The Teacher-Student Chatroom Corpus

Cambridge Language Sciences

Andrew Caines¹, Helen Yannakoudakis^{2,1}, Helena Edmondson³,
Helen Allen⁴, Pascual Pérez-Paredes⁵, Bill Byrne⁶, Paula Buttery¹

¹ ALTA Institute & Computer Laboratory, University of Cambridge

² King's College London ³ Theoretical & Applied Linguistics ⁴ Cambridge Assessment

⁵ Faculty of Education ⁶ Department of Engineering

A new resource:

- One-to-one language lessons in an online chatroom
- Between experienced teachers and intermediate/advanced learners of English
- So far: 102 written lessons between 2 teachers and 8 students, 13.5K conversation turns, 133K word tokens
- L1s: Japanese, Spanish, Italian, Ukrainian, Mandarin Chinese, Thai
- Anonymised & available for research use; supported by Cambridge Language Sciences Research Incubator Fund, the Isaac Newton Trust & Cambridge Assessment

The chatroom:

ShinyChat
For teacher-student conversations

Welcome to the ShinyChatRoom! Maintained by researchers at the University of Cambridge.
By chatting here you agree that these conversations will be recorded and used for research.
For help please contact chat.corpus@cl.cam.ac.uk

Thank you for taking part in this study. Please go ahead and chat here, and remember that all conversations are recorded.

User29599 entered the room.
"User29599" -> "Teacher"
User71199 entered the room.
"User71199" -> "Student"
Teacher : Hello, Student!
Student : Hello, Teacher!

Please set your username here:

Teacher

Connected Users

- Teacher
- Student

Type your response here (press the enter key or send button)

Send

Annotation:

- Grammatical error correction
- 'Responding to' threading of non-linear conversation
- Sequence type: conversation analysis model of action sequences in lesson structure
- Teaching focus & use of external materials
- CEFR level assessment of student contribution

Enables:

- Analysis of discourse and interaction in teacher-student one-to-one chat
- Analysis of pedagogical strategy in lesson management, language teaching feedback, and exercise setting
- Analysis of individual and (with more data) group-level variation
- Eventual use for human-in-the-loop dialogue system development (with more data)

Publication at NLP4CALL 2020

Corpus application form: <https://forms.gle/oW5fwTTZfZcTkp8v9>