Style-sensitive Sentence Embeddings for Evaluating Similarity in Speech Style of Japanese Sentences by Contrastive Learning

Yuki Zenimoto   Shinzan Komata   Takehito Utsuro
Degree Programs in Systems and Information Engineering,
Graduate School of Science and Technology, University of Tsukuba
{s2220753, s2320742}_@_u.tsukuba.ac.jp
utsuro_@_iit.tsukuba.ac.jp

Abstract

Since dialogue systems are required to keep its speech style consistency, evaluating the similarity of speech styles is an important task. However, the Japanese language has a wide variety of speech styles, and also for each speech style, huge variety of vocabulary and word usage characteristics are observed, making it difficult to evaluate the similarity of speech styles. This study proposes a speech style embedding model that produces a style-sensitive sentence embedding of Japanese sentences. The speech style embedding model is constructed by fine-tuning a pre-trained BERT model. Here, sentence pairs with similar/dissimilar speech styles are automatically collected on a large scale using a sequence of sentences in web novels, with which constrastive learning is performed for the training of the speech style embedding model. Using the Ward's hierarchical clustering method, we also analyze the clusters of speech styles and the characteristic vocabulary/word usage of each speech style. Finally, we focus on the variation in speech styles of each person depending on the situation, and further analyze the variation in style-sensitive embeddings of each character in the novel.

1 Introduction

The speech style of a dialogue system plays a crucial role in user interaction, and dialogue systems are expected to keep its speech style consistency (Zhou et al., 2020). Therefore, a mechanism for evaluating the similarity of speech styles across entire utterances is necessary. However, the Japanese language has a diverse range of speech styles (Kinsui, 2003; Akama et al., 2018), and a vast variety of characteristic expressions exist for each speech style, making it difficult to evaluate the similarity of speech styles across entire utterances.

In this study, we propose a simple speech style sentence embedding method, which can produce embeddings capable of evaluating the similarity of speech styles by fine-tuning a pre-trained BERT model (Devlin et al., 2019) using contrastive learning (Gao et al., 2021). The training dataset consists of positive instances collected from pairs of utterances estimated to be by an identical character and negative instances from pairs of utterances estimated to be by distinct characters. This data collection method is based on the observation that the utterances by an identical character have a consistent speech style, whereas those from different characters have distinct speech styles. Through this training, it is expected that the embeddings can be obtained for evaluating the similarity of speech styles rather than the content similarity of sentences. Furthermore, we revealed characteristic words of various speech styles through unsupervised clustering of style-sensitive sentence embeddings. Finally, we focused on utterances by specific characters within a novel and conducted an analysis of variations in the speech styles of an identical character.

The results contribute to confirming the following insights:

1. The consecutive utterances in novels are effective in training the embeddings of speech style by the proposed approach using contrastive learning.
2. The style-sensitive sentence embeddings correctly capture various speech styles including those representing politeness, gender, and typical fictional character.
3. Even for an identical character, the speech style can vary significantly depending on conversation partners and surrounding situations, where this variation in speech styles constitutes the characteristic of the character.

Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop, pages 32–39
November 1–4, 2023.

2 Related Work

Numerous prior studies have targeted English speech styles, investigating various aspects such as politeness (Rao and Tetreault, 2018; Danescu-Niculescu-Mizil et al., 2013) and sarcasm (Khodak et al., 2018). Previous research has been conducted on a diverse range of speech styles, and datasets have also been constructed. In contrast to previous research, Kang and Hovy (2021) proposed a novel approach to comprehensively grasping the phenomenon of cross-stylistic language variation. The primary emphasis lies in examining the interdependence of diverse styles within written text, elucidating the interplay between these styles, and systematically deconstructing their composition in text generation.

Japanese speech styles rely on various characteristics of the entire sentence, such as first-person pronouns and sentence-ending expressions (Kinsui, 2003; Matsuyoshi et al., 2006; Miyazaki et al., 2016). Consequently, it is desirable to evaluate the similarity of speech styles across an entire sentence. Akama et al. (2018) proposed an unsupervised learning method of style-sensitive word vector that evaluates the similarity of speech styles. However, Akama's method focuses on the speech style of individual words and does not evaluate the similarity of speech styles across an entire sentence. To evaluate the similarity of speech styles across an entire sentence, Miyazaki et al. (2021a); Zenimoto and Utsuro (2022) proposed a method that utilizes a speech style classification model built with training data containing sentences in specific speech styles. However, these methods are not adaptable to the speech styles of unknown characters. In contrast, our method can be adapted to the speech styles of unknown characters. Furthermore, prior researches utilize undisclosed datasets for training and evaluation, making it difficult to conduct comparative experiments.

3 Japanese Speech Style

This section describes the characteristics of Japanese speech styles, which the proposed model are expected to capture. In the Japanese language, first-person pronouns, particles, and auxiliary verbs in utterances differ depending on the speaker's trait, such as gender, age, and role (Kinsui, 2003; Matsuyoshi et al., 2006; Miyazaki et al., 2016). For example, the utterance "俺はこれが好きだぜ (I like it)" is reminiscent of a masculine person because the characteristic words "俺 (I)" and "だぜ (*daze*)" are used predominantly by males. In contrast, the utterance "私はこれが好きですわ (I like it)", which has the same meaning as the utterance "俺はこれが好きだぜ (I like it)", is reminiscent of a female person because the characteristic words "私 (I)" and "ですわ (*desuwa*)" are used predominantly by females. As mentioned above, the Japanese language has numerous expressions that, while conveying the same meaning, are reminiscent of significantly different persons. The proposed model is expected to capture the speech style of utterances that elicit such specific personal associations.

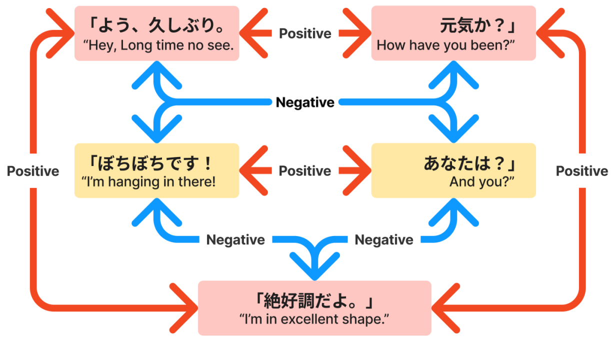

Figure 1: Procedures of Creating Positive and Negative Instances from Consecutive Utterances ("Positive" indicates the pairs of utterances estimated to be by an identical character. "Negative" indicates the pairs of utterances estimated to be by distinct characters.)

Data type	Positive	Negative	Total
Training	51.3M	53.2M	105M
Validation	259K	269K	528K
Test	259K	269K	528K
Total	51.8M	53.8M	105M

Table 1: Statistics of Experimental Data

4 Dataset Construction

4.1 The Procedure

For the training of speech style embedding model using contrastive learning, it is required to prepare pairs of sentences with similar speech styles as positive instances and pairs of sentences with dissimilar speech styles as negative instances. This study proposes a method to automatically collect a large amount of positive and negative instances from consecutive utterances in novels. In general, speakers alternate in consecutive utterances in novels. Therefore, in consecutive utterances, the following pairs of sentences can be considered as positive and negative instances:

Positive instance 1 Pairs of the n-th and n+2-th utterances in consecutive utterances.

Positive instance 2 Pairs of sentences in an utterance obtained by splitting an utterance with specific symbols (“!”, “?” and “。”).

Negative instance Pairs of the n-th and n+1-th utterances in consecutive utterances.

Figure 1 shows examples of positive and negative instances. As the resource for the dataset, we collected about 9,000 novels published on the novel posting site called “小説家になろう (Aim to be a novelist)”[1], and collected approximately 50 million pairs of positive and negative instances. Of this total of approximately 100 million pairs dataset, 99% is used as training data, and the remaining 0.5% each as validation and test data. The statistics of positive and negative instances are shown in Table 1.

4.2 Evaluation of the Dataset

To verify the correctness of the automatically collected positive and negative instances, we manually annotated the 5-level scale of similarity grade from -2 to 2 to each of randomly selected utterance pairs (250 pairs each)[2]. Table 2 shows that 62.4% of the automatically collected positive instances correctly have *similar* (i.e., the grades of 2 and 1) speech styles, while 58.8% of the negative instances correctly have *dissimilar* (i.e., the grades of -2 and -1) speech styles. When we consider the pairs of the similarity grade 0 as correct instances, the proportion of data that can be properly used for the training exceeds 70% for both positive and negative instances, indicating that these automatically collected data are sufficiently useful for constructing a speech style embedding model. The inter-annotator agreement of the annotation was evaluated using the Quadratic Weighted Kappa (QWK) score (Cohen, 1968), which ranges from 0 to 1, with a higher value indicating better agreement. Our annotation achieved a QWK score of 0.763, suggesting that there were no significant disagreement between the two annotators.

[1] https://syosetu.com/

[2] The annotation work was done by the first and second authors, where each pair is annotated with “—” when either of the two sentences is not an utterance, but an emphasis or a quotation.

sim. grade	Comparison with それは安心ね / sore wa anshin ne / That’s a relief (female speech style)	avg. sim. grade of two annotators	ratio (%)	
			pos.	neg.
2	The two sentences are completely with an equivalent speech style, containing identical characteristic words. すごいわね / sugoi wa **ne** / That’s great (female speech style)	2.0, 1.5	34.8	12.4
1	The two sentences are with an equivalent speech style, but containing distinct characteristic words. どうかしら / do **kashira** / I’m not sure (female speech style)	1.0, 0.5	27.6	12.4
0	Either of the two sentences is an utterance that could be uttered by anyone. それは... / sore wa / That is ... (common speech style)	0	15.6	14.4
-1	The two sentences are not equivalent speech style, but are utterances that could be used by an identical person in some situations. すごいよ / sugoi **yo** / It’s great (kind male speech style)	−0.5, −1.0	7.2	18.0
-2	The two sentences are completely dissimilar. 俺の版だぜ / ore no ban **daze** / It’s my turn (masculine speech style)	−1.5, −2.0	7.2	40.8
—	Either of the two sentences is not an utterance, but an emphasis or a quotation. 洞窟 / dokutsu / The Cave	—	4.8	2.0

Table 2: Evaluation of Automatically Collected Positive/Negative Instances

5 Speech Style Embedding Model

5.1 Model Configuration

For the construction of the speech style embedding model, we utilize the Sentence-BERT architecture (Reimers and Gurevych, 2019), and Tohoku University’s Japanese version of BERT-base[3] as pre-trained BERT model. The batch size is set to 128 sentences, and the maximum input token length is set to 64 tokens. The utilized loss function is defined by the Contrastive Loss equation (Hadsell et al., 2006) as presented below:

$$L = \frac{1}{2} Y D^2 + (1 - Y) \max(margin - D, 0)^2$$

Here, Y represents the label where 1 indicates a positive instance and 0 indicates a negative instance. Following (Gao et al., 2021), we use D as the cosine distance between the two utterances of positive/negative instances, and the $margin$ is set to 1. Through this training process, we anticipate that the speech style of the input utterance will be embedded in the 768-dimensional output vector for the CLS token of the speech style embedding model.

[3] https://github.com/cl-tohoku/bert-japanese

Target Utterance	Example Utterances to Compare Similarities	Similarity
(1) 俺はこれが好きだぜ / *ore wa kore ga suki* ***da ze*** / I like it / (a masculine speech style utterance)	いいじゃねえの / *ii* ***ja ne no*** / That's good, isn't it? / (masculine)	0.839
	ちょうどよかったぜ / *cho do yokatta* ***ze*** / Just in time / (masculine)	0.770
	すごいだろ！ / *sugoi* ***da ro*** / It's great! / (masculine)	0.502
	すごいだろ**?** / *sugoi* ***da ro*** / It's great? / (masculine)	0.390
	すごいだろ**...** / *sugoi* ***da ro*** / It's great... / (masculine)	0.367
	すごいだろ / *sugoi* ***da ro*** / It's great / (masculine)	0.345
(2) どういたしましょうか / *do* ***itashi masyo ka*** / what should I do? / (a strong polite speech style utterance reminiscent of a maid or servant)	いかがしますか / ***ikaga shimasu ka*** / What would you like? / (strong polite)	0.924
	申し訳ありません / ***moshi wake ari mase n*** / I'm so sorry / (strong polite)	0.830
	お願い致します / *onegai* ***itashi masu*** / I'm begging you / (strong polite)	0.456
	お願いします / *onegai* ***shi masu*** / I'm begging you / (polite)	0.480
	お願いだよ / *onegai* ***da yo*** / I'm begging you / (casual)	-0.514
	田中様 / *tanaka* ***sama*** / Mr.Tanaka / (strong polite)	0.844
	田中殿 / *tanaka* ***dono*** / Mr.Tanaka / (classical polite)	0.536
	田中 / tanaka / *Tanaka* / (informal)	-0.116
(3) ふなっしーはこれが好きなっしー / *funassyi wa kore ga suki* ***nassyi*** / I like it / (a speech style used only by the fictional mascot character "ふなっしー (Funassyi)"	お疲れ様なっしー / *otsukare sama* ***nassyi*** / You must be tired / (Funassyi)	0.758
	よろしくなっしなー / *yoroshiku* ***nassyina*** / Nice to meet you / (Funassyi)	0.451
	危なかったー / *abuna katta* / That was close / (childish)	0.756
	誰だろー / *dare daro* / I wonder who it is / (childish)	0.670
	私はこれが好きです / *watashi wa kore ga suki* ***desu*** / I like it / (formal)	0.259
	俺はこれが好きだぜ / *ore wa kore ga suki* ***daze*** / I like it / (masculine)	0.096
	儂はこれが好きなんじゃ / *watashi wa kore ga suki* ***desu*** / I like it / (classical)	0.039

Table 3: Examples of Comparing Speech Style Similarities

5.2 Dimensionality Reduction

Using 768-dimensional vectors directly as style-sensitive sentence embeddings could be considered excessive in terms of the complexity of speech styles. Therefore, we attempt to convert them into smaller-dimensional vectors through dimensionality reduction using Principal Component Analysis (PCA).

The result of PCA shows that the cumulative proportion exceeds 99.8% with the first 32 principal components, indicating that 32 dimensions are sufficient to retain speech style information. Consequently, we treat the vectors with their dimensionality reduced to 32 dimensions as the final style-sensitive sentence embedding.

6 Analysis of Style-sensitive Sentence Embeddings

6.1 Comparison of Similar Utterances

We verify whether the speech style embedding model appropriately captures speech style expressions. Table 3 shows examples of comparing speech style similarities. In the comparison between (1) a masculine speech style utterance "俺はこれが好きだぜ (I like this)" and other utterances, it can be seen that the similarities with other masculine speech style utterances such as with " ねえの (*ne no*)" and "よかったぜ (*yokatta ze*)" are appropriately high. Moreover, it is evident that the similarity varies significantly depending on the specific symbols ("!", "?", and "⋯"). Furthermore, in the comparison between (2) a strong polite speech style utterance reminiscent of a maid or servant "どういたしましょうか (What should I do?)" and other utterances, it can be seen that the similarities are higher with utterances containing strong polite expressions like "致します (*itashi masu*)" and "様 (*sama*)" whereas they decrease significantly for casual speech styles and utterances without honorifics. These results indicate that the speech style embedding model can effectively distinguish distinct speech styles and understand various nuances in the Japanese language.

Finally, we conduct a comparison between (3) a unique speech style utterance "ふなっしーはこれが好きなっしー (I like this)" and other utterances. This speech style is used only by the fictional mascot character "ふなっしー (Funassy)"[4]. This character, "ふなっしー (Funassy)" usually appends words such as "なっしー (*nassyi*)" or its variant "なっしなー (*nassyina*)" at the end of sentences. From the comparison results, the similarity with the utterances ending with "なっしー (*nassyi*)" is appropriately high, and the utterances with distinctly different speech styles, such as the formal speech style "私はこれが好きです (I like this)" or the masculine speech style "俺はこれが好きだぜ (I like this)", are appropriately low. It is important to note that the training data does not include the speech style of "ふなっしー (Funassy)", suggesting that the speech style embeddings model can

[4] https://274ch.com/

ID speech style	Top 5 bi-grams by tf-idf
46 housemaid	(まし た / *mashi ta*), (ん です / *n desu*) (ませ ん / *mase n*), (の です / *no desu*) (し て / *shi te*)
35 ninja	(で ござる / *de gozaru*), (ござる よ / *gozaru yo*) (と は / *to wa*), (です な / *desu na*) (ござる か / *gozaru ka*)
31 boss	(よう に / *yo ni*), (し て / *shi te*) (て いる / *te iru*), (だ から / *da kara*) (し なさい / *shi nasai*)
12 king	(て おる / *te oru*), (し て / *shi te*) (で は / *de wa*), (て いる / *te iru*) (お 主 / *o nushi*)
13 female	(の よ / *no yo*), (わ よ / *wa yo*) (し て / *shi te*), (ない わ / *nai wa*) (た の / *ta no*)
33 masculine	(ん だ / *n da*), (か ？ / *ka ?*) (の か / *no ka*), (ぜ ！ / *ze !*) (だ ぜ / *da ze*)
10 cat	(にゃ ！ / *nya !*), (にゃ ？ / *nya ?*) (た にゃ / *ta nya*), (か にゃ / *ka nya*) (に よ / *ni yo*)
22 kind male	(だ ね / *da ne*), (かい ？ / *kai ?*) (ん だ / *n da*), (だ よ / *da yo*) (の かい / *no kai*)

Table 4: Example Clusters and their Characteristic Expressions

correctly evaluate unknown speech styles. However, the similarity with the utterance containing "なっしなー (*nassyina*)" as well as with the variant of "なっしー (*nassyi*)" is incorrectly low, suggesting that the speech style embeddings model cannot correctly evaluate the variant of unknown speech styles. In addition, the similarity with the different speech style utterance containing only a prolonged sound "ー" at the end of sentence is incorrectly high. While the model can correctly identify and evaluate distinct speech styles, there is room for improvement in capturing the variants of unknown speech styles.

6.2 Analyzing Speech Styles through Clustering

To analyze the clusters of speech styles and their characteristics, we conduct clustering on approximately 820,000 unique utterances in the test data, which have been converted into style-sensitive sentence embeddings. To analyze the clustering process, we attempt hierarchical clustering using the Ward's method (Ward Jr., 1963). Due to computational limitation, it is not feasible to cluster all 820,000 style-sensitive sentence embeddings using the Ward's method, so we first classify them into 10,000 clusters using the k-means method, and then cluster the centroid vectors of k-means clusters using the Ward's method[5].

Figure 2 shows the dendrogram of the uppermost 50 clusters in the Ward's method clustering results. Next, we extract bi-grams for each sentence within a cluster and calculate their tf-idf values. Then, we extract the top-5 bi-grams with the highest tf-idf values as the characteristic expressions of each cluster. Table 4 shows the characteristic expressions for the sample 8 clusters. From Figure 2 and Table 4, it is evident that various interesting clusters are observed, such as "masculine" speech styles, "housemaid" speech styles, and "cat" speech styles. In the cluster "ID=33", expressions like "んだ (*n da*)" and "だぜ (*da ze*)" are ranked at the top, suggesting that "masculine" speech styles are grouped together. On the other hand, clusters around "ID=26" and "ID=21" also contain "masculine" speech styles, indicating that the utterances with closely related speech styles distributed in distinct clusters. Furthermore, within the clusters grouped together around "boss" ("ID=31") speech styles, there are clusters of more unique "king" speech styles that use expressions like "ておる (*te oru*)" and "お主 (*o nushi*)". Therefore, it is observed that, while the expressions that should ideally be in an identical cluster are somehow closely distributed but still dispersed, clusters that should be farther apart are relatively close together. In order to handle these issues, it is necessary to devise ways to lower the similarity between different speech styles during the learning process or to remove wrong data from the training data.

7 Analysis of Variation in Speech Styles of an Identical Character

It can be assumed that even for an identical person, their speech style may change depending on conversation partners and surrounding situations. Therefore, we analyze the variation in speech styles embeddings of the utterances of three main characters in a romance novel[6] published on "小説家になろう (Aim to be a novelist)". Table 5 shows the character names, speech styles, and number of utterances of the three characters.

Figure 3 shows the scatter plot of the first

[5]While our style-sensitive embeddings are based on cosine distance, the k-means and the Ward' methods use euclidean distance. Therefore, we normalized the speech style embeddings prior to the application of k-means and Ward's methods.

[6]https://ncode.syosetu.com/n1860fv/

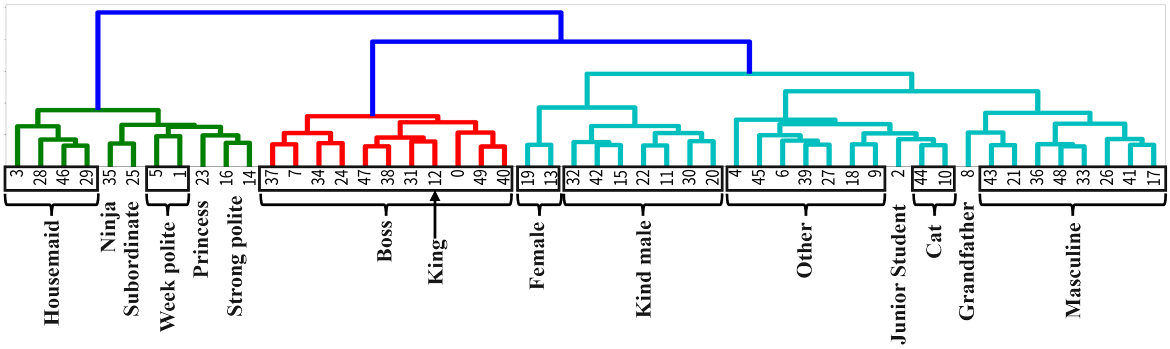

Figure 2: Dendrogram of the Results of Clustering Speech Style Vectors using the Ward's Method.

Character (traits)	Speech Style	#Utterance
Marie (aristocratic woman)	casual female, polite	1,177
Kyuros (aristocratic man)	only male	1,030
Mio (housemaid)	only housemaid	450

Table 5: Speech Style and Number of Utterances of the Three Characters in a Novel

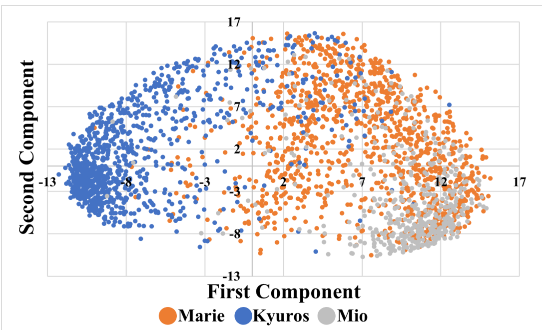

Figure 3: Scatter Plot of the Style-sensitive Sentence Embeddings of the Three Characters in a Novel

and second principal components[7] of the style-sensitive sentence embeddings for all utterances of three characters. While utterances from the identical character tend to cluster together, it is evident that the distribution of the embeddings is scattered to some extent even for an identical character.

Focusing on the variation of the distribution for each character, it is evident that the variation in Mio's style-sensitive sentence embeddings, who always uses typical polite speech reminiscent of a housemaid's speech, is small, while the variation in Marie's style-sensitive sentence embeddings, who switches between "casual female" and "polite" depending on the conversation partner, is larger. In other words, the distribution of the style-sensitive sentence embeddings can be considered as a characteristic of the character's speech style.

8 Conclusion

In this study, we proposed a speech style embedding model that produces style-sensitive sentence embeddings capable of evaluating the similarity of speech styles. The speech style embedding model was constructed using contrastive learning with training data consisting of pairs of utterances with similar/dissimilar speech styles collected from consecutive dialogues in novels. We demonstrated that this speech style embedding model not only captures the similarity of speech styles, but also the strength of politeness, masculinity, and femininity. Furthermore, we confirmed the formation of characteristic speech style clusters such as female and ninja speech styles through clustering of style-sensitive sentence embeddings using the Ward's method. In addition, we analyzed the variation of style-sensitive sentence embeddings across the entire utterances of all the characters in a novel.

Future challenges include constructing a dataset for evaluating speech style similarity with multiple annotators and generating style-sensitive sentence embeddings that take into account the conversation partners and surrounding situations. Additionally, it is necessary to incorporate training methodologies such as the triplet objective function (Reimers and Gurevych, 2019) and in-batch negatives (Gao et al., 2021) to improve model performance. It is equally essential to conduct performance comparison experiments with preceding studies (Akama et al., 2018; Miyazaki et al., 2021b) and verify the usefulness of style-sensitive embeddings in downstream tasks such as controlling the speech style of dialogue systems.

[7]The cumulative proportion up to the second principal component is 41.9%.

References

Reina Akama, Kento Watanabe, Sho Yokoi, Sosuke Kobayashi, and Kentaro Inui. 2018. Unsupervised learning of style-sensitive word vectors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 572–578, Melbourne, Australia. Association for Computational Linguistics.

Jacob Cohen. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70:213–220.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 250–259, Sofia, Bulgaria. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742.

Dongyeop Kang and Eduard Hovy. 2021. Style is NOT a single variable: Case studies for cross-stylistic language understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2376–2387, Online. Association for Computational Linguistics.

Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2018. A large self-annotated corpus for sarcasm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Satochi Kinsui. 2003. *Vaacharu nihongo: yakuwari-go no nazo (In Japanese)*. Iwanami, Tokyo, Japan.

Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. 2006. Compilation of a dictionary of japanese functional expressions with hierarchical organization. In *Proceedings of the 21st International Conference on Computer Processing of Oriental Languages: Beyond the Orient: The Research Challenges Ahead*, ICCPOL'06, page 395–402, Berlin, Heidelberg. Springer-Verlag.

Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, and Yoshihiro Matsuo. 2016. Towards an entertaining natural language generation system: Linguistic peculiarities of Japanese fictional characters. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 319–328, Los Angeles. Association for Computational Linguistics.

Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono, and Hiromi Wakaki. 2021a. Fundamental exploration of evaluation metrics for persona characteristics of text utterances. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 178–189, Singapore and Online. Association for Computational Linguistics.

Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono, and Hiromi Wakaki. 2021b. Fundamental exploration of evaluation metrics for persona characteristics of text utterances. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 178–189, Singapore and Online. Association for Computational Linguistics.

Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 129–140, New Orleans, Louisiana. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Joe H. Ward Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Yuki Zenimoto and Takehito Utsuro. 2022. Speaker identification of quotes in Japanese novels based on gender classification model by BERT. In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pages 126–136, Manila, Philippines. De La Salle University.

Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93.