

A Language Tutoring Tool based on AI and Paraphrase Detection

Anas Basalamah
Computer Engineering Department
Umm Al-Qura University

Abstract—A language tutoring tool (LTT) helps learning a language through casual human-like conversations. Natural language understanding (NLU) and natural language generation (NLG) are two key components of an LTT. In this paper, we propose a paraphrase detection algorithm that is used as the building block of the NLU. Our proposed tree-LSTM with a self-attention method for paraphrase detection shows accuracy of 87% with a lower parameter of 6.5m, which is much robust and lighter than the other existing paraphrase detection algorithms. Furthermore, we discuss an LTT prototype using the proposed algorithm with having some featured components like- message analysis, grammar detection, dialogue management, and response generation component. Each component is discussed in detail in the methodology section of this paper.

Keywords—LTT; NLG; NLU; paraphrase detection; LSTM

I. INTRODUCTION

The conversation is an effective technique for learning a second language [1]. When people try to learn a new language, they learn it through verbal and written communication, including reading, writing, listening, and speaking skills. The former includes listening and speaking, and the latter includes reading and writing. A partner is mandatory for whatever conversation a learner wants to make. It is easy these days to make friends from different countries, cultures, and tribes with the rise of the internet and social media. People can easily share their cultures and values and can teach their languages to others. However, learning a second language requires regular practice for a long time. It isn't easy to find someone available for live conversation every day for a long time because of different time-zone, work schedules, and many other reasons. Although movies, prerecorded videos, or online courses that provide video lessons can be useful equipment to some extent for understanding a language but not adequate for making conversations.

In recent times, smartphones and computers with internet connection are available to everyone from developed to emerging countries. What if someone develops an LTT that can act as a language expert and help the learner with grammar errors, spelling or syntax errors, dictionary and sentence corrections. People can speak and chat with it at their convenient time to learn a new language. For example, if a traveler visits a country without knowing its languages, it becomes difficult to communicate with the local people. This LTT can help the traveler in such a case. Many researchers have already researched on various conversation applications for both pedagogical and industrial purpose [2], [3], [4] but the idea of language specific

LTTs (language other than English) is relatively new in this domain.

The concept of an LTT comes from the *chatbots* providing 24-hours customer support or technical support on online platforms [2]. *Chatbots* are widely used to serve as a customer service agent in many web-platforms. However, an AI-based LTT is different in functionality comparing to a *chatbot*. A *chatbot* is only capable of conversation related to customer service management whether an AI-based LTT performs conversation related to language tutoring. The LTT consistently learns through conversations with the user. It is a game changer as a language tutor and impacts significantly in learning a second language effectively.

Generally, human agents provide service to the customers in online platforms, but recently they are getting replaced by *chatbots*. Because a *chatbot* saves time, money, energy, and physical costs for the company. Besides, it generates responses with fewer grammatical errors and typos. Since the text-based conversation part is also a feature of the conversation agent, it is important to adapt the mechanism of a *chatbot* in the development of a conversation agent. However, it is never easy to develop such a *chatbot* system to continue a meaningful conversation with customers and solve their problems. The critical issue here is to detect different sentences with the same meaning, i.e., paraphrase detection. Naturally, people use different expressions and structures of sentences to express the same thing when they make conversation in different circumstances. For example-

Sentence-1: "I have to learn this language very quickly."

Sentence-2: "I need to pick up this language as soon as possible."

Clearly, 'Sentence-1' and 'Sentence-2' have two distinct structures but express the same meaning. Paraphrase detection and its natural language understanding (NLU) component play an essential role in detecting these sentences. Therefore, paraphrase detection is considered a vital element of a conversation agent or a dialogue system. The next important step is natural language generation (NLG), as the system has to respond to customers after understanding their problems. This paper follows these NLU and NLG steps for making a conversation agent that acts as a language learning tool to the user.

We develop an AI-based human-like interactive system that can understand the context of natural language sentences and help learn a new language. When the system is trained on a massive dataset of semantic relatedness, it retains the relative behavior of words and vocabulary in the sentence.

The developed application is efficient for people learning or exploring a second language. A novel tree-LSTM with self-attention is proposed that is implemented as the system core of the developed application. The proposed tree-LSTM with self-attention shows moderate accuracy than the existing models but offers the best efficacy as a lightweight and robust model. Performance comparison of the proposed and existing methods are demonstrated in Section III.

The remaining part of the paper is structured as follows. The methodology of our proposed LTT with its demonstration, components, architecture, and workflow is discussed in Section II. The performance of the proposed method is analyzed and compared in section III. Previous research works on language learning agent and paraphrase detection are studied in Section IV, and finally, the paper is concluded in Section V.

II. METHODOLOGY

This section describes the components, architecture, and workflow of the LTT. In learning, tourism, visiting, hosteling, and exposing to new places, LTTs are very efficient as the way of interaction is human-like. Additionally, the mapping of sentences to intent makes it more flexible for users to interact by using any phrases. Contextual conversation makes the user or learner feel more active and realistic. A user-LTT sample communication is depicted in Fig. 1. Fig. 1 illustrates how the grammar detection component of an LTT corrects grammatical errors and helps with dictionary.

The primary goal is to create a task-oriented AI-based contextually aware LTT which can interact with humans naturally. This research proposes and trains the system to learn a second language through chat or voice LTT.

Instead of traditional language learning platforms or apps, the LTT will communicate with the learner. It is inspired by the natural phenomenon that humans can learn a new language better when he visits a new place and communicates with natives. In this situation, the primary concern of the visitor is to convey his intention through a message. This case study motivates to create a contextually aware platform that can talk, understand and reply to the user more personally and intelligently. Unlike Google assistant or Alexa, the system does not need to be a knowledge-based assistant; it will be solely trained for language practicing. User will share his feeling: e.g., "how was the day, what he took in the lunch, what he observed today" in second language, and the bot would extract the context, understand the user and reply appropriately. The proposed system can recognize the same sentence in multiple ways as it is trained to paraphrase detection. The language learning contextual system would teach the language to the learner by analyzing the input sentence and informing the user about sentence accuracy.

Fig. 2 represents the machine learning-based architecture of the LTT. The message analysis component (as NLU) extracts entities and classifies intents. This module recognizes the intention of the user and the semantic of the sentence. The grammar detection component checks grammar and punctuation errors, correct sentences, and acts as a dictionary. The dialogue management module tracks the text, implements relevant policy, and triggers the appropriate action based on the database. Finally, the Response generation component provides

the corresponding response of the query based on the context. This system communicates with user training on the proposed architecture based on tree-LSTM and attention.

RvNNs are efficient for natural language tasks but require structural input, which is hard for data preparation and implementation. This paper proposes a natural language task-specific model that can compose a tree structure of a plain text. The architecture of the system consists of Tree Long-Short term memory and Attention mechanism. Tree-LSTM is a refined version of RvNN and LSTM that manages the information flow from child to parent more structurally than LSTM. In tree-LSTM, the cell state computes parent representation by utilizing cell dependencies at a vertical distance.

Tree-LSTM extracts information from unstructured sentences by creating a structure of input dynamically. The information about the structure of the input is not provided to the Tree-LSTM model. Therefore, an extra process is required to compose task-specific tree structures of the input sentence. In the equation, $q.h$ express the score of representation validity $r = [h; c]$.

The trainable composition query vector is required to measure the representation validity for building the tree structure from the unstructured sentence. At the first layer, the candidacy for the parent representation is computed using equations 1 to 3. Query vector calculates the validity score for each candidate. Then, the validity of each candidate is prioritized based on their score. During training, the softmax estimator samples the parent node among candidates weighted by the query vectors, and for the testing, the model itself selects the candidate with the highest validity.

$$\begin{bmatrix} i \\ f_l \\ f_r \\ o \\ g \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(w_{comp} \begin{bmatrix} h_l \\ h_r \end{bmatrix} + b_{comp} \right) \quad (1)$$

$$c_p = f_l \odot c_l + f_r \odot c_r + i \odot g \quad (2)$$

$$h_p = \mathbf{0} \odot \tanh(c_p) \quad (3)$$

$$v_i = \frac{\exp(q \cdot \tilde{h}_i^{t+1})}{\sum_{j=1}^{M_{t+1}} \exp(q \cdot \tilde{h}_j^{t+1})} \quad (4)$$

Additional components are also added to encode the input sentences into vectors dynamically using the bottom-up tree structure technique. We focused on a type-specific attention mechanism that aims to encourage the model to focus on salient latent information of the composition or constituency tree relevant to the classification decision. We denoted the output of the intermediate nodes and leaf nodes in the layers of the tree-LSTM as $\rightarrow_{h_1}, \rightarrow_{h_2}, \dots, \rightarrow_{h_{2n-1}}$ where 'n' is the sentence length. The information of the architecture is as follows:

$$\tilde{a}_i = \exp(W_a \times h_i) \quad (5)$$

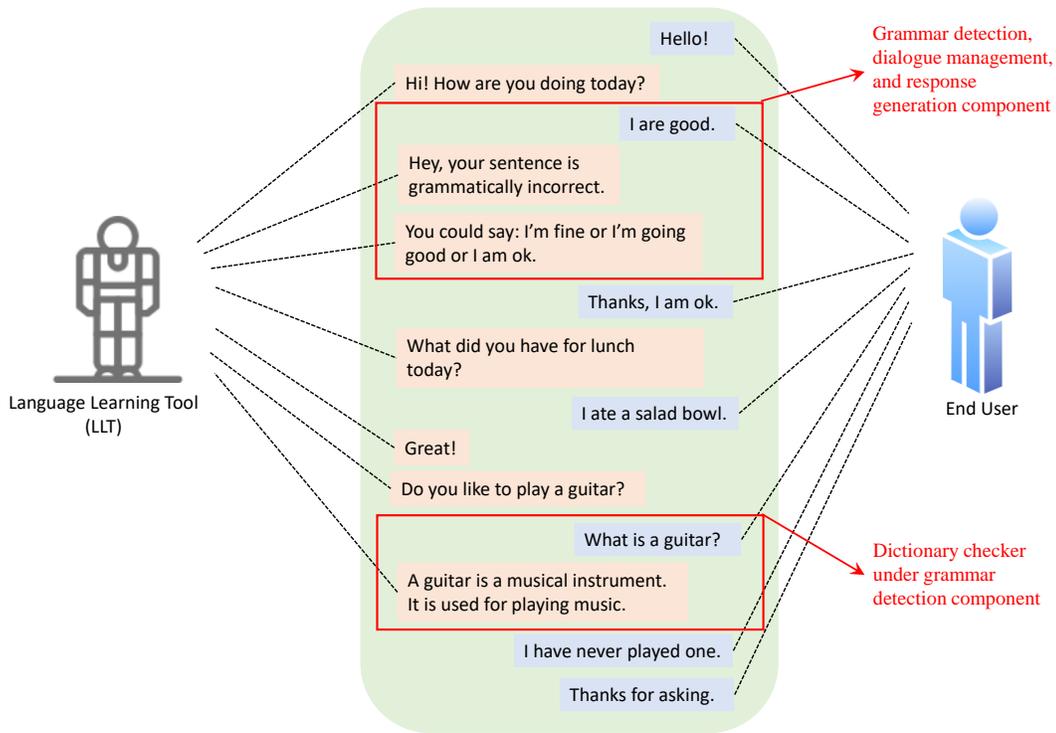


Fig. 1. A Sample Chat Window of End user with the LTT.

$$a_i = \frac{\tilde{a}}{\sum_{i=1}^{2n-1} (\tilde{a})} \quad (6)$$

$$\vec{H}_c = \sum_{i=1}^{2n-1} (a_i \vec{h}_i) \quad (7)$$

In our architecture, the context vector H_c is learned using attention-weighted contextual representation by passing Node vectors to the fully connected layer with D_a hidden units in a hidden layer. For each such hidden state vector e_i , a scalar unit a_i is learnt with the weight matrix $W_a \in R^{1 \times D_a}$. These scalar units are then normalized to 1, and the whole formulation of the context word vector is defined in equation 7. The process is visually summarized in the diagram of the architecture in Fig. 3.

III. EXPERIMENTAL RESULTS

The section describes a complete experimental setup, hyperparameters, and dataset to analyze and evaluate the proposed model for paraphrasing and contextual similarity in conversation agents. The natural language inference (NLI) task is conducted to investigate the semantic relationship between two sentences. The proposed model is trained on the standard SNLI dataset that contains 570K pairs of English sentences.

The SNLI dataset* and Glove embedding* are available at resources. The model is trained on a large dataset containing sentences and learns from vector representation considering grammar and language constraints. The GPU Geforce RTX

2080 ti with 11GB memory and Cuda version 10.1 was used to train the model. Experimental results are evaluated in the table. The accuracy matrix in the table is employed to assess the model's efficiency and compare the results. The proposed model is reached based on pre-trained vectors' dimensions and several parameters used in training.

The presented table shows that Bi-LSTM with inner Attention is one of the stable models that use fewer parameters, 2.8m trained on 600D word vectors. The model achieved 84.2% testing accuracy, which is reasonably comparable with other models. The DiSAN uses 300D word vectors and achieves 85.6% accuracy with 2.4m parameters. This model is time efficient and consumes fewer parameters, but the proposed approach would only work for context-aware representations with temporal information encoded. Gumbel tree-LSTM consists of tree-LSTM based on 600D word vectors and attains 86% inadequacy, but the deficiency is that it utilizes 10M parameters. BERT base model is proficient at reaching 90% testing accuracy. However, it stands among the list of large models as it utilizes 335m parameters.

TABLE I. DIFFERENT MODEL RESULTS

Model	Emb Dim	Params	Testing Accuracy
Bi-LSTM with inner Attention	600D	2.8m	84.20%
DiSAN	300D	2.4m	85.6%
Gumbel Tree-LSTM	600D	10m	86%
Tree-LSTM with attention	300D	6.5m	87%
BERT BASE		335m	90.7%

The above discussed models can be utilized for the same task, but they are not entirely appropriate in terms of task,

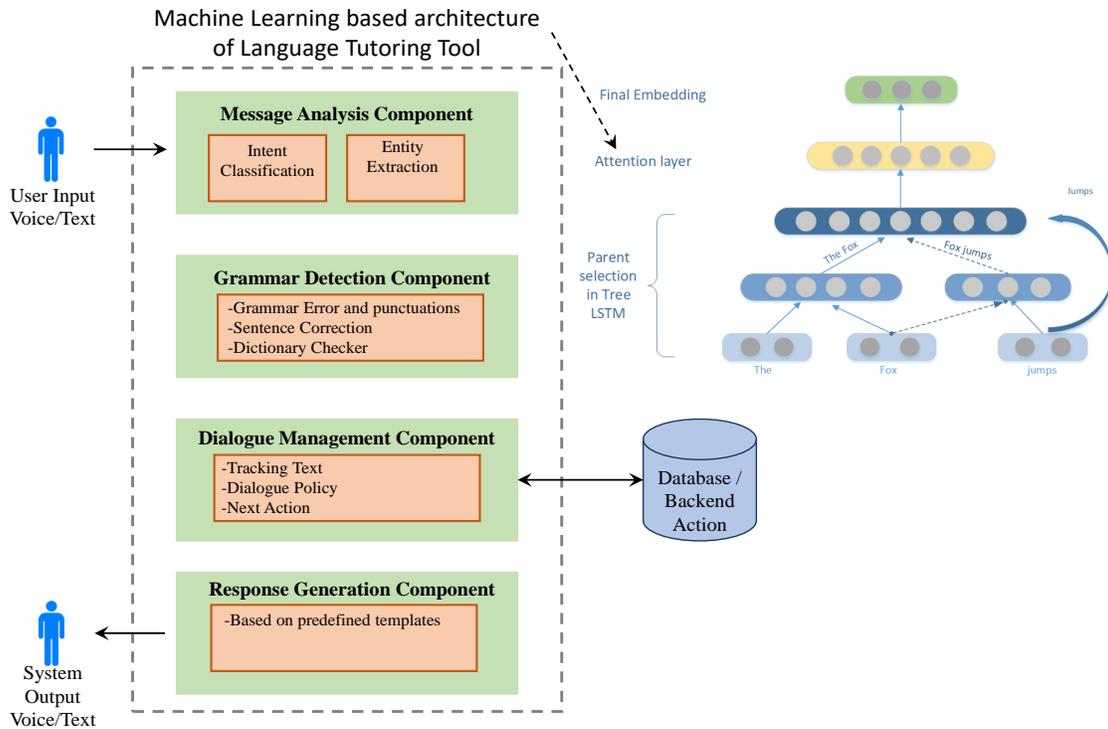


Fig. 2. Primary Components of the LTT based on the Proposed Architecture.

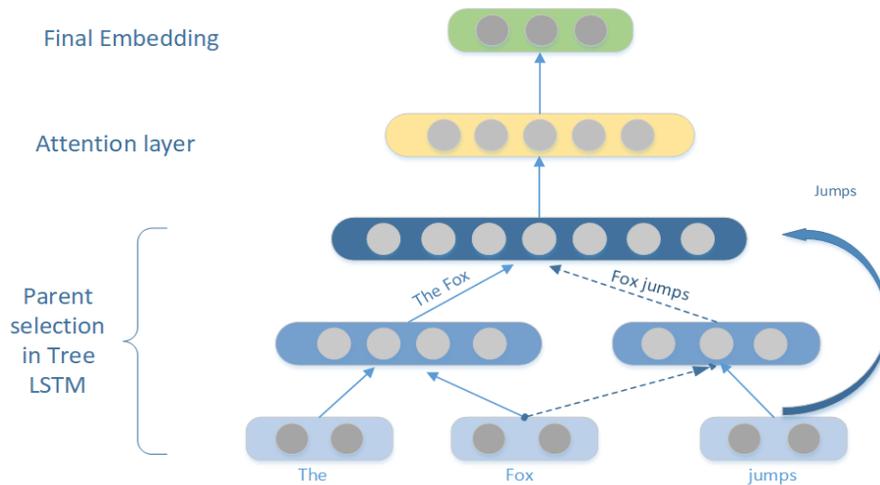


Fig. 3. Proposed Architecture for the LTT.

resources, and parameters. The proposed model is more robust, lightweight, and remembers the hidden states that help it in the question answering system. Therefore, our proposed model is efficient to be used in an LTT.

IV. RELATED WORK

AI-based LTTs have the potential for human-like conversation. They work well for companies that have a lot of data. Despite the fact they need colossal time to train, they do save many in the long run. Based on previous information, they gather the information that learns new patterns from the input

and improves the system after iterative training. Assistants understand patterns and behavior, and have a broader range of decision-making than any person, and over time they get to be very sophisticated. Many researchers investigated foreign language learning using a *chatbot*, where the *chatbot* mainly helps to learn english as a foreign language [5], [6], [7], [8], [9], [10]. Our proposed model is generic and can be used for learning any language if the model is trained with respective languages' data set.

Our proposed platform comes under the task-oriented dialogue system in which end users will learn second language

through conversation. The proposed platform is initially trained on English language which will be extended further using datasets of other languages. However, the NLU and NLG are two must-have components that we needed to develop in the LTT. In the past decades, these modules are implemented individually and optimized using statistical models[11], [12]. Due to rapid and growing advancements in technology and algorithms, deep learning and reinforcement learning have reached the stage of producing human-like conversation agents. The trend to talk with a machine is not as new as more than 50 years ago; an agent is released that tries to maintain a conversation as mentioned in [13]. In recent years, messaging platforms offer developers to custom their chats and develop AI-based LTTs. The integration of AI modules to task oriented *chatbot* can be complicated.

After the vector representation to evaluate the semantics of natural language text, sentence representation came into the ground to perform NLP tasks. The LSTM and its variants (RNN) sequentially process sentences to extract sentence meaning where the output depends on all previous hidden states. Recursive NN is the generalized version of RNN based on the structured input for sentence encoding as it extracts semantics through a hierarchical structure. The number of rvNNs were proposed in Socher et al. [14] which consider phrases instead of the complete sentence.

Matrix-Vector RvNN was reported as the best model that represents the word as vector and matrix at the same time. After that, Tai et al. [15] proposed two variants of standard LSTM (Child sum and N-ary tree-LSTM) that perform better for hierarchical and structural data. Additionally, hidden and cell states do not depend on the entire previous sequences in these variants as only the children's hidden and cell conditions contribute to the parent's states.

The Tree RNNs were extended by Zhou et al. [16] into attention-based tree RNN for semantic relatedness. In this model, semantic similarity is extracted by encoding attention in the tree structure of one sentence with vector representation of other sentences. Attention is used in machine translation by aligning the source or target sentence. It can focus on the most concerning sentences. The pioneers of the machine translation [4], [17] developed a model based on RNN variants that use attention for source words to generate target words.

A task-specific conversational agent is a kind of *chatbot* that tries to imitate a human-like response bypassing the Turing test. Unlike pattern matching and simple algorithm-based *chatbot*, AI-based conversation agents contain complex algorithms and structures. Whether, we present a system in which an AI-based voice bot as a tutor provides services to learn a second language. A person or a student or a learner can speak a sentence multiple-way regardless of the vocabulary's dimension, as the system would efficiently detect the paraphrased sentences. At the beginning stage, when a person is learning the second language, it demands freedom to speak sentences and require hint to improve himself. This system considers the language constraints, grammar knowledge, and contextually aware algorithms trained on paraphrase detection.

V. CONCLUSION

This paper proposed an AI-based system that can perform casual conversation with a user in the second language for teaching and training purposes. The system based on the tree-LSTM with self-attention efficiently extracts contextual information from unstructured sentences and long conversations. The system is trained on a standard SNLI dataset that has been used widely for contextual understanding and paraphrasing detection systems. The experimental results expressed that the proposed model has achieved efficient results compared with state-of-the-art models.

REFERENCES

- [1] A.-M. Barraja-Rohan, "Using conversation analysis in the second language classroom to teach interactional competence," *Language Teaching Research*, vol. 15, no. 4, pp. 479–507, 2011.
- [2] M. Adam, M. Wessel, and A. Benlian, "Ai-based chatbots in customer service and their effects on user compliance," *Electronic Markets*, vol. 31, no. 2, pp. 427–445, 2021.
- [3] N. Haristiani, "Artificial intelligence (ai) chatbot as language learning medium: An inquiry," in *Journal of Physics: Conference Series*, vol. 1387, no. 1. IOP Publishing, 2019, p. 012020.
- [4] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [5] J. Jia, "The study of the application of a web-based chatbot system on the teaching of foreign languages," in *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 2004, pp. 1201–1207.
- [6] L. Fryer and R. Carpenter, "Bots as language learning tools," *Language Learning & Technology*, vol. 10, no. 3, pp. 8–14, 2006.
- [7] J. Jia, "Csiec: A computer assisted english learning chatbot based on textual knowledge and reasoning," *Knowledge-Based Systems*, vol. 22, no. 4, pp. 249–255, 2009.
- [8] Y. Goda, M. Yamada, H. Matsukawa, K. Hata, and S. Yasunami, "Conversation with a chatbot before an online efl group discussion and the effects on critical thinking," *The journal of information and systems in education*, vol. 13, no. 1, pp. 1–7, 2014.
- [9] L. K. Fryer, M. Ainley, A. Thompson, A. Gibson, and Z. Sherlock, "Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners," *Computers in Human Behavior*, vol. 75, pp. 461–468, 2017.
- [10] D. Coniam, "Evaluating the language resources of chatbots for their potential in english as a second language," *ReCALL*, vol. 20, no. 1, pp. 98–116, 2008.
- [11] G. Tur and R. De Mori, *Spoken language understanding: Systems for extracting semantic information from speech*. John Wiley & Sons, 2011.
- [12] S. Young, M. Gašić, B. Thomson, and J. D. Williams, "Pomdp-based statistical spoken dialog systems: A review," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013.
- [13] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
- [14] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011, pp. 151–161.
- [15] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.
- [16] Y. Zhou, C. Liu, and Y. Pan, "Modelling sentence pairs with tree-structured attentive encoder," *arXiv preprint arXiv:1610.02806*, 2016.
- [17] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.