An Overview of Natural Language State Representation for Reinforcement Learning

Brielen Madureira [1] **David Schlangen** [1]

Abstract

A suitable state representation is a fundamental part of the learning process in Reinforcement Learning. In various tasks, the state can either be described by natural language or be natural language itself. This survey outlines the strategies used in the literature to build natural language state representations. We appeal for more linguistically interpretable and grounded representations, careful justification of design decisions and evaluation of the effectiveness of different approaches.

1. Introduction

In any typical Reinforcement Learning (RL) scenario, there is an agent in an environment following a policy to take actions which make it transition through states and receive rewards (Sutton & Barto, 1998). Each of these elements must be modeled according to the purpose of the task and their representation can influence the outcome of RL algorithms. Natural Language Understanding can be integrated into these components, both in *language-conditional* RL (when natural language is inherent to the task) and in *language-assisted* RL (when natural language is an additional tool), as discussed in Luketina et al. (2019).

The reward function is usually under the spotlight because the agent's goal is maximizing the expected long-term reward. But the state representation is no less vital. Based on it, the agent perceives the environment and comes to decisions on how to act. While in robotics the agent observes a spatial environment and in arcade games the state may be composed of a sequence of images, natural language is as well a common part of RL states, as illustrated in Figure 1.

The choice of state representation is a problem on its own. It directly affects the learning process (Jones & Canas, 2010), so if applications neglect this, the agent may be prevented from accessing key information for decision-making.

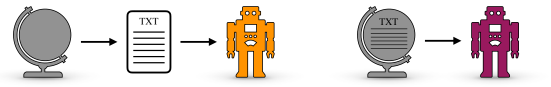

Figure 1. Natural language can describe the agent's state (*e.g.* text-based games) or be part of it (*e.g.* dialogue or text summarization).

There is ongoing research on the use of natural language to model the reward and the actions (He et al., 2016; Feng et al., 2018; Goyal et al., 2019b), but we are not aware of any recent systematic survey about the various possibilities of using natural language to model the state representation and how to do that effectively. We aim at filling this gap by providing an overview of previous work in which the state is based on text, delineating the main approaches.

We dive into a wide range of NLP papers that apply RL methods and whose state representations have to capture linguistic features that influence decision-making. Findings about state representations in this area may potentially be extrapolated to other language-informed RL tasks. We thus hope this overview aids those seeking the objective of having RL agents capable of understanding natural language. Since we notice there is no consensus on how to design natural language state representations, we conclude with some concrete recommendations for future research.

2. State Representation in RL

The construction of suitable state representations for RL has been assessed by several works, for example in the field of robotics and autonomous cars, where modeling the state is particularly hard due to very high dimensional data coming from multiple sensors (Jonschkowski & Brock, 2013; De Bruin et al., 2018). Lesort et al. (2018), for instance, presented a review on State Representation Learning algorithms and evaluation methods for control scenarios.

By definition, the state in a Markov Decision Process, which underlies the formalization of RL, needs to be Markovian; *i.e.* all that needs to be known about the past must be contained in the current state. Jonschkowski & Brock (2013) summarized other desirable properties of state representations proposed by many authors: It should provide good features for learning the value function; be compact but still allowing the original observations to be reconstructed;

[1] University of Potsdam, Potsdam, Germany. Correspondence to: Brielen Madureira <madureiralasota@uni-potsdam.de>.

Accepted to the 1^{st} Workshop on Language in Reinforcement Learning, ICML 2020. Copyright 2020 by the author(s).

change slowly over time; be useful for predicting future observations and rewards given future actions; and ideally be shared by various similar tasks. Although their focus was robotics, these properties can be extended to textual states.

Many other aspects of state representation have been investigated.[1] In early works, states were handcrafted by the system designer using features. Although it remains a common practice, feature engineering has evident drawbacks: it needs an experienced designer, is tedious and time-consuming and does not generalize (Böhmer et al., 2015). The popularization of deep learning can to some extent spare us part of these shortcomings, as raw data can potentially be used as input to end-to-end models that build state representations implicitly. De Bruin et al. (2018), for instance, analyzed methods for integrating state representation learning into deep RL. Still, choosing which information, data structure and model to use can be regarded as a modern form of feature engineering. Several options have been applied to textual data, as we describe in the next section.

3. State Representation in NLP tasks

When RL methods are adopted to solve NLP tasks, one common characteristic is that the state S is represented as a function of texts T, *i.e.* $S = f(T)$. As we will see, f may take a variety of forms and additional inputs. We begin by reviewing the development of natural language state representation in dialogue, instruction following, text-based games and text summarization, NLP areas in which RL methods have been used more often. With the consolidation of deep RL methods, more studies tried to solve several other tasks, which we aggregate in the last part.[2]

Dialogue is likely the NLP area with the most substantial number of papers that adopt RL, with research going on for more than two decades, *e.g.* Levin et al. (1998). In dialogue, the state representation is usually a mapping from text to an abstract or compact format, such as a template, a belief state or an embedding vector, used by the agent to generate the next utterance. A Natural Language Understanding component between the text and the state representation is common, as illustrated, for example, in Lipton et al. (2018).

Delineating the state space was at first mostly a manual task, because considering the entire dialogue used to be impractical (Singh et al., 2000a). Keeping the state space small was a common concern in order to avoid the curse of dimensionality (Levin et al., 2000; Singh et al., 2000a). Therefore, system designers picked a set of variables or features based on their experience, usually resulting in application-dependent representations that did not generalize (Walker, 2000; Levin et al., 2000; Frampton & Lemon, 2005; English & Heeman, 2005; Mitchell et al., 2013; Khouzaimi et al., 2015). The slots and values of the dialogue state were used to represent the environment state (Papangelis, 2012) or the Information States approach was incorporated (Georgila et al., 2005; Henderson et al., 2005; Heeman et al., 2012). It was soon clear that modeling the state space is a fundamental aspect of RL for dialogue, as it has direct impact on the dynamics of the system, and attention was drawn to the fact that the research community had neither established best practices for modeling the state nor agreed upon domain-independent variables (Paek, 2006).

Some studies tried to examine which state representation was more effective for the main task (Scheffler & Young, 2002), constructing them with different feature combinations. There was also discussion on ensuring the Markovian property (Singh et al., 2000b), which means that the "representation must encode everything that the system observed about everything that has happened in the dialogue so far" (Walker, 2000). Ablation studies and metrics to evaluate the effect of each feature were proposed (Frampton & Lemon, 2005; Tetreault & Litman, 2008; Heeman, 2009; Mitchell et al., 2013).

Once neural network models started to be employed, new ways to represent states and to integrate representation learning into the agent's learning were enabled. States could be more easily represented by a vector and fed directly into a parametrized value or policy function. A common approach was building a belief state (a distribution over possible dialogue states, usually represented by a fixed set of slot-value pairs) when dialogue tasks are modeled as POMDPs (Su et al., 2016; Fatemi et al., 2016; Wen et al., 2017b; Weisz et al., 2018). In Wen et al. (2017b), a sequence of free form text was mapped into a fixed set of slot-value pairs by an RNN. Dhingra et al. (2017) compared a handcrafted and a neural belief tracker that uses a GRU over turns. The belief state was sometimes combined with other representations. Wen et al. (2017a) modeled the dialogue state vector as a concatenation of user input (encoded by a BiLSTM), a belief vector (probability distributions over domain specific slot-value pairs, extracted by RNN-CNN belief trackers) and the degree of matching in a knowledge base.

Note that, even when the designer does not have to build the representation manually, the selection of slots and value ranges, the choice of input features and the NLU representations still require human intervention. Features can still be used as input to deep learning models. Williams & Zweig (2016) represented the dialogue history with features (such as input and output entities) as input to an LSTM, which in-

[1] See *e.g.* McCallum (1996); Kaelbling et al. (1998); Finney et al. (2002); Van Otterlo (2002); Morales (2004); Roy et al. (2005); Frommberger (2008); Mahmud (2010); Maillard et al. (2011); Ortiz et al. (2018); François-Lavet et al. (2019).

[2] LSTM (Hochreiter & Schmidhuber, 1997) and GRU (Cho et al., 2014) are references to models mentioned throughout.

ferred the state representation and mapped the input directly to actions. The NLU component played a key role in the state representation *e.g.* in Manuvinakurike et al. (2017).

Cuayáhuitl et al. (2016) pointed out that it is typically unclear what features to incorporate in a multi-domain dialogue and proposed applying deep RL directly to texts, so that the agent learns feature representation and the policy together, bypassing NLU components. Similarly, Li et al. (2016) used the two previous dialogue turns, transformed into a continuous vector representation by an LSTM encoder, to build representations directly from raw text.

A common strategy has been to use the hidden vector of RNNs to represent and update the environment state (Zhao & Eskenazi, 2016; Williams et al., 2017; Liu & Lane, 2017). Multimodal state representation has also become frequent, combining encoded text with images (Das et al., 2017), knowledge base queries (Li et al., 2017; Liu & Lane, 2017) or an embedding of a knowledge graph (Yang et al., 2020).

In **Instruction Following**, deciding the next action means directly interpreting a textual instruction. In this task, states were first designed as tuples of world and linguistic features: words and documents (Branavan et al., 2009; 2010) or cardinal directions and utterances (Vogel & Jurafsky, 2010). We then observed the use of sentence embeddings (usually by an LSTM or GRU) combined with other sources such as image embeddings (Misra et al., 2017; Hermann et al., 2017; Kaplan et al., 2017; Fu et al., 2019) or instruction memory (Oh et al., 2017). A state processing module was introduced in Chaplot et al. (2018) to create a joint representation of the image and the instruction. Janner et al. (2018) converted the instruction text into a real-valued vector and used it both to build a global map-level representation and as a kernel in a convolution operation to obtain a local representation of the state.

In the related vision-language navigation task, the state was modeled by combining visually grounded textual context and textually grounded visual context (Wang et al., 2019; Wang et al., 2020) or by using an attention mechanism over the representations of the instruction (Anderson et al., 2018).

Text-Based Games pose a related, but more general challenge, as here the text describing the current game state must be integrated into the agent state and the best action must be inferred. The work of Narasimhan et al. (2015) was a reference for modeling the state representation in text-based games, which is jointly learned together with action policies in their setting. They used an LSTM over textual data with a mean pooling layer on top, in an attempt to capture the semantics of the game states. Subsequent works followed the same approach (Ansari et al., 2018; Yuan et al., 2018; Jain et al., 2020) or used variants of continuous vectors built by neural networks (He et al., 2016; Foerster et al., 2016). Other elaborate ideas appeared subsequently. Narasimhan et al. (2018) used a factorized state representation, concatenating an object embedding with its textual specification embedding (LSTM or bag-of-words). Ammanabrolu & Riedl (2019) combined the embedded text (by a sliding BiLSTM) with an embedded knowledge graph built by the agent throughout the game. Their concatenation served as input to an MLP that outputted state representations. Murugesan et al. (2020) used embeddings of a local belief graph and a global common sense graph of entities. Zhong et al. (2020) proposed building representations that capture interactions between the goal, a document describing environment dynamics, and environment observations.

Text Summarization is another flourishing area for RL. In this task, an agent processes a text and either pick key sentences to compose a summary or use Natural Language Generation to output its own words. Initial works applying RL used tuples of features to represent the state, composed of the summary at each time step, a history of actions and a binary variable indicating the terminal state (Ryang & Abekawa, 2012; Rioux et al., 2014; Henß et al., 2015).

RL was really consolidated for text summarization after deep learning methods became available. In most cases, the state representation was, as in other tasks, the hidden state of an RNN generating the summary (Ling & Rush, 2017; Pasunuru & Bansal, 2018), also together with an encoding of a candidate sentence (Lee & Lee, 2017). Paulus et al. (2018), a reference for subsequent works (Kryściński et al., 2018), used context vectors with intra-temporal attention over the hidden states of the encoder (which process the original text) concatenated to the hidden state of the decoder (which generates the summary).

Hierarchical approaches are common. Wu & Hu (2018) built a document encoding with a CNN operating at word level and a BiGRU at sentence level, whereas Narayan et al. (2018) and Chen & Bansal (2018) used a CNN at sentence level and an LSTM or BiLSTM at document level to capture global information. Likewise, Yao et al. (2018) employed an RNN or CNN sentence encoder, together with a representation of the current summary and the document content.

Other NLP Tasks comprise each a currently smaller number of studies using RL, so we group them here into main general choices of state design. Some tasks used ordered words (Li et al., 2018, paraphrase generation), predicted words (Grissom II et al., 2014, simultaneous machine translation), a vocabulary set and a taxonomy (Mao et al., 2018, taxonomy induction) or a bag of sentences (Zeng et al., 2018, relation extraction). Coreference resolution adopted partially formed coreference chains (Stoyanov & Eisner, 2012) or word embeddings and features (Clark & Manning, 2016). For syntactic or semantic parsing, the parser configuration was used as set of discrete variables describing the state of a parse structure (Zhang & Chan, 2009; Jiang et al., 2012; Lê & Fokkens, 2017), a concatenation of their representation built by an LSTM (Naseem et al., 2019) or a query for a knowledge base (Liang et al., 2017).

A usual strategy to represent the state was handcrafting vectors composed of selected variables or metrics, for example, similarity scores and number of words (Ling et al., 2017, text-based clinical diagnosis), confidence scores, td-idf and one-hot encoding of matches (Narasimhan et al., 2016; Taniguchi et al., 2018, information extraction), situational and linguistic information (Dethlefs & Cuayáhuitl, 2011, NLG), entitites and relations (Godin et al., 2019, question answering), probability distribution over a set of objects (Hu et al., 2018, question selection) or values generated by a parser (Wang et al., 2018, math word problem).

There has been a myriad of creative efforts to build representations using neural networks, especially RNNs, as one can seamlessly regard the RNN as an agent and its hidden state as the environment state. The internal state (hidden vector and/or cell vector in LSTMs) was set as the environment state in language modeling (Ranzato et al., 2015), image captioning (Rennie et al., 2017), math word problem (Huang et al., 2018) and NLG (Yasui et al., 2019), sometimes with attention over hidden states of a sequence in grammatical error correction (Sakaguchi et al., 2017) or a concatenation of hidden states further encoded by another neural network in sentence representation (Yogatama et al., 2017). A Transformer's hidden states (Vaswani et al., 2017) were also used in machine translation (Wu et al., 2018). The encoding provided by the output layer was used for semantic parsing (Guu et al., 2017) and also with location-based attention for text anonymization (Mosallanezhad et al., 2019). Hierarchical network representations combining word level and sentence level encodings were proposed, as in text classification (Zhang et al., 2018). Both the hidden state of the encoder and the decoder comprised the state in sentence simplification (Zhang & Lapata, 2017). CNNs representations combined with predictive marginals were used in active learning for NER (Fang et al., 2017).

4. Concluding Discussion

The prolific literature combining NLP and RL reveals a noticeable synergy between them. While RL methods have extended frontiers of NLP research, natural language can be a valuable source of information for RL agents, in particular to model the state signal. An abundance of tailored natural language state representations has been explored and there is a current trend to opt for end-to-end models, encoding or decoding text with the aid of RNNs and for multimodal or hierarchical representations, as shown in Figure 2.

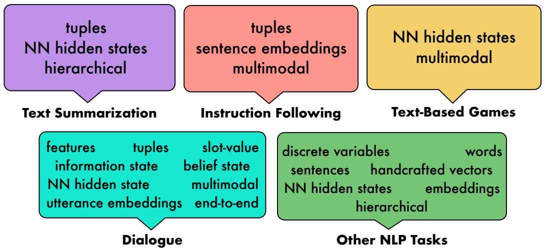

Figure 2. Key concepts of state representation in each area.

More insight is needed on how natural language state representations should be put together. Unfortunately, it is not uncommon to find papers with unclear or missing definitions of their state signal or without much justification of their choices, which hinders comparison and reproducibility. We thus encourage researchers to carefully describe the state representation, provide justification for state design modeling and to also report attempts that had negative impact on the agent's learning process.

Studies to validate the effectiveness of NL state representation used to be common when engineering features was a trend. Deep learning methods may give us the illusion that almost no human intervention or manual feature selection takes place, but they still play a relevant role in state signal design in terms of choice of architecture, model and input data. Adjusting evaluation and ablation methods to new approaches is thus necessary. Goyal et al. (2019a), for instance, experiment with three different natural language instructions representations, to test the effect of language-based reward. Similar analyses can also be done with language-based states, as in Narasimhan et al. (2018).

There is cutting-edge research being conducted about the interpretability of neural NLP models and their linguistic representational power (Belinkov & Glass, 2019). Natural language state representation would thus profit from studies about interpretability and also from diagnostic research (Hupkes et al., 2018) on their abilities of distilling, composing and retaining semantic information throughout the agent's steps, usually expressed by recurrence in the neural networks.

In addition, grounding the meaning of texts to the dynamics of the environment, as discussed in Narasimhan et al. (2018), is a promising area of future research. The authors delineate two needs that must be met in state design with language grounding: the representation should fuse different modalities and capture the compositional nature of language in order to map semantics to the agent's world.

Finally, Narasimhan et al. (2015; 2018) discuss the benefits of representations that are effective across different games and that enable policy transfer. Therefore, another desirable property of natural language state representation is cross-domain validity, so that their encoding of world knowledge can be exploited in various RL scenarios.

Acknowledgments

We thank the two anonymous reviewers for their feedback and suggestions.

References

Ammanabrolu, P. and Riedl, M. Playing text-adventure games with graph-based deep reinforcement learning. 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)*, pp. 3557–3565, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1358. URL https://www.aclweb.org/anthology/N19-1358.

Anderson, P., Wu, Q., Teney, D., Bruce, J., Johnson, M., Sünderhauf, N., Reid, I., Gould, S., and van den Hengel, A. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3674–3683, 2018.

Ansari, G. A., Chandar, S., Ravindran, B., et al. Language expansion in text-based games. *arXiv preprint arXiv:1805.07274*, 2018.

Belinkov, Y. and Glass, J. Analysis methods in neural language processing: A survey. *Transactions of the Association for Computational Linguistics*, 7:49–72, March 2019. doi: 10.1162/tacl_a_00254. URL https://www.aclweb.org/anthology/Q19-1004.

Böhmer, W., Springenberg, J. T., Boedecker, J., Riedmiller, M., and Obermayer, K. Autonomous learning of state representations for control: An emerging field aims to autonomously learn state representations for reinforcement learning agents from their real-world sensor observations. *KI-Künstliche Intelligenz*, 29(4):353–362, 2015.

Branavan, S., Chen, H., Zettlemoyer, L., and Barzilay, R. Reinforcement learning for mapping instructions to actions. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 82–90, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P09-1010.

Branavan, S., Zettlemoyer, L., and Barzilay, R. Reading between the lines: Learning to map high-level instructions to commands. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1268–1277, 2010.

Chaplot, D. S., Sathyendra, K. M., Pasumarthi, R. K., Rajagopal, D., and Salakhutdinov, R. Gated-attention architectures for task-oriented language grounding. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Chen, Y.-C. and Bansal, M. Fast abstractive summarization with reinforce-selected sentence rewriting. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 675–686, 2018.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.

Clark, K. and Manning, C. D. Deep reinforcement learning for mention-ranking coreference models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2256–2262, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1245. URL https://www.aclweb.org/anthology/D16-1245.

Cuayáhuitl, H., Yu, S., Williamson, A., and Carse, J. Deep reinforcement learning for multi-domain dialogue systems. *arXiv preprint arXiv:1611.08675*, 2016.

Das, A., Kottur, S., Moura, J. M. F., Lee, S., and Batra, D. Learning cooperative visual dialog agents with deep reinforcement learning. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

De Bruin, T., Kober, J., Tuyls, K., and Babuška, R. Integrating state representation learning into deep reinforcement learning. *IEEE Robotics and Automation Letters*, 3(3): 1394–1401, 2018.

Dethlefs, N. and Cuayáhuitl, H. Hierarchical reinforcement learning and hidden Markov models for task-oriented natural language generation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 654–659, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P11-2115.

Dhingra, B., Li, L., Li, X., Gao, J., Chen, Y.-N., Ahmad, F., and Deng, L. Towards end-to-end reinforcement learning of dialogue agents for information access. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 484–495, 2017.

English, M. and Heeman, P. Learning mixed initiative dialog strategies by using reinforcement learning on both conversants. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp. 1011–1018, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/H05-1127.

Fang, M., Li, Y., and Cohn, T. Learning how to active learn: A deep reinforcement learning approach. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1063. URL https://www.aclweb.org/anthology/D17-1063.

Fatemi, M., El Asri, L., Schulz, H., He, J., and Suleman, K. Policy networks with two-stage training for dialogue systems. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 101–110, 2016.

Feng, W., Zhuo, H. H., and Kambhampati, S. Extracting action sequences from texts based on deep reinforcement learning. In *27th International Joint Conference on Artificial Intelligence, IJCAI 2018*, pp. 4064–4070. International Joint Conferences on Artificial Intelligence, 2018.

Finney, S., Gardiol, N. H., Kaelbling, L. P., and Oates, T. The thing that we tried didn't work very well: Deictic representation in reinforcement learning. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pp. 154–161, 2002.

Foerster, J., Assael, I. A., de Freitas, N., and Whiteson, S. Learning to communicate with deep multi-agent reinforcement learning. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 2137–2145. Curran Associates, Inc., 2016.

Frampton, M. and Lemon, O. Reinforcement learning of dialogue strategies using the user's last dialogue act. In *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2005.

François-Lavet, V., Bengio, Y., Precup, D., and Pineau, J. Combined reinforcement learning via abstract representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3582–3589, 2019.

Frommberger, L. Learning to behave in space: A qualitative spatial representation for robot navigation with reinforcement learning. *International Journal on Artificial Intelligence Tools*, 17(03):465–482, 2008.

Fu, J., Korattikara, A., Levine, S., and Guadarrama, S. From language to goals: Inverse reinforcement learning for vision-based instruction following. In *International Conference on Learning Representations*, 2019.

Georgila, K., Henderson, J., and Lemon, O. Learning user simulations for information state update dialogue systems. In *Ninth European Conference on Speech Communication and Technology*, 2005.

Godin, F., Kumar, A., and Mittal, A. Learning when not to answer: A ternary reward structure for reinforcement learning based question answering. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*, pp. 122–129, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-2016. URL https://www.aclweb.org/anthology/N19-2016.

Goyal, P., Niekum, S., and Mooney, R. J. Using natural language for reward shaping in reinforcement learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 2385–2391. International Joint Conferences on Artificial Intelligence Organization, 7 2019a. doi: 10.24963/ijcai.2019/331. URL https://doi.org/10.24963/ijcai.2019/331.

Goyal, P., Niekum, S., and Mooney, R. J. Using natural language for reward shaping in reinforcement learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 2385–2391. AAAI Press, 2019b.

Grissom II, A., He, H., Boyd-Graber, J., Morgan, J., and Daumé III, H. Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1342–1352, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1140. URL https://www.aclweb.org/anthology/D14-1140.

Guu, K., Pasupat, P., Liu, E., and Liang, P. From language to programs: Bridging reinforcement learning and maximum marginal likelihood. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1051–1062, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1097. URL https://www.aclweb.org/anthology/P17-1097.

He, J., Chen, J., He, X., Gao, J., Li, L., Deng, L., and Ostendorf, M. Deep reinforcement learning with a natural language action space. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1621–1630, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1153. URL https://www.aclweb.org/anthology/P16-1153.

Heeman, P. A. Representing the reinforcement learning state in a negotiation dialogue. In *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, pp. 450–455. IEEE, 2009.

Heeman, P. A., Fryer, J., Lunsford, R., Rueckert, A., and Selfridge, E. Using reinforcement learning for dialogue management policies: Towards understanding mdp violations and convergence. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

Henderson, J., Lemon, O., and Georgila, K. Hybrid reinforcement/supervised learning for dialogue policies from communicator data. In *IJCAI workshop on knowledge and reasoning in practical dialogue systems*, pp. 68–75, 2005.

Henß, S., Mieskes, M., and Gurevych, I. A reinforcement learning approach for adaptive single-and multi-document summarization. In *GSCL*, pp. 3–12, 2015.

Hermann, K. M., Hill, F., Green, S., Wang, F., Faulkner, R., Soyer, H., Szepesvari, D., Czarnecki, W. M., Jaderberg, M., Teplyashin, D., et al. Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*, 2017.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Hu, H., Wu, X., Luo, B., Tao, C., Xu, C., Wu, W., and Chen, Z. Playing 20 question game with policy-based reinforcement learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3233–3242, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1361. URL https://www.aclweb.org/anthology/D18-1361.

Huang, D., Liu, J., Lin, C.-Y., and Yin, J. Neural math word problem solver with reinforcement learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 213–223, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/C18-1018.

Hupkes, D., Veldhoen, S., and Zuidema, W. Visualisation and 'diagnostic classifiers' reveal how recurrent and recursive neural networks process hierarchical structure. *Journal of Artificial Intelligence Research*, 61:907–926, 2018.

Jain, V., Fedus, W., Larochelle, H., Precup, D., and Bellemare, M. G. Algorithmic improvements for deep reinforcement learning applied to interactive fiction. In *Thirty-Fourth AAAI Conference on Artificial Intelligence (forthcoming)*, 2020.

Janner, M., Narasimhan, K., and Barzilay, R. Representation learning for grounded spatial reasoning. *Transactions of the Association for Computational Linguistics*, 6:49–61, 2018.

Jiang, J., Teichert, A., Eisner, J., and Daume, H. Learned prioritization for trading off accuracy and speed. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 25*, pp. 1331–1339. Curran Associates, Inc., 2012.

Jones, M. and Canas, F. Integrating reinforcement learning with models of representation learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32, 2010.

Jonschkowski, R. and Brock, O. Learning task-specific state representations by maximizing slowness and predictability. In *6th international workshop on evolutionary and reinforcement learning for autonomous robot systems (ERLARS)*, 2013.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

Kaplan, R., Sauer, C., and Sosa, A. Beating Atari with natural language guided reinforcement learning. *arXiv preprint arXiv:1704.05539*, 2017.

Khouzaimi, H., Laroche, R., and Lefèvre, F. Optimising turn-taking strategies with reinforcement learning. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 315–324, Prague, Czech Republic, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-4643. URL https://www.aclweb.org/anthology/W15-4643.

Kryściński, W., Paulus, R., Xiong, C., and Socher, R. Improving abstraction in text summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1808–1817, 2018.

Lê, M. and Fokkens, A. Tackling error propagation through reinforcement learning: A case of greedy dependency parsing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 677–687, Valencia, Spain, April 2017. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/E17-1064.

Lee, G. H. and Lee, K. J. Automatic text summarization using reinforcement learning with embedding features. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 193–197, 2017.

Lesort, T., Díaz-Rodríguez, N., Goudou, J.-F., and Filliat, D. State representation learning for control: An overview. *Neural Networks*, 108:379–392, 2018.

Levin, E., Pieraccini, R., and Eckert, W. Using Markov decision process for learning dialogue strategies. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, volume 1, pp. 201–204. IEEE, 1998.

Levin, E., Pieraccini, R., and Eckert, W. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on speech and audio processing*, 8(1):11–23, 2000.

Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M., and Gao, J. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1192–1202, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1127. URL https://www.aclweb.org/anthology/D16-1127.

Li, X., Chen, Y.-N., Li, L., Gao, J., and Celikyilmaz, A. End-to-end task-completion neural dialogue systems. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 733–743, 2017.

Li, Z., Jiang, X., Shang, L., and Li, H. Paraphrase generation with deep reinforcement learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3865–3878, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1421. URL https://www.aclweb.org/anthology/D18-1421.

Liang, C., Berant, J., Le, Q., Forbus, K. D., and Lao, N. Neural symbolic machines: Learning semantic parsers on Freebase with weak supervision. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 23–33, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1003. URL https://www.aclweb.org/anthology/P17-1003.

Ling, J. and Rush, A. M. Coarse-to-fine attention models for document summarization. In *Proceedings of the Workshop on New Frontiers in Summarization*, pp. 33–42, 2017.

Ling, Y., Hasan, S. A., Datla, V., Qadir, A., Lee, K., Liu, J., and Farri, O. Learning to diagnose: Assimilating clinical narratives using deep reinforcement learning. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 895–905, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing. URL https://www.aclweb.org/anthology/I17-1090.

Lipton, Z., Li, X., Gao, J., Li, L., Ahmed, F., and Deng, L. BBQ-networks: Efficient exploration in deep reinforcement learning for task-oriented dialogue systems. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Liu, B. and Lane, I. Iterative policy learning in end-to-end trainable task-oriented neural dialog models. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 482–489. IEEE, 2017.

Luketina, J., Nardelli, N., Farquhar, G., Foerster, J., Andreas, J., Grefenstette, E., Whiteson, S., and Rocktäschel, T. A survey of reinforcement learning informed by natural language. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 6309–6317. AAAI Press, 2019.

Mahmud, M. Constructing states for reinforcement learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 727–734, 2010.

Maillard, O.-A., Ryabko, D., and Munos, R. Selecting the state-representation in reinforcement learning. In *Advances in Neural Information Processing Systems*, pp. 2627–2635, 2011.

Manuvinakurike, R., DeVault, D., and Georgila, K. Using reinforcement learning to model incrementality in a fast-paced dialogue game. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 331–341, Saarbrücken, Germany, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-5539. URL https://www.aclweb.org/anthology/W17-5539.

Mao, Y., Ren, X., Shen, J., Gu, X., and Han, J. End-to-end reinforcement learning for automatic taxonomy induction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2462–2472, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1229. URL https://www.aclweb.org/anthology/P18-1229.

McCallum, R. A. Hidden state and reinforcement learning with instance-based state identification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(3):464–473, 1996.

Misra, D., Langford, J., and Artzi, Y. Mapping instructions and visual observations to actions with reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1004–1015, 2017.

Mitchell, C., Boyer, K., and Lester, J. Evaluating state representations for reinforcement learning of turn-taking policies in tutorial dialogue. In *Proceedings of the SIGDIAL 2013 Conference*, pp. 339–343, Metz, France, August 2013. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W13-4052.

Morales, E. F. Relational state abstractions for reinforcement learning. In *Proceedings of the ICML'04 workshop on Relational Reinforcement Learning*, 2004.

Mosallanezhad, A., Beigi, G., and Liu, H. Deep reinforcement learning-based text anonymization against private-attribute inference. 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. 2360–2369, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1240. URL https://www.aclweb.org/anthology/D19-1240.

Murugesan, K., Atzeni, M., Shukla, P., Sachan, M., Kapanipathi, P., and Talamadupula, K. Enhancing text-based reinforcement learning agents with commonsense knowledge. *arXiv preprint arXiv:2005.00811*, 2020.

Narasimhan, K., Kulkarni, T., and Barzilay, R. Language understanding for text-based games using deep reinforcement learning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1–11, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1001. URL https://www.aclweb.org/anthology/D15-1001.

Narasimhan, K., Yala, A., and Barzilay, R. Improving information extraction by acquiring external evidence with reinforcement learning. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2355–2365, Austin, Texas, November 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1261. URL https://www.aclweb.org/anthology/D16-1261.

Narasimhan, K., Barzilay, R., and Jaakkola, T. Grounding language for transfer in deep reinforcement learning. *Journal of Artificial Intelligence Research*, 63:849–874, 2018.

Narayan, S., Cohen, S. B., and Lapata, M. Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1747–1759, 2018.

Naseem, T., Shah, A., Wan, H., Florian, R., Roukos, S., and Ballesteros, M. Rewarding Smatch: Transition-based AMR parsing with reinforcement learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4586–4592, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1451. URL https://www.aclweb.org/anthology/P19-1451.

Oh, J., Singh, S., Lee, H., and Kohli, P. Zero-shot task generalization with multi-task deep reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 2661–2670. JMLR.org, 2017.

Ortiz, J., Balazinska, M., Gehrke, J., and Keerthi, S. S. Learning state representations for query optimization with deep reinforcement learning. In *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning*, pp. 1–4, 2018.

Paek, T. Reinforcement learning for spoken dialogue systems: Comparing strengths and weaknesses for practical deployment. In *Proc. Dialog-on-Dialog Workshop, Interspeech*. Citeseer, 2006.

Papangelis, A. A comparative study of reinforcement learning techniques on dialogue management. In *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 22–31, Avignon, France, April 2012. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/E12-3003.

Pasunuru, R. and Bansal, M. Multi-reward reinforced summarization with saliency and entailment. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 646–653, 2018.

Paulus, R., Xiong, C., and Socher, R. A deep reinforced model for abstractive summarization. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=HkAClQgA-.

Ranzato, M., Chopra, S., Auli, M., and Zaremba, W. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*, 2015.

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., and Goel, V. Self-critical sequence training for image captioning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

Rioux, C., Hasan, S. A., and Chali, Y. Fear the reaper: A system for automatic multi-document summarization with reinforcement learning. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 681–690, 2014.

Roy, N., Gordon, G., and Thrun, S. Finding approximate POMDP solutions through belief compression. *Journal of artificial intelligence research*, 23:1–40, 2005.

Ryang, S. and Abekawa, T. Framework of automatic text summarization using reinforcement learning. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 256–265. Association for Computational Linguistics, 2012.

Sakaguchi, K., Post, M., and Van Durme, B. Grammatical error correction with neural reinforcement learning. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 366–372, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing. URL https://www.aclweb.org/anthology/I17-2062.

Scheffler, K. and Young, S. Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning. In *Proceedings of the second international conference on Human Language Technology Research*, pp. 12–19. Citeseer, 2002.

Singh, S., Kearns, M., Litman, D. J., Walker, M. A., et al. Empirical evaluation of a reinforcement learning spoken dialogue system. In *AAAI/IAAI*, pp. 645–651, 2000a.

Singh, S. P., Kearns, M. J., Litman, D. J., and Walker, M. A. Reinforcement learning for spoken dialogue systems. In *Advances in Neural Information Processing Systems*, pp. 956–962, 2000b.

Stoyanov, V. and Eisner, J. Easy-first coreference resolution. In *Proceedings of COLING 2012*, pp. 2519–2534, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL https://www.aclweb.org/anthology/C12-1154.

Su, P.-H., Gasic, M., Mrksic, N., Rojas-Barahona, L., Ultes, S., Vandyke, D., Wen, T.-H., and Young, S. Continuously learning neural dialogue management. *arXiv preprint arXiv:1606.02689*, 2016.

Sutton, R. and Barto, A. *Reinforcement Learning: An Introduction*. A Bradford book. MIT Press, 1998. ISBN 9780262193986.

Taniguchi, M., Miura, Y., and Ohkuma, T. Joint modeling for query expansion and information extraction with reinforcement learning. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pp. 34–39, Brussels, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5506. URL https://www.aclweb.org/anthology/W18-5506.

Tetreault, J. R. and Litman, D. J. A reinforcement learning approach to evaluating state representations in spoken dialogue systems. *Speech Communication*, 50(8-9):683–696, 2008.

Van Otterlo, M. Relational representations in reinforcement learning: Review and open problems. In *Proceedings of the ICML*, volume 2, 2002.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

Vogel, A. and Jurafsky, D. Learning to follow navigational directions. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 806–814. Association for Computational Linguistics, 2010.

Walker, M. A. An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artificial Intelligence Research*, 12: 387–416, 2000.

Wang, L., Zhang, D., Gao, L., Song, J., Guo, L., and Shen, H. T. MathDQN: Solving arithmetic word problems via deep reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Wang, X., Huang, Q., Celikyilmaz, A., Gao, J., Shen, D., Wang, Y.-F., Wang, W. Y., and Zhang, L. Reinforced cross-modal matching and self-supervised imitation learning for vision-language navigation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6629–6638, 2019.

Wang, X., Huang, Q., Celikyilmaz, A., Gao, J., Shen, D., Wang, Y., Wang, W., and Zhang, L. Vision-language navigation policy learning and adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

Weisz, G., Budzianowski, P., Su, P.-H., and Gašić, M. Sample efficient deep reinforcement learning for dialogue systems with large action spaces. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26 (11):2083–2097, 2018.

Wen, T.-H., Miao, Y., Blunsom, P., and Young, S. Latent intention dialogue models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 3732–3741. JMLR. org, 2017a.

Wen, T.-H., Vandyke, D., Mrkšić, N., Gasic, M., Barahona, L. M. R., Su, P.-H., Ultes, S., and Young, S. A network-based end-to-end trainable task-oriented dialogue system. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 438–449, 2017b.

Williams, J. D. and Zweig, G. End-to-end LSTM-based dialog control optimized with supervised and reinforcement learning. *arXiv preprint arXiv:1606.01269*, 2016.

Williams, J. D., Atui, K. A., and Zweig, G. Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 665–677, 2017.

Wu, L., Tian, F., Qin, T., Lai, J., and Liu, T.-Y. A study of reinforcement learning for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3612–3621, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1397. URL https://www.aclweb.org/anthology/D18-1397.

Wu, Y. and Hu, B. Learning to extract coherent summary via deep reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Yang, K., Kong, X., Wang, Y., Zhang, J., and De Melo, G. Reinforcement learning over knowledge graphs for explainable dialogue intent mining. *IEEE Access*, 8:85348–85358, 2020.

Yao, K., Zhang, L., Luo, T., and Wu, Y. Deep reinforcement learning for extractive document summarization. *Neurocomputing*, 284:52–62, 2018.

Yasui, G., Tsuruoka, Y., and Nagata, M. Using semantic similarity as reward for reinforcement learning in sentence generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 400–406, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-2056. URL https://www.aclweb.org/anthology/P19-2056.

Yogatama, D., Blunsom, P., Dyer, C., Grefenstette, E., and Ling, W. Learning to compose words into sentences with reinforcement learning. In *5th International Conference on Learning Representations (ICLR 2017)*. International Conference on Learning Representations, 2017.

Yuan, X., Côté, M.-A., Sordoni, A., Laroche, R., Combes, R. T. d., Hausknecht, M., and Trischler, A. Counting to explore and generalize in text-based games. *arXiv preprint arXiv:1806.11525*, 2018.

Zeng, X., He, S., Liu, K., and Zhao, J. Large scaled relation extraction with reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Zhang, L. and Chan, K. P. Dependency parsing with energy-based reinforcement learning. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, pp. 234–237, Paris, France, October 2009. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W09-3838.

Zhang, T., Huang, M., and Zhao, L. Learning structured representation for text classification via reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Zhang, X. and Lapata, M. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 584–594, 2017.

Zhao, T. and Eskenazi, M. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 1–10, 2016.

Zhong, V., Rocktäschel, T., and Grefenstette, E. RTFM: Generalising to new environment dynamics via reading. In *International Conference on Learning Representations*, 2020.