Learn to Speak Like A Native: AI-powered Chatbot Simulating Natural Conversation for Language Tutoring

Jianhong Tu
Guangdong Experimental High School, Guangzhou, Guangdong, 510375, China.
tovitusaikou@gmail.com, sheppaj5849@student.laccd.edu

Abstract. In a globalizing world, the demand for acquiring a new language is expanding at a fast pace. However, language learners who have little opportunity to use the targeted language in their daily life are unable to increase their familiarity with the language given that contextual conversation is proven helpful in language acquisition. The current study dedicates to address this problem by proposing a language learning chatbot with which one can practice in a simulated setting. The chatbot is based on the deep neural network and is trained by MultiWOZ and personaChat collected in previous works. One chatbot structure is proposed to optimize the performance of the DNN chatbot. In comparison with the conventional chatbot, the new structure features a higher accuracy rate of generating appropriate utterance and suffers less from overfitting. Both machine and human evaluation show that the chatbot is able to produce transactional conversations that cover multiple domains of daily life conversations and incorporates a large vocabulary base as well as diverse sentence structures, which are beneficial for students.

1. Introduction
As globalization proceeds, English, as a language, has rapidly gained importance and popularity in a global range. Unlike other languages which are barely used outside their original country, English continued to serve as a tool with unifying force in an ever globalizing and diversifying world, and thus a “global language” [1]. John Knagg, a leading representative of the British Council, reported that the English as Second Language (ESL) learners across the world amounts to 1.5 billion and growing in the “Teaching English to Speakers of Other Languages 2014”, thus showing the great demand of English learning. However, all language educators still face difficulties. Among all, the lack of an English learning environment outweighs others and remains the hardest to solve. Educators in the pedagogy field demonstrate frustration that ESL learners are by no means exposed to interactive activities including listening or speaking [2]. Students with no assess to English speakers not only fail to master their skills through need-based conversation, but also gradually lose motivation in language learning for its seemingly trivial practical value. The more goal-oriented teaching style in countries like China and Saudi Arabia, which aim to accomplish a certain task such as examination, jeopardizes students’ ability to use the language [3]. The disproportional time distributed in practicing reading, speaking, listening, or writing results in stronger reading and writing abilities but much weaker speaking and listening skills [4]. A means to foster communicational skills such as providing a proper English-speaking environment is needed by the current pedagogy system. The positive effect of conversational interactions has been supported by numerous study findings. Exposing students to natural conversation benefits them by giving more chances to assimilate and make utterances and at the same time strengthening their motivation of learning the language [4]. Given that, teachers are suggested that English itself should be
part of the instructions in classes and they should facilitate English speaking by creating an appropriate environment where they address each other in English [5]. Acquiring background knowledge about the culture as a by-product of learning language awards students with a deeper understanding of the language and in turn faster their learning progress.

Lately, chatbots powered by artificial intelligence technology underwent rapid development. Artificial Neural Networks (ANN) as the lying-behind mechanisms of sophisticated chatbots take in sample utterances along with target response and try making links between them through mathematical computation. The model generated by automated programs then are capable of reverse-engineering new inputs and producing relevant answers or follow-up sentences like a question-answer (QA) system. Based on the different fundamental architectures of the chatbots, they are classified into Deep Neural Network, Sequence to Sequence (Seq2Seq), and other models. Chatbots can be used in a wide variety of applications. Currently, taking advantage of demonstrating more patience over human and having the ability to interact with human beings using natural languages, transactional chatbots are deployed in social media platforms, such as twitter, to provide customer services [6]. The fine-tuned AI models pay attention to people’s diction and tone to adjust their responses to various perceived sentiments [7]. Their speech abilities are so well recognized that they are used to train people on interviewing strategies [8].

In addition, the value of ANN is not restricted to mimic speaking solely. Speaking in an abstract way, the power of neural networks lies in the ability to learn from abundant existing patterns, infer from information, and adapt on a larger scale. The case where the same seq2seq model is used to discover new medicine further proves its flexibility to be applied in different working contexts [9].

In this study, we decide to leap forward and utilize the natural language processing technique to build a chatbot specifically used as a robot tutor for ESL learners in a non-English-speaking setting. The bot aims to provide contextual conversations, which are an essential part of language teaching that current schools fail to include. The rationale behind this is to cultivate students’ ability to instantly form a sentence in response to a random question in a specific scenario through simulating a daily context with detailed information such as real-life location, a goal of the conversation and unexpected problems. The mechanism of the chatbot will be discussed in section 4. The proposed chatbot replies and actively engages with the learner by implying the theme of the next utterance together with its response. In order to ensure the maximum freedom of topics, the dataset used embeds multiple domains and contexts. Meanwhile, the role of the context is reinforced so that any utterance irrelevant to the setting would be simply ignored, and a warning indicating confusion is returned, thus to some extent limiting the scope of conversations. The answers from the model are designed to incorporate many utterances with the same meaning, allowing a diversity of word choices, common phrases, and sentence structures. Through multiple rounds of interaction with the chatbot, the learner is expected to comprehend and use any new words or phrases they encounter, thus improving their English. As will be presented in section 5, two architectures, “integrated structure” and “network structure”, are proposed, and their effectiveness is compared. The former features one single neural network model, while the latter is a combination of multiple interconnected models. The final evaluation tests their ability to formulate a fluent conversation. Also, the accuracy of the next utterance generated is considered. Both automated metrics and human evaluations are then conducted to analyze its effectiveness.

1.1. Related Works

Traditional chatbots can be divided into two separate categories according to their function—transactional and conversational. Two examples of the former are Alexa and Siri, which are developed by Amazon and Apple, respectively. The primary duty of these chatbots is to carry out a specific action according to its understanding of users’ demands. Such assistant work includes sending a message to a designated user. In these cases, despite that the agents are capable of formulating human-like sentences, they are majorly serving as a notification reporting the status of their commands, whether they are successful or not, thus with little intention to interact with users. Contexts and variation of responses tend to be sacrificed for convenience to use. For many transactional chatbots, the dedicated dataset used
to feed the models are prone to be confined to a small scale, enlarging the discrepancy between perceived chatbot speeches and real human responses [10].

Conversational chatbots are generally considered to be more complex because they are meant to generate smooth natural conversation, which drives it to approach human forms of speech and pay more attention to the nuanced points like opinions as well as variations in tones, knowledge, speaking context, and memory. To increase interactions among users, chatbots are even equipped with a persona core, which provides views from all aspects of an imagined character and shifts their replies to questions. Eliza bot, as an early successful case, was created to conduct mental therapy through caring conversation, similarly to traditional human mental therapists [11]. However, one noticeable disadvantage of it or its later counterparts is that they are hand-coded rather than program-generated, restricting the scalability of the chatbot.

Apart from aiding therapy services, conversational chatbots also have begun to assume the role of teaching. Alicebot, an ALML structure chatbot based on a corpus, achieved to animate or visualize a corpus for the purpose of learning language [12]. A similar approach is also seen in other bots dedicated to language teaching. These bots have demonstrated the characteristics of machine-driven chatbot and have several advantages over human beings. For example, they are patient enough to repeat the same utterance over and over and are accessible at anytime and anywhere, totally eliminating physical blocks. The study also shows that, in comparison to human tutors, chatbots are more welcome by students who thus become more motivated due to the novelty effect [13]. At the same time, the flaws of these bots with similar ALML format are also observable. They are unaware of the actual meaning of the conversations [14], and predesigned conversations give little freedom for students to freely express themselves [15]. In particular, the chatbot fails to accomplish the teaching quest because of their insensitivity to any grammatical errors in students’ sentences (14). Other chatbot using Chatfuel, a market chatbot deployed in Facebook, report the same problem of strict conversation flow that it refuses to move on or hints the users but repeat the same sentence until they closely focus on the topic [16]. In order to create a smooth context-directed conversation for students to learn from, utterances from both sides of the speech need to be carefully structured and guided. Otherwise, unintended flaws that may impair the authenticity of the communication would distract students from properly studying the essential points. The purpose of our chatbot is to engage the users in a daily setting with conversations of multiple themes. Additionally, a need for persona to be added to chatbots is expressed and valued in its ability to engage with users [7].

2. Method

2.1. Datasets
The model is applied on two separate databases: the personaChat dataset and the MultiWOZ dataset. In combination, they make up the inner personality core and a knowledge base, respectively. Users are expected to perceive the distinctive personality of the chatbot in each round and fulfill a specific task with the bot. The entire system is designed to be dynamic but consistent throughout each conversation until its end.

The MultiWOZ dataset is a multi-domain dialogue corpus that covers 7 distinct fields and contains more than 10,000 dialogues, collected by a team led by Mihail Eric [17]. The latest 2.1 version, or MultiWOZ2.1, is available at www.repository.cam.ac.uk/bitstream/handle/1810/294507/MULTIWOZ2.1.zip?sequence=1&isAllowed=y. The dataset is collected from human-to-human conversations, resolving daily issues, or accidental emergencies. The subjects included in each conversation are required to perform a task chosen from one of the seven topics. The topics include asking for help from a police officer, booking a hotel room, calling a taxi, enquiring information about local attractions, and looking for a restaurant. One plays the enquirer, while the other serves as the helper. The former role always starts the conversation by explaining his or her need and requirements. The other would then reply with detailed information, which includes postcode, phone number, information of the target service, and realistic names. Once the
enquirer feels satisfied with the outcome or given information, he or she simply exit the conversation with a brief greeting as an ending sign. Despite that the general tasks to accomplish are common, the sequence and whether a request ends successfully are random, such that conversations vary from each other. Each turn of conversation is labeled with occurring scenarios and requirements, which are useful for constructing the model.

Table 1. Examples of conversations from MultiWOZ2.1

| Scenarios | Conversation | requirements |
|-----------|-------------|--------------|
| hotel     | User: I am looking for a place to stay that has a cheap price range it should be in a type of hotel | Hotel, cheap |
|           | Agent: Okay, do you have a specific area you want to stay in? | Parking |
|           | User: no, I just need to make sure it's cheap. oh, and I need parking | |
|           | Agent: I found 1 cheap hotel for you that includes parking. Do you like me to book it? | |
|           | User: Yes, please. 6 people 3 nights starting on Tuesday | Tuesday, 6 people, 3 days |
|           | Agent: I am sorry, but I wasn't able to book that for you for Tuesday. Is there another day you would like to stay or perhaps a shorter stay? | |
|           | User: how about only 2 nights | 2 nights |
|           | Agent: Booking was successful. Your Reference number is: 7GAWK763. Anything else I can do for you? | |

The PersonaChat dataset is collected by Zhang et al, in 2018 [18]. The full dataset is available at convai.io/data/. The personaChat dataset is collected from random pairs of subjects from a crowd, via the Amazon Mechanical Turk. In each interaction, the turkers are first informed of a given profile, and the conversation centers around the subject, which means that they are exploring the personalities of each other. The assigned personalities are reflected through their opinions on some topics, like “whether you like playing video games”, and direct description, like “I am a programmer”. The dataset has crowdsourced a list of over 1100 probable characteristics regarding a human being. For each set of traits, a description of no lesser than 5 sentences is provided. The profiles are meant to be descriptive and arouse interest in a topic. The length of each sentence is limited for the sake of computation. The result of the collecting process is a dataset consisting of 162,064 sentences and more than 10,000 conversations.

Table 2. Samples of profiles from personaChat

| Profiles |
|----------|
| I like to talk but people have a hard time understanding. |
| I like to look at blocks and sing about letters. |
| When I grow up, I want to be a dog. |
| I am a clean eater. |
| my parents were both very athletic. |
| I love running and preparing for marathons. |
| I am a cancer survivor. |

2.2. Model

In the current paper, the deep neural network model is applied as the base of the chatbot. The user behaviors are identified and classified into several intents that correlate to one of the topics from the MultiWOZ dataset and the PersonaChat dataset. The utterances together with the labels, or resolved intents, are fed into the neural network. As a result, the chatbot builds links between a target input and its intent. The outcome of the model is assessed by the accuracy rate recorded as learning proceeds and the final data of the average correct rate.
2.2.1. Word-embedding

Since language is a complicated system with complex grammar rules and a large vocabulary base that require years for a language learner to master, how a machine can comprehend human texts in a short period of time is a major issue. Therefore, texts need to be prepared and transformed into numerals and binaries that our computers can process. During the process of implementing neural networks, the word embedding mechanism is used to represent each word as a vector with a fixed size. To apply word embedding, it is trained on a large quantity of natural conversations. During this process, the words are gradually simplified to capture their essential meaning. For instance, the program would eradicate the variation of words due to tense and plurality. The automated program would build the specific vector of each word, and meanwhile, the semantic meaning behind the word is also considered. This would help the model to make connections between words with a similar meaning but in different forms. According to the “Distributional Hypothesis”, words with a similar meaning often appear in a similar context, and thus their similarity can be discerned by calculating their frequency in each context [19]. After resolved into vectors, the words are learned by neural networks.

2.2.2. DNN

In the case of building a chatbot, we applied a DNN model through which the natural language data propagates. Mimicking the way human brains function, a DNN is structured with multiple layers of neurons, where computation takes place.

Data is feedforwarded through each layer. As for one neuron, the activation function generates the output of its computation. The type of activation function applied here is the SoftMax function, which takes in a vector and normalizes it into a probability distribution within the interval of (1,0). By applying the activation function, each node generates a result that will be passed to and processed by the next layer of nodes.

\[
\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for } i = 1, \ldots, K \text{ and } z = (z_1, \ldots, z_K) \in \mathbb{R}^K
\]  

(1)

Between each layer of nodes, each two from different layers are connected with a given weight, which determines the extent of influence on the next neuron. Though the initial weight is randomly assigned, the model corrects it to match the right prediction in each loop.

The entire dataset is fed into the model multiple times, defined by the parameter “epoch”. Based on the internal computation, the model fine-tunes itself through regression. The model after multiple rounds of training is then ready to make predictions. Sample utterance outside the original dataset is fed into the model, which automatically computes the possibilities of it falling into each predefined intent. Though the model doesn’t provide a firm answer, it is able to generate satisfactory outcome through filtering the options by setting a threshold.

\[
\frac{d\text{Error}}{d\text{weight}} = \frac{d\text{Error}}{d\text{activation}} * \frac{d\text{activation}}{d\text{weight}}
\]

Responses randomly chosen from the previously-built answer library for each intent are then returned to the users, fulfilling the user’s demand and simultaneously prompting him or her to continue asking. Such information retrieval structure of the chatbot allows engineers to have more control over the learning materials intended to teach.
3. Results

In this study, we constructed a chatbot using deep neural network (DNN). To ensure the correctness of users’ utterance, a grammarBot is embedded into the system to check the grammars before it is processed by the model [20]. To evaluate the effectiveness of the bot and the extent of its ability to correctly identify with various intent, both automated metrics of accuracy in terms of original texts and human evaluation are conducted. As an alternative to training one dialogue agent, we created a network constituted of 8 models trained separately; each model is specialized in a specific context. Their performances are then compared with each other to determine the better architecture in this case as well.

By feeding the MultiWOZ dataset and the personaChat dataset, we achieved the chatbot’s essential function of comprehending the meaning, or intent, or the user. Each intent corresponds to a field of information required by the subjects in the collecting phase and is thus general enough to incorporate several utterances with noticeable differences in structure and word choice. For example, one data field required by a subject in the context of “booking a hotel room” is asking for the “phone number”. Given that performing the same tasks in distinct scenarios exist, they are further classified based on their outside context. For instance, the action of requesting address is divided into the context of “hotel” and the context of “police” in an effort to avoid unintended confusion.

3.1. Integrated Structure

In this work, we propose the concept of “integrated structure” to distinguish the conventional chatbot structure from the novel one devised to optimize its performance for the language acquisition purpose. The definition of “integrated structure” is that a DNN chatbot relies on one model trained on a set of intents, each of which corresponds to a group of utterances with similar meaning. Therefore, it generates predictions based on the highest probability calculated by the model.

As for our implementation, collected utterances together with resolved intents from both datasets are indiscriminately fed into the model. Though the model is also context-sensitive, it leaves most of its decision making to the model prediction phase. As the chatbot tries to verify the bigger picture of the conversation, it is also trying to classify each intent within the context. Given that the data collection from conversations that perform the same task in different scenarios to a large extent overlaps with the other, we hypothesize that the chance is higher that a given utterance falls into a wrong intent category.

3.2. Network Structure

Unlike the conventional architecture, the proposed “network structure” separates the process of identifying each intent from identifying context. As indicated in Figure 1, the chatbot is a system made up of several trained models, and each is used when the topic of the conversation shifts. After the user sends a message, the chatbot would check the message content and automatically load the suitable model according to which intent the message falls into. Rather than feeding all data to train one single model, in this case, the intent belonging to a context is first categorized into “starter prompts” or “follow-ups”. The former type consists of only one intent, which is defined as “to start up a conversation” or “to begin a request”. For the case of seeking help from a police officer, the starter prompt is generally like “I have been robbed”, “I encountered a car accident” and “I would like to know where the police department is.” Utterances like these work as an indication of the context where the conversation occurs. The model receiving this piece of information identifies the starter prompt and relevant context, and the chatbot automatically loads the “follow-up” intents from the corresponding model. “The follow-up” intents are grouped based on the context, and they generally occur after the user begins a request, or after a “starter prompt”. As a continuation of the example of police office, after the user expresses his will of getting help from the police, a proper follow-up set of intents to be loaded into the chatbot would be “asking for road rescue”, “requiring postcode of the police office” and “express gratitude”. From a perspective of natural conversation, these speeches in usual times occur after “calling for help” and only in the context of the police office. Thus, it is more logical to identify them as “follow-ups”.
Figure 1. The graphic visualization of network structure. Elliptical shapes represent a neural network model, and rectangular boxes represent an individual intent. Each intent is embedded in the model it is attached to.

However, apart from the transactional intents, the other intents mostly for conversational purposes like the personalities of the chatbot are regarded as inherited intents and are still attached to model 1 as the graph indicates. User can start a free chat, or “small talk” by provoking these intents. The other models based on “follow-up” intents are solely connected to model 1, making it impossible to shift the topic in-between for such behavior is considered “perplexing” in a natural language perspective. At the end of each model, besides model 1, an intent is specifically signed as the end of the conversation. Such intent tends to be an expression of gratitude or “goodbye”. As a signal of ending session is received, the model resets itself and restart the conversation from model 1. The various models are interconnected to form a network. Our expectation for this structure is that the chatbot can have some control over the conversation, otherwise the users may abuse it by shifting context as they please, which is thought to be a threat to the learning effect.
3.3. Integrated Chatbot VS. Chatbot Network

The intents designed are essential to its function as we intend. We adopted both automatic metrics and human evaluation to evaluate the effectiveness of the chatbot. The two models are compared in terms of their rate of successfully identifying with each intent. In total, 3 parameters are compared: the maximum change of accuracy rate during training, the average accuracy rate after training, and the accuracy rate in human random tests after training.

The first parameter assessed is the change in the accuracy rate during training. For the chatbot network structure, the overall accuracy rate of 8 different models is rather smooth throughout and relatively high in values. The largest difference of accuracy rates is smaller than 5% if the first 5000 runs are not considered. However, the situation is considerably worse on the integrated structure side. Ignoring the first 5000 runs, the accuracy rate keeps fluctuating with a maximum difference of 20%. As Figure 2 indicates, the accuracy rate bounces between 0.5 to 0.8 in a regular pattern and shows no sign of strict improvements, implying potential outfitting of the model. The chatbot network outperforms the integrated chatbot. Provided that the models are well-trained after the first 500k steps, both models exhibit the potential to improve training efficiency by reducing the “epoch”, or the times the data is fed into the model and total training time on a device to avoid overfitting [21].

![Figure 2. The real-time accuracy rate against training steps. The data is logged during training and plotted in the graph using “tensorboard” built into Tensorflow. The curve with stronger color is smoothed out to show a general trend, while the curve with fainter color is the raw data. A surge in accuracy is observed from the beginning to 200k, then approaches 0.66 after 600k.](image)

The Second parameter to compare is the average accuracy rate after training. For the chatbot network, 7 out of 8 achieved an average accuracy rate of over 96%, with one achieving a rate of 87%. The integrated chatbot is again defeated this time with an accuracy rate of poor 68%, which is considerably lower than the other structure. A lower average accuracy rate indicates that the model experiences a high risk of making an error of wrongly identifying intents.

The last parameter considered is an accuracy rate of identifying intents of randomly human-generated utterances. This part of the research includes human evaluation. To better assess the effectiveness of the model in a practical sense, two human operators are told to interact with the chatbot using utterances covering all 8 models, making sure every intent and context are equally tested. In total, 100 interactions are conducted by both testers. The “network structure” speculates the context correctly for over 80 percent of the times, and almost 97 percent correct for the intents. Moreover, the model tends to confuse the context of “hotel” with “attraction” but works well when other contexts are tested. The “Integrated structure” features a correct rate of approximately 33 percent. According to testers’ feedback, it has a hard time identifying the correct context even though it gets the intent right. If the tester expresses the will of getting the contact number of the specific place, the model struggles to fit the utterance into the correct context but does provide a phone number. In tests with both models, this type of error is found to be prevalent. The overlapping patterns of vocabularies in these contexts account for the failure. In the previous case of “hotel” and “attraction”, “finding a hotel” is one of the requests seen in “attraction” setting aside from other intents like “finding a tourist information center”, causing the language pattern in “attraction” setting somewhat similar to that in “hotel” setting. Due to the model’s inability to
recognize the nuanced details in sentences, which may indicate the user’s will more precisely, the chatbot fails to recognize the exact intention by the user.

4. Discussion

In this research, a chatbot based on the technique of deep neural networks is proposed. It’s capable of conversation in multiple domains for the purpose of better educating students in contextual talking. The chatbot accomplishes the task of formulating replies and guiding the user through a conversation with an awareness of the context. This study has tested two distinct structures with the same deep neural network. Two datasets, one as the external knowledge regarding task performances and context, and the other as the internal knowledge dealing with personalities and ideas, are implemented to the model to mimic the behavior of a human talker. Through our examination, the “network structure” embeds 8 disjoint models, each of one context has the advantage of accuracy over the “integrated structure” solely relying on one single model.

The benefits of this work are threefold. The chatbot provides a novel way of learning a foreign language in the absence of a real living human tutor. Its ability to reconstruct contextual conversation and thus help to learn languages can overcome the predicament faced by millions of ESL learners in a non-English-speaking setting [22]. For language learning, the size of the vocabulary size is crucial for student’s development of word base. The current built-in dataset provides a vocabulary base as large as 10000 words. With this base, the chatbot is capable of diversifying user’s learning experience and familiarizes them with a wide variety of words in daily lives, for the average English vocabulary base of a 12-grade Chinese student contains only 4600 words, less than the half of our base [23]. The Deep Neural Network model approach ensures the scalability of the learning material. Teaching faculties are given the flexibility to customize a teaching bot based on their needs by feeding the model with resources they prefer. It is also relatively easy to scale up the bot to cover more domains, enriching the experience of learning. The inclusion of a grammar bot adds an element of surveillance to help students form a sound habit of using English. In addition, the Chatbot Network architecture can significantly optimize the reliability of chatbot. Meanwhile, it gives more control over the design of the course, making sure the student is strictly applying logic in his or her answering to the bot and is not distracted by small flaws.

Even though the chatbot has shown its obvious advantages in many aspects of real application, there are several noticeable limitations that we plan to solve in a future study. The chatbot is relatively easily distracted when encountering specific contexts. This issue is associated with overlapping vocabulary and overly similar speaking patterns in various intents. In future work, we suggest that the dataset should be more fine-tuned to discard inappropriate and misleading samples and carefully design and define each intent set to exaggerate its uniquity. Furthermore, in the current stage, the primary way to interact with the bot is through typing. In the future, learners should be able to interact with the chatbot through voice, which is a more direct and more intuitive way. The intelligent input method can be accomplished by adding text-to-speech and speech-to-text functions into the system. Another issue is that the current model of the neural network is incompetent to fully comprehend forming sentences. Though it may properly respond to an intent, it neglects minor details in a request, like the actual number of people booking the hotel. A mechanism of recognizing variables in a request is to be added to make the chatbot more responsive.

5. Conclusion

In conclusion, the current study aims to apply chatbot into a real-life setting and become a helpful aid for ESL learners. Since a dataset that covers many fields is applied, it can generate responses across multiple domains. We believe that in the future study, its dataset should be further expanded to include more fields relevant to our daily use of language to drive the chatbot to become more capable of natural conversations. Also, an actual experiment that tests the effectiveness, regarding students’ engagingness and learning efficiency should be conducted to confirm its potential in the field of education.
Reference

[1] Crystal, D. (2003). English as a Global Language. doi:10.1017/cbo9780511486999

[2] Khan, I. (2011). Learning Difficulties in English: Diagnosis and Pedagogy in Saudi Arabia. Educational Research. *International Research Journals, 2*, 1248-1257.

[3] Derakhshan, A., & Shirmohammadi, M. (2015). The Difficulties of Teaching English Language: The Relationship between Research and Teaching. *International Journal of Linguistics, 7*(1), 102. doi:10.5296/ijl.v7i1.6648

[4] Zhang, Y. (2007). Communication Strategies and Foreign Language Learning. *US-China Foreign Language, 5*(4), 43-48.

[5] Bashir, M. (2011). Factors Effecting Students’ English Speaking Skills. *British Journal of Arts and Social Science, 2*(1), 48.

[6] Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., . . . Akkiraju, R. (2018). Touch Your Heart. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*. doi:10.1145/3173574.3173989

[7] Schreuder, A. (2017). Customer Centric Artificial Intelligence – Using Text and Sentiment Analysis & Deep Neural Network Learning to make Chabot Reply in a more Customer Centric Fashion. *SAMRA, 1*-34.

[8] Su, M., Wu, C., Huang, K., Hong, Q., & Huang, H. (2018). Follow-up Question Generation Using Pattern-based Seq2seq with a Small Corpus for Interview Coaching. *Interspeech 2018*. doi:10.21437/interspeech.2018-1007

[9] Xu, Z., Wang, S., Zhu, F., & Huang, J. (2017). Seq2seq Fingerprint. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. doi:10.1145/3107441.3107424

[10] Serban, Iulian Vlad, et al. “A Survey of Available Corpora for Building Data-Driven Dialogue Systems.” *ArXiv:1512.05742 [Cs, Stat]*, Mar. 2017. arXiv.org, http://arxiv.org/abs/1512.05742.

[11] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36-45. doi:10.1145/365153.365168

[12] Abushawar, B., & Atwell, E. (2015). ALICE Chatbot: Trials and Outputs. *Computación Y Sistemas, 19*(4). doi:10.13053/cys-19-4-2326

[13] Berghe, R. V., Verhagen, J., Oudgenoeg-Paz, O., Ven, S. V., & Leseman, P. (2018). Social Robots for Language Learning: A Review. *Review of Educational Research, 89*(2), 259-295. doi:10.3102/0034654318771726

[14] Jia, Jiyou. “The Study of the Application of a Keywords-Based Chatbot System on the Teaching of Foreign Languages.” *ArXiv:Cs/0310018*, Oct. 2003. arXiv.org, abs/cs/0310018.

[15] Fryer, L., & Carpenter, R. (2006). Emerging technologies- bots as learning tool. *Language Learning & Technology, 10*(4), 8-14.

[16] Huang, J., Lee, K., Kwon, O., & Kim, Y. (2017). A chatbot for a dialogue-based second language learning system. *CALL in a Climate of Change: Adapting to Turbulent Global Conditions – Short Papers from EUROCALL 2017*, 151-156. doi:10.14705/rpnet.2017.eurocall2017.705

[17] Eric, Mihail, et al. “MultiWOZ 2.1: A Consolidated Multi-Domain Dialogue Dataset with State Corrections and State Tracking Baselines.” *ArXiv:1907.01669 [Cs]*, Dec. 2019. arXiv.org, abs/1907.01669.

[18] Zhang, Saizheng, et al. “Personalizing Dialogue Agents: I Have a Dog, Do You Have Pets Too?” *ArXiv:1801.07243 [Cs]*, Sept. 2018. arXiv.org, http://arxiv.org/abs/1801.07243.

[19] Harris, Z. S. (1954). Distributional Structure. *WORD, 10*(2-3), 146-162. doi:10.1080/00437956.1954.11659520

[20] Mangukiya, M. (2018, December 14). Grammarbot. Retrieved July 5, 2020, from pypi.org/project/grammarbot/
[21] Hardt, M. (2016). Train faster, generalize better: Stability of stochastic gradient descent. *ICML*, 48, 1125-1234.

[22] Y. J. A., Barnes, D. W., & Cochran, S. W. (1988). Schema and purpose in reading comprehension and learning vocabulary from context. *Reading Research and Instruction*, 28(2), 16-28. doi:10.1080/19388078909557965

[23] Jia, G., & Wang, Y. (2015). A Study on the Reasonability of the Provided Vocabulary Base by Standard English Course ---- an Investigation and Analysis of Middle School Students’ English Vocabulary. *China Fundamental Education Yearbook*