A Phrase-Based Opinion List for the German Language

Sven Rill, Sven Adolph,
Johannes Drescher, Dirk Reinel,
Jörg Scheidt, Oliver Schütz,
Florian Wogenstein

University of Applied Sciences Hof
Hof, Germany

(srill,sadolph,jdrescher,dreinel,jscheidt,
oschuetz,fwogenstein)@iisys.de

Roberto V. Zicari, Nikolaos Korfiatis

Goethe University Frankfurt
Frankfurt am Main, Germany

zicari@informatik.uni-frankfurt.de
nikos@dbis.cs.uni-frankfurt.de

Abstract

We present a new phrase-based generated list of opinion bearing words and phrases for the German language. The list contains adjectives and nouns as well as adjective- and noun-based phrases and their opinion values on a continuous range between -1 and $+1$. For each word or phrase two additional quality measures are given. The list was produced using a large number of product review titles providing a textual assessment and numerical star ratings from Amazon.de. As both, review titles and star ratings, can be regarded as a summary of the writers opinion concerning a product, they are strongly correlated. Thus, the opinion value for a given word or phrase is derived from the mean star rating of review titles which contain the word or phrase. The paper describes the calculation of the opinion values and the corrections which were necessary due to the so-called "J-shaped distribution" of online reviews. The opinion values obtained are amazingly accurate.

1 Introduction

The amount of textual data increases rapidly in the World Wide Web, so does the need of algorithms enabling an efficient extraction of information from this data. Namely the extraction of opinions (opinion mining) gained increasing attention in the research community.

Many opinion mining algorithms and applications need text resources like polarity lexicons / opinion lists consisting of words or phrases with their opinion values. These opinion values either classify words in positive, neutral or negative words or give opinion values in a continuous range between -1 and $+1$ providing a finer resolution in the measure of their opinion polarities. Polarity lexicons usually are derived from dictionaries or text corpora. The quality of the used lexical resources is of utmost importance for the quality of the results obtained from opinion mining applications. In (Oelke et al., 2009) the authors applied opinion mining on customer feedback data. They analyzed error sources and came to the conclusion that about 20% of the errors occurred due to faults in the opinion list. In addition, most of the other error sources were related to the opinion list.

Using opinion lists in applications raises the question, how to handle phrases containing opinion words plus one or several valence shifters (intensifiers or reducers) or negation words or combinations of both. One could assume that negation words just change the sign of the opinion value and valence shifters change its absolute value by a defined step. In some cases, however, this is not correct. For example, "gut" - *'good'* and "perfekt" - *'perfect'* are positive opinion words. The negation of "gut", "nicht gut" - *'not good'* can be regarded as a negative phrase, although "nicht perfekt" - *'not perfect'* cannot.

Similar effects occur for valence shifter words. In the field of sentiment composition, the handling of these valence shifters and negation words is discussed in several papers (Choi and Cardie, 2008; Klenner et al., 2009; Liu and Seneff, 2009; Moilanen and Pulman, 2007).

An alternative way is the inclusion of intensifiers, reducers and negation words directly in the opinion list. We follow this approach and use an algorithm presented in (Rill et al., 2012) to generate a list containing opinion bearing phrases together with their opinion values for the German language.

2   Related Work

The automatic extraction of opinions and sentiments has gained interest as the amount of textual data increases permanently. Thus, a lot of research work has been done in the area of opinion mining. An overview of the whole topic recently has been given in (Liu and Zhang, 2012).

Text resources, namely lists of opinion bearing words together with an assessment of their subjectivity, have been provided for several languages.

For the English language publicly available word lists are SentiWordNet (Baccianella et al., 2010; Esuli and Sebastiani, 2006a; Esuli and Sebastiani, 2006b), Semantic Orientations of Words (Takamura et al., 2005), the Subjectivity Lexicon (Wilson et al., 2005) and two lists of positive and negative opinion words provided by (Liu et al., 2005).

Also for the German language lists of opinion bearing words already exist. In (Clematide and Klenner, 2010) the authors described a polarity lexicon (PL) listing the opinion values for about 8,000 German nouns, adjectives, verbs and adverbs. The words are classified into negative and positive words with opinion values in six discrete steps. In addition, PL includes some shifters and intensifiers. The list was generated using *GermaNet*, a German lexicon similar to *WordNet*, which already was used to derive English polarity lexicons.

In (Waltinger, 2010) *GermanPolarityClues* (GPC) was introduced. It consists of more than 10,000 German nouns, adjectives, verbs and adverbs classified as positive, neutral or negative opinion words. For a part of these words, also the probabilities for the three classes are given. GPC also lists about 300 negation words. *GermanPolarityClues* was not derived directly from German text data but was produced using a semi-automatic translation approach of English-based

sentiment resources.

PL and GPC will be used as benchmarks for our list (see Section 4.3).

In (Rill et al., 2012), the authors proposed a generic algorithm to derive opinion values from online reviews taking advantage of the fact that both, star ratings and review titles can be regarded as a short summary of the writer's opinion and therefore are strongly correlated. The authors use this algorithm to derive a list of opinion bearing adjectives and adjective-based phrases for the English language. In this work, we use this algorithm to produce a new opinion list consisting of adjectives and nouns as well as adjective- and noun-based phrases for the German language.

In addition and in contrast to the previous work, corrections, which are necessary due to the "J-shaped distribution" of online reviews, are applied. Reasons and implications of this "J-shaped distribution" are discussed in several publications (Hu et al., 2007; Hu et al., 2009).

Online reviews are used for several other research projects, for an overview see (Tang et al., 2009).

3   Generation of the Opinion List

3.1   General Approach

On a typical review platform of an online shop, user-written product reviews include a title and a numerical evaluation among other information. Amazon uses a star rating with a scale of one to five stars. Both, title and star rating can be regarded as a summary of the user's opinion about the product under review. Thus, the opinion expressed in the title, using opinion bearing words and phrases, is strongly correlated to the star rating. This leads to the conclusion that opinion values for words or phrases occurring in the titles of reviews can be generated by taking advantage of this correlation.

The calculation of opinion values is performed in several steps, described in the subsequent sections. Figure 1 depicts the whole process.

3.2   Data Retrieval and Preprocessing

Crawling and Language Detection

Basis of this work are review titles and star ratings crawled from the German Amazon site
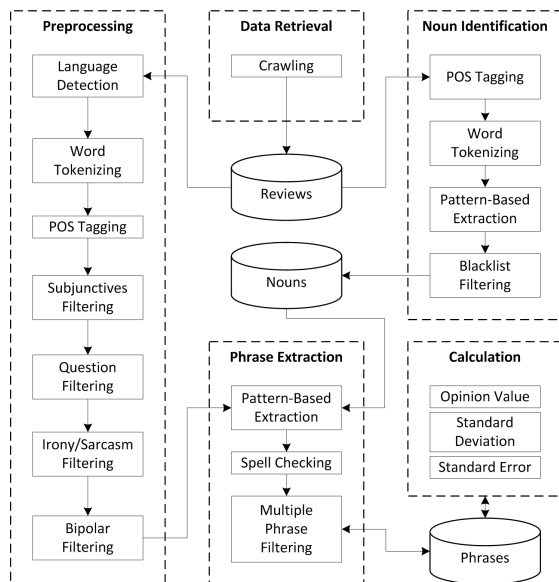
Figure 1: Overview of the opinion list generation.

(Amazon.de). The review texts and the additionally available information like product information, helpfulness count and the comments on the reviews are of no interest for this project. Thus, they were excluded from the crawling. The data set consists of about 1.16 million pairs of review titles and star ratings.

A language detection was performed using the Language Detection Library for Java by S. Nakatani[1].

Word Tokenizing and Part-of-Speech Tagging

The word tokenizing and part-of-speech (POS) tagging are the next preprocessing steps to be performed.

In some cases, the POS tagging is quite difficult for review titles as they sometimes consist only of some words instead of a complete sentence and therefore words are quite often mistagged. Especially, if the title starts with a capitalized adjective, it is often mistagged as a noun or a named entity, e.g., "Gutes Handy!" - *'Good mobile phone!'* "Gutes" - *'Good'* has to be tagged as an adjective. We used the *Apache OpenNLP POS Tagger*[2], with the maximum entropy model which was trained using the *TIGER treebank* (Brants et al., 2002). The POS tags obtained are given using

[1] http://code.google.com/p/language-detection/
[2] http://opennlp.apache.org/

the *Stuttgart-Tübingen Tagset* (STTS) (Schiller and Thielen, 1995).

To improve the POS tagging in respect to the above-mentioned problem, we converted each first word of a review title to small letters, if it is tagged as a noun, and repeated the POS tagging. If the probability for being an adjective exceeds the noun probability after the conversion to small letters, the word is taken to be an adjective.

Filtering

In some cases star rating and textual polarity are not correlated. Therefore, we perform the same filtering steps like proposed in (Rill et al., 2012):

- Subjunctives often imply that a statement in a review title is not meant as the polarity of the word or phrase indicates, e.g., "Hätte ein guter Film werden können" - *'Could have been a good film'*. In the German language "hätte" - *'could'*, "wäre" - *'would be'*, "könnte" - *'might be'* and "würde" - *'would'* are typical words indicating a subjunctive. Hence, review titles containing one of these words are omitted.

- Some titles are formulated as questions. Many of them are not useful as they often express the opposite opinion compared to the star rating, e.g., "Warum behaupten Leute, das sei gut?" - *'Why do people say that this is good?'*. Therefore, titles are excluded if they contain an interrogative and a question mark at the end.

- Some review titles are meant ironically. Irony cannot be detected automatically in most of the cases (Carvalho et al., 2009) but exceptions exist. Sometimes, for example, emoticons like ";-)" can be regarded as signs of irony. Also, quotation marks are sometimes used to mark a statement as ironic, e.g., "Wirklich ein 'großartiger' Film!" - *'Really a "great" movie!'*. Thus, titles containing emoticons or quotation marks are excluded from the data set.

- The words "aber" - *'but'*, "jedoch" - *'however'*, "sondern" - *'but'* and "allerdings" - *'though'* are indicators for a bipolar opinion, e.g., "scheint gut zu sein, aber ..." - *'seems*

307

good, but ...'. Again, the star rating does not correspond to the opinion value of the opinion word so titles containing one of these words are omitted.

3.3 Opinion Word and Phrase Extraction

In this work we extract opinion words and phrases based on opinion bearing adjectives and nouns like "absolut brilliant" - *'absolutely brilliant'*, "nicht sehr gut" - *'not very good'*, "exzellent" - *'excellent'* (adjective-based) and "totaler Müll" - *'complete rubbish'* (noun-based). Verb-based phrases like "enttäuscht (mich) nie" - *'never disappoints (me)'* are not regarded. Opinion phrases consist of at least one opinion bearing word. In addition, they might contain shifters and/or negation words and/or other words like adverbs or adjectives.

Opinion Bearing Nouns

The first step in the construction of opinion bearing phrases is the identification of candidates for opinion bearing nouns like "Meisterwerk" - *'masterpiece'*, "Enttäuschung" - *'disappointment'* or "Schrott" - *'dross'*. To create a list consisting of such words, we look at the review titles considering the nouns of following patterns as candidates for opinion bearing nouns:

1. A single noun or a single noun with an exclamation sign, e.g., "Frechheit" - *'Impudence'* or "Wahnsinn!" - *'Madness!'*.

2. A single noun with an article in front and an exclamation mark (optional), e.g., "Der Alleskönner! - *'The all-rounder!'*.

3. A noun with a form of "sein" - *'to be'* or "haben" - *'have'*, e.g., "Der Service ist eine Frechheit!" - *'The support is a cheek!'* or "Die Kamera hat ein Problem" - *'There is a problem with the camera'*.

4. A noun with "mit" - *'with'*, or "in" - *'in'* in front, e.g., "Karte mit Macken" - *'Card with faults'* or "Trio in Höchstform" - *'Trio in top form'*.

5. A noun with a following "bei" - *'during'*, e.g., "Tonstörung bei Wiedergabe" - *'Sound problem during playback'*.

Afterwards, some nouns are removed from the list according to a manually created blacklist. This is necessary as the list still contains some mistagged words, e.g., adjectives or named entities.

Noun-Based Opinion Phrases

Every opinion bearing noun in a review title is a candidate for an opinion phrase. We start at the end of each review title. For each candidate the phrase is extended to the left as long as it fulfills one of the patterns below.

1. Single noun, e.g., "Enttäuschung" - *'disappointment'*.

2. A noun with an adjective, e.g., "absoluter Mist" - *'absolute rubbish'*.

3. A noun with one or more adverbs, adjectives and/or indefinite pronouns, e.g., "Keine absolute Kaufempfehlung" - *'Not an absolute recommendation to buy'*.

As a last step, the nouns of noun-based opinion phrases have to be lemmatized. This means that each noun has to be reduced to its canonical form. For nouns, the several plural forms as well as forms of several cases are changed to the nominative singular, e.g., "des Meisterwerks" and "Meisterwerke" to "Meisterwerk" - *'masterpiece'*. For the lemmatizing, we used the Web service provided by the *Deutscher Wortschatz* project of the University Leipzig (Quasthoff, 1998).

Adjective-Based Opinion Phrases

In contrast to the construction of noun-based phrases, single opinion bearing adjectives and adjective-based phrases can be retrieved in one step. They are extracted according to the following patterns:

1. Single adjective, e.g., "Großartig!" - *'Great!'*.

2. One or more adverbs, particles or past participles and an adjective, e.g., "Sehr guter (Film)" - *'Very good (movie)'*, "Nicht gut (für ein iPad)" - *'Not good (for an iPad)'*, "Gewohnt gut" - *'Good as usual'*, "Gut verarbeitet" - *'Well processed'* .

3. Like pattern number 2 but with one or more adverbs replaced by adjectives, e.g., "Sehr schöner kleiner (Bildschirm)" - *'Very nice little (screen)'*.

As for noun-based opinion phrases a lemmatizing of the base adjectives has to be performed. For German adjectives, for example, the forms "großer" and "großes" have to be reduced to "groß" - *'big'*.

Filtering of Opinion Bearing Phrases

At this stage of the algorithm a spell checker is applied to identify misspelled words in the phrases. We used the *Hunspell Spell Checker*[3] with the *de-DE frami* word list[4]. In cases where the spell checker marks a word as misspelled, this review title is omitted.

In addition, only titles with exactly one opinion phrase are accepted at this point. The reason is that titles containing more than one opinion phrase have the problem that phrases normally have different opinion values. In extreme cases they are contradicting, e.g., "Gute Geschichte, schlecht geschrieben" - *'Good story, badly written'*. Therefore, titles containing two or more opinion phrases are discarded.

3.4 Calculation of Opinion Values

After the preselection steps described in the sections before, the data set consists of about 420,000 review titles each having one opinion phrase and a star rating between one and five. For each phrase occurring frequently in the review titles, the opinion value is calculated by transposing the mean star rating of all review titles having this phrase to the continuous scale $[-1, +1]$, assuming that a three star rating represents a neutral one:

$$OV = \left(\frac{\sum_{s=1}^{5} n_s \cdot s}{n} - 3 \right) : 2 \qquad (1)$$

Here, s is the number of stars (one to five), n_s the number of review titles with s stars and n the total number of review titles for the given phrase. Frequently at this stage means that a phrase has to occur at least ten times in the preselected review titles.

In addition to the opinion value, two quality measures are calculated. The first one is just the standard deviation σ_{OV} of the opinion value. It

is a measure of how much the star rating spreads for a given opinion phrase. The second one is the standard error (SE) calculated by dividing the standard deviation by the square root of the number n_i of review titles having phrase i. In addition to the spread of the stars, it indicates on how many review titles the opinion value of a given phrase is based.

These quality measures can be helpful in the usage of our list. If the opinion value of a phrase is near zero, the σ_{OV} indicates whether the phrase is really used mainly in neutral reviews (small σ_{OV}) or in both very positive and negative reviews (large σ_{OV}). For example, the word "fassungslos" - *'stunned'* with an opinion value of 0.05 is used as a positive word for book reviews, and also as a negative word for quality features. This results in a large σ_{OV}.

Table 1 shows some opinion phrases together with their opinion values and the two quality measures at this point of the algorithm.

Phrase	OV	σ	SE
großartig - *great*	0.95	0.23	0.01
einfach gut - *just good*	0.93	0.19	0.01
sehr gut - *very good*	0.90	0.22	0.00
nur Schrott - *total dross*	-0.85	0.52	0.10
nur schlecht - *very bad*	-0.97	0.16	0.01

Table 1: Some words and phrases with their opinion values and two quality measures.

3.5 Correction of the Opinion Values

Most opinion values already look reasonable at this point, but some are not yet satisfactory. Especially, the values for some single adjectives, expected to carry no opinion, are shifted to positive values. The reason is that samples of online reviews often show a so-called "J-shaped distribution". This means that for a big collection of reviews on a one to five scale, the star distribution has a parabolic shape with a minimum at about two stars. In our sample of single adjectives, we find about 7% 1-star, 5% 2-star, 6% 3-star, 17% 4-star and 65% 5-star review titles. For an adjective, expressing no opinion and therefore being distributed equally over all review titles, this means that it will receive an opinion value accord-

ing to this "J-shaped distribution". The mean star rating of this distribution at 4.28 stars corresponds to an opinion value of 0.64, so many neutral adjectives get an opinion value in this region. Thus, a correction is necessary. We proceed in the following way. In a first step, we classify all single adjectives into the two classes "Neutrals Following J-shaped Distribution" and "Others". For words in the first class we require their star distribution to follow the "J-shaped distribution" quantitatively and qualitatively. The quantitative deviation from this distribution we estimate by calculating the measure

$$S_{J_1} = 1 - \sqrt{\sum_{s=1}^{5} \left(\frac{n_s}{n} - a_s \right)^2} \qquad (2)$$

where a_s is the relative frequency of s-star review titles in the whole sample. We found out that a value of S_{J_1} greater than 0.85 indicates that the star distribution for a word is quite similar to the "J-shaped distribution" of the whole set of review titles. In addition, we set the measure S_{J_2} to TRUE (T) if $n_1 \geq n_2$ and $n_4 \leq n_5$, otherwise S_{J_2} is set to FALSE (F). In this way, we make sure that the correction of the opinion value is only applied to words also following the "J-shaped distribution" qualitatively.

In the second step, the opinion values get corrected for words fulfilling both conditions by weighting the star rating frequencies n_s with the relative frequencies of the whole sample (a_s):

$$OV_c = \left(\frac{\sum_{s=1}^{5} \frac{n_s \cdot s}{a_s}}{\sum_{s=1}^{5} \frac{n_s}{a_s}} - 3 \right) : 2 \qquad (3)$$

Table 2 lists some words and the effect of the "J-shaped distribution" correction to them.

The same correction procedure has to be applied for single nouns. The corresponding fractions of s-star ratings are 13% for 1-star, 6% for 2-star, 8% for 3-star, 15% for 4-star and 58% for 5-star review titles containing the single nouns. The correction is calculated in the same way as described above. Table 3 gives some examples for nouns with corrected opinion values.

Adjective	OV	S_{J_1}	S_{J_2}	OV_c
schnell - *fast*	**0.69**	0.96	F	—
gut - *good*	**0.67**	0.75	F	—
hübsch - *nice*	**0.56**	0.71	T	—
jung - *young*	0.66	0.93	T	**0.12**
schwarz - *black*	0.67	0.97	T	**0.08**
andere - *other*	0.57	0.88	T	**0.01**

Table 2: Some adjectives and the effect of the "J-shaped distribution" correction. Bold opinion values mark the values entering the final list.

Noun	OV	S_{J_1}	S_{J_2}	OV_c
Traum - *dream*	**0.91**	0.67	T	—
Gefühl - *feeling*	**0.69**	0.88	F	—
Gehirn - *brain*	0.59	0.93	T	**0.08**
Zeit - *time*	0.57	0.96	T	**0.06**
Kunst - *art*	0.56	0.96	T	**0.05**

Table 3: Some nouns and the effect of the "J-shaped distribution" correction. Bold opinion values mark the values entering the final list.

4 Results and Discussion

4.1 Statistical Summary

After these steps, our list consists of 3,210 words and phrases. Table 4 summarizes the frequencies of adjectives, adjective phrases, nouns and noun phrases as well as the number of weak and strong subjective words and phrases. We regard a word or phrases as weak subjective, if the opinion value lies between 0.33 and 0.67 or -0.33 and -0.67. Having an opinion value greater than 0.67 or smaller than -0.67, a word or phrase is assumed to be strong subjective. Words and phrases with an opinion value between -0.33 and $+0.33$ we regard as being neutral.

	total	n	ws	ss
Adjectives	1,277	390	400	487
Adjective phrases	938	135	170	633
Nouns	502	124	112	266
Noun phrases	493	51	88	354
Sum	**3,210**	**700**	**770**	**1,740**

Table 4: Number of neutral (**n**), weak subjective (**ws**) and strong subjective (**ss**) words and phrases.

310

More than half of the words and phrases are strong subjective while less than one fourth are neutral.

4.2 Examples of Opinion Values for Words and Phrases

Table 5 lists some adjectives and adjective phrases, Table 6 some nouns and noun phrases together with their opinion values.

Adjectives	OV
großartig - *great*	+0.95
exzellent - *excellent*	+0.91
optimal - *optimal*	+0.87
amüsant - *amusing*	+0.71
gut - *good*	+0.67
brauchbar - *useful*	+0.38
durchschnittlich - *average*	−0.02
mies - *crummy*	−0.44
schlecht - *bad*	−0.56
enttäuschend - *disappointing*	−0.65
unbrauchbar - *unusable*	−0.86
grottenschlecht - *abysmal*	−0.96
Adjective Phrases	**OV**
extrem gut - *extremely good*	+1.00
sehr sehr gut - *very very good*	+0.97
sehr gut - *very good*	+0.90
sehr praktisch - *very useful*	+0.87
gewohnt gut - *good as usual*	+0.78
recht gut - *quite good*	+0.48
nicht schlecht - *not bad*	+0.38
nicht optimal - *not optimal*	−0.05
eher schwach - *rather weak*	−0.26
ziemlich langweilig - *rather boring*	−0.61
nicht gut - *not good*	−0.64
sehr schlecht - *very bad*	−0.83
ganz mies - *just crummy*	−1.00

Table 5: Examples of adjectives and adjective phrases with their opinion values.

The opinion values look plausible. The whole range of possible opinion values is covered.

Interesting is a look at the role of valence shifters and negation words.

The valence shifter "sehr" - *'very'* shifts the opinion values in the expected direction. Also a multiple usage ("sehr sehr gut" - *'very very good'*)

leads to a further increase of the opinion value. Also visible is the fact that a valence shifter can not be described with a common factor which can be applied to all adjectives.

The results for the negation word "nicht" - *'not'* show that it does not always change the sign of the opinion value (as for "gut" - *'good'*) leaving its absolute value nearly unchanged. In fact, for many words the negation changes a strong polarity to a weak or neutral one, e.g., for "schlecht" - *'bad'* or "optimal" - *'optimal'*.

Nouns	OV
Weltklasse - *world class*	+0.98
Superprodukt - *super product*	+0.97
Meisterwerk - *masterpiece*	+0.94
Durchschnitt - *average*	−0.05
Unsinn - *nonsense*	−0.52
Enttäuschung - *disappointment*	−0.71
Zumutung - *impertinence*	−0.77
Frechheit - *impudence*	−0.89
Zeitverschwendung - *waste of time*	−0.93
Noun Phrases	**OV**
absolutes Muss - *absolute must*	+0.97
sehr gute Qualität - *very good quality*	+0.94
großer Spaß - *great fun*	+0.75
nur Durchschnitt - *only average*	−0.02
mangelnde Qualität - *lack of quality*	−0.55
absoluter Fehlkauf - *absolute bad buy*	−1.00

Table 6: Examples of nouns and noun phrases with their opinion values.

Also for the nouns the opinion values obtained look good. Again, the valence shifters change the opinion values in the expected way.

4.3 Comparison to Existing Lists

To compare our list with the two existing polarity lexicons (PL and GPC, see Chapter 2), we compare some single opinion words, which are frequently used in text sources. The result of this comparison is given in Table 7. For both, PL and GPC, the polarity (positive - P and negative - N) is given. For GPC, the attached numbers give the probability of the word having this polarity. For PL, the number indicates the strength of the opinion. We can see that the values in the three lists

agree in the sense that the classification into positive and negative words is consistent in all cases.

Adjectives	OV	PL	GPC
toll - *great*	+0.87	P 1.0	P 0.43
zufrieden - *satisfied*	+0.82	P 1.0	P 0.49
spannend - *exciting*	+0.78	P 1.0	P 0.17
gut - *good*	+0.67	P 1.0	P 0.32
schwach - *weak*	−0.18	N 0.7	N 0.61
schlecht - *bad*	−0.56	N 1.0	N 0.61

Table 7: Opinion values taken from our list (OV) compared to values from the polarity lexicon (PL) and the GermanPolarityClues (GPC).

Quantitatively, each of the two benchmark lists contain more than twice as many words than our list.

However, we see an advantage of our approach in the fact that our list not only contains single words but also phrases, so the treatment of negation and intensification gets much easier. Furthermore, instead of only classifying the words into a few polarity classes, we calculated opinion values in a continuous range between −1 and +1 which leads to a more precise assessment when applying the list, e.g., in the field of aspect-based opinion mining.

4.4 Shortcomings and Future Work

In Section 3.5 we discussed a correction necessary due to the "J-shaped distribution" of online reviews. This special distribution could cause another effect on the opinion values of single adjectives carrying a negative opinion. Some of these words sometimes are used ironically or in an idiomatic expression and therefore do not express a negative opinion. If these cases are rare and equally distributed over all review title, this effect would not lead to shifts in the opinion values. However, as these "misusages" cannot be regarded as distributed equally due to the "J-shaped distribution" of reviews, this can lead to a shift of the opinion values to positive values for some opinion words.

As stated in Section 3.4 we assumed a three star rating to be the expression of a neutral opinion about a product. However, this has not to be exactly the case. In fact, reviews with a three star rating are sometimes regarded as slightly negative ones. This could lead to a shift of opinion values to positive values. If the mean star rating for neutral reviews is shifted by 0.1 stars to 3.1 stars for example, the resulting opinion values would be shifted by 0.05 to the positive side. This systematic uncertainty should be kept in mind for phrases with opinion values having a small statistical error and being close to zero.

A general problem of the approach results from the fact, that online reviews were used as a text resource to derive the opinion values. Thus, the vocabulary used in these reviews determines the content of the opinion list. This leads to the conclusion, that opinion lists produced using this approach may be suitable for analyzing user generated content from Web 2.0 sources but may not be applicable for other text resources.

To measure the quality of the list, we intent to perform benchmark tests using manually created opinion lists. Later we will perform also a quantitative evaluation.

For a later version of the list, we want to enrich it with values for opinion bearing phrases based on verbs.

5 Conclusion

In this paper we calculated opinion values for German adjective- and noun-based phrases with a fine granularity using the titles of Amazon reviews together with the star rating. Necessary corrections were applied.

It seems as if we got astonishingly good results for more than 1,700 single words and 1,400 phrases, which are to evaluate in detail. The list obtained will be made available to the community soon.

Acknowledgments

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-10)*, pages 2200–2204.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER

treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*, pages 24–41.

Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. 2009. Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement (TSA-09)*, pages 53–56.

Yejin Choi and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-08)*, pages 793–801.

Simon Clematide and Manfred Klenner. 2010. Evaluation and Extension of a Polarity Lexicon for German. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA-10)*, pages 7–13.

Andrea Esuli and Fabrizio Sebastiani. 2006a. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, pages 193–200.

Andrea Esuli and Fabrizio Sebastiani. 2006b. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-06)*, pages 417–422.

Nan Hu, Paul A. Pavlou, and Jennifer Zhang. 2007. Why do Online Product Reviews have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication. *Marketing Science*, 198:7.

Nan Hu, Jie Zhang, and Paul A. Pavlou. 2009. Overcoming the J-shaped Distribution of Product Reviews. *Communications of the ACM*, 52:144–147.

Manfred Klenner, Stefanos Petrakis, and Angela Fahrni. 2009. Robust Compositional Polarity Classification. In *In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-09)*, pages 180–184.

Jingjing Liu and Stephanie Seneff. 2009. Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*, pages 161–169.

Bing Liu and Lei Zhang. 2012. A Survey of Opinion Mining and Sentiment Analysis. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 415–463. Springer US.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion Observer: Analyzing and Comparing

Opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference (WWW-05)*, pages 342–351.

Karo Moilanen and Stephen Pulman. 2007. Sentiment Composition. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-07)*, pages 378–382.

Daniela Oelke, Ming Hao, Christian Rohrdantz, Daniel A. Keim, Umeshwar Dayal, Lars-Erik Haug, and Halldór Janetzko. 2009. Visual Opinion Analysis of Customer Feedback Data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST-09)*, pages 187–194.

Uwe Quasthoff. 1998. Projekt Deutscher Wortschatz. In Gerhard Heyer and Christian Wolff, editors, *Linguistik und neue Medien. Proc. 10. Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung.*, pages 93–99. DUV.

Sven Rill, Johannes Drescher, Dirk Reinel, Jörg Scheidt, Oliver Schütz, Florian Wogenstein, and Daniel Simon. 2012. A Generic Approach to Generate Opinion Lists of Phrases for Opinion Mining Applications. In *Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM-12)*.

Anne Schiller and Christine Thielen. 1995. Ein kleines und erweitertes Tagset fürs Deutsche. In *Tagungsberichte des Arbeitstreffens "Lexikon + Text", 17./18. Februar 1994, Schloß Hohentübingen*, Lexicographica Series Maior, pages 193–203. Niemeyer, Tübingen.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting Semantic Orientations of Words using Spin Model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-05)*, pages 133–140.

Huifeng Tang, Songbo Tan, and Xueqi Cheng. 2009. A survey on sentiment detection of reviews. *Expert Systems with Applications*, 36:10760–10773.

Ulli Waltinger. 2010. GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-10)*, pages 1638–1642.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Human Language Technology Conference (HLT-05)*, pages 347–354.