DeMINT: Automated Language Debriefing for English Learners via AI Chatbot Analysis of Meeting Transcripts

Miquel Esplà-Gomis, Felipe Sánchez-Martínez,
Víctor M. Sánchez-Cartagena, Juan Antonio Pérez-Ortiz

Dep. de Llenguatges i Sistemes Informàtics, Universitat d'Alacant
E-03690 Sant Vicent del Raspeig (Spain)

https://github.com/transducens/demint
{mespla,fsanchez,vmsanchez,japerez}@dlsi.ua.es

Abstract

The objective of the DeMINT project is to develop a conversational tutoring system aimed at enhancing non-native English speakers' language skills through post-meeting analysis of the transcriptions of video conferences in which they have participated. This paper describes the model developed and the results obtained through a human evaluation conducted with learners of English as a second language.

1 Project Overview

DeMINT (Automated Language Debriefing for English Learners via AI Chatbot Analysis of Meeting Transcripts) was developed by the Transducens Research Group at Universitat d'Alacant from January to September 2024. It is funded under the UTTER project,[1] a collaborative Research and Innovation project under Horizon Europe (grant agreement ID: 101070631), via *financial support to third parties*.

Conversational intelligent tutoring systems (ITS) are poised to revolutionize education by providing personalized, interactive, and inclusive one-on-one learning. The objective of this project is to develop an educational chatbot aimed at leveraging large language models (LLMs) to improve speakers' language skills through interactive error-driven conversations. A full-length description of our model is described by Pérez-Ortiz et al. (2024). Although not centered on machine translation (MT), our chatbot relies on components common to many MT-related tasks, such as speech transcription or grammatical error correction. Moreover, a similar ITS could also support translation tasks by helping translators improve text quality.

The architecture of our system is described in Figure 1, which reflects the modules that interact to provide a comprehensive tutoring experience:

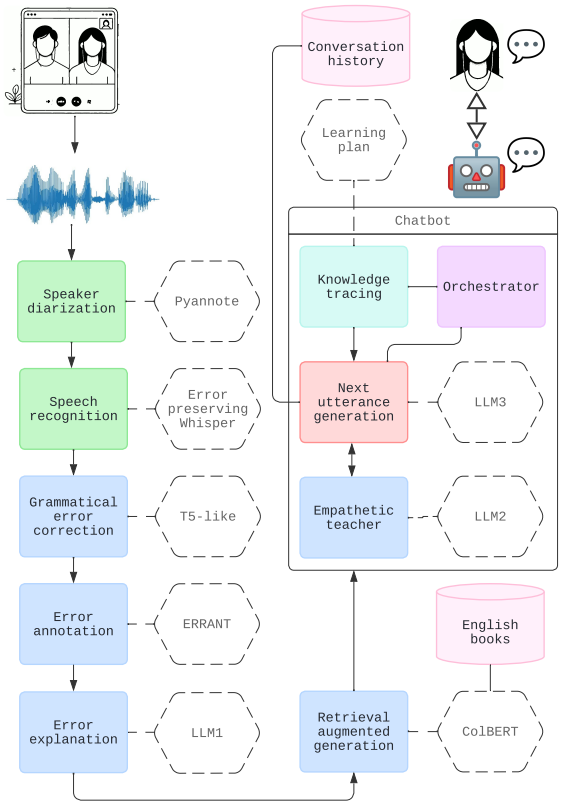

Figure 1: Main components of the DeMINT system.

1. *Diarization*: The pipeline starts by processing the audio recorded during an online meeting with the library `pyannote.audio`,[2] to identify the segments corresponding to each speaker.

2. *Speech Recognition*: Audio fragments are processed by a speech recognition model built by fine-tuning Whisper[3] on a custom dataset of spoken sentences with grammatical errors.[4]

3. *Grammatical Error Correction*: For this task, we employ a T5 model (Raffel et al., 2020) fine-tuned on the JF-LEG dataset.[5]

[1] https://he-utter.eu

[2] https://pyannote.ai/

[3] https://openai.com/index/whisper/

[4] https://huggingface.co/blog/asr-diarization

[5] https://huggingface.co/vennify/

4. *Error annotation*: Given the original and the corrected version of each sentence, we use the ERRANT toolkit (Bryant et al., 2017) to annotate the edits necessary to transform one sentence version into the other.

5. *Error explanation*: An LLM is used to generate finer-grained natural-language explanations from the high-level errors annotated with ERRANT via few-shot in-context learning.

6. *Retrieval from Textbooks*: This module uses retrieval-augmented generation (RAG) to get relevant information from English learning textbooks and provide it to the chatbot's *next-dialog-line generator* module.

7. *Empathetic Teacher*: The Llama-3.1-8B[6] model was fine-tuned with real-life, ideally-empathetic teacher-student conversations. This model processes the recent conversation history and provides guidance on how a teacher might respond.

8. *Orchestrator*: The orchestrator is a simple Python program that iterates through the different errors and sentences building the complex prompt that will be used to guide the interaction with the user.

9. *Next-Dialog-Line Generator*: GPT-4 is used to generate the next line of the conversation based on the informative prompt from orchestrator. This step also aims at modeling the learning process of the student; the information is included in the prompt, and is presented in the diagram as *knowledge tracing*.

10. *Chatbot Interface*: The interface is a simple web app built with `gradio`.[7] It shows the chatbot conversation in one column and the transcription, centered on the current sentence, in another. The user types their input, and the machine responds accordingly on the screen.

2 Results of the project

As the results of this project, a prototype of our ITS has been implemented. All the models and the datasets used to build them have been released and are available on the project repository.[8]

The prototype developed has been evaluated through interactions between the chatbot and L1-Spanish/L2-English students. Seven students with B2/C1 English levels (according to the Common European Framework of Reference for Languages) participated in 15 video-calls. These calls were recorded and processed by our ITS, and students spent about 75 minutes interacting with it. Students were then surveyed on overall user experience and chatbot's effectiveness as an English tutor, using a 1–5 Likert scale. In response to "Did you enjoy interacting with the chatbot?", all gave positive feedback (scores of 4 or 5). Fluency was identified as the main area for improvement, with an average score of 3. Regarding the chatbot's role as a tutor, the main concern was its accuracy in identifying speech errors (average score: 3). Still, most students felt it helped improve aspects of their English, with five out of seven giving a score of 4. When asked about future use of similar tutoring tools in video conferences, all but one rated their interest as 4 or 5, showing overall enthusiasm for such tools.

Acknowledgements

DeMINT was funded under the UTTER project (Horizon Europe, GA: 101070631), via *financial support to third parties*. In addition to the researchers signing this paper, Roman Chernysh, Gabriel Mora-Rodríguez and Lev Berezhnoy were also part of the work team behind this project.

References

Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th ACL Conference (Volume 1)*, pages 793–805, Vancouver, Canada.

Juan Antonio Pérez-Ortiz, Miquel Esplà-Gomis, Víctor M. Sánchez-Cartagena, Felipe Sánchez-Martínez, Roman Chernysh, Gabriel Mora-Rodríguez, and Lev Berezhnoy. 2024. A conversational intelligent tutoring system for improving English proficiency of non-native speakers via debriefing of online meeting transcriptions. In *Proceedings of the 13th Workshop NLP4CALL*, pages 187–198, Rennes, France.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

`t5-base-grammar-correction`

[6] https://huggingface.co/meta-llama/Meta-Llama-3.1-8B

[7] https://github.com/gradio-app/gradio

[8] https://github.com/transducens/demint