Table 7
Ten most frequent positive words

BPW_O			BPW_N		
Word	**Number**	**cumulative %**	**Word**	**Number**	**cumulative %**
erfolgreich	2,143	2.73%	erfolgreich	2,143	2.65%
erfolg	2,015	5.29%	erfolg	2,015	5.13%
erreicht	1,624	7.36%	erreicht	1,624	7.14%
erreichen	1,566	9.35%	erreichen	1,566	9.07%
großen	1,546	11.31%	großen	1,546	10.98%
besser	1,515	13.24%	besser	1,515	12.85%
positiv	1,157	14.71%	große	1,209	14.34%
stärker	1,089	16.10%	positiv	1,157	15.77%
positive	1,040	17.42%	stärker	1,089	17.11%
stärken	1,035	18.74%	positive	1,040	18.40%

Source: Author's calculation.

As Table 8 illustrates, the adjustment in the parsing process for the words "betrug" and "sorgen" leads to an extensive decrease of those words, to the extent to which they do not appear in the ten most frequent negative words.

Table 8
Ten most frequent negative words

BPW_O			BPW_N		
Word	**Number**	**cumulative %**	**Word**	**Number**	**cumulative %**
herausforderungen	1,019	3.03%	herausforderungen	1,019	3.13%
betrug	876	5.64%	krise	845	5.72%
krise	845	8.15%	schwierigen	792	8.15%
schwierigen	792	10.51%	rückgang	728	10.39%
rückgang	728	12.67%	gegen	650	12.38%
gegen	650	14.60%	minus	483	13.86%
minus	483	16.04%	verfügung	476	15.33%
verfügung	476	17.46%	wider	415	16.60%
wider	415	18.69%	leider	356	17.69%
sorgen	398	19.87%	finanzkrise	330	18.71%

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

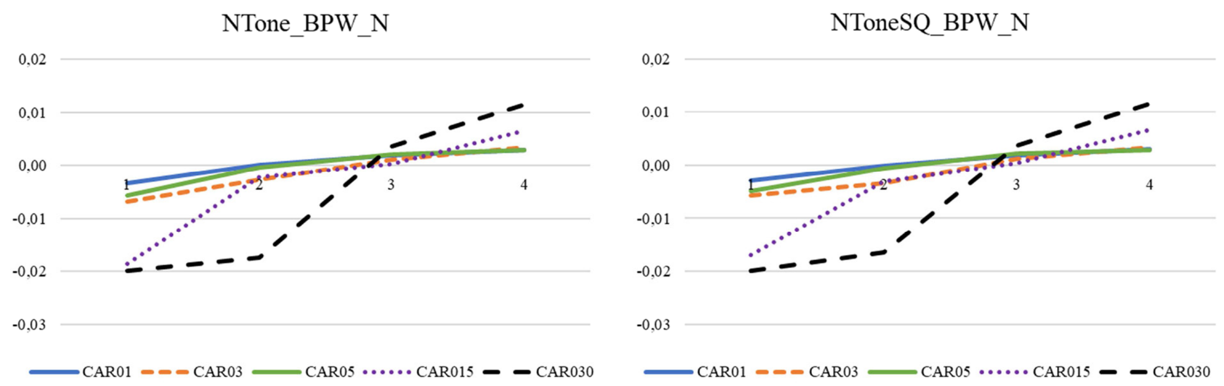
An English translation of all words listed in Table 7 and Table 8 is given in the appendix. Note that an important distinction of German words through small and capital letters is not possible due to the nature of the parsing procedure and structure of the dictionaries. Because of their impact, I only considered this distinction for the words "betrug" and "sorgen."

Of the 2,223 (BPW_N: 2,849) positive words available, I only found 1,123 (BPW_N: 1,131) words. A comparably small fraction of those words found is able to account for 18.74% (BPW_N: 18.40%). The same applies to the more extensive list of 10,147 (BPW_N: 12,661) negative words. Of this list, I only found 2,180 (BPW_N: 2,183) words in the speeches, with ten words accounting for 19.87% (BPW_N: 18.71%) of all negative words found. These results clearly indicate that the correct words are more important than the mere extent of the used list.

5.2. Sentiment Measurement

Following Loughran and McDonald (2011, pp. 50f.), the assumption that the sentiment of certain texts is relevant leads in the case of CEO speeches to the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns and higher abnormal trading volumes. By dividing all texts into quartiles based on the different sentiment measures[4] and analyzing the median CARs and CAVs, a visual examination can be conducted. Figure 1 gives the only two measurements that meet the stated assumptions. Using the sentiment measures *NTone* and *NToneSQ*, it is possible to have ascending quartile medians for all five event windows.

Figure 1
CARs by quartiles (sufficient)



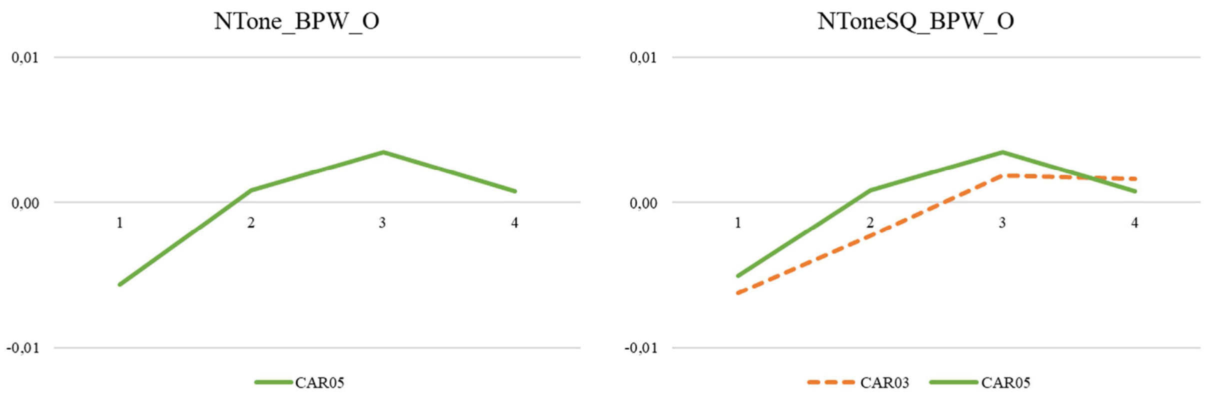
Source: Author's calculation.

The equivalent measures for the BPW_O cannot provide comparable sufficient results for all analyzed event windows. The affected windows and the not sufficient results for the associated quartiles are given in Figure 2. Here the window CAR [0,5] does not meet the assumptions for the sentiment measurement *NTone*. The same applies to the two windows CAR [0,3] and CAR [0,5] for *NToneSQ*. Other measurements of sentiment using the BPW_O or BPW_N do not meet this assumption either and therefore are not discussed further.

[4] Note that only the share of negative words (*N*) was sorted in the descending order. All other sentiment measures are sorted in the ascending order.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Figure 2
CARs by quartiles (not sufficient)

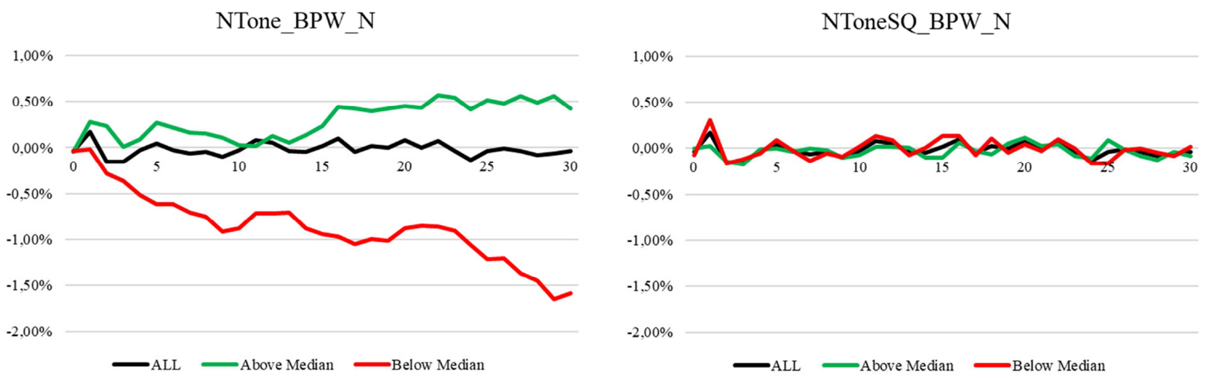


Source: Author's calculation.

With regard to the visual examination of the CAVs for different sentiment measures, no measure meets the above stated assumptions. Therefore, I excluded those figures.

Another essential assumption independent of certain event windows is the separation of above and below average abnormal returns through the use of sentiment measures as precisely as possible. Therefore, following Bannier et al. (2019a, pp. 17f., 37) and Price et al. (2012, pp. 1001f.), Figure 3 gives the average cumulative abnormal returns for up to 30 days following the AGM, divided by the above and below median sentiment measures *NTone* and *NToneSQ*. Additionally, the average CARs for all days are given[5].

Figure 3
CARs over time



Source: Author's calculation.

The accumulation of abnormal returns in Figure 3 for up to 30 days following the AGM shows that the average CARs are close to zero. By dividing the different observations into above and below median *NTone*, it is possible to separate positive and negative CARs. This is in accordance with the results of Bannier et al. (2019a, pp. 17f., 37). This separation can only be conducted using *NTone*. The same analysis using *NToneSQ* allows no distinction of positive and negative CARs using above and below median *NToneSQ*.

It therefore can be stated as an interim result that only the usage of the reformed and extended BPW_N dictionary with *NTone* as a sentiment measure is able to meet one of the central assumptions stated in the pioneer paper by Loughran and McDonald (2011, pp. 50f.) and the additional assumption of distinction.

[5] Due to the results stated in Figure 1 and Figure 2, only the results for *NTone* and *NToneSQ* calculated using the BPW_N are given.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

5.3. Significance of Results

Based on the preceding results, this section examines the relation between *NTone* and CARs for different event windows in a multivariate context using the control variables that I described above. Table 9 reports the regression results for *NTone* using the BPW_N and the five different event windows for CARs.

Table 9
Regression of NTone_BPW_N and CARs

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
NTone_BPW_N	0.014***	0.018***	0.018***	0.035***	0.064***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_N	0.009**	0.004	0.011**	0.011	0.014
	(0.004)	(0.005)	(0.005)	(0.008)	(0.012)
IND_BPW_N	0.071***	0.045	0.041	0.042	0.056
	(0.027)	(0.032)	(0.036)	(0.055)	(0.082)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0002	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.028	-0.091	-0.144	-0.679**	-1.162**
	(0.201)	(0.247)	(0.210)	(0.326)	(0.499)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.084***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00001	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
DIV_SPP	-0.0002	0.003	0.007**	0.007	0.018***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.119**	-0.064	-0.115*	-0.128	-0.152
	(0.046)	(0.052)	(0.062)	(0.097)	(0.141)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.032	0.050	0.053	0.073	0.121
Adjusted R2	0.004	0.022	0.026	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.149	1.826**	1.977***	2.747***	4.800***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: Author's calculation.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

The results show a high statistical significance of the coefficient of the sentiment measurement *NTone* that I calculated using the BPW_N and the five different CARs as dependent variables. Thus, more positive speeches of CEOs can be associated with higher abnormal returns. An increase in *NTone* by the interquartile change of 0.326 leads to a minor increase of 0.42% in CAR [0,1], but a major increase of 1.53% in CAR [0,30]. This role as a key factor in the market reaction to AGMs becomes more interesting, when other variables, based on the performance or the dividend policy are considered. The ROA negatively relates to all five event windows and is only significant for the first three windows. The dividend surprise can only partially account for the significance of the longer event windows. I could verify only a significant association with individual event windows for the analyzed control variables. None of the variables are able to explain all windows.

Regarding the significant relation of *NTone* as a relative measurement of sentiment and short- and long-term event windows, the results are consistent with Price et al. (2012, pp. 1004f.) and Bannier et al. (2017, p. 37, 2019a, p. 34).

Despite the insufficient fulfillment of the assumption that speeches with a more positive measurement of sentiment lead to higher abnormal returns for *NTone* using the BPW_O, Table 10 shows that the positive relation between this measurement and the different CARs is almost as significant as the usage of BPW_N. Only for the event windows CAR [0,1] and CAR [0,5], the coefficient is significant at a 5% level. Due to the smaller interquartile change of 0.323, a change in *NTone* by this change leads to a 0.39% higher CAR [0,1] and a 1.45% higher CAR [0,30]. Interestingly, these results show higher significance than Bannier et al. (2019a, p. 34), where maximum significance at the 5% level was achieved (CAR [0,30]: 10%).

Table 10
Regression of NTone_BPW_O and CARs

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
NTone_BPW_O	0.012**	0.016***	0.017**	0.034***	0.062***
	(0.005)	(0.006)	(0.007)	(0.011)	(0.017)
LN_COUNT_BPW_O	0.008**	0.003	0.010*	0.009	0.011
	(0.004)	(0.005)	(0.005)	(0.009)	(0.012)
IND_BPW_O	0.075**	0.036	0.037	0.022	0.032
	(0.031)	(0.037)	(0.042)	(0.064)	(0.097)
LN_SIZE	0.001	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
M2B	-0.0001	-0.0003	0.0004	0.001	0.001
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.002)
LEV	-0.002	-0.006	-0.007	-0.003	-0.007
	(0.005)	(0.006)	(0.007)	(0.010)	(0.016)
VOLA	0.022	-0.098	-0.149	-0.688**	-1.172**
	(0.202)	(0.248)	(0.210)	(0.324)	(0.496)
LN_VOL	-0.001	-0.001*	-0.001	-0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ROA	-0.044**	-0.080***	-0.083***	-0.035	-0.054
	(0.021)	(0.024)	(0.025)	(0.039)	(0.063)
EPS_SP	0.00002	0.00002	0.00002	-0.0001	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 10 (cont.)

	Dependent variable:				
	CAR01 (1)	**CAR03** (2)	**CAR05** (3)	**CAR015** (4)	**CAR030** (5)
DIV_SPP	-0.00001	0.003	0.007**	0.008	0.018***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)
DIV_SPN	-0.003	-0.002	0.002	-0.003	-0.022**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.010)
Constant	-0.110**	-0.043	-0.104	-0.096	-0.111
	(0.048)	(0.054)	(0.065)	(0.098)	(0.145)
Observations	829	829	829	829	829
Year Fixed Effects	YES	YES	YES	YES	YES
R2	0.029	0.047	0.052	0.072	0.120
Adjusted R2	0.001	0.020	0.025	0.046	0.095
Residual Std. Error (df = 805)	0.026	0.031	0.036	0.057	0.082
F Statistic (df = 23; 805)	1.042	1.736**	1.937***	2.721***	4.790***

Significance levels are based on robust standard errors (given in parentheses) and are indicated by * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Source: Author's calculation.

Based on the already stated results for the necessary assumptions of the cumulative abnormal trading volumes under 5.2, I will not discuss those regressions further.

6. CONCLUSION

This paper focuses on textual analysis as an important part of accounting and finance research using the dictionary-based approach with the first available finance-related dictionary for the German language (BPW_O). Due to the novelty of this dictionary, the aim of this paper is to propose several reforms and extensions (BPW_N) to improve its performance and to find the most appropriate measurement of sentiment.

Based on the visual examination of the two central assumptions that speeches with a more positive measurement of sentiment lead to higher abnormal returns and that it is possible to separate above and below average abnormal returns through the use of sentiment measures, the use of the measurement *NTone* calculated using the BPW_N should be preferred. Additionally, I was able to supplement the significance of these results by several regressions. Here the use of *NTone*, calculated by using the BPW_N, could provide highly statistically significant results for all five analyzed event windows. Thus, more positive speeches of CEOs can be associated with higher abnormal returns following the Annual General Meeting. Based on the event window, an increase in *NTone* by the interquartile change of 0.326 leads to an increase in cumulative abnormal returns ranging from 0.42% (CAR [0,1]) to 1.53% (CAR [0,30]).

Using the most comprehensive collection of German CEO speeches so far, this paper is able to give two contributions to the literature on textual analysis of German texts. Through implementing reforms and extensions, I improved the results of the original BPW_O and confirmed the stated hypothesis. Additionally, the combination of the BPW_N and the relative measurement of sentiment *NTone* has proven to be the most suitable one for measuring business texts and therefore answers the additional research question.

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Due to the results of the proposed adjustments on the newly developed BPW_O, additional improvements should be considered and tested. Moreover, this new version of the BPW (BPW_N) should be compared to old and new versions of general German dictionaries. As there is a wide range of publicly available textual data, the BPW_N should be used to analyze other types of corporate disclosures.

References

Ahmed, Y., & Elshandidy, T. (2016). The effect of bidder conservatism on M&A decisions: Text-based evidence from US 10-K filings. International Review of Financial Analysis, 46, 176–190. https://doi.org/10.1016/j.irfa.2016.05.006

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. Retrieved on 02.04.2020 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2652876. https://doi.org/10.1111/joes.12370

Allee, K. D., & Deangelis, M. D. (2015). The structure of voluntary disclosure narratives: Evidence from tone dispersion. Journal of Accounting Research, 53(2), 241–274. https://doi.org/10.1111/1475-679X.12072

Ammann, M., & Schaub, N. (2016). Social interaction and investing: Evidence from an online social trading network (Working Paper). Retrieved on 11.07.2018 from https://www.rsm.nl/fileadmin/home/Department_of_Finance__VG5_/PAM2016/Final_Papers/Nic_Schaub.pdf

Apel, M., & Blix Grimaldi, M. (2012). The information content of central bank minutes. Sveriges Riksbank Working Paper Series, (261). Stockholm: Sveriges Riksbank. Retrieved on 13.02.2020 from http://archive.riksbank.se/Documents/Rapporter/Working_papers/2012/rap_wp261_120426.pdf. https://doi.org/10.2139/ssrn.2092575

Bannier, C. E., Pauls, T., & Walter, A. (2017). CEO-speeches and stock returns (Working Paper). Retrieved on 15.08.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3051151_code1882913.pdf?abstractid=3051151&mirid=1

Bannier, C. E., Pauls, T., & Walter, A. (2019a). The Annual General Meeting revisited: The role of the CEO speech (Working Paper). Retrieved on 11.12.2021 from https://ssrn.com/abstract=2869785

Bannier, C. E., Pauls, T., & Walter, A. (2019b). Content analysis of business specific text documents: Introducing a German dictionary. Journal of Business Economics, 89(1), 79–123. https://doi.org/10.1007/s11573-018-0914-8

Boudt, K., & Thewissen, J. (2019). Jockeying for position in CEO letters: Impression management and sentiment analytics. Financial Management, 48(1), 77115. https://doi.org/10.1111/fima.12219

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. Review of Accounting Studies, 20(2), 639–673. https://doi.org/10.1007/s11142-014-9309-4

Doran, J. S., Peterson, D. R., & Price, M. S. (2012). Earnings conference call content and stock price: The case of REITs. Journal of Real Estate Finance and Economics, 45(2), 402–434. https://doi.org/10.1007/s11146-010-9266-z

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending: Evidence from two leading European platforms. Journal of Banking & Finance, 64, 169–187. https://doi.org/10.1016/j.jbankfin.2015.11.009

Ferguson, N. J., Philip, D., Lam, H. Y. T., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. Multinational Finance Journal, 19(1), 1–31. https://doi.org/10.17578/19-1-1

Franke, B. (2018). Qualitative information and loan terms: A textual analysis (Working Paper). Retrieved on 15.09.2019 from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3209201_code1660824pdf?abstractid=3152458&mirid=1

Fritz, D., & Tows, E. (2018). Text mining and reporting quality in German banks: A cooccurrence and sentiment analysis. Universal Journal of Accounting and Finance, 6(2), 54–81. https://doi.org/10.13189/ujaf.2018.060204

Garcia, D. (2013). Sentiment during recessions. The Journal of Finance, 68(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535–574. https://doi.org/10.1257/jel.20181020

González, M., Guzmán, A., Téllez, D. F., & Trujillo, M. A. (2019). What you say and how you say it: Information disclosure in Latin American firms. Journal of Business Research, 127(3), 427 443. https://doi.org/10.1016/j.jbusres.2019.05.014

Gurun, U. G., & Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. The Journal of Finance, 67(2), 561–598. https://doi.org/10.1111/j.1540-6261.2012.01725.x

Hart, R. P. (2000). DICTION 5.0. Retrieved on 09.06.2020 from https://rhetorica.net/diction.htm

Henry, E. (2006). Market reaction to verbal components of earnings press releases: Event study using a predictive algorithm. Journal of Emerging Technologies in Accounting, 3, 1–19. https://doi.org/10.2308/jeta.2006.3.1.1

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Henry, E. (2008). Are investors influenced by how earnings press releases are written?. Journal of Business Communication, 45(4), 363–407. https://doi.org/10.1177/0021943608319388
Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. The Accounting Review, 91(1), 153–178. https://doi.org/10.2308/accr-51161
Jandl, J.-O., Feuerriegel, S., & Neumann, D. (2014). Long- and short-term impact of news messages on house prices: A comparative study of Spain and the United States. Paper presented at Thirty Fifth International Conference on Information Systems, Auckland. Retrieved on 15.09.2019 from https://aisel.aisnet.org/icis2014/proceedings/DecisionAnalytics/17/
Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. Journal of Financial Economics, 110(3), 712–729. https://doi.org/10.1016/j.jfineco.2013.08.018
Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171–185. https://doi.org/10.1016/j.irfa.2014.02.006
Kim, Y. H., & Meschke, F. (2014). CEO interviews on CNBC (Working Paper). Retrieved on 12.02.2020 from http://dx.doi.org/10.2139/ssrn.1745085. https://doi.org/10.2139/ssrn.1745085
Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. Accounting and Business Research, 49(5), 587–615. https://doi.org/10.1080/00014788.2019.1611730
Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. Journal of Accounting Literature, 29, 143–165.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x
Loughran, T., & McDonald, B. (2015). The use of word lists in textual analysis. Journal of Behavioral Finance, 16(1), 1–11. https://doi.org/10.1080/15427560.2015.1000335
Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 54, 1187–1230. https://doi.org/10.1111/1475-679X.12123
Loughran, T., & McDonald, B. (2020). Stop words. Retrieved on 21.01.2021 from https://drive.google.com/file/d/0B4niqV00F3mseWZrUk1YMGxpVzQ/view?usp=sharing
Loughran, T., McDonald, B., & Yun, H. (2009). A wolf in sheep's clothing: The use of ethics-related terms in 10-K reports. Journal of Business Ethics, 89(1), 39–49. https://doi.org/10.1007/s10551-008-9910-1
Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. The Journal of Finance, 67(1), 1–43. https://doi.org/10.1111/j.1540-6261.2011.01705.x
Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). »LIWC auf Deutsch«: The development, psychometrics, and introduction of DE-LIWC2015. Retrieved on 08.03.2019 from https://osf.io/tfqzc/. https://doi.org/10.31234/osf.io/uq8zt
Mengelkamp, A., Wolf, S., & Schumann, M. (2016). Data driven creation of sentiment dictionaries for corporate credit risk analysis. Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). Retrieved on 10.07.2018 from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1058&context=amcis2016
Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. Journal of International Money and Finance, 79, 136–156. https://doi.org/10.1016/j.jimonfin.2017.09.005
Price, M. S., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking & Finance, 36(4), 992–1011. https://doi.org/10.1016/j.jbankfin.2011.10.013
Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - A publicly available German-language resource for sentiment analysis. Proceedings of the 7th International Language Ressources and Evaluation (LREC'10) (pp. 1168–1171). Retrieved on 19.12.2018 from http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf
Röder, F., & Walter, A. (2019). What drives investment flows into social trading portfolios?. The Journal of Financial Research, 42(2), 383–411. https://doi.org/10.1111/jfir.12174
Schmeling, M., & Wagner, C. (2016). Does central bank tone move asset prices?. Paper presented at the 77th Annual Meeting of American Finance Association (AFA 2017). Retrieved on 29.06.2018 from https://research.cbs.dk/en/publications/does-central-bank-tone-move-asset-prices(c6401864-a921-401c-90db-57d42d6b5022).html
Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). The General Inquirer: A computer approach to content analysis, Cambridge, Mass.: The M.I.T. Press.
Tillmann, P., & Walter, A. (2018). ECB vs Bundesbank: Diverging tones and policy effectiveness. Joint Discussion Paper Series in Economics, (20). Retrieved on 13.02.2020 from https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/paper_2018/20-2018_tillmann.pdf
Tillmann, P., & Walter, A. (2019). The effect of diverging communication: The case of the ECB and the Bundesbank. Economics Letters, (176), 68–74. https://doi.org/10.1016/j.econlet.2018.12.035
Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. Diagnostica, 54(2), 85–98. https://doi.org/10.1026/0012-1924.54.2.85

APPENDIX

Table 11
Description of variables

Variable	Description
SIZE	Firm Size: Daily market value
M2B	Market to Book Value: Ratio of the market value of the ordinary (common) equity to the balance sheet value of the ordinary (common) equity
LEV	Leverage: Ratio of the total liabilities to the total assets
VOLA	Volatility: Standard deviation of the daily returns for the ninety trading-day window ending ten days prior to the AGM
VOL	Volume: Number of shares traded on the day of the AGM
COUNT	Total number of Words. Due to different stop word lists calculated individually for BPW_O and BPW_N
IND_NUM	Number of individual words. Due to different stop word lists calculated individually for BPW_O and BPW_N.
IND	Individual Words: *IND_NUM* divided by *COUNT*
ROA	Return on Assets: Net income divided by total assets
EPS_SP	Earnings Surprise: Calculated according to Bannier et al., 2017: The difference between the last reported earnings per share at time t minus the latest reported earnings per share in the year prior to date t, divided by the stock price one year before t times 100 $EPS_{SP} = \frac{EPS_t - EPS_{t-1}}{Price_{t-1}} \cdot 100$
DIV_SPP	Dividend Surprise Positive: Calculated according to Bannier et al., 2017: *DIV_SPP* equals one if the dividend per share is increased compared to the previous year, zero otherwise
DIV_SPN	Dividend Surprise Negative: Calculated according to Bannier et al., 2017: *DIV_SPN* equals one if the dividend per share is decreased compared to the previous year, zero otherwise
P_NUM	Number of positive words
N_NUM	Number of negative words

DOI: 10.7172/2353-6845.jbfe.2021.2.1

Table 12
Translation of ten most frequent words

positive words		negative words	
German	**English**	**German**	**English**
besser	better	betrug	fraud, amounted
erfolg	success	finanzkrise	financial crisis
erfolgreich	successful	gegen	against
erreichen	achieve	herausforderungen	challenges
erreicht	achieved	krise	crisis
große	large	leider	unfortunately
großen	large	minus	minus
positiv	positive	rückgang	decline
positive	positive	schwierigen	difficult
stärken	strenghten	sorgen	sorrow, care
stärker	stronger	verfügung	decree
		wider	against

Note that the listed translations represent only one of several possibilities. Due to the nature of the parsing procedure and structure of the dictionaries, an important distinction of German words through small and capital letters is not possible.

DOI: 10.7172/2353-6845.jbfe.2021.2.1