Use of Negation Markers in German Customer Reviews

Amelie I. Metzmacher[a], Verena Heinrichs[a], Björn Falk[a], Robert H. Schmitt[a]
[a] Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen, Germany
E-mail: A.Metzmacher@wzl.rwth-aachen.de, V.Heinrichs@wzl.rwth-aachen.de, B.Falk@wzl.rwth-aachen.de, R.Schmitt@wzl.rwth-aachen.de

***Abstract*—The research presented examines the use of negation markers in German customer reviews. The objective is to identify differences, as well as similarities in the use of language for reviews rating material products and services. Therefore, in an annotation study, negation markers and sentiment values of customer reviews rating these product categories had to be assigned. The results obtained confirm the hypothesis that customer reviews relating to services contain more negation markers than customer reviews rating material products. However, there exists no significant difference in token distances between the negation marker and the sentiment decisive part of speech (POS). Finally, the findings should be applied in a machine learning algorithm for extracting relevant information from German customer reviews.**

Keywords-Social Media; Customer Reviews; NLP; Sentiment Analysis.

I. Introduction

More than 80% of all customers express their experiences with products and services in social media [1]. Being publicly available, information provided by customer reviews in social media is important for both potential customers and companies. 80% of all potential customers base their purchase decision on the experiences of other customers trusting the judgment of strangers more than the recommendations of family and friends [1]. For companies, customer reviews provide insights with respect to not only the product, its functionalities or the service experience but also regarding the person who is the customer, his needs, wishes and persona. In analyzing customer feedback, companies are able to gain up-to-date and authentic knowledge about the product, the service and the customer.

The steadily growing number of customer reviews available in social media requires text mining and machine learning techniques such as sentiment analysis for a detailed understanding of the information provided by customer reviews. A major issue in sentiment analysis is the identification and handling of negation markers [2] - [6]. Negation markers cause valence shifts in customer reviews, e.g., shifting a positive statement into a negative statement and vice versa [7]. Since the use of negation markers is often very language-specific, language dependent approaches and algorithms are needed to analyze the sentiment of customer reviews correctly.

The German language shows a strong tendency for conventional indirectness by using syntactic downgraders. Syntactic downgraders modify the intended illocutionary act, i.e., the meaning conveyed, of the speaker towards the audience [8]. Examples of syntactic downgraders are modal verbs, tense or negation markers [9].

There exist different types of negation markers relating to different language levels. Examples given are firstly listed in the English translation and secondly in the German original word. On the morphological level, negation is expressed using distinct prefixes (e.g., “unhappy”, “unglücklich”) or suffixes (e.g., “senseless”, “sinnlos”). Discrete meaning bearing units in place of morphemes (e.g., “not”, “nicht”) represent the class of syntactic negations. In addition, the use of diminishers (e.g., “hardly”, “kaum”) negates the inherent polarity of an expression [2] [3] [9]. Negation markers either apply to words directly carrying a certain sentiment (local negation) or relate to words that do not carry a sentiment (long distance negation; indirect) [4].

Following Giora et al. [10], negation markers function as an instruction from a speaker to an audience to suppress the negated information. By using negation markers to suppress the negated information for the hearer, the speaker saves his face and remains polite when communicating a negative statement [11]. Politeness is a distinct feature of civilized societies and thus seen as an important social value guiding social interactions [12].

As opposed to evaluating a material product during application, the evaluation of a service includes the evaluation of the person of the service provider during service provision. The customer’s evaluation then applies directly to the professional and personal behavior of the service provider throughout social interaction.

Considering the aspect of politeness and keeping one`s face in social interaction, we hypothesize that customer reviews relating to services contain more negation markers than customer reviews rating material products. Finding significant differences might ameliorate algorithms for extracting relevant information out of social media content. In addition, it could help to identify automatically if a customer judges a product or a service. In particular, this would be of interest for companies providing both services and products.

To examine the above stated hypothesis, an annotation study was conducted. The database consisted of 3,767 German customer reviews in total extracted from different social media platforms. The customer reviews related to material products, as well as services. Three subjects were asked to annotate randomly chosen sentences of the reviews. The subjects' task was to annotate the sentiment of each given sentence and explicitly reference whether there is a negation marker present or not.

In the following, section 2 deals with related literature covering similar topics. In section 3, the annotation study is described, whereas in section 4 the results are analyzed and discussed. Finally, section 5 finishes this contribution with a brief conclusion and information on future working steps.

II. Related Work

Related work for the presented approach can be divided into at least two research areas: sentiment analysis as part of natural language processing and linguistic research analyzing customer reviews in social media with respect to material products and services. However, there are often overlaps, particularly if the objective has a clear focus on negations. Moreover, there are only very few contributions dealing with German customer reviews in general.

Wiegand et al. [13] give an overview about the role of negations, as well as about different approaches to include negations into sentiment analysis. They state that although integrating negations is very difficult, contributions dealing with this topic generally agreed on its high relevance for sentiment analysis. For instance, Kennedy and Inkpen [7] point out that considering the effects of valence shifters has a generally positive effect on all classification methods for reviews.

Furthermore, Asmi & Ishaya [5] integrate negation calculation rules into the general framework of a rule-based polarity classification. Therein, they defined these rules based on part of speech (POS). One main finding indicates that most negation words are classified as adverbs, suffixes, prefixes or verbs. Using this information, a dependency tree is developed. Its output is the scope of negation, which indicates how negation is interacting with other words in the sentence. Although, their approach already improved the polarity classification as there was a strong correlation between the classification results of the algorithm and those of humans, the authors point out the importance of additionally implementing prepositional negations.

In a rather current contribution, Diamantini et al. [3] apply a dependency-based parse tree to investigate the scope of negation. Implementing the negation handling component just before the sentiment calculation, meaning after all other pre-processing steps have been conducted, increases the accuracy from 64.4% to 67%. In their approach, they also regard a three-class-problem as the sentiment is distinguished between negative, positive and neutral. The authors hypothesize that those samples calculated wrong imply irony. Thus, as a necessity for future work, they suggest to extend their system to consider effects of irony.

In summary, independently from the chosen method, in most cases integrating a negation model should ameliorate the accuracy of the sentiment analysis. Concerning the use of negations in German, Wiegand et al. [13] point out the requirement for more complex processing as the negated expression either precedes or follows the actual statement. Therefore, not all findings discovered e.g., in English texts, can be applied in German texts, at least not without adjustments.

As Wiegand et al. [13] point out, not all negations indicate a negative sentiment. Thus, it is important to use syntactic knowledge and regard the context. Most approaches dealing with sentiment analysis, use reviews, which are not domain-independent, e.g., a collection of reviews of several products found on google.com or movie reviews [14]. Within the group working with corpora built from movie reviews, the classification usually follows the star ratings of the authors of these reviews [15].

Although, there are some contributions, which conducted annotation studies to produce corpora from German customer reviews, they usually aim at sentiment analysis in general and do not consider the usage of negation markers. Moreover, many do not address different domains, and in particular services, as well as material goods. For instance, Boland et al. [16] conducted a study, which focuses on different domains, but does not address the use of negations.

In summary, no study has yet been conducted in which a text corpus of domain-independent customer reviews in German is annotated with regard to sentiment and particularly negations.

A prior study by the authors indicates an influence of personal commitment on customers' writing styles while formulating a product review. This is particularly the case while rating services. On the one hand, it seems that the writing becomes more precise [17]. On the other, one might argue that the human interaction required in services leads to more polite formulations.

Thus, based on the literature review and this prior study, an investigation of the amount of negations in services compared to material products is intended. Thereby, we look for hints for the application of a more polite form of criticism within German service reviews.

III. Annotation Study

The objective of the study was to examine whether customer reviews relating to services contain more negation markers than customer reviews rating material products. The products selected are accessible to the German end-consumer. The two classes contain three product types each. For products a shoe, a hazelnut spread, and a smartphone were chosen, whereas the services contained a hotel, a financial service for online businesses, and a car service station with several stations across Germany.

Altogether 38 different social media platforms, including German discussion forums and shopping sites with user comments, were chosen as data sources. To this end, 3,767 German customer reviews relating to both, material products and services, have been extracted with the open-source Java library jsoup [18]. Prior to annotation, the reviews were parsed into single sentences using the Stanford Parser [19]. The annotation was carried out on sentence level. 1,200

 ISBN: 978-1-61208-600-2

sentences, 600 for products and services each, were randomly chosen for annotation and annotated by three subjects. The subjects were German native-speakers and familiar with the process of annotating. Each subject had to annotate 200 sentences, whereby each sentence was annotated by three subjects.

Subjects were asked to identify the sentiment of each sentence while assigning the POS, which induces the negativity, positivity or neutrality of the given statements, i.e., the level of sentiment. Herein, the opinionated words are called attributes. Moreover, the subjects were asked to determine negation markers if present in the sentence, e.g., mark the indefinite pronoun "no" ("kein") or the particle "not" ("nicht"). In addition, the negation markers were assigned to the attribute the negation is associated with. For instance, the subjects had to indicate that the negation "not" ("nicht") is associated with the attribute "good" ("gut"). Thereby, it was possible to mark more than one attribute, aspect and/or negation marker per sentence, e.g., if a conjunction was present.

The annotation process was explained to the subjects with three exemplary sentences. The sentences were chosen to show different characteristics, which influence the grade of simplicity or complexity of identifying the sentiment of a sentence and its possible negation marker. For instance, one sample contained two attributes in one sentences ("The shoe looks nice, but is too heavy") or another one included only an implicit product review, meaning an attribute without an aspect ("Too heavy."). The examples ensure that the annotation process was carried out consistently.

IV. RESULTS AND DISCUSSION

First, the interrater reliability was investigated. Fleiss' Kappa values for assigning aspects, attributes, sentiment, and negation markers between 0.5 and 0.8 are located within a moderate level of agreement [20].

The observed frequencies of the labelled negation markers were displayed and analyzed for material products and services. The frequencies of negation markers were computed based on statements within a sentence, i.e., based on attributes within a sentence.

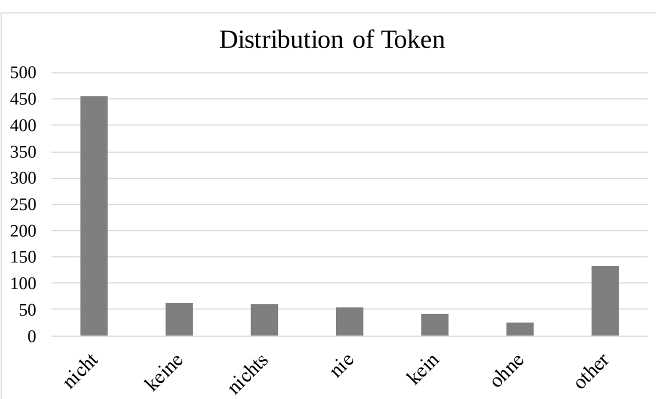

Figure 1: Distribution of token used as negation markers on sentence level

Figure 1 depicts the most used negation markers for both categories. The most frequent negation marker is the participle "not" ("nicht"). In comparison with the other frequently appearing negation markers, the participle "not" ("nicht") is not as harsh as the negation marker "any" ("keine"), "nothing" ("nichts") or "never" ("nie"), for the latter have a clear excluding character. The last item "other" includes all other negation markers, which only appear rarely. However, comparing the distribution of negation markers between material products and services, no significant differences were found.

TABLE 1: MEAN TOKEN DISTANCE BETWEEN ATTRIBUTE AND NEGATION MARKER

Product Categories	**Mean Token Distance**
Services	1,7918
Material Products	1,9234
All	1,8576

In the following, the mean token distance between attribute and negation marker was examined (see TABLE 1). The results show that the distance between attribute and negation marker tends to be shorter within customer reviews relating to services than within customer reviews rating material products. However, the Kruskal-Wallis-Test

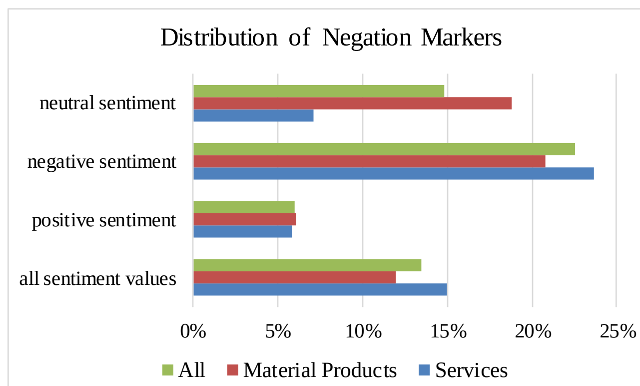

Figure 2: Distribution of negation markers

revealed that with a p-value > 0.05 the token distances are equally distributed. Therefore, the differences in the token distances are not significant.

Third, the frequency of negation marker use was examined between both categories. Figure 2 illustrates the percentages of negation markers used within the different product categories, as well as negation marker used in total. The use of negation markers is displayed with respect to the sentiment value of the sentence. On average, each sentence contained more than 1.7 marked statements respectively attributes evaluating (a part of) a product. Approximately 52% statements were assigned as positive, 41% as negative, and 7% as neutral. The values indicate a difference between the use of negation markers in customer reviews for material products and services. Regarding the negation markers used within all sentiment values, as well as negation markers used

 ISBN: 978-1-61208-600-2

within negative sentiment, customer reviews relating to services contain more negation markers. The results of the distributions also reveal that there seem to be differences in the use of negation markers between customer reviews about services and material products in general. However, a significance test was applied to test whether differences across sentiment values are significant.

For applying an appropriate significance test, we assume that the decisions, whether to use a negation marker associated with an attribute or not, are independent from one another and are the result of a Bernoulli distribution as the decision is either “yes” or “no”. This includes a distributed random variable with success probability p, where p in {p_service, p_materialproduct} may or may not be different from customer reviews about services and those about material products. We test the null hypothesis, that p_service = p_materialproduct, with a binomial test using the statistical software R.

For the test, we consider the binomial distributed variable X_service that counts the number of negation markers used in the attributes in service reviews, and construct a confidence interval on a confidence level of 98% based on the annotated customer reviews. The confidence interval consists of the 0.01- and 0.99-quantiles of the distribution of X_service ~ B(p_service, n_materialproduct).

As a result, we observe, that p_materialproduct*n_ materialproduct is not located inside the confidence interval. We repeat the test for X_materialproduct ~ B(p_materialproduct, n_service) and find, that p_service*n_service is outside the obtained confidence interval.

Therefore, we postulate with a certainty of 98% that p_service != p_materialproduct. Concomitantly, the p-value <0.05 states that a significant difference between these two categories exists. As a consequence, we can confirm our hypothesis that customer reviews relating to services contain more negation markers than customer reviews rating material products.

Additionally, regarding the distributions of negations within neutral sentiment values (see Figure 2), there is a rather large difference in frequency recognizable. Although only 7% of the statements were marked as neutral, we conducted the significance test in the same way as for all sentiment values, but only for neutral sentiment. We receive a p-value <0.05 stating that there exist a significant difference. As an explanation, one might argue that customers judge products in a much less euphoric way than services. Thus, if the sentiment is neutral, for products, people might use formulations like “not so bad”. In contrast, for services they preferably use e.g., “okay”. However, as these are assumptions, it is necessary to examine these in more detail.

V. CONCLUSION

The results obtained in our study confirm our hypothesis about a difference in the use of negation markers in customer reviews rating services compared to customer reviews rating material products.

Human social interaction, as well as the personal commitment towards the person providing the service leads to a more polite writing style. When rating services customers rate the executing individuals and thus, are more moderate and polite in their judgement using negation markers instead of words containing a negative polarity value on a lexical basis. To prove this concept, in our future work we aim to investigate this assumption in more detail.

As our analysis also showed a difference in the use of negation markers in neutral sentiment, but with a reverse distribution, it would be interesting to examine these findings in detail as well. However, as neutral sentiment seems to be not that numerous in customer reviews, a special corpus for this issue needs to be compiled.

In addition, we examined the mean token distance between the attribute and the associated negation marker within the two product categories. In contrast to other features of language use, significant differences in the use of language could not be proven here. However, the mean token distance could still be a useful input variable for sentiment analysis of German customer reviews.

Generally, we strive to use our findings in the analysis of complaints from German customer reviews. In our future work, we aim to filter relevant information about products or services. If a company provides as well services as products, it would be very beneficial to identify automatically if customers speak about the product or about the service. Thus, the information could be allocated directly towards the right product type.

ACKNOWLEDGMENT

The support of the German National Science Foundation (DFG) through the funding of the research project “Entwicklung eines Sensors für ungerichtete Beschwerden aus Online Foren” (SCHM 1856/69-1) is gratefully acknowledged.

REFERENCES

[1] HolidayCheck Group, “Psychologie des Bewertens”, 2016. [Online] Available from: https://www.holidaycheckgroup.com/wp-content/uploads/2016/07/TOMORROW-FOCUS-AG_Umfrage_Die-Psychologie_des_Bewertens_Pr%C3%A4sentation-1.pdf. [Retrieved: September, 2017].

[2] U. Farooq, H. Mansoor, A. Nongaillard, Y. Ouzrout, and M. A. Qadir, 2017, “Negation Handling in Sentiment Analysis at Sentence Level”, JCP, 12(5), 2017, pp. 470-478.

[3] C. Diamantini, A. Mircoli, and D. Potena, “A Negation Handling Technique for Sentiment Analysis”. In Collaboration Technologies and Systems (CTS), 2016 International Conference on, IEEE, October, 2016, pp. 188-195.

[4] A. Hogenboom, P. Van Iterson, B. Heerschop, F. Frasincar, and U. Kaymak,, “Determining negation scope and strength in sentiment analysis”. In Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, October, 2011, pp. 2589-2594.

[5] A. Asmi and T. Ishaya, “Negation identification and calculation in sentiment analysis”. In The Second International Conference on Advances in Information Mining and Management, 2012, pp. 1-7.

 ISBN: 978-1-61208-600-2

[6] I.G. Councill, R. McDonald, and L. Velikovich, “What's great and what's not: learning to classify the scope of negation for improved sentiment analysis”. In Proceedings of the workshop on negation and speculation in natural language processing. Association for Computational Linguistics, July, 2010, pp. 51-59.

[7] A. Kennedy and D. Inkpen, “Sentiment classification of movie reviews using contextual valence shifters”. Computational intelligence, 22(2), 2006, pp. 110-125.

[8] E. Ogiermann, 2009, “Politeness and in-directness across cultures: A comparison of English, German, Polish and Russian requests.” Journal of Politeness Research, 5, 2009, pp. 189-216.

[9] Stickel, G., “Untersuchungen zur Negation im heutigen Deutsch”, Vol. 1, Springer-Verlag, 2013.

[10] R. Giora, N. Balaban, O. Fein, and I. Alkabets, “Negation as positivity in disguise”. Figurative language comprehension: Social and cultural influences, 2005, pp. 233-258.

[11] E. Goffman, “Interaction ritual: Essays in face to face behavior.” AldineTransaction, 2005.

[12] S. Rogatcheva, 2004, “Politeness in English and German: a contrastive study” GRIN Verlag.

[13] M. Wiegand, A. Balahur, B. Roth, D. Klakow, and A. Montoyo, “A survey on the role of negation in sentiment analysis”. In Proceedings of the workshop on negation and speculation in natural language processing, Association for Computational Linguistics, July, 2010, pp.60-68.

[14] G. Vinodhini and R. M. Chandrasekaran, “Sentiment analysis and opinion mining: a survey”. International Journal, 2(6), 2012, pp. 282-292.

[15] B. Pang and L. Lee, “Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales”. In Proceedings of the 43rd annual meeting on association for computational linguistics, Association for Computational Linguistics, June, 2005, pp. 115-124.

[16] K. Boland, A. Wira-Alam, and R. Messerschmidt, “Creating an annotated corpus for sentiment analysis of german product reviews.” GESIS - Leibniz-Institut für Sozialwissenschaften. Mannheim, GESISTechnical Reports 2013/05, May, 2013.

[17] A. I. Metzmacher, V. Heinrichs, B. Falk, and R. H. Schmitt, “Customer Language Processing“. In Proceedings of the 12th International Workshop on Semantic and Social Media Adaptation and Personalization, IEEE, July, 2017, pp.32-33.

[18] J. Hedley “jsoup: Java HTML Parser”, 2009. [Online] Available from: https://jsoup.org/ [Retrieved: September, 2017].

[19] A. N. Rafferty and C. D. Manning, “Parsing three German treebanks: Lexicalized and unlexicalized baselines”. In Proceedings of the Workshop on Parsing German. Association for Computational Linguistics, June 2008, pp.40-46.

[20] M. L. McHugh, “Interrater reliability: the kappa statistic”, Biochemia medica, 22(3), 2012, pp. 276-282.