

Analyzing Syntactic Generalization Capacity of Pre-trained Language Models on Japanese Honorific Conversion

Ryo Sekizawa and Hitomi Yanaka

The University of Tokyo

{ryosekizawa,hyanaka}@is.s.u-tokyo.ac.jp

Abstract

Using Japanese honorifics is challenging because it requires not only knowledge of the grammatical rules but also contextual information, such as social relationships. It remains unclear whether pre-trained large language models (LLMs) can flexibly handle Japanese honorifics like humans. To analyze this, we introduce an honorific conversion task that considers social relationships among people mentioned in a conversation. We construct a Japanese honorifics dataset from problem templates of various sentence structures to investigate the syntactic generalization capacity of GPT-3, one of the leading LLMs, on this task under two settings: fine-tuning and prompt learning. Our results showed that the fine-tuned GPT-3 performed better in a context-aware honorific conversion task than the prompt-based one. The fine-tuned model demonstrated overall syntactic generalizability towards compound honorific sentences, except when tested with the data involving direct speech.

1 Introduction

The correct use of Japanese honorifics is difficult because it requires both the knowledge of grammatical rules (i.e., verb conjugation) and contextual information (i.e., social relationships among the speaker, the hearer, and the people mentioned in a conversation) (Harada, 1976). We expect this syntactic and pragmatic ability for pre-trained large language models (LLMs), as they have shown high performance on natural language tasks (Brown et al., 2020, Ouyang et al., 2022, *inter alia*). However, it remains unclear whether LLMs can handle Japanese honorifics in a similar manner to humans, based on sentence structures and social context.

Several studies proposed datasets of Japanese honorifics for classification (Liu and Kobayashi, 2022; Someya and Oseki, 2022) and generation (Matsumoto et al., 2022). Liu and Kobayashi

(2022) introduced a task in which a model takes an honorific sentence as input and classifies its honorific level or the types of honorifics used in the sentence. Someya and Oseki (2022) provided a Japanese acceptability classification dataset called JCoLA. In JCoLA, subject honorifics are categorized as sub-categories of subject-verb agreement tasks. However, these datasets aim to evaluate the syntactic performance of language models, and they do not analyze their pragmatic ability to understand honorifics by considering social relationships behind sentences. Matsumoto et al. (2022) introduced an evaluation dataset for an honorific conversion task in which the input was a non-honorific sentence, and the output was an honorific sentence. Matsumoto et al. (2022) mentioned the necessity of considering the information on social relationships among people in honorific conversion but did not clarify how such information should be processed in the task. In summary, the existing benchmark datasets of Japanese honorifics focus on either syntactic or pragmatic knowledge required for honorific understanding, not both (Appendix A). Additionally, none of these existing studies discusses the generalization capacity toward various syntactic structures of honorific sentences.

In this research, we introduce a new honorific conversion task that uses information on person’s social relationships as additional input. In Matsumoto et al. (2022)’s proposed honorific conversion, the input was only a non-honorific sentence. In our task, social relationships are expressed as a sentence and concatenated into an input sentence (Section 2). This enables us to analyze whether LLMs could consider information on social relationships when executing honorific conversion. We also construct a dataset to investigate the syntactic generalizability of LLMs for this honorific task. We create hand-crafted templates and generate problems for the task by filling in the placeholders (Section 3). By focusing on the syntactic generalization

Type	Target of respect and deference	Example
Subject honorifics (SH)	Agent	<u>Sensei-ga</u> Hanako-o homete-irasshatta . Teacher-NOM Hanako-ACC praised-SH
Object honorifics (OH)	Patient	(Watashi-ga) <u>sensei-no-tokoro-ni</u> ukagau . (I-NOM) teacher-POSS-place-LOC visit-OH

Table 1: Types of Japanese honorifics with conjugation rules. The underlined part is a person to whom the speaker should show respect or deference. The bolded parts are conjugated verbs. A verb *hometa* (praised) conjugates to its subject honorific form *homete-irasshatta* by attaching a suffix *irasshatta* and *tazuneru* (visit) conjugates to its object honorific form *ukagau*.

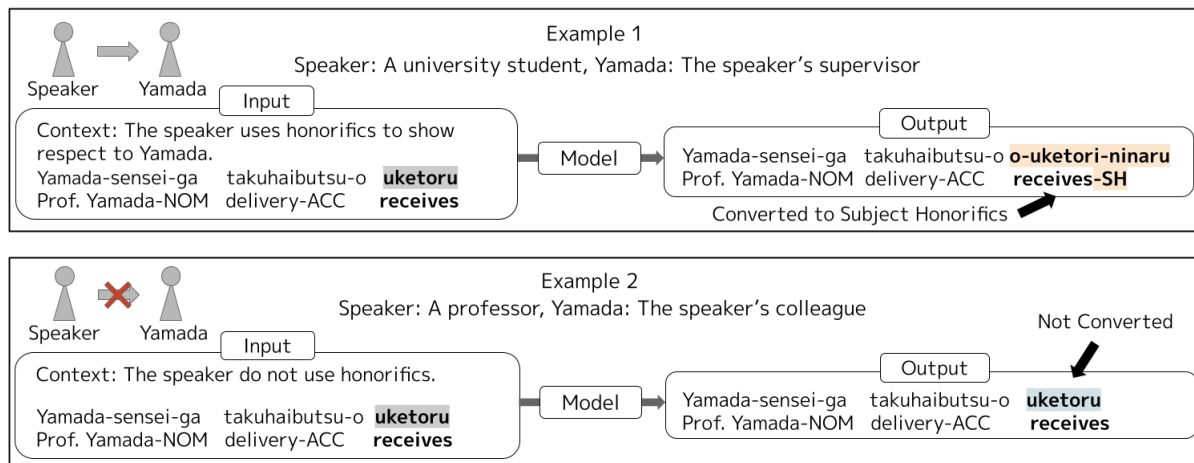


Figure 1: Examples of an honorific conversion task that considers contextual information. The bolded verbs conjugate to their honorific form if needed, considering the context.

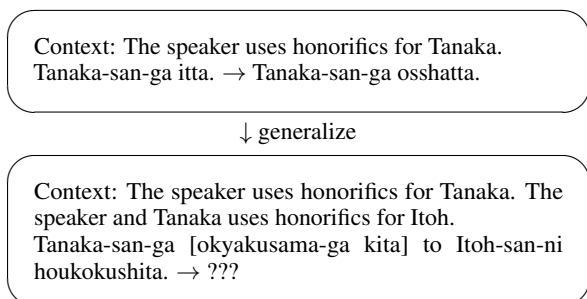


Figure 2: An example of the syntactic generalization of honorifics. The lower problem is made of a more complex sentence structure (center embedding and indirect speech) than the upper problem.

capacity, we analyze how flexibly LLMs can apply the grammatical rules of honorifics. Using our dataset, we then fine-tune and evaluate the performance of GPT-3 on the task (Section 4). Additionally, we evaluate the models using zero-shot learning to determine how well these models perform for honorific conversion using the prompt-based method. Our experiments indicate that the fine-tuned GPT-3 successfully generalizes to sentences with more complex structures, such as scrambling, but not to those involving direct speech. We also

show that the model with prompt learning demonstrates much lower performance than that with fine-tuning.

Our dataset will be publicly available at https://github.com/ynklab/japanese_honorifics.

2 Task Overview

Japanese Honorifics Japanese honorifics are based on various linguistic phenomena (Council for Cultural Affairs, 2007; Nihongo Kijutsu-bunpou Kenkūkai, 2009); some have grammatical rules of conjugation. We target Subject Honorifics (SH) and Object Honorifics (OH). As shown in Table 1, these honorifics are applied depending on the grammatical position of *sensei* (a teacher) so that the speaker can express respect or deference towards the teacher. SH is applied to the predicate when *agent* has a higher social status than the speaker, and OH is applied when the *patient* has a higher social status than the speaker.

Honorific Conversion The existing research proposed honorific conversion (Matsumoto et al.,

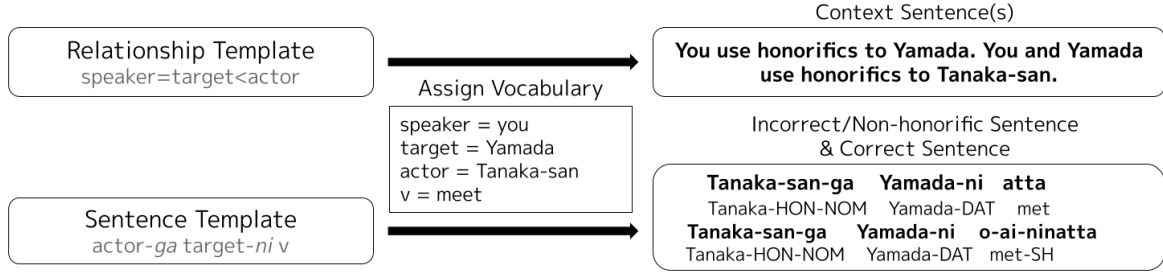


Figure 3: Overview of the process of dataset construction. The bolded sentences are used in the task. The verb *met* has to conjugate to its subject honorific form *met-SH* (*o-ai-ni-natta*) since the speaker is supposed to use honorifics for *Tanaka-san*.

2022). We extend this task to include sentences explaining social relationships as input. In the upper example of Figure 1, the speaker is talking about supervisor Yamada’s actions, so the verb *uketoru* (receive) should be converted into the subject honorific form. In the lower example, the speaker and Yamada are in a casual relationship because they are colleagues; therefore, the model should output the same sentence as the input without honorific conjugation.

Syntactic Generalization We focus on the models’ syntactic generalization ability to capture whether models can flexibly use honorific rules. In this paper, syntactic generalization refers to a model’s ability to use honorific rules for not only simple syntactic structures but also complex syntactic structures (see Figure 2).

3 Dataset Construction

We construct a Japanese honorific dataset by manually creating problem templates and filling their placeholders with vocabulary using dictionaries to evaluate LLMs’ performance on the honorific task. Our dataset construction method is shown in Figure 3. We take this approach instead of automatically collecting data from corpora for two reasons. First, it is difficult to create sentence data with complex structures, such as scrambling, from corpora in a controlled manner. This possibly makes it easier for a model to do honorific conversion than when the information is implicit. Second, we need to prepare controlled settings for social relationship information to evaluate whether LLMs utilize it in honorific conversion; however, such information does not appear explicitly in the corpora.

The second problem is related to the fact that words in argument positions are often dropped in Japanese, especially in dialogue sentences. (1) and

(2) display a conversation between a junior worker and their boss.

- (1) *A boss asks a question to a junior*
Okashi-tte mada nokotteru ?
snack-TOP, still remain ?
‘Are there any snacks left?’
- (2) *The junior answers*
 ϕ_i nokori wa itadakimashita;
 ϕ remained TOP had-OH
‘I/We had them all.’

In this conversation, two ambiguous points must be clarified to determine the honorific relationship behind the conversation. The first point is that the junior answers with object honorifics to show deference in (2), but the target of deference is ambiguous without additional context. If snacks are something the boss originally brought to their office, the boss is the target of the junior’s deference. However, if the snacks are prepared by some third person with a higher rank of job position than the junior, the deference must be towards them instead of the boss. The second point is that the subject is dropped in (2) (pro-drop), but we cannot determine whether ϕ refers to the junior or to a group of workers containing the junior. Considering such language-specific phenomena, we take a template-based approach instead of a corpus-based approach for dataset construction.

3.1 Templates

We create 39 problem templates based on the literature on Japanese linguistics (Council for Cultural Affairs, 2007; Nihongo Kijutsu-bunpou Kenkūkai, 2009). A graduate student with a linguistic background created the templates by consulting a linguistics researcher. Each problem template has three elements for generating input and output data for honorific conversion: the relationship template, sentence template, and honorific type.

Relationship Templates Relationship templates represent social relationships among a speaker, a person who makes an action (agent), and one who is the target of the action (patient) in an equation-like format. For example, *speaker=actor<target* means that the speaker and actor do not use honorifics for each other and should use honorifics for the target.

Sentence Templates Sentence templates have placeholders for person’s names and verbs. Based on their structural complexities, we prepare two types of sentence templates: SIMPLE and COMPLEX. SIMPLE is a template that has one clause and S(O)V structure, and COMPLEX is a template that has more complex syntactic structures: scrambling (SC), center embedding (CE), direct speech (DS), and indirect speech (IS). The first two structures change the argument positions within a sentence, potentially posing challenges for the model in capturing subject-verb agreement. The last two are related to honorific application, depending on whether the sentence has quotation marks (brackets) or not (see Appendix B). A COMPLEX template may contain multiple structures (e.g., IS & CE). See Appendix B for further details.

3.2 Problem Generation

We create problem data for training and evaluating models by filling in placeholders of the templates for verbs and person’s names. From the relationship template, context sentences are generated that explain the social relationships between the speaker and the people mentioned in the input sentence. In addition, from the sentence template, we create an incorrect or non-honorific sentence and a correct honorific sentence. The verb conjugates according to the honorific type given when its placeholder is being filled. We used 23 verbs and 19 names in this experiment. We chose the verbs which are commonly used in daily conversation. We also avoid verbs such as *nusumu*(steal) because honorifics cannot usually be applied to disrespectful actions. Regarding the names of people, we used the 19 most common family names in Japan in 2022¹. Finally, a set of the following data is generated from each problem template: context sentences, an incorrect or non-honorific sentence, and a correct sentence.

¹<https://myoji-yurai.net/prefectureRanking.htm>

4 Experiments

4.1 Experimental Setup

We evaluate GPT-3 models on the proposed honorific conversion task under two different experimental settings: fine-tuning and prompt learning. Despite the general expectation of the superior performance of fine-tuning compared to zero-shot prompt learning, no prior research has aimed to evaluate the performance of LLM on honorific conversion in a prompt-based method. Thus, we compare the scores of these two methods to validate whether the same goes true for honorific conversion. For the two settings accordingly, we use davinci (Brown et al., 2020) and text-davinci-003 (Ouyang et al., 2022), which are available in the OpenAI API (see Appendix C for details including hyperparameter settings).

Fine-tuning We fine-tune two models that differ in the training dataset’s size to measure how much data are needed to generalize the problems. SIMPLE_TRAIN is used for training and SIMPLE_TEST and COMPLEX_TEST for evaluation. **3_times** is a model trained with 117 problems we prepare by generating three data from each problem template, and in the same way, **7_times** is trained with 273 problems. Although our dataset has relatively little data, we consider it enough for the experiments because the minimum dataset size for fine-tuning GPT-3 is “a few hundred.”² As shown in Figure 1, the input is a concatenation of condition sentences and an incorrect sentence, and the output is a proper honorific sentence.

Prompt Learning GPT-3 is known for zero-shot learning, solving some tasks given only a natural language description as a prompt. In addition to the input text used for fine-tuning, we include a task description in the input prompt (see Appendix E).

Evaluation We manually calculate the percentage of correct sentences generated by a model. In this experiment, we regard the output as correct if the verb conjugates to one of the possible honorific forms. We also ignore mistakes unrelated to verb conjugation (e.g., adding a comma in a natural position). We create test datasets using the same problem templates and vocabulary as the training datasets. SIMPLE_TEST contains 108 examples, and COMPLEX_TEST has 408 examples (see Appendix D).

²<https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset>

Context: The speaker uses honorifics for Kimura. The speaker and Kimura use honorifics for Takahashi-san.
 →Speaker<Kimura<Takahashi

Translation	<i>Takahashi says “Kimura is going home.”</i>
Source	Takahashi-san ga “Kimura ga okaerininaru (=go-home-SH)” to ossharu.
Target	Takahashi-san ga “Kimura ga kaeru (=go-home)” to ossharu.
Model’s Prediction	Takahashi-san ga “Kimura ga <u>okaerininaru</u> (=go-home-SH)” to ossharu. (Not converted)

Table 2: An example of the errors regarding direct speech. The speech within brackets is made by Takahashi. The verb *kaeru* should not be in a subject honorific form *okaerininaru* because Takahashi does not use honorifics for Kimura, given their relationships.

	Setting	Simple	Complex			
			CE	SC	IS	DS
FT	3_times	.889	.230	.297	.081	.368
	7_times	.990	.326	.452	.231	.293
PL	zero-shot	.212	.115	.174	.168	.100

Table 3: Evaluation results of the models on our test dataset through honorific conversion. FT refers to fine-tuning, and PL to prompt learning.

4.2 Results

Table 3 shows the scores under all settings of our experiments on the honorific conversion task. Overall, the fine-tuning scores surpass those of the prompt-based method.

4.2.1 Fine-tuning

The scores plummeted when the models were tested with COMPLEX_TEST compared to SIMPLE_TEST. When we increased the data size, the scores increased in most cases, except when tested for problems with direct speech sentences. In Table 2, the model failed to convert a direct speech sentence (*Takahashi-san ga “Kimura ga okaerininaru” to ossharu*). The verb *kaeru* should not be in a subject honorific form (*okaerininaru*) because Takahashi does not use honorifics for Kimura, given their relationships. However, if the brackets (quotation marks in Japanese, see Appendix B) are removed, the sentence (*Takahashi-san ga Kimura ga okaerininaru to ossharu*) becomes an indirect speech sentence and thus becomes proper honorifics. Based on this characteristic, we suppose that the model applied the same honorific knowledge as indirect speech to direct speech, ignoring the role of brackets.

4.2.2 Prompt Learning

The scores were relatively higher when tested with SIMPLE_TEST than with COMPLEX_TEST, but the scores under all of our settings were lower than 25%. We found that the models transferred non-

honorific sentences to polite forms in almost all cases by simply changing the last letters of the verbs that end *-suru* into *-shimasu* instead of applying SH or OH. This conversion is possibly caused by our prompt instructing the models to “convert to the proper honorific sentence,” which may include polite forms too. To validate whether the models use contextual information, we need to construct a prompt that can differentiate SH and OH from polite speech because polite forms are less restricted to social relationships.

5 Conclusions and Future Work

In this paper, we introduced an honorific conversion task that requires not only syntactic knowledge but also pragmatic knowledge, such as social relationships among people. We constructed a Japanese honorific dataset using problem templates created manually and evaluated the syntactic generalization capacity of GPT-3 models on the task using our dataset. The experiments showed that the fine-tuned models could solve problems with simple structures but failed to generalize to problems with more complex structures, particularly with direct speech. Regardless of the sentence structure, the prompt-based models did not successfully solve the problems with our current prompt setting.

In future work, we plan to expand our dataset to include more diverse Japanese honorific expressions, such as predicates other than verbs or honorific prefixes attached to nouns. For the prompt-based experiments, we evaluated the models using zero-shot learning. It would be valuable to test them using few-shot learning by including simple examples in the prompts.

We conducted experiments by explicitly providing information about social relationships. We will also continue to seek data construction methods to extract such information from the corpora, although we did not apply these corpora-based methods in this paper.

Limitations

We discuss two limitations of this research in this section. First, this research focuses on Japanese honorifics with grammatical rules of verb conjugation, which we can judge whether the honorific conversion is correct based on social relationships and sentence structures created in a controlled manner. Japanese honorifics have more expressions based on linguistic phenomena that we did not include in our templates, such as noun honorifics (e.g., *ofutagata*, a polite and formal way of saying “the two people”). Creating templates for noun honorifics requires more detailed settings because they are based on information on context other than social relationships. Second, GPT-3 is the only language model evaluated on our honorific conversion task. This research aims to analyze how capable the well-known, high-performing GPT-3 is of generalizing Japanese honorific sentences and not to explore which existing LLM can achieve the best performance in honorific conversion.

Acknowledgements

We thank three anonymous reviewers for their helpful comments and suggestions. We would also like to thank Anirudh Reddy Kondapally, Tomoki Sugimoto, Tomoya Kurosawa, and the other laboratory members for their helpful advice. This work was supported by JSPS KAKENHI Grant Number JP20K19868.

References

- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. [Language Models are Few-Shot Learners](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Council for Cultural Affairs. 2007. Keigo no Shishin (Guidelines of Honorifics) (in Japanese). online. http://www.bunka.go.jp/kokugo_nihongo/bunkasingi/pdf/keigo_tousin.pdf, accessed: 2023-03-01.
- Shin-ichi Harada. 1976. Honorifics. In Masayoshi Shibatani, editor, *Syntax and Semantics: Japanese Generative Grammar*, volume 5, pages 499–561. New York: Academic Press.
- Muxuan Liu and Ichiro Kobayashi. 2022. [Construction and Validation of a Japanese Honorific Corpus Based on Systemic Functional Linguistics](#). In *Proceedings of the Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, pages 19–26, Marseille, France. European Language Resources Association.
- Yuta Matsumoto, Yu Hayashizaki, Kotaro Kitayama, Hiroaki Funayama, Masato Mita, and Kentaro Inui. 2022. [Proposal of Honorific Style Transfer Task for Supporting Japanese Language Learners \(in Japanese\)](#). In *the proceedings of the 36th Annual Conference of the Japanese Society for Artificial Intelligence*.
- Nihongo Kijutsu-bunpou Kenkukai. 2009. *Danwa, Taigu Hyogen (Discourse, Attitudinal Expression) (in Japanese)*, volume 7 of *Gendai Nihongo Bunpou (Modern Japanese Grammar)*. Kuroshio Shuppan.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#).
- Taiga Someya and Yohei Oseki. 2022. [Nihongoban CoLA no Kouchiku \(Constructing Japanese CoLA Dataset\) \(in Japanese\)](#). In *the proceedings of the 28th Annual Conference of the Association for Natural Language Processing*.

A Existing Datasets

Table 4 shows examples of existing Japanese honorifics datasets.

B Templates

Table 5 shows examples of templates we created. Table 6 shows examples of indirect and direct speech in Japanese.

C Model Details

davinci is the largest model among the ones provided for fine-tuning, and text-davinci-003 is trained by reinforcement learning on human feedback and aimed at being used with prompt learning.

Hyperparameters For fine-tuning GPT-3, n_epochs is set to 2. For text generation, we set max_tokens to 50 and temperature to 0.

Table 4: Examples from the existing honorific datasets.

Original	Converted	Label
朝ごはんはトーストにバターとベジマイトを薄くぬって食べました。 (I had toast for breakfast with a thin layer of butter and Vegemite.)	朝ごはんはトーストにバターとベジマイトを薄くぬっていただきました。 (I had_OH toast for breakfast with a thin layer of butter and Vegemite.)	変換:謙讓語 (Converted: OH)
そして10時くらいに、喫茶店でレーシャルとジョノサンとベルに会いました。 (Then, at around 10:00, I met Rachel, Jonathan, and Belle at a coffee shop.)	そして10時くらいに、喫茶店でレーシャルとジョノサンとベルに会いました。 (Then, at around 10:00, I met Rachel, Jonathan, and Belle at a coffee shop.)	無変換 (Not converted)

(Matsumoto et al., 2022)

Sentences from KeiCO corpus	Honorific level	Respectful 尊敬語	Humble 謙讓語	Polite 丁寧語	Field
本日は、かねてより相談したいことがあり、参上しました。(I have come here today to discuss something that I have been wanting to discuss for some time.)	1	0	1	0	相談 consult
今日は、折り入ってご相談したいことがあって伺ったのですが。(I came here today because I wanted to ask you about something.)	2	0	1	0	相談 consult
今日は相談したいことがあったため、来ました。(I came here today because I had something I wanted to discuss.)	3	0	0	1	相談 consult
今日はずっと相談したいことがあって来た。(I came here today to consult with you about something.)	4	0	0	0	相談 consult
⋮	⋮	⋮	⋮	⋮	⋮

(Liu and Kobayashi, 2022)

Pre-training Data of GPT-3 GPT-3 can input and output Japanese texts because some of its pre-training datasets (Common Crawl, WebText) contain Japanese texts, although the proportion of Japanese texts is not clarified.

D Test Dataset

Within the COMPLEX_TEST dataset, 156 data have center embedding, 252 for scrambling, 160 for indirect speech, and 160 for direct speech. Scrambling and center embedding can not appear in one problem data; the same goes for indirect and direct speech.

E Prompt Example

Figure 4 shows an example of our prompt used for prompt learning.

以下の文はあなたの発言です。人物間の敬語の条件を踏まえて、敬語が不十分かそれらを誤って使っている場合は正しい敬語に変換してください。**(The following sentence is your speech. Given the condition of usage of honorifics between people, convert the sentence to the proper honorific one if it contains wrong or insufficient honorifics.)**

===
敬語の条件:あなたは田中に敬語を使います。
(Condition: You use honorifics for Tanaka.)
田中が受け取る (Tanaka receives) ->

Correct output: 田中がお受け取りになる (Tanaka receives-SH)

Figure 4: An example of the prompt used for zero-shot learning. The bold text is a task description.

Relationship template	Honorific type	Sentence template and an example of created correct sentence	Structure type
speaker=actor_1=target_1	v_ni_1 → NA	actor_1 <i>ga</i> target_1 <i>ni</i> v_ni_1. actor_1 NOM target_1 DAT v_ni_1. Sasaki <i>ga</i> Saito <i>ni</i> <i>au</i> . (Sasaki meets Saito.)	SIMPLE
speaker=target_1<actor_1	v_ni_1 → SH	actor_1 <i>ga</i> target_1 <i>ni</i> v_ni_1. actor_1 NOM target_1 DAT v_ni_1. Takahashi-kyoju <i>ga</i> Kimura <i>ni</i> o-ai-ninaru. (Prof. Takahashi meets Kimura.)	SIMPLE
speaker=target_1<actor_1	v_o_1 → SH	actor_1 <i>ga</i> target_1 <i>o</i> v_o_1. actor_1 NOM target_1 ACC v_o_1. Kimura-hakase <i>ga</i> Yamada <i>o</i> shokai-nasaru. (Dr. Kimura introduces Yamada.)	SIMPLE
speaker=actor_1=target_1	v_ni_1 → NA	target_1 <i>ni</i> actor_1 <i>ga</i> v_ni_1. target_1 DAT actor_1 NOM v_ni_1. Kimura <i>ni</i> Yamamoto <i>ga</i> kanshasuru. (Yamamoto thanks Kimura.)	COMPLEX (SC)
speaker=actor_1=actor_2	v_to_1 → NA v_single_2 → NA	actor_1 <i>ga</i> “actor_2 <i>ga</i> v_single_2” <i>to</i> v_to_1. actor_1 NOM “actor_2 NOM v_single_2” CITE v_to_1. Itoh <i>ga</i> “Matsumoto <i>ga</i> iku” <i>to</i> <i>iu</i> . (Itoh says “Matsumoto goes.”)	COMPLEX (DS, CE)
speaker<actor_2<actor_1	v_to_1 → SH v_single_2 → NA	“actor_2 <i>ga</i> v_single_2” <i>to</i> actor_1 <i>ga</i> v_to_1. “actor_2 NOM v_single_2” CITE actor_1 NOM v_to_1. “Kimura-sensei <i>ga</i> uketoru” <i>to</i> Kato-hakase <i>ga</i> o-kangae-ninaru. (Dr. Kato considers, “Kimura-sensei will receive it.”)	COMPLEX (DS, SC)
speaker<actor_2<actor_1	v_to_1 → SH v_single_2 → SH	actor_2 <i>ga</i> v_single_2 <i>to</i> actor_1 <i>ga</i> v_to_1. actor_2 NOM v_single_2 CITE actor_1 NOM v_to_1. Kimura-sensei <i>ga</i> o-uketori-ninaru <i>to</i> Kato-hakase <i>ga</i> o-kangae-ninaru. (Dr. Kato considers that Kimura-sensei will receive it.)	COMPLEX (IS, SC)

Table 5: Examples of problem templates. NA in the honorific type section means no honorific needs to be applied to a verb. SC=scrambling, CE=center embedding, DS=direct speech, IS=indirect speech

Social relationships: Speaker<Taro=Hanako

Indirect speech	Taro- san - <i>ga</i> irasshatta <i>to</i> Hanako-san- <i>ga</i> itta. Taro-HON-NOM came -SH CITE Hanako-HON-NOM said.
Direct speech	「Taro- <i>ga</i> kita 」 <i>to</i> Hanako-san- <i>ga</i> itta. Taro-NOM came CITE Hanako-HON-NOM said.

Table 6: Examples of indirect speech and direct speech in Japanese. Indirect speech is the citation of someone’s speech without quotation marks (brackets), and direct speech is the one with them. In the example of indirect speech, subject honorifics are applied to Taro’s name (-*san*) and his action (*irasshatta*) to express the speaker’s respect for him. In contrast, the sentence within brackets is written without any honorifics in direct speech. Hanako does not use honorifics for Taro’s actions according to their social relationships, so the quoted sentence is what Hanako said, and no honorifics from the speaker’s view of the entire sentence are reflected.