Generating Dialog Responses with Specified Grammatical Items for Second Language Learning

Yuki Okano[1], **Kotaro Funakoshi**[1], **Ryo Nagata**[2,3], **Manabu Okumura**[1,3]
[1]Tokyo Institute of Technology
[2]Konan University
[3]RIKEN Center for Advanced Intelligence Project
{okano, funakoshi, oku}@lr.pi.titech.ac.jp
nagata-bea2023@ml.hyogo-u.ac.jp.

Abstract

This paper proposes a new second language learning task of generating a response including specified grammatical items. We consider two approaches: 1) fine-tuning a pre-trained language model (DialoGPT) by reinforcement learning and 2) providing a few-shot prompt to a large language model (GPT-3). For reinforcement learning, we examine combinations of three reward functions that consider grammatical items, diversity, and fluency. Our experiments confirm that both approaches can generate responses including the specified grammatical items and that it is crucial to consider fluency rather than diversity as the reward function.

1 Introduction

The use of dialog systems for language learning has attracted attention. Many studies have introduced dialog systems as training partners for language learners and verified their effectiveness. According to previous studies (Kim, 2016; Tegos et al., 2014; Ruan et al., 2019), the advantages of using dialog systems in language education include: they can be used regardless of time, i.e., are more available for learners, they can be easily integrated into chat-based applications that many people are familiar with, i.e., are more user-friendly, and they can be adapted to each learner using various information from chit-chat, i.e., are more supportive.

Needless to say, experiencing a substantial amount of production is critical in language acquisition. Nagata et al. (2020) showed that even a very primitive rule-based chatbot like ELIZA has the potential to increase learner's sentence production. Their experiments also revealed that learners adopted words that appeared in the chatbot's responses, suggesting that the expressions used by the dialog system had a positive impact on learners and that the system was effective in helping them learn unfamiliar words.

Considering these results, we propose a task of generating a response including the specified grammatical items. Here, grammatical items refer to such as to the present perfect, subjunctive, and relative clauses. Usually, they are gradually covered in a language learning course, typically through a school curriculum. Such responses can naturally expose learners to a variety of uses of a specific item and can give them experience of how to use the item in a variety of topics and situations, based on their own past experiences evoked in the conversation. In turn, we expect the learners to use the exposed constructions in their own production more as the exposed uses are linked tightly to their memories by encountering usage examples through dialog based on their own experiences.

The proposed task is formalized as follows. Given $\boldsymbol{C} = [c_1, c_2, ..., c_n]$, a dialog context that is a sequence of n utterances between two interlocutors (the system and the learner), and I, a set of grammatical items specified to be included, the task is to generate r, a natural response that follows c_n, on the condition that r includes an expression corresponding to each item $i \in I$. To the best of our knowledge, this is the first work to tackle this generation task for language learning.

To generate text that satisfies particular conditions, Lin et al. (2021) propose using auxiliary modules to guide pre-trained language models. Keskar et al. (2019) propose training language models with control code. Since these methods are based on supervised learning, they require annotated datasets. However, there is a lack of large labeled dialog datasets for grammatical items.

In this paper, we examine two approaches for generating responses containing the specified grammatical items without a large labeled dataset: 1) RL-based generation: fine-tuning a pre-trained language model using reinforcement learning (RL), and 2) Prompt-based generation: providing a large language model with prompt text with a task in-

Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), pages 184–194

struction and a few examples. The experiments confirm both approaches are promising.

2 Related Work

2.1 Dialog systems and adaptation in language learning

According to Xiao et al. (2023), there are three main uses for dialog systems in language learning.

One way is language learning through general communication. As one of the educational applications of dialog systems, there is a growing body of research on introducing dialog systems in second language learning through free interaction with dialog systems. Alexa (Moussalli and Cardoso, 2020; Dizon, 2017; Dizon and Tang, 2020) and Google Assistant (Tai, 2022) were used. In most studies, learners favorably accepted the system as a dialog partner.

Another way is task-based language learning. The introduction of a dialog system into a task allows for more content-focused learning. Tasks can be varied, such as asking for the time of day at a particular location or ordering at a coffee shop (Wu et al., 2020; Timpe-Laughlin et al., 2020). Learners are allowed to interact and receive feedback throughout the task, which contributes to second language acquisition.

The third way is language learning based on structured pre-programmed dialog. To create a dialog on a specific topic, researchers design their system, rather than adapting a general dialog system. Many studies have been conducted with children. Some had three to six-year-olds learn to read through questions (Xu et al., 2021a,b), and had nine-year-olds answer their questions (Lee and Jeon, 2022). Another related survey is (Huang et al., 2022).

A further related area is user adaptation to difficulty in language tutoring. Pandarova et al. (2019) worked on predicting the difficulty of fill-in-the-blank questions in which the words to be entered were specified.

Our study proposes a new task not addressed in these studies and provides new insights into methods for this task.

2.2 Reinforcement Learning

Reinforcement learning is a machine learning framework that acquires an optimal action policy based on non-instantaneous evaluations given by a reward function for a set of actions. By considering the output tokens as actions, language generation can be treated as a reinforcement learning problem. Given an appropriate reward function, policy gradient methods such as REINFORCE (Williams, 1992) can fine-tune a pre-trained generative neural language model without a training dataset. In this paper, we adopt self-critical sequence training (SCST) (Rennie et al., 2017). SCST is proposed for image caption generation and is known for its simplicity and effectiveness.

The design of the reward function varies from task to task, but unlike the loss function in supervised learning, it allows the use of non-differentiable functions including the evaluation metrics used in text generation tasks such as BLEU and ROUGE (Paulus et al., 2018; Wu et al., 2018; Narasimhan et al., 2016).

Language generation based on deep learning generally uses cross-entropy as the loss function, which means that the objective function and the evaluation measure will be different. By incorporating the evaluation measure in the reward function, the gap can be alleviated.

2.3 Large-scale Pre-trained Language Model

In recent years, many researchers have studied methods for controlling the output of generative language models by providing prompts containing task instructions and examples as input (Li et al., 2022; Reynolds and McDonell, 2021; Dou et al., 2022).

In particular, GPT-3 (Brown et al., 2020) has achieved significant performance comparable to or better than other fine-tuned models in CoQA and TriviaQA in few-shot settings.

3 Method

For the sake of simplicity, in this paper, we assume context $\boldsymbol{C}$ contains only the immediately previous utterance ($n = 1$). We also limit the number of specified items to 1 ($|I| = 1$).

3.1 RL-based generation

For simplicity again, we train a different model for each grammatical item. In applications, we assume the models are to be switched given a learner's need. For example, when a learning partner chatbot finds that the learner tends to make errors with a particular item, the chatbot can increase the frequency of opting the generation model for the item than the vanilla generation model.

We consider three sub-functions for the reward, R_g for inclusion of grammatical items, which is the main objective, R_d for greater diversity, and R_f for higher fluency. The latter two are to mitigate learning bias towards including grammatical items. When only R_g is used, the model easily starts to exploit a fixed utterance against any input context. We will examine several combinations of these functions in our experiment in the next section.

Reward on grammatical items Let $F_i(s) \in [0, 1]$ be a soft classifier that evaluates whether a given sentence s contains a specified grammatical item i. When we train a response generation model for item i, we set $R_g(s) = F_i(s)$.

For $F_i(s)$, we use BERT (Devlin et al., 2019). We obtain hidden representation $\boldsymbol{h}_{[\text{CLS}]}$ of the [CLS] token from the final layer of a pre-trained BERT model. $F_i(s)$ is formulated as follows: $F_i(s) = \sigma(\boldsymbol{w}^\top \boldsymbol{h}_{[\text{CLS}]} + b)$, where $\sigma()$ is the sigmoid function and $\boldsymbol{w}, b$ are the learnable parameters. In training, the BERT model is not frozen and fine-tuned together with the parameters.

Although $F_i(s)$ is trained in a supervised manner, the necessary data for this training is much more affordable than that for training a generation model. We will revisit this point in the next section.

Rewards on diversity and fluency We use Distinct-N (Li et al., 2016), an n-gram based diversity metric, as R_d. As R_f, we use the likelihood of the output r conditioned on the input, i.e., the dialog context $\boldsymbol{C}$. The likelihood is computed by a pre-trained dialog model.

3.2 Prompt-based generation

In the same way with the RL-based approach, we prepare a prompt template for each item i. The templates are to be switched by applications.

Figure 1 shows a prompt template used in this study, which consists of an instruction indicating what the task is, some examples (called shots) and a query at the end. $< c >$ in Figure 1 is replaced with an input context utterance. Given an input prompt, a left-to-right generative language model outputs a sentence r that follows the prompt.

4 Experiment

We verified the effectiveness of both RL-based and prompt-based approaches using three items in the SCoRE corpus (Chujo et al., 2015): the present perfect, relational clause, and subjunctive.

A and B are speaking. Create B's response using the present perfect.
===
A: Good morning, how are you doing today?
B: I have been feeling pretty good, Dr. Smith.
===
A: What's your plan for your future?
B: I'd like to work in a law firm to enrich my experience and put what I've learned into practice.
===
A: I'm going to Japan this year on vacation.
B: Have you ever been to America?
===
A: $< c >$
B:

Figure 1: Prompt template for the present perfect tense

4.1 Datasets

In accordance with the assumption of $n = 1$, we extracted only the first utterance pair of each dialog from the Daily Dialog corpus[1] (Li et al., 2017) to compose our dataset. The first utterance of each pair was used as a context $\boldsymbol{C}$, and the second was used as a reference (used for analysis purposes). We split the pairs into three subsets: 10,618 for training, 500 for development, and 1,000 for test.

We used the SCoRE corpus to build $F_i(\cdot)$. We built a classifier for each of the three items above. Appendix A gives the details of the SCoRE dataset, classifier training, and performance. Note that the required data for training here need not be dialog data and can be much smaller than that for supervised training of a dialog language model.

4.2 Evaluation metrics

We used three metrics for our evaluation. First, we defined the function $\delta_i(s)$, which returns 1 or 0 for sentence s by using $F_i(s)$ with a threshold of 0.5.

As the first metric, we introduced G-ratio to measure the capability of the model to generate responses that include the specified grammatical item. G-Ratio indicates the percentage of outputs containing the item and can be automatically measured by using $\delta_i(s)$.

Considering our aim of exposing learners to various uses of grammatical items in dialog, the model should be able to return diverse responses. We adopted Distinct-N (N=2) as the second metric.

Finally, we defined GOAL (Grammar Oriented Average Likelihood), which measures the fluency of only the generated sentences that contain the specified item using the output likelihood based on

[1] https://huggingface.co/datasets/daily_dialog

a dialog language model P_m as follows:

$$H_i^T = \{s \in G_i^T | \delta(s) = 1\},$$

$$\text{GOAL}(H_i^T; P_m) = \frac{\sum_{s \in H_i^T} P_m(s|c(s))}{|H_i^T|},$$

where G_i^T the set of the generated responses given test set T in terms of item i, and H_i^T is the set of responses in G_i^T that $F_i(\cdot)$ evaluated as containing the grammatical item. $c(s)$ denotes the input context for output s.

4.3 Experimental setups

For the RL-based approach, we used DialoGPT (Zhang et al., 2020), a GPT-2 based dialog language model trained on a Reddit corpus, as the initial model in SCST, the main body of R_f, and P_m. For decoding, we used top-k sampling (Fan et al., 2018) ($k = 50$). The model was evaluated every 10 batches using the development data, and training was stopped with a patience of 3. As training progressed , the number of sentences containing the target grammatical item increased, but many similar sentences were generated, resulting in a loss of diversity. Therefore, as we observed a trade-off between G-Ratio and diversity, we adopted the product of the two as an indicator of early stopping.

For the prompt-based approach, we used GPT-3 davinci. We set the sampling temperature to 1 for GPT-3. Other settings are detailed in Appendix B.

4.4 Evaluation

For the RL-based approach, ten sentences were generated using beam search with a beam width of 10 for each test case. Out of the ten, the sentence with the highest likelihood and the specified item is chosen as the output. If no sentence included the item, the first one was chosen. We compare the following five combinations of the reward functions: R_g, $R_g + R_d$, $R_g \times R_d$, $R_g + R_f$, and $R_g \times R_f$.

For the prompt-based approach, ten sentences were generated thorough the web API using a prompt for each test case, from which one was picked as above. We compared the following five variations, which combines 0, 1, and 3 task examples (called shots) and with/without task instructions: instr., 1-shot, 3-shots, instr.+1-shot, and instr.+3-shots. For example, "instr." means 0-shot with instructions. "1-shot" means 1-shot without instructions. "instr.+3-shot" means 3-shot with instructions.

All metrics were applied to 1,000 outputs.

5 Results

Table 1 shows the results for each grammatical item. Example outputs are shown in Appendix C.

$R_g \times R_f$ showed the highest GOAL for the present perfect and the subjunctive, while $R_g + R_f$ showed the highest GOAL for the relative clause.

The RL-based approach successfully improved G-Ratio in all cases. Although the Dist.-2 values got lower than before training (Baseline), this was expected in advance as the result of introducing a grammatical constraint in generation.

In the RL-based approach, a higher Dist.-2 tended to be obtained with the fluency reward function R_f than with the diversity reward function R_d except for the subjunctive, suggesting that the effect of R_d was limited. The reasons for this may be as follows. Even if sentences with a high Dist.-2 are more likely to be generated, it does not necessarily reflect the diversity of the model overall, and if the input sentences in the batch are similar, Dist.-2 in the output will naturally decrease, but the current reward function does not fully take this into account. In addition, taking fluency into account suppresses the abuse of fixed patterns (fixed patterns increase R_g but decrease diversity). For all grammatical items tested, GOAL improved when the reward function for fluency, R_f, was applied.

In the prompt-based approach, G-Ratio tended to be higher for inputs with both task instruction and shots. However, 3-shots sometimes gave worse results than 1-shot. This suggests that task instructions should be included in the input, but that increasing the number of shots may add noise or unintended bias to the language model, making it more difficult to obtain the desired output.

Comparing the two approaches, the prompt-based one demonstrated higher diversity than the RL-based one, and a comparable G-Ratio. Though the GOAL scores for the RL-based approach were higher than those for the prompt-based approach, we must note that GOAL is favorable to the RL-based approach that, in this paper, uses the same DialoGPT model as GOAL. As far as we manually compared the concrete responses from GPT-3 and DialoGPT for a small number of randomly picked cases, we did not find significant differences.

6 Discussion

Even though we want to expose more instances of a particular item to a learner, it is not natural to include the item in every dialog response. Therefore,

Table 1: Generation results of plain DialoGPT, DialoGPT fine-tuned by RL, and GPT-3 with prompts.

Model	Method	Present perfect			Relative clause			Subjunctive		
		G-Ratio	Dist.-2	GOAL	G-Ratio	Dist.-2	GOAL	G-Ratio	Dist.-2	GOAL
DialoGPT	Baseline	0.145	0.588	0.096	0.822	0.426	0.103	0.037	0.755	0.084
DialoGPT (RL)	w/ R_g	0.789	0.264	0.088	0.911	0.388	0.124	0.860	0.197	0.114
	w/ $R_g + R_d$	0.781	0.121	0.120	0.888	0.355	0.119	0.566	0.182	0.101
	w/ $R_g \times R_d$	0.789	0.265	0.093	0.854	0.411	0.096	0.794	0.207	0.091
	w/ $R_g + R_f$	0.792	0.290	0.107	0.896	0.386	**0.139**	0.941	0.095	0.214
	w/ $R_g \times R_f$	0.603	0.186	**0.147**	0.833	0.420	0.110	0.949	0.036	**0.241**
GPT-3 (prompt)	w/ instr.	0.735	0.681	0.014	0.996	0.682	0.017	0.279	0.737	0.014
	w/ 1-shot	0.493	0.701	0.016	0.992	0.575	0.041	0.568	0.512	0.036
	w/ 3-shots	0.514	0.666	0.027	0.997	0.563	0.038	0.359	0.593	0.033
	w/ instr. + 1-shot	0.901	0.511	0.035	0.997	0.588	0.034	0.721	0.484	0.031
	w/ instr. + 3-shots	0.753	0.594	0.033	0.997	0.571	0.036	0.535	0.539	0.031

we do not need to pursue 100% for G-Ratio.

We presented GOAL as a primary metric candidate for the proposed task. However, as noted in the previous section, it is not reliable when one wants to compare two results based on different language models. Taking the similarity to the reference sentences into account is one direction to mitigate this issue. Another strategy is combining GOAL with reference-free unsupervised dialog evaluation methods using follow-ups such as FULL (De Bruyn et al., 2022). Unlike GOAL, these evaluation methods do not measure the likelihood of the target utterances directly; they, however, still rely on a particular language model. A simple way to make this issue easier would be an ensemble approach using multiple language models or majority voting.

Considering the high diversity and the nature of training-free, so far the prompt-based approach seems to be advantageous, assuming the availability of a huge pre-trained model such as GPT-3. However, the RL-based approach may have merits in terms of its fine-grained, delicate, and implicit control than the prompt-based approach. (Besides, DialoGPT and GPT-2 did not work in the prompt-based approach. See Appendix C.)

7 Conclusion

We have proposed a new task of generating a response including the specified grammatical items for language learners. We examined two approaches and found that both are feasible.

Future directions include the expansion of the grammatical items. To push this task to practical use, locating appropriate places in conversations to include the items is also important.

This paper aimed to increase learners' exposure to specific grammatical items, but another interesting direction is generating preceding utterances that encourage or facilitate learners to use specific grammatical items in their next utterances.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared 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 M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher 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. *CoRR*, abs/2005.14165.

Kiyomi Chujo, Kathryn Oghigian, and Shiro Akasegawa. 2015. A corpus and grammatical browsing system for remedial EFL learners. *Multiple affordances of language corpora for data-driven learning*, pages 109–130.

Maxime De Bruyn, Ehsan Lotfi, Jeska Buhmann, and Walter Daelemans. 2022. Open-domain dialog evaluation using follow-ups likelihood. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 496–504, Gyeongju, Republic of Korea. International Committee on 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.

Gilbert Dizon. 2017. Using intelligent personal assis-

tants for second language learning: A case study of alexa. *TESOL Journal*, 8:811–830.

Gilbert Dizon and Daniel Tang. 2020. Intelligent personal assistants for autonomous second language learning: An investigation of alexa. *The JALT CALL Journal*, 16.

Yao Dou, Maxwell Forbes, Rik Koncel-Kedziorski, Noah A Smith, and Yejin Choi. 2022. Is gpt-3 text indistinguishable from human text? scarecrow: A framework for scrutinizing machine text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7250–7274.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.

Weijiao Huang, Khe Hew, and Luke Fryer. 2022. Chatbots for language learning-are they really useful? a systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*.

Yasutake Ishii and Yukio Tono. 2018. Investigating japanese efl learners' overuse/underuse of english grammar categories and their relevance to cefr levels. In *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference*, pages 160–165.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Na-Young Kim. 2016. Effects of voice chat on efl learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning*, 19(4):63–88.

Seongyong Lee and Jaeho Jeon. 2022. Visualizing a disembodied agent: young efl learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*, 0(0):1–26.

Jiaoda Li, Ryan Cotterell, and Mrinmaya Sachan. 2022. Probing via prompting. *arXiv preprint arXiv:2207.01736*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California. Association for Computational Linguistics.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957*.

Zhaojiang Lin, Andrea Madotto, Yejin Bang, and Pascale Fung. 2021. The adapter-bot: All-in-one controllable conversational model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16081–16083.

Souheila Moussalli and Walcir Cardoso. 2020. Intelligent personal assistants: can they understand and be understood by accented l2 learners? *Computer Assisted Language Learning*, 33(8):865–890.

Ryo Nagata, Tomoya Hashiguchi, and Driss Sadoun. 2020. Is the simplest chatbot effective in english writing learning assistance? In *Computational Linguistics*, pages 245–256, Singapore. Springer Singapore.

Karthik Narasimhan, Adam Yala, and Regina Barzilay. 2016. Improving information extraction by acquiring external evidence with reinforcement learning. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2355–2365, Austin, Texas. Association for Computational Linguistics.

Irina Pandarova, Torben Schmidt, Johannes Hartig, Ahcene Boubekki, Roger Jones, and Ulf Brefeld. 2019. Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education*.

Romain Paulus, Caiming Xiong, and Richard Socher. 2018. A deep reinforced model for abstractive summarization. In *International Conference on Learning Representations*.

Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. 2017. Self-critical sequence training for image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7008–7024.

Laria Reynolds and Kyle McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Sherry Ruan, Angelica Willis, Qianyao Xu, Glenn M Davis, Liwei Jiang, Emma Brunskill, and James A Landay. 2019. Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. In *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, pages 1–4.

Tzu-Yu Tai. 2022. Effects of intelligent personal assistants on efl learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 0(0):1–30.

Stergios Tegos, Stavros Demetriadis, and Thrasyvoulos Tsiatsos. 2014. A configurable conversational agent to trigger students' productive dialogue: a pilot study in the call domain. *International Journal of Artificial Intelligence in Education*, 24(1):62–91.

Veronika Timpe-Laughlin, Tetyana Sydorenko, and Phoebe Daurio. 2020. Using spoken dialogue technology for l2 speaking practice: what do teachers think? *Computer Assisted Language Learning*, 35:1–24.

Ronald J. Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229—256.

Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018. A study of reinforcement learning for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3612–3621, Brussels, Belgium. Association for Computational Linguistics.

Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020. See what i'm saying? comparing intelligent personal assistant use for native and non-native language speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20, New York, NY, USA. Association for Computing Machinery.

Feiwen Xiao, Priscilla Zhao, Hanyue Sha, Dandan Yang, and Mark Warschauer. 2023. Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*.

Ying Xu, Joseph Aubele, Valery Vigil, Andres Bustamante, Young-Suk Kim, and Mark Warschauer. 2021a. Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development*, 93.

Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021b. Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, 161:104059.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, Online. Association for Computational Linguistics.

Supportive Material (Appendices)

A Classifier for grammatical items

We used a classifier that determines whether a grammatical item is included or not as a reward function for RL. The structure of the classifier is as described in §3.1, where the input sentences to be judged are estimated to determine whether they contain grammatical items or not by a linear layer and a sigmoid function based on the embedding of BERT's [CLS] tokens.

The classifier requires a dataset for training. However, the required data need not be interactive, and can be smaller than for supervised learning of a language model. When data is not available, regular expression-based classification can be used as a substitute.

In this section, we describe the dataset used to train the classifier and the settings. The performance of the classifier is compared with rule-based classification using regular expressions. The regular expressions were created on the basis of the CEFR-J regular expression list (Ishii and Tono, 2018).

A.1 SCoRE Corpus

The SCoRE corpus, in which grammatical items are manually assigned to sentences, was used to train the classifier. Therefore, the grammatical items were those included in the SCoRE corpus. The SCoRE corpus contains approximately 20 grammatical items, and Table 2 shows the number of data corresponding to the grammatical items used in this study. For example, in the subjunctive, I wish, if I were, if + verb past tense, if + had + verb past participle, etc. are included in the data.

In addition to positive examples with the target grammar item, negative examples without the item are required to train the classifier. Therefore, for the negative examples, we use sentences in the SCoRE corpus that are assigned grammatical items that are not the target ones.However, if all sentences that do not have the target grammar item are used as negative examples, there is a possibility that unsuitable data will be included, and the proportion of unsuitable data will be greatly biased. We constructed a dataset for training by extracting data from the negative examples in the dataset in such a way that there is no bias in the number of positive examples.

From the data obtained, 80% was split into training data and 20% into test data. Finally, for the present perfect, the training data and test data were 1,222 and 306, respectively, and for the relational clause and hypothetical, the training data and test data were 1,977 and 495, respectively.

Table 2: SCoRE dataset statistics

Grammatical item	# of text
Present Perfect	547
Relative Clause	1,142
Subjunctive	783

Table 3: Results: Classifier Accuracy

Grammatical item	BERT	Regular expression
Present Perfect	0.9902	0.9641
Relative Clause	0.9879	0.6909
Subjunctive	0.9919	0.7394

A.2 Hyperparameter for training the classifier

We used BERT (bert-large-uncased) to set the initial values for the classification model. Parameters were optimized by AdamW during training. The learning rate was set to $2e^{-5}$ and the coefficient of L2 regularization to $1e^{-2}$. The batch size was set to 10 and the number of epochs was set to 10. In this experiment, the classifier is the model that performed best on the test data.

A.3 Classification Performance

Table 3 shows the classification performance of the classifiers for each grammar item. The evaluation was conducted using the percentage of correct answers between the correct and predicted labels as the evaluation measure. In the experiment, the BERT-based classifier was used as the reward function for the other items because BERT had better classification performance than the regular expression.

B Hyperparameter in the experiment

In top-k sampling in SCST, we set k to 50. For Distinct-N in R_d, $N = 2$. The parameters were optimized by AdamW during training, with a learning rate of $2e^{-5}$ and a coefficient of L2 regularization of $1e^{-2}$. The minimum output length was set to 10 in order to properly compute Distinct-N. The batch size was set to 10, with a maximum of 1100 iterations. For GPT-3, we set engine to davinci, max_tokens to 20, temperature to 1, n to 10, and stop to "\n".

Table 4: Generation results for DialoGPT, GPT-2, and GPT-3 with prompts.

Model	Method	Present perfect			Relative clause			Subjunctive		
		G-Ratio	Dist.-2	GOAL	G-Ratio	Dist.-2	GOAL	G-Ratio	Dist.-2	GOAL
DialoGPT	w/ instr. + 1-shot	0.065	0.182	0.108	0.953	0.096	0.091	0.235	0.073	0.081
	w/ instr. + 3-shots	0.569	0.051	0.040	0.960	0.237	0.013	0.049	0.292	0.026
GPT-2	w/ instr. + 1-shot	0.753	0.131	0.012	0.943	0.191	0.029	0.201	0.163	0.007
	w/ instr. + 3-shots	0.638	0.071	0.015	0.955	0.276	0.022	0.253	0.211	0.008
GPT-3	w/ instr. + 1-shot	0.901	0.511	0.035	0.997	0.588	0.034	0.721	0.484	0.031
	w/ instr. + 3-shots	0.753	0.594	0.033	0.997	0.571	0.036	0.535	0.539	0.031

C Examples

In this section, we provide generated sentences of compared methods. First, we discuss additional smaller models we experimented with in addition to the GPT-3. Next, we show samples of outputs for two inputs for several RL-based and prompt-based methods.

C.1 Other Models in the Prompt-based Approach

We also tested the performance of GPT-2 and DialoGPT in the same settings as GPT-3. Table 4 shows the results. Comparing the performance of the three models in terms of G-Ratio, GPT-3, which has the largest model size, shows the best performance, while GPT-2 tends to perform better than DialoGPT. In GOAL, GPT-3 showed consistently high, but DialoGPT also showed high values in some settings. Note, however, that DialoGPT was used in the GOAL calculations and is a favorable indicator for this model. Also, GPT-2 and DialoGPT did not seem to produce higher quality responses than GPT-3, as far as we could visually confirm. (See Appendix C.2) Therefore, GPT-3 is superior to the other models in terms of both the G-Ratio and GOAL value, regardless of the grammatical items, and in terms of the quality of the response sentences.

C.2 Samples

Table 5, 6 show examples of output in the present perfect tense with different input contexts. Compared with the Daily Dialog corpus and DialoGPT, after learning, the response sentences are in the present perfect tense, and the responses of the method that performed well in our experiments are not too broken to be used as a dialog response. However, some of the methods showed unstable output, such as repetition of similar sentences or very few words.

Input context	Look at the show on TV. I am watching a food show at a very famous seafood restaurant. I really want to eat at that restaurant. I am a seafood lover.
Daily dialog (reference)	Speaking of seafood , my mouth is watering. Let's go to the seafood restaurant in our neighborhood.
DialoGPT	I love seafood!
DialoGPT w/ $R_g + R_d$	**I've been** there. **I've been** there. **I've been** there. **I've been** there. **I've been** there. ... **I've been** there. **I've been** there. **I've been** there. **I've been** there. **I've been** there. ... **I've been** there. **I've been** there. **I've been** there. **I've been** there! **I've been** there! ... **I've been** there. **I've been** there. **I've been** there. **I've been** there. **I've been** there! ... **I've been** there. **I've been** there. **I've been** there. **I've been** there. **I've been** there. ...
DialoGPT w/ $R_g \times R_d$	**I've never been** to a seafood restaurant, but **I've heard** good things! **I've never been** to a seafood restaurant, but **I've always wanted** to go to one. **I've never been** to a seafood restaurant, but **I've always wanted** to. **I've never been** to a seafood restaurant, but **I've always wanted** to try one. **I've never been** to a seafood restaurant, but **I've heard** good things about them.
DialoGPT w/ $R_g + R_f$	**I've never been** to a seafood restaurant, but **I've heard** good things. **I've never been** to a seafood restaurant, but **I've heard** good things about them. **I've never been** to a seafood restaurant, but **I've always wanted** to go there. **I've never been** to a seafood restaurant, but **I've always wanted** to try one. **I've never been** to a seafood restaurant, but **I've always wanted** to go to one.
DialoGPT w/ $R_g \times R_f$	**I've never been** to a seafood restaurant. I love seafood. I want to go to that restaurant. **I've never been** to a seafood restaurant, but I've always wanted to. **I've never been** to a seafood restaurant, but I've heard good things.
DialoGPT w/ instr. + three-shots	I That You Weird H
GPT-2 w/ instr. + three-shots	I'm going to Japan this year. Do you have any plans to go to Japan this year? I want to eat at that restaurant. Do you have a plan for your future? I want to eat at that restaurant.
GPT-3 w/ instr.+ one-shots	I think it is kind of expensive to eat at that restaurant . I want to go to that I see, the seafood restaurant's food is always delicious here. I am very hungry right now I **have never heard** of that restaurant . It must be very famous . Yes, that is a very famous food and show but I Love shark fin soup I kind of miss home cooking . I really want to make some of my favorite easy dishes.
GPT-3 w/ instr.+three-shots	**I've never been** to that restaurant for a lunch buffet . I **have been** wanting to have sea food, too . I will take the test . Is there any part that you want to eat first ? I want to go to Japan someday. That sounds really delicious . I love ALL foods, but I like seafood too .

Table 5: Output samples of considered methods and DialoGPT. (1)

Input context	Do you have a package tour to Hollywood?
Daily dialog (reference)	Yes, here is the table, which plans do you like best?
DialoGPT	I don't have a package tour.
DialoGPT w/ $R_g + R_d$	**I've been** there, done that, and done that. **I've been** there, done that, and done that. ... **I've been** there, done that, and done that. **I've been** there, done that. **I've been** there, ... **I've been** there, done that, and done that. **I've been** there, done that. **I've been** there, ... **I've been** there, done that, and done that. **I've been** there, done that, and done that. ... **I've been** there, done that, and done that. **I've been** there, done that. **I've been** there, ...
DialoGPT w/ $R_g \times R_d$	No, **I've never been** to Hollywood. **I've never been** to Hollywood, but **I've heard** good things about it. Not yet, but **I've been** thinking about it. **I've never been** to Hollywood, but **I've heard** good things. **I've never been** to Hollywood, but **I've heard** it's great!
DialoGPT w/ $R_g + R_f$	No, **I've never been** to Hollywood. Not yet, but **I've been** thinking about it. No, **I've never been** to Hollywood before. Not yet, but **I've been** thinking about it! **I've never been** to Hollywood, but **I've heard** good things about it.
DialoGPT w/ $R_g \times R_f$	I don't, but **I've heard** of it. No, **I've never been** to Hollywood. No, but **I've been** thinking about it. No, but **I've been** thinking about doing one. No, **I've never been** to Hollywood, but I've heard good things about it.
DialoGPT w/ instr. + three-shots	I **have not** I **'ve never** I don't Haha, I'd like
GPT-2 w/ instr. + three-shots	I'd like to work in a law firm to enrich my experience and put what **I've learned** into practice. I don't have a package tour. Yes. I would love to. I don't have a package tour to Hollywood.
GPT-3 w/ instr.+ one-shots	No , I don't. However , I do have individual support to resit your broken leg I don't have any right now , I just got back yesterday from a vacation in Hawaii . After taking several Korean movies , I decided I didn't need to go . No , but I just returned from my vacation to Hollywood and Yellowstone Park last night. Yes , I do . And I completely planned on it ,too.
GPT-3 w/ instr. + three-shots	Definitely . In fact, I **have been** living in Hollywood for around twenty days . I'm not sure, but I will contact our office about it. Yes , I do. Would you like to book? Traveling by yourself is more fun than traveling in a group . No, but we have a tour to San Francisco .

Table 6: Output samples of the considered methods and DialoGPT. (2)