A corpus of spontaneous dialogues in L2 English by French and Japanese L1 speakers for automated assessment of fluency

Sylvain COULANGE[1,2], Takayuki KONISHI[3], Tsuneo KATO[2], Mariko SUGAHARA[2], Solange ROSSATO[1], Monica MASPERI[1]

[1] Grenoble Alpes Univ.; [2] Doshisha Univ.; [3] Waseda Univ.

Context:

- CAPT tools rarely deal with spontaneous speech, and even more rarely with speech in real discussion situation.
- Lack of L2 spontaneous speech corpus.
- Lack of speech in peer dialogue situations.

Creation of a speech corpus:

- 2- or 3-student role play type argumentative discussion on a contentious topic.
- Different topics were used, such as security cameras, animal testing, the use of technology in classrooms, part-time jobs...
- Each candidate assumed a specific given role, either advocating for or against the subject.
- 2~5 minutes of preparation before the talk.
- Objective: negotiate, exchange viewpoints, and eventually work towards a compromise.

Processing Pipeline

- We released an open-source automated processing pipeline specifically design for processing multi-speaker spontaneous L2 English speech. [3]
- The processing steps are as follows:

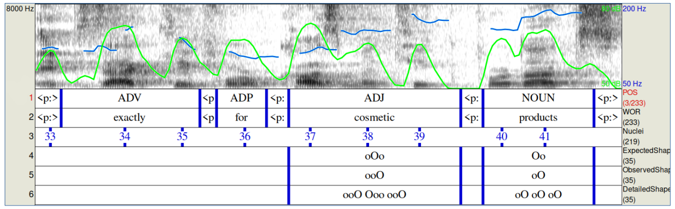
- Speech detection and neural speaker diarization (Pyannote)
- ASR & word-level alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)

- **Localisation of pauses** with POS context and constituency analysis (Benepar)

- **Syllable** nuclei detection [4]
- Syllabic **parameter extraction** (intonation, intensity, duration ; speaker norm.)
- **Comparison** of prosodic shape of words with a reference dictionary

pipeline
GitLab

- Insight of a TextGrid output:



Data Available for Academic Research

CLES French-L1 corpus

(Public portion of the CLES corpus of spontaneous L2 English) [1,2]

- 128 speakers
- French as mother tongue: 93%
  (other: Albanese, Arabic, Chinese, Georgian, Indonesian, Latvian, Persian, Spanish, Ukrainian)
- 48% F, 52% M
- 62 groups (3-speaker: 4, 2-speaker: 58)
- Proficiency: B1~B2
- Speech duration: 10 hours
  (mean 9'35'', min 5'12'', max 14'30'')

Data available for academic research: coordination-nationale@certification-cles.fr

Waseda-Doshisha Japanese-L1 corpus

- 30 speakers
- 100% Japanese-L1
- 60% F, 40% M
- 16 groups
  (16 pairs, including 2 with a Japanese-L1 English teacher)
- Proficiency: B1~C2
- Speech duration: 4 hours
  (mean 16:39, min 9:54, max 33:52)

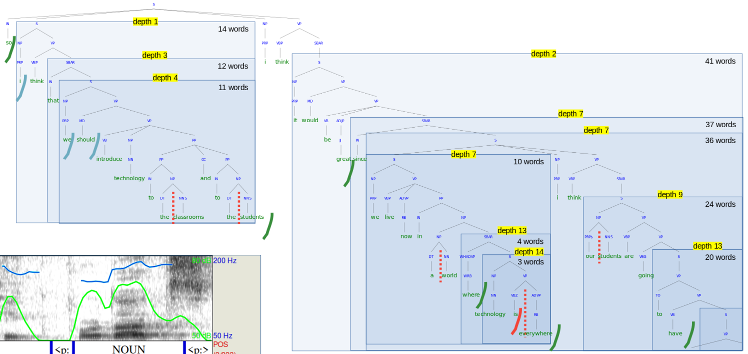
Data available for academic research: msugahar@mail.doshisha.ac.jp

Doshisha English-L1 corpus

- 14 speakers
- 100% English-L1
- 64% F, 36% M
- 7 groups
- Speech duration: 2 hours
  (mean 17:24, min 13:20, max 21:18)

Data available for academic research: msugahar@mail.doshisha.ac.jp

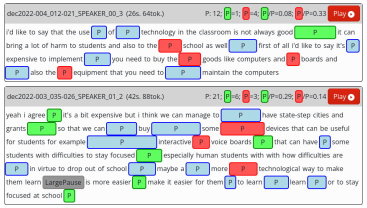
Pause Position Analysis



- Pauses are categorized into inter-clause, inter-phrase and intra-phrase pauses, along with POS context, syntactic depth and nb. of words of adjacent constituents.

Visualisation Platform

- A server-based visualisation tool allows to easily view the processing outputs.



dec2022-004_012-021_SPEAKER_00_3 (26s. 64tok.) P: 12; P=1; P=4; P/P=0.08; P/P=0.33 Play

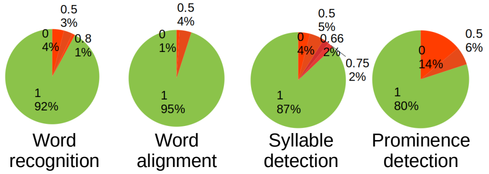
i'd like to say that the use P of P technology in the classroom is not always good P it can bring a lot of harm to students and also to the P school as well P first of all i'd like to say it's P expensive to implement P you need to buy the P goods like computers and P boards and P also the P equipment that you need to P maintain the computers

dec2022-003_035-026_SPEAKER_01_2 (42s. 88tok.) P: 21; P=6; P=3; P/P=0.29; P/P=0.14 Play

yeah i agree P it's a bit expensive but i think we can manage to P have state-step cities and grants P so that we can P buy P some P devices that can be useful for students for example P interactive P voice boards P that can have P some students with difficulties to stay focused P especially human students with with how difficulties are P in virtue to drop out of school P maybe a P more P technological way to make them learn LargePause is more easier P make it easier for them P to learn P learn P or to stay focused at school P

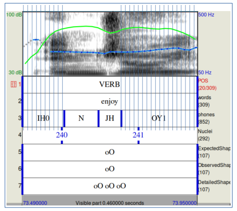
Pipeline Evaluation

- Evaluation of 100 random target words on the French-L1 corpus, manual verification:



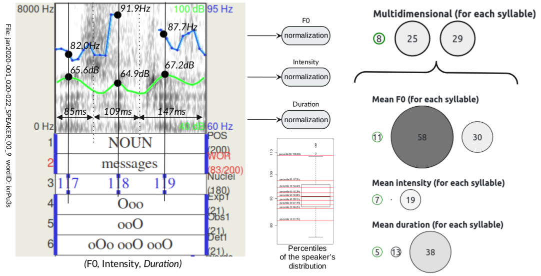
New Version of the Pipeline

- A new version of the pipeline is currently being developed. It now uses a **phoneme-level** alignment step to measure prominency at vowel intervals, as well as F0 dynamics.



Lexical Stress Analysis

- Lexical stress is estimated from prosodic prominency of syllables, based on measures of F0, intensity and duration.



References:

[1] *CLES official website: https://www.certification-cles.fr/english/*

[2] *Coulange, S., Fries, M.-H., Masperi, M., Rossato, R. (submitted). A corpus of spontaneous L2 English speech for real-situation speaking assessment. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 20-25 May, Torino, Italy.*

[3] *Coulange, S., Kato, T., Rossato, R., Masperi, M. (in press). Enhancing Language Learners'Comprehensibility through Automated Analysis of Pause Positions and Syllable Prominence. In Mairano, P. & Schwab, S (eds.) Languages, Special Issue "Speech Analysis and Tools in L2 Pronunciation Acquisition".*

[4] *De Jong, N. H., Pacilly, J., Heeren, W. (2021) "Praat scripts to measure speed fluency and breakdown fluency in speech automatically." Assessment in Education: Principles, Policy & Practice, 28, 456-476.*

financé par
IDEX Université Grenoble Alpes

This research is supported by Kakenhi n°23H00648, n°20K13145 and n°22K00627

A corpus of spontaneous dialogues in L2 English by French and Japanese L1 speakers for automated assessment of fluency

Sylvain COULANGE[1,2], Takayuki KONISHI[3], Tsuneo KATO[2], Mariko SUGAHARA[2], Solange ROSSATO[1], Monica MASPERI[1]

[1] Grenoble Alpes Univ.; [2] Doshisha Univ.; [3] Waseda Univ.

Context:

- CAPT tools rarely deal with spontaneous speech, and even more rarely with speech in real discussion situation.
- Lack of L2 spontaneous speech corpus.
- Lack of speech in peer dialogue situations.

Creation of a speech corpus:

- 2- or 3-student role play type argumentative discussion on a contentious topic.
- Different topics were used, such as security cameras, animal testing, the use of technology in classrooms, part-time jobs…
- Each candidate assumed a specific given role, either advocating for or against the subject.
- 2~5 minutes of preparation before the talk.
- Objective: negotiate, exchange viewpoints, and eventually work towards a compromise.

Processing Pipeline

- We released an open-source automated processing pipeline specifically design for processing multi-speaker spontaneous L2 English speech.
- The processing steps are as follows:

- Speech detection and neural speaker diarization (Pyannote)
- ASR & word-level alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)

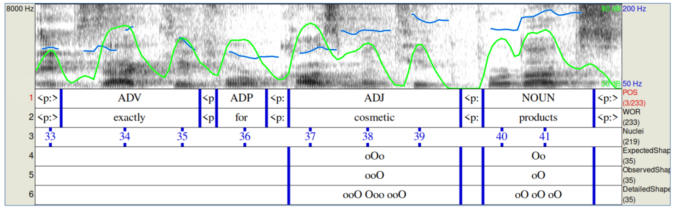
- **Localisation of pauses** with POS context and constituency analysis (Benepar)

- **Syllable** nuclei detection [3]
- Syllabic **parameter extraction** (intonation, intensity, duration ; speaker norm.)
- **Comparison** of prosodic shape of words with a reference dictionary

pipeline

GitLab

- Insight of a TextGrid output:



Data Available for Academic Research

CLES French-L1 corpus

(Public portion of the CLES corpus of spontaneous L2 English) [1,2]

- 128 speakers
- French as mother tongue: 93% (other: Albanese, Arabic, Chinese, Georgian, Indonesian, Latvian, Persian, Spanish, Ukrainian)
- 48% F, 52% M
- 62 groups (3-speaker: 4, 2-speaker: 58)
- Proficiency: B1~B2
- Speech duration: 10 hours (mean 9'35'', min 5'12'', max 14'30'')

Data available for academic research: coordination-nationale@certification-cles.fr

Waseda-Doshisha Japanese-L1 corpus

- 30 speakers
- 100% Japanese-L1
- 60% F, 40% M
- 16 groups (16 pairs, including 2 with a Japanese-L1 English teacher)
- Proficiency: B1~C2
- Speech duration: 4 hours (mean 16:39, min 9:54, max 33:52)

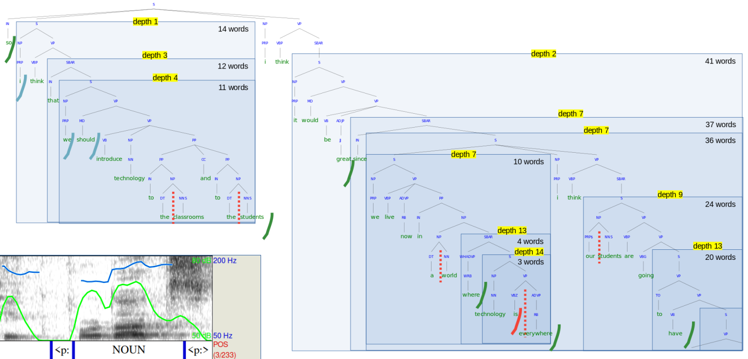
Data available for academic research: msugahar@mail.doshisha.ac.jp

Doshisha English-L1 corpus

- 14 speakers
- 100% English-L1
- 64% F, 36% M
- 7 groups
- Speech duration: 2 hours (mean 17:24, min 13:20, max 21:18)

Data available for academic research: msugahar@mail.doshisha.ac.jp

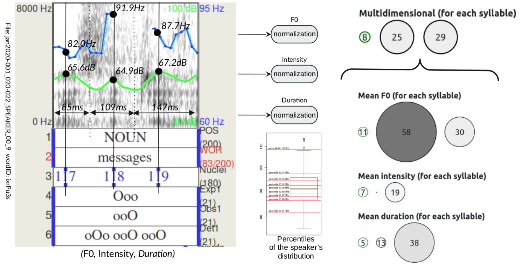
Pause Position Analysis



- Pauses are categorized into inter-clause, inter-phrase and intra-phrase pauses, along with POS context, syntactic depth and nb. ff words of adjacent constituents.

Lexical Stress Analysis

- Lexical stress is estimated from prosodic prominency of syllables, based on measures of F0, intensity and duration.



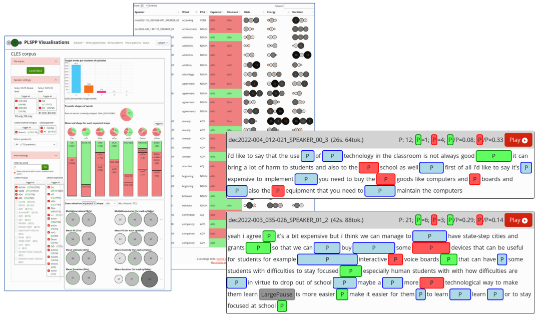
References:

[1] *CLES official website: https://www.certification-cles.fr/english/*

[2] *Coulange, S., Fries, M.-H., Masperi, M., Rossato, R. (submitted). A corpus of spontaneous L2 English speech for real-situation speaking assessment. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 20-25 May, Torino, Italy.*

[3] *De Jong, N. H., Pacilly, J., Heeren, W. (2021) "Praat scripts to measure speed fluency and breakdown fluency in speech automatically." Assessment in Education: Principles, Policy & Practice, 28, 456-476.*

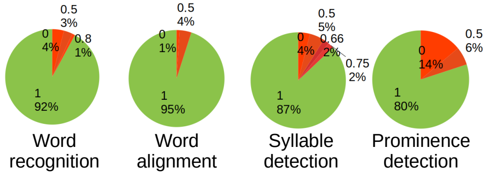
Visualisation Platform



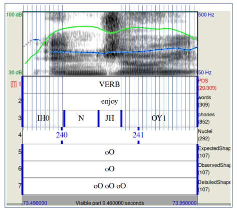
- A server-based visualisation tool allows to easily view the processing outputs.

Pipeline Evaluation

- Evaluation of 100 random target words on the French-L1 corpus, manual verification:



New Version of the Pipeline



- A new version of the pipeline is currently being developed. It now uses a **phoneme-level** alignment step to measure prominency at vowel intervals, as well as F0 dynamics.

financé par IDEX Université Grenoble Alpes

This research is supported by Kakenhi n°23H00648, n°20K13145 and n°XXXXXXX