How Much Sentiment Can Be Carried by Modal Particles?
- A Comparative Analysis Using BERT, Hugging Face, and FastText Models

Hyeyeon Park
Term Paper for ”Deep Learning in NLP”
Summer Semester 2024

15. 4. 2025

1 Abstract

This paper explores the emotional tendencies of German modal particles based on real-world sentence data. Building on a previous study that compared modal particles with expressive emojis through a survey, this research uses sentiment analysis to calculate average emotional values for selected particles (ja, doch, schon, etc.). Three models were tested to compare their performance. The goal of this study is to quantify the sentiment carried by German modal particles and examine to what extent they encode affective meaning. By comparing three sentiment analysis models(FastText, Hugging Face pipeline, and BERT), this research investigates whether similar sentimental patterns are detected across models, which model responds most sensitively to subtle sentiment shifts, and which model's output aligns most closely with human evaluation from a previous study. The findings are intended to inform future strategies for mapping sentiment-driven features in multilingual translation contexts.

2 Introduction

2.1 Background

Modal particles are small words in German that do not change the literal meaning of a sentence but instead influence its subtle nuances and convey emotions or attitudes. They are frequently used in everyday German speech, but according to Möllering (2001)[4], they can be challenging for non-native speakers due to their homonyms. In some cases, German modal particles cannot even be directly translated into other languages, which adds to the confusion for German language learners. Therefore, it may be helpful to examine how much emotional meaning these particles can carry.

This study builds on my previous experiment, conducted as part of the term paper for the course ”The Meaning of Gestures (and more)” during the Winter Semester 2023–4[6], in which the emotional impact of modal particles was analyzed through a questionnaire-based survey. I explored the limitations of translating German modal particles into Korean and proposed the use of expressive emojis to retain emotional nuances. In this study, I tried to analyze the emotional tendencies of modal particles directly through sentiment scores based on real-world German texts. In my previous experiment, particles such as ”ja”, ”doch”, and ”schon” were examined separately in terms of their use in positive or negative contexts. In contrast, in this study I would like to calculate the average emotional value of each modal particle based on real-world sentence data, without dividing them by polarity. The goal is to explore the overall emotional tendency of modal particles in practice by comparing three sentiment analysis models.

2.2 Previous Study

In my previous study [6], I selected ten commonly used modal particles in German: *mal*, *ruhig*, *doch*, *halt*, *eben*, *nun einmal*, *eigentlich*, *wirklich*, *ja*, and *schon*.

These particles were categorized based on their typical emotional or pragmatic functions in my previous study [6]:

- *halt*, *eben*: impatience, frustration
- *eigentlich*, *wirklich*: skepticism, doubt, surprise, disbelief
- *mal*, *ja*: acceptance, permission, politeness
- *ruhig*, *nun einmal*: permission or obligation
- *doch*, *schon*: impatience or permission

Each particle was then evaluated using a five-point scale ranging from very negative to very positive. Based on this scale, a survey was conducted in which participants rated the emotional tone of each particle according to its perceived sentiment. The results are summarized in Table 1. Please refer to the appendix for the full text, which includes the detailed calculation method.

Table 1: Sentiment Scores of Modal Particles from Previous Study

Modal Particle (Context)	**Sentiment Score**
halt	-0.1571
eben	-0.0143
eigentlich	-0.1214
wirklich	0.1929
ruhig	0.6071
nun einmal	0.6357
ja (when positive)	0.8710
mal (when *ja* positive)	0.7147
ja (when negative)	-0.1268
mal (when *ja* negative)	-0.5018
schon (when positive)	0.1587
doch (when *schon* positive)	0.4921
schon (when negative)	-0.1050
doch (when *schon* negative)	0.1450

Note: Survey conducted as part of a previous study [6]. 2024.02.07-13, $N = 35$.

2.3 Models for Sentiment Analysis

I selected 3 differnent language models(see: Table 2) for comparing the sentiment analysis' scores, while they can show different performances for analysing subtle sentiments of modal particles.

Table 2: Comparison of the Three Sentiment Analysis Models Used in This Study

Aspect	**Hugging Face Sentiment BERT**	**Multilingual BERT (Generic)**	**FastText + Custom Classifier**
Training Data	German sentiment-labeled data, mostly from reviews	Mixed-domain multilingual data including reviews, news, and general texts	Pretrained word vectors; classifier trained using labels from this study's dataset
Language Focus	German only	Multilingual	Supports German words, but embedding is shallow and not language-specific
Task Specialization	Fine-tuned for sentiment classification	General-purpose classification model that includes sentiment tasks	Requires separate classifier training
Context Sensitivity	Likely to capture nuance and emotion well due to review-based training	Less sensitive to subtle emotional cues	No context awareness, predictions are based on word-level averages only

Note: Information about model is based on the official model documentation and relevant papers [1, 2, 5].

According to Devlin et al.(2018)[1], ”BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers.” This allows the BERT model to be aware of context in sentiment analysis tasks. However, the BERT model is not specialized in German sentiment analysis. It is trained based on multilingual texts, while it can be limited in analysing subtle sentiments.

The Hugging Face model[2], by contrast, is ”fine-tuned specifically for sentiment classification on German texts.” Therefore, I compare it with the Hugging Face BERT model, which is a model trained to be more sensitive to German sentiment changes.

FastText, on the other hand, ”offers a lightweight embedding-based approach that enables fast computation but with limited contextual sensitivity.” Joulin et al.(2017)[3] suggests that ”FastText does not use any word ordering or structure, but instead represents a document by averaging word vectors." Thus, it can be limited in performance in sentiment analysis tasks, however, comparing the performance of these three models would be still meaningful.

By including these three models, this study aims to evaluate whether nuanced emotional expressions, such as those conveyed by modal particles, are equally detectable across different types of sentiment analysis models. This also enables a comparison between transformer-based and embedding-based architectures in the specific context of German sentiment detection.

3 Methods

The Hugging Face model used in this study [2] is based on the BERT architecture [1], but has already been pre-trained and fine-tuned for sentiment analysis on large German datasets. Since the model was specifically trained for this task, I was able to use it directly without additional fine-tuning.

FastText provides word embeddings at the word level, but does not generate sentence embeddings by default. In this study, I followed the common practice of averaging the word vectors to obtain a sentence-level representation [3]. Given a sentence with n tokens and their corresponding word vectors $w_1, ..., w_n$, the sentence embedding s is computed as:

$$s = \frac{1}{n} \sum_{i=1}^{n} w_i$$

This embedding was then used as input to a logistic regression classifier, following the standard linear classification setup:

$$\hat{y} = \sigma(Ws + b)$$

where W and b are learned parameters, and σ is the sigmoid activation function. The model was trained on labeled sentiment data from the Leipzig corpus.

3.1 Dataset

For this study, I used the *Filmstarts* dataset, which consists of German movie reviews annotated with sentiment scores. Unlike other available German corpora such as the Leipzig corpus—which lacks explicit sentiment annotations—the Filmstarts dataset includes a wide range of emotional expressions, making it more suitable for sentiment analysis.

Since the focus of this research is on analyzing sentiment shifts caused by modal particles, a dataset that reflects emotional variation was essential. Therefore, Filmstarts dataset was selected as the sole source of both training and evaluation data because it has rich emotional expressions as samples.

This dataset was previously used in the work of Guhr et al. [2], and is a part of the training resources for the BERT-based sentiment model available on Hugging Face. As it is already labeled and preprocessed, it was particularly well-suited for direct use in this experiment without requiring additional annotation or data cleaning. The dataset contains more than 70,000 samples, which is very large, so I divided the dataset into four chunks. For this analysis, I only used the first chunk, collecting 100 samples for each modal particle and extracting the scores.

3.2 Method for Filtering Modal Particles

In order to analyze the emotional contribution of each modal particle, I firstly made a list of ten commonly used modal particles in German based on the previous study: *halt*, *eben*, *eigentlich*, *wirklich*, *mal*, *ja*, *ruhig*, *nun einmal*, *doch*, and *schon*.

Identifying modal particles can be tricky due to the presence of homophones. Initially, I tried identifying modal particles by simple string matchings. But this approach quickly showed its limits, while it should be possible that the word actually not functioning as a modal particle.

To improve accuracy, I developed a lightweight filtering process as follows.

- **Grammatical awkwardness**: Does the sentence become grammatically awkward when the candidate word is removed?
- **POS Tags**: Does the word match typical POS tags associated with modal particles?
  - `PART` (particle): Words that don't carry semantic meaning on their own but influence the tone or mood
  - `ADV` (adverb): Often used for modal nuances (e.g., *ja*, *doch*)
  - `INTJ` (interjection): For particles that convey emotional emphasis.
- **Contextual similarity**: Is the word semantically consistent with its surrounding context, based on cosine similarity from BERT embeddings?

Using this 3-step method, I built a comparison set of sentences where modal particles are more likely to be used correctly. It was possible to collect 100 sample data for each of 9 modal particles out of a total of 10 modal particles, excluding 'nun einmal'. Later in the Results section, I compare the sentiment analysis results between this method and the basic string-matching approach, to see if this extra filtering makes a real difference.

3.3 Method for Extracting Sentiment Scores of Modal Particles

For each modal particle, 100 sentences from the data set that contained the particle are sampled. In each case, the sentiment score of the original sentence is calculated using three different models, BERT, Hugging face-BERT and FastText(score a). Then, I removed the modal particle from the sentence and recalculated the sentiment score(score b).

To quantify the sentiment impact of the modal particle itself(Δs), the original score(score a) is subtracted from the modified sentences' score(score b):

$$\Delta s = b - a$$

Δs represents the average change in sentiment when the particle is removed. A positive value indicates that the particle tends to decrease the sentiment score, and a negative value suggests that the particle adds positive emotional value.

4 Results

The results obtained for each modal particle are presented below.

Table 3: Sentiment Delta Scores After Removing Modal Particles

Particle	**HuggingFace**	**Multilingual BERT**	**FastText**
doch	1.98	-0.42	-0.15
schon	2.06	0.51	0.69
mal	1.60	1.32	-0.38
eigentlich	-2.10	0.49	-0.16
halt	1.88	1.58	-0.34
eben	1.86	-1.41	0.84
ja	2.36	-0.36	0.71
wirklich	-2.44	-1.70	-0.38
ruhig	1.42	-0.63	-0.13

These scores are also compared with the average scores from my previous study based on a questionnaire survey in Figure 1.

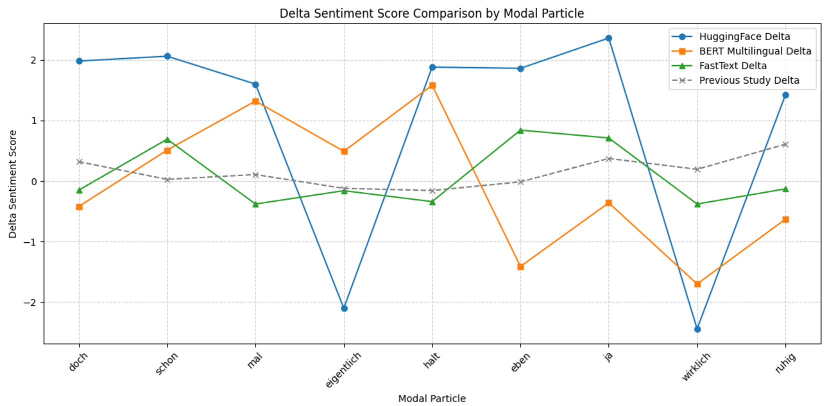

Figure 1: Comparison of Sentiment Score Changes (Δ) by Modal Particle across Models and Previous Study

Table 4 shows the comparison of modal particles' scores retrieved from my last experiment based on survey and delta scores from three different language models.

Table 4: Comparison of Delta Scores and Previous Study Scores by Modal Particle

Modal Particle	**HF**	**BERT**	**FastText**	**Previous study**
doch	1.98	-0.42	-0.15	0.3185
schon	2.06	0.51	0.69	0.0269
mal	1.60	1.32	-0.38	0.1065
eigentlich	-2.10	0.49	-0.16	-0.1214
halt	1.88	1.58	-0.34	-0.1571
eben	1.86	-1.41	0.84	-0.0143
ja	2.36	-0.36	0.71	0.3721
wirklich	-2.44	-1.70	-0.38	0.1929
ruhig	1.42	-0.63	-0.13	0.6071

What can be seen from the table is that, first of all, the sentiment scores extracted from the previous study are very conservative, and FastText model's conservative sentiment scores are closest to the sentiment scores of this previous study, even more than those of the Hugging Face or BERT models. The Hugging Face model allows us to clearly see the emotional differences brought about by modal particles, but the emotional scores of this model may be somewhat overestimated.

5 Conclusion

This study examined how German modal particles influence sentiment in sentences using three different sentiment analysis models. The goal was to explore whether the subtle emotional effects of these particles, which are often lost in translation, can be measured in a way that helps preserve their originally intended nuances.

Basically, this experiment was inspired by the idea that sentiment values could be matched with expressive elements such as emojis. If the emotional contribution of a modal particle can be quantified, then in cases where direct translation is difficult, such as in German-Korean, emojis or similar symbols could serve as tools to retain the speaker's intended tone.

Interestingly, the results showed significant differences across models. The Hugging Face model produced large sentiment shifts when modal particles were removed, suggesting a strong sensitivity to emotional cues (though possibly exaggerated). FastText, by contrast, showed more conservative changes (closer to the previous study's human-evaluated scores).

These findings suggest a potential approach: even if sentiment scores are exaggerated, they still provide directional information. By aligning these scores with emoji sentiment ranges and adjusting them accordingly, it may be possible to develop a translation method that captures the emotional nuance of modal particles. This could enhance machine translation outputs, especially when translating into languages like Korean, where one-to-one equivalents for such modal particles often do not exist.

Future work could explore how to adjust these scores and map them to expressive indicators such as emojis. This might help improve emotionally aware translations across different languages.

6 Discussion

6.1 Performance Limitations of FastText

There are a few interesting points that came up during this experiment. First of all, the FastText model showed very low sensitivity in sentiment change. Since FastText only works with word-level embeddings and doesn't really understand the context or tone of a sentence, it's not surprising that the sentiment scores barely shifted even when a modal particle was removed. In that sense, FastText seems too simplistic for this kind of subtle nuanced task.

Table 5: Classification Report of FastText Classifier on Filmstarts Sample (Test Size = 200)

Class	Precision	Recall	F1-Score	Support
0.0	0.00	0.00	0.00	8
1.0	0.00	0.00	0.00	13
2.0	0.00	0.00	0.00	23
3.0	0.36	0.09	0.14	47
4.0	0.31	0.70	0.43	60
5.0	0.25	0.29	0.27	49
Accuracy		0.30 (out of 200)		
Macro Avg	0.16	0.18	0.14	200
Weighted Avg	0.24	0.30	0.23	200

Table 5 shows the performance of the FastText classifier. The lack of samples in lower sentiment ranges (0.0 to 2.0) likely affected the model's output, making it difficult to determine whether the conservative delta scores are due to the model's limitations or simply a reflection of the dataset's imbalance. It is possible that FastText was more sensitive to the dominant sentiment distribution in the sample rather than underperforming in itself.

6.2 Possible Overestimation in Hugging Face Model's Delta Scores

On the other hand, the Hugging Face model showed significant differences in sentiment before and after removing modal particles. This might suggest that it's quite sensitive to emotional cues, but at the same time, the magnitude of the changes seems a bit too strong. It's possible that because this model was trained on review-based data, where emotions tend to be more explicit, it overreacts to certain phrases or expressions, especially when it comes to modal particles. This raises questions about how much of that shift is truly reflecting emotional change, and how much is just noise or exaggeration from the model.

6.3 Limitations of the Current Study

One another limitation of this study is that only one chunk of the Filmstarts dataset was analyzed due to time and computational constraints. Although the sample size was large enough to perform meaningful analysis, using the full dataset would likely give a more robust picture. Also, the sentences were sampled randomly, which means some particles may have been over- or under-represented depending on their frequency in the dataset.

Despite these limitations, the results still offer useful insight into how different models interpret emotional nuance in German modal particles. In the future research, it might be interesting to study how to better adjust overly sensitive models, or how to incorporate these scores into practical machine translation systems, such as ones that use emojis to reflect tone of the source texts.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[2] Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. Training a broad-coverage German sentiment classification model for dialog systems. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 1627–1632, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://aclanthology.org/2020.lrec-1.202.

[3] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759, 2017.

[4] M. Mollering. Teaching german modal particles: A corpus-based approach. In Language Learning Technology, 5(3), pages 130–151, 2001. URL https://dx.doi.org/10125/44570.

[5] NLPTown. Multilingual bert for sentiment analysis. https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment, 2021.

[6] Hyeyeon Park. Modalpartikeln: Can sentiments survive translation with emojis? Unpublished term paper, The Meaning of Gestures (and more), Winter Semester 2023–4, 2024. URL https://github.com/bakeyeon/Modalparticles_Emojis.

Appendix

A Previous Experiment

- `Modalpartikeln_Can_Sentiments_Survive_Translation_with_Emojis.pdf`: Full Text of the previous paper.
- `Modalpartikeln_and_Emojis.R`: Code for calculating scores for modal particles based on the surveys.
- `Modalpartikeln.csv`: Raw result of modal particle sentiment survey.

B Python Code

- `modal_particle_final_chunk0.json`: Chunk with filtered samples only containing modal particles.
- `dataset_chunk.ipynb`: code for generating dataset chunks. https://colab.research.google.com/drive/1zmQJ-rClkL8jK9RBthElISOklkUagoxT?usp=sharing
- `modalparticle_final.ipynb`: code for calculating delta scores. https://colab.research.google.com/drive/18g3R6Rc-4S4MutUHBiL_Hlmpt_E-6mBm?usp=sharing

Acknowledgements

This paper was written as a term project for the course Deep Learning in NLP (SoSe 2024). For technical assistance during the development of this research, resources such as OpenAI's ChatGPT, the Scikit-learn documentation, and Stack Overflow were consulted, particularly for debugging and improving code implementations. Additionally, I benefited from taking several KAIST online courses on Deep Learning and Natural Language Processing through K-MOOC (https://www.kmooc.kr/).