Effective Use of Japanese Dictionary Definition Sentences in Learning Hierarchical Embedding of Dictionaries

Yuki Ishii
Ibaraki University

22nm702y@vc.ibaraki.ac.jp

Minoru Sasaki
Ibaraki University

minoru.sasaki.01@vc.ibaraki.ac.jp

Abstract

Existing knowledge graph-based learning methods for word sense disambiguation use word-to-word relations to learn models, but do not learn using the hierarchical relations of word senses in a single word. In addition, even when defining sentences in a Japanese dictionary are applied, the accuracy of judging the hierarchical relationship of word senses is poor, and the effect of knowledge graph embedding learning is not fully achieved. This study analyzes how to edit dictionary descriptions to improve the accuracy of models that judge the hierarchical relationship between senses in a Japanese dictionary. The results of the analysis showed that the accuracy of the unedited dictionary was 60.9%, while the accuracy of the edited dictionary was 83.3%, confirming the improved performance of the model.

1 Introduction

Word sense disambiguation (WSD) is the task of identifying which sense is used in a sentence for a polysemous word in the sentence. In the last few years, many knowledge graph-based approaches have been studied that do not require the cost of word sense labeling. Related works include GlossBERT (Huang et al., 2019) and BEM (Blevins and Zettlemoyer, 2020), which combine word sense definition sentences with supervised learning, and EWISE (Kumar et al., 2019) and EWISER (Bevilacqua and Navigli, 2020), which use word-to-word relations in knowledge data. However, these methods do not learn models using the hierarchical relationship of word senses in words. When the EWISE system learned the hierarchical relationship of word senses using the Japanese dictionary definition sentences in their original form, it was found that the relationship judgment accuracy of the generated models was low, and the knowledge information was not sufficiently learned. In the Iwanami Japanese Dictionary, there are definition sentences that only describe usage expressions and part-of-speech expressions such as "《attached to noun》" and "((adjective))". In Japanese, it is expressed as "《名詞に付けて》" and "((形))". Such definitions are likely to exist in multiple dictionaries and overlap in content, thus inhibiting learning of the hierarchical relationship between senses. Therefore, in addition to the existing relations between words, we will learn the hierarchical relations of word senses and analyze how the contents of the Japanese dictionary can be edited and modified to improve the accuracy of the hierarchical relation judgment model of word senses. Learning a model for judging the hierarchical relationship of word senses is an initial task to improve the performance of the knowledge graph embedding system.

To investigate how well the information in the knowledge graph can be embedded, we use the knowledge graph embedding learning of the EWISE system, which provides accuracy in judging the relationship between triples of data in the knowledge data.

2 Related Work

Many approaches based on knowledge graphs have been studied.

In GlossBERT, a pre-trained BERT encoder is given a context and a word definition sentence, and is trained to judge whether the word definition sentence correctly describes the usage of the target word. In BEM, two encoders are used for the GlossBERT approach to learn the distributed representations of the context and the word definition sentences separately. These studies combine word definition sentences with supervised learning.

In EWISE, distributed representations of word senses are learned from WordNet word-word relations and incorporated into the WSD task. In EWISER, the weights given to words are learned using WordNet word-to-word relationships, and the resulting matrices are incorporated into the WSD task. These studies use word-to-word relationships in knowledge data.

However, in these related works, no model learning has been conducted using the hierarchical relationship of word senses in a single word. In this study, the model is trained using a single word hierarchical relationship.

3 Hierarchical relation judgment system

The knowledge graph embedding system of the EWISE system is used to judgment the hierarchical relationship between senses in a Japanese dictionary. The system is shown in Figure 1.

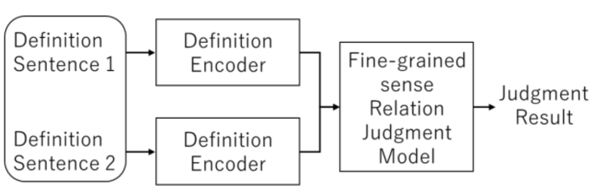

Figure 1: Hierarchical relation judgment system

3.1 Definition Sentence Encoder

The definition sentence encoder uses the BiLSTM Max encoder (Conneau et al, 2017). The fixed-length representation obtained by inputting definition sentences to BiLSTM and Max Pooling the output is the output of the definition sentence encoder.

3.2 Fine-grained Sense Relation Judgment Model

The fine-grained sense relation judgment model takes as input a distributed representation of two Japanese definitions and judges the hierarchical relation between the definitions. ConvE (Dettmers et al., 2013) is used for this fine-grained sense relation judgment model. A knowledge graph usually consists of a set K of N triples (h, l, t) consisting of two entities (h, t) and one relation (l). h is the head entity and t is the tail entity. ConvE formulates the scoring function $\psi_l(e_h, e_t)$ for a triple (h, l, t) as:

$$\psi_l(e_h, e_t) = f\big(\mathrm{vec}\big(f([\overline{e_h}; \overline{e_l}] * w)\big)W\big)e_t \qquad (1)$$

where e_h and e_t are the parameters of the entity, e_l is the parameter of the relation, $\bar{x}$ is a 2-dimensional deformation of x, w is a 2-dimensional convolution filter, $\mathrm{vec}(x)$ is a vectorization of x, w is a linear transformation, and f is a normalized linear unit. For the target head entity h, we compute the score $\psi_l(e_h, e_t)$for each entity in the graph as a tail entity. Probability estimates for the validity of a triple (h, l, t) are obtained by applying a sigmoid function to the scores:

$$p = \sigma\big(\psi_l(e_h, e_t)\big) \qquad (2)$$

3.3 Model Learning

Figure 2 shows the training flow of the definition sentence encoder and the fine-grained sense relation judgment model. The training data are knowledge data describing the hierarchical relationship of word meanings and definition sentences associated with the definition sentence IDs. The definition sentence is decomposed into morphemes, and the distributed representation of each morpheme is input to the definition sentence encoder. The distributed representation of morphemes is pre-trained using fastText (Bojanowski et al., 2016) and GloVe (Pennington et al., 2014). The word sense relations are transformed into embedding vectors by the embedding layer in the inter-sense relation judgment model. The parameters of the fine-grained sense relation judgment model are initially set to the parameters of the initial model trained with only triples (h, l, t) of word sense hierarchical relations. Equation (1) is modified to Equation (3) to learn the definition sentence encoder.

$$\psi_l(e_h, e_t) = f\left(\mathrm{vec}\left(f\left(\left[\overline{q(h)}; \overline{e_l}\right] * w\right)\right) W\right) e_t \quad (3)$$

where $q(.)$ is the definition sentence encoder and the head entity is the encoded definition of the entity. The parameters of the model are updated based on the estimates p and the labels. The loss function uses the binary cross-entropy:

$$L_C = -\frac{1}{N} \sum_i (t_i \cdot log(p_i) + (1 - t_i) \cdot log(1 - p_i)) \quad (4)$$

where t_i is 1 if the triple (h, l, t) is appropriate, 0 otherwise. p_i is an estimate of the score shown in Equation (3).

Section 5.1 shows the training data, the distributed representation of morphemes, and Section 5.3 details the evaluation method of the model.

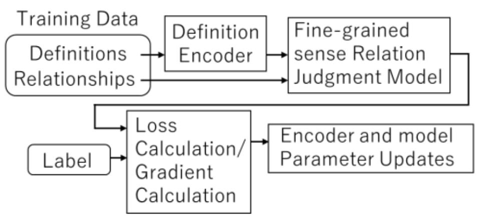

Figure 2: Model Learning Flow

4 Editing and modifying the definition text

4.1 Editing Definition Sentences

Edit the Iwanami Japanese Dictionary to correspond to the WordNet (Miller, 1995) triple data used in EWISE.

4.1.1 Assigning IDs to Definition Sentences

The Iwanami Japanese Dictionary's definition sentence IDs are assigned by combining the headword number (1 to 5 digits), major category number (1 digit), middle category number (1 digit), and minor category number (1 to 2 digits) in this order, excluding the compound word number.

The hierarchical relationship of word senses is learned on the two relationships of hypernym and hyponym from the classification method of the Iwanami Japanese Dictionary. The hierarchical relationship is a tree structure as shown in Figure 3. The parent of a tree structure is a hypernym of its children, and its children are hyponym of the parent. When recording triples, only the relationships between nodes connected by edges are recorded.

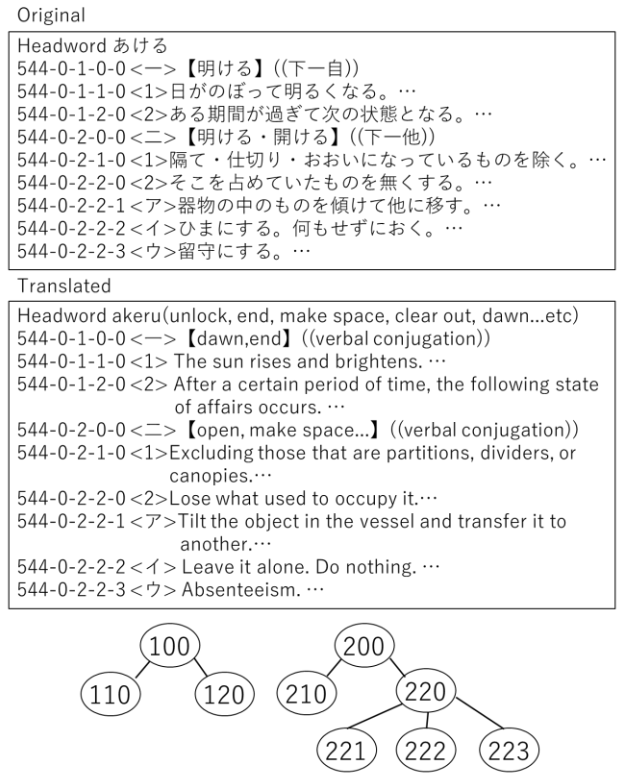

Figure 3: Hierarchical relationship of word senses (The number in the node corresponds to the right three digits of the dictionary ID.)

4.2 Change Definition Sentence

To improve the accuracy of judging the relationship between triples of knowledge data, we propose the following settings A, B, ① to ⑤.

4.2.1 Setting A and B

There are word definitions in which the definition is written only with the double parentheses "(( ))", double mountain brackets "《》", and turtle-shell bracket "[ ]", which do not explain the meaning of the word, so the definition itself is deleted. Deleting a definition sentence affects the total number of triples of knowledge data differently from changes in subsequent clauses. Experiments were conducted for setting A, in which the Iwanami Japanese Dictionary was not edited, and setting B, in which definition sentences that do not explain word meanings were deleted, and the setting with the highest accuracy was carried over to the changes in subsequent clauses.

4.2.2 Setting ① and ①*.

The three types of expressions, double parentheses "(( ))", double parentheses "《》", and parentheses "< >", indicate the attributes, uses, and classification of the headword, and were deleted from the definition text. For the turtle-shell bracket "[ ]" expression, a similar expression exists in the WordNet definition text, but the setting ①* is deleted to check whether the category name is necessary. Symbols were removed and replaced because they were superfluous expressions. The symbols deleted were "↓", "△", and "×", and the symbols replaced were "▽" with "。" and "【】" with "『』". In addition, full-width alphabetic characters have been changed to half-width characters. Example sentences are described in the dictionary, but the headwords included in the example sentences are replaced by hyphens "-". The hyphen "-" is replaced by the headword (hiragana).

4.2.3 Setting ②

In the case of hiragana, it is preferable to convert the words to kanji to obtain an accurate representation of the variance because of the presence of words with homonymous meanings. Since many of the headwords have more than one Kanji character, those that can be replaced one-to-one were replaced with Kanji characters.

4.2.4 Setting ③

There are many expressions in the definitions that cannot be the direct meaning of a word. The following is a list of those that were deleted.

- The sentences following "▽" provide explanations that go beyond the meaning of the target word.
- The sentence followed by "派生｜" ("Derivatives|") indicates the derivation of the declension.
- To remove reading kana, delete the content enclosed in full-width parentheses "( )" if it is in hiragana only.
- The content enclosed by single square brackets "( )" indicates the figure number and annotation number in the dictionary.

4.2.5 Setting ④

In setting ④, delete the contents enclosed in full-width parentheses "( )" regardless of the contents.

4.2.6 Setting ⑤

In the Iwanami Japanese Dictionary, some example sentences surrounded by "「」" are sentences without headwords. In such cases, it is difficult to complete the sentence and incomplete example sentences remain. Therefore, if a single sentence delimited by a punctuation mark is surrounded by key brackets "「」" in the definition sentence, the entire example sentence is deleted.

5 Experiment

5.1 Experimental Data

- The hierarchical relationship of word senses
  Triples of training data from the Iwanami Japanese Dictionary, in which the IDs of definitions (h, t) and the hierarchical relationship of meanings (l) are recorded in the order h, l, t, are used. Setting A with no editing obtained 24040 triples, while setting B, which removed definition sentences without word descriptions, obtained 10842 triples. The triples of training data obtained are divided into training, development, and test data in the ratio of 8:1:1.
- Definition sentences
  Avoiding duplicate definition sentences may have improved accuracy. The definition sentences of the word senses associated with the definition sentence IDs of the knowledge data are used as experimental data. MeCab and mecab-ipadic-NEologd were used to separate words in the definition sentences. The distributed representation of morphemes in the definition sentences is based on a corpus of dumped data from the full-page articles of Japanese Wikipedia updated on October 10, 2021, which was pre-trained using fastText and GloVe.

Table 1: Experimental results
(Total vocabulary (total number of words in the definition sentence), Words Vector Number (number of words for which a distributed representation was obtained), Ratio (number of vocabulary vectors/total vocabulary))

Setting/ Result Detail	fastText				GloVe			
	Words Vector Number	Total vocabulary	Ratio	MRR	Words Vector Number	Total vocabulary	Ratio	MRR
A	60067	72619	0.8272	0.60853	60066	72618	0.8272	0.60957
B	60147	72752	0.8267	0.82286	60146	72751	0.8267	0.82203
B①*	66813	81039	0.8245	**0.83420**	66812	81038	0.8245	0.82640
B①②	68311	79958	0.8543	0.82890	68310	79957	0.8543	0.82590
B①②③	62275	71860	0.8666	0.83137	62274	71859	0.8666	0.82874
B①②③④	61449	70805	0.8679	0.83177	61449	70805	0.8679	0.82928
B①②③④⑤	43267	48612	0.8900	0.82251	43266	48611	0.8900	0.82098
B①*③④	60417	72181	0.8370	0.83341	60416	72180	0.8370	**0.83135**

5.2 Hyperparameter

Below are the hyperparameters for each study, modified from the default values.

When training the definition encoder and the fine-grained sense relation judgment model, the batch size was changed to 64, and the number of epochs was set to 160. The number of epochs for the initial model was set to 300, while for GloVe pre-training, the number of surrounding words considered for training was set to 10, and the dimensionality of the distributed representation was set to 300.

5.3 Evaluation Method

The MRR (Mean Reciprocal Rank) is used as the evaluation method for the fine-grained sense relation judgment model. MRR is one of the ranking evaluation indexes, which evaluates a model with a target of the relationship estimates output from the fine-grained sense relation judgment model. In this case, MRR is given as:

$$MRR = \frac{1}{|U|}\sum_{u \in U}\frac{1}{k_u} \tag{5}$$

where u is the target triple, U is the total triple, and k_u is the order in which the entities whose relations were correctly judged for the target triple appeared.

6 Results

Table 1 shows the experimental results. Deletion of definitions that do not explain word meanings was effective, while deletion of expressions in the dictionary and deletion of additional information slightly improved the accuracy of judgments.

The accuracy improvement was significantly greater when definition sentences that did not explain the word meanings were removed. Most of these definitions are conjugational, which means that there are multiple definitions that are identical word-for-word.

Changing the definition sentence affects the number of words in the entire definition sentence and the number of words from which the initial value of the distributed expression can be obtained. Regardless of the number of word vectors or the number of words in the vocabulary, it is acceptable to remove expressions that are deemed unnecessary to explain the meaning of the target word.

In setting ⑤, it is difficult to remove example sentences that cannot be complemented, which may have reduced the accuracy of the results.

In addition, training with fastText showed a slight improvement in accuracy compared to GloVe. Distributed representations of words learned using subword information were shown to improve decision accuracy compared to using global co-occurrence information.

7 Conclusion and Future Work

Experimental results showed that learning by removing or changing expressions that do not explain the meaning of a definition sentence is beneficial for improving the performance of the hierarchical relation judgment model. The hierarchical relationship judgment model obtained in this study is expected to improve the performance of knowledge graph embedding systems. In the future, we plan to study the effectiveness of this method by conducting word sense disambiguation experiments using the knowledge graph embedding system.

References

Huang, L., Sun, C., Qiu, X. and Huang, X. “GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge.” 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), pp. 3509-3514, 2019.

Blevins, T. and Zettlemoyer, L. “Moving Down the Long Tail of Word Sense Disambiguation with Gloss Informed Bi-encoders”, In Proceedings of the 58th Association for Computational Linguistics (ACL2020), pp. 1006-1017, 2020.

Sawan, K. Sharmistha. J. Karan, S. and Partha T."Zero-shot word sense disambiguation using sense definition embeddings." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 5670-5681, 2019.

Bevilacqua M. and Navigli R. “Breaking Through the (80%) Glass Ceiling: Raising the State of the Art in Word Sense Disambiguation by Incorporating Knowledge Graph Information.” In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL2020), pp. 2854-2864, 2020.

Alexis, C. Douwe, K. Holger S. Loïc, B. and Antoine, B. “Supervised learning of universal sentence representations from natural language inference data.” In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 670–680, Copenhagen, Denmark. Association for Computational Linguistics, 2017.

Tim, D. Pasquale, M. Pontus, S. and Sebastian, R. “Convolutional 2d knowledge graph embeddings.” In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

Piotr, B. Edouard, G. Armand, J. and Tomas, M. “Enriching Word Vectors with Subword Information.” In Transactions of the Association for Computational Linguistics, Volume 5, pp. 135–146, 2016.

Jeffrey, P. Richard, S. and Christopher, M. “GloVe: Global Vectors for Word Representation.” In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, 2014.

George, M. “Wordnet: a lexical database for english.” Communications of the ACM, 38(11):39– 41, 1995.

Appendix: Specific examples of the settings

Specific examples of the settings shown in section 4.2 of the main text are given below. These definitions are taken from the Iwanami Japanese Dictionary, 5th edition.

Original

Headword あかず【飽かず】
288-0-0-0-0 ((連語))《副詞的に》
288-0-0-1-0 <1>あきもせずに。「一見入る」
288-0-0-2-0 <2>そうするのはあきたりないのに、やむなく。「一別れる」

Translated

Headword akazu【Without being bored, be forced to do】
288-0-0-0-0 ((phrase))《Adverbially》
288-0-0-1-0 <1> Without giving up. 「I'll never — looking at it. 」
288-0-0-2-0 <2> It's not enough to do so, but we have no choice. 「be — to leave 」

Figure A: Example of specific changes to setting B (removing red text)

Original

Headword こぼれる
18050-0-0-1-0 <1>【△零れる・×溢れる】((下一自))余って漏れ出る。
18050-0-0-1-1 <ア>液体や粒状のものなどが、あふれて落ちる。「涙が—」。…
18050-0-0-1-2 <イ>あり余って外に出る。あふれる。…
18050-0-0-2-0 <2>【×毀れる】((下一自))欠けたりくずれたりして、完全な姿を失う。「刃が—」▽(1)に対する他動詞は「こぼす」、…

Headword こぼれる
18050-0-0-1-0 『零れる・溢れる』余って漏れ出る。
18050-0-0-1-1 液体や粒状のものなどが、あふれて落ちる。「涙がこぼれる」。…
18050-0-0-1-2 あり余って外に出る。あふれる。…
18050-0-0-2-0 『毀れる』欠けたりくずれたりして、完全な姿を失う。「刃がこぼれる」。(1)に対する他動詞は「こぼす」、…

Translated

Headword koboreru
18050-0-0-1-0 <1>【△fall, flood... ・ ×fall, flood...】((verbal conjugation)) Excess leaks out.
18050-0-0-1-1 <ア>Liquid or granular material overflows and falls.「a tear begins to —」。…
18050-0-0-1-2 <イ>Go out in abundance. Overflowing.…
18050-0-0-2-0 <2>【×spills over】((verbal conjugation)) Chipped or crumbled, it loses its integrity. lose their integrity.「Knife blade —」▽ The transitive verb for (1) is "to spill." 、…

Headword koboreru
18050-0-0-1-0 『fall, flood... ・ fall, flood...』 Excess leaks out.
18050-0-0-1-1 Liquid or granular material overflows and falls. 「a tear begins to fall」。…
18050-0-0-1-2 Go out in abundance. Overflowing. …
18050-0-0-2-0 『spills over』 Chipped or crumbled, it loses its integrity. lose their integrity. 「Knife blade spills over」。 The transitive verb for (1) is "to spill." …

Figure B: Examples of specific changes to setting ①

Original

Before
Headword しだい【次第】
…
21323-0-1-3-0 <3>経過。成行き。
21323-0-1-3-1 <ア>物事の事情。「こういう—だ」「事と—によっては」
21323-0-1-3-2 <イ>由来。「—書（がき）」
…

After
Headword しだい【次第】
…
21323-0-1-3-0 <3>経過。成行き。
21323-0-1-3-1 <ア>物事の事情。「こういう次第だ」「事と次第によっては」
21323-0-1-3-2 <イ>由来。「次第書（がき）」
…

Translated

Before
Headword shidai【comes down, Circumstances, Program】
…
21323-0-1-3-0 <3>Progress. Accomplishment.
21323-0-1-3-1 <ア> Circumstances of things.「It — to this.」「under certain —」
21323-0-1-3-2 <イ> origin.「printed — 」
…

After
Headword shidai【 comes down, Circumstances, Program 】
…
21323-0-1-3-0 <3>Progress. Accomplishment.
21323-0-1-3-1 <ア> Circumstances of things. 「 It comes down to this. 」「under certain circumstances」
21323-0-1-3-2 <イ> origin.「printed program」
…

Figure C: Specific examples of changes in setting ②

Original

Headword けっこう【結構】
…
14890-0-0-2-2 <イ>気だてがよい。「一な御仁（ごじん）」▽古風な言い方。
14890-0-0-2-3 <ウ>これ以上は、いらない。「もう一（です）」▽ウは、主として言い切りの形で使う。派生|さ
…
Headword こて【鏝】
…
17794-0-0-2-0 <2>こて(1)に形の似た、熱して使う道具。
…

Translated

Headword kekkou【tolerable】
…
14890-0-0-2-2 <イ> gracious.「man of — character（gojin）」▽ Old-fashioned way of saying.
14890-0-0-2-3 <ウ> don't want any more. 「already —（desu）」▽ウ is used primarily in the form of an alliteration. Derivative|sa
…
Headword kote【trowel】
…
17794-0-0-2-0 <2> A tool similar in shape to an iron (1) that is heated to be used. …

Figure D: Example of specific change in setting ③ (removing red text)

Original

Headword いたむ【痛む・傷む・悼む】
…
2216-0-1-2-1 <ア>（食品が）くさる。「りんごが―」
2216-0-1-2-2 <イ>（器物・建物などが）破損する。「ペン先が―」
…

Translated

Headword itamu【 Hurts, Wounds, Mourns 】
…
2216-0-1-2-1 <ア>（Food）stinks.「The apple ―」
2216-0-1-2-2 <イ> Damage (to property, buildings, etc.).「The nib ―」
…

Figure E: Example of a specific change in setting ④ (removing red text)

Original

Headword あ【亜】
…
1-0-0-2-1 <ア>「亜細亜（アジア）」の略。「亜欧・東亜」
…
Headword いたい
…
2161-0-1-1-0 <1>…外力・病気で肉体や精神が苦しい。「くもない腹をさぐられる」（思ってもみない事で邪推される）「そんな事はくもかゆくもない（＝少しもこたえない。↓いたしか ゆし）」
…

Translated

Headword a【a】
…
1-0-0-2-1 <ア> abbreviation of「Asia（Asia）」.「Asia-Europe and East Asia.」
…
Headword itai
…
2161-0-1-1-0 <1>… Physical or mental suffering due to external forces or illness.
「be suspected without」（You'll get the wrong idea by something you don't expect.）
「That won't anything（= not be daunted at all . ↓ itashika yushi）」
…

Figure F: Example of a specific change in setting ⑤ (removing red text)

The Wikipedia corpus used for pre-training of word distributed representations was obtained from

https://ja.wikipedia.org/wiki/Wikipedia:%E3%83%87%E3%83%BC%E3%82%BF%E3%83%99%E3%83%BC%E3%82%B9%E3%83%80%E3%82%A6%E3%83%B3%E3%83%AD%E3%83%BC%E3%83%89