# A corpus of spontaneous L2 English speech for real-situation speaking assessment

Sylvain COULANGE[1,2], Marie-Hélène FRIES[3], Monica MASPERI[1], Solange ROSSATO[2]

1. Univ. Grenoble Alpes, Laboratory of Linguistics and Didactics of Foreign and Mother Tongues (LIDILEM)
38000 Grenoble, France

2. Univ. Grenoble Alpes, CNRS, Institute of Engineering, Grenoble Computer Science Laboratory (LIG)
38000 Grenoble, France

3. National Coordination for the Certificate of language skills in French higher education (CLES)

{ sylvain.coulange, monica.masperi, solange.rossato }@univ-grenoble-alpes.fr, coordination-nationale@certification-cles.fr

financé par IDEX Université Grenoble Alpes

## Context:

- Computer Assisted Pronunciation Training tools rarely deal with **spontaneous speech**
- Lack of **L2 spontaneous** speech corpus.
- Lack of speech in **peer dialogue situations**.

## Creation of a speech corpus:

- We started gathering **L2 spontaneous speech** data recorded in exam situations.
- Our first aim is to **train score prediction models** based on near-real-situation L2 speech, but this corpus can also serve other purposes in L2 acquisition, teaching, testing, or L2 speech processing.

## Automated file processing:

- We made a dedicated **speech processing pipeline** to annotate this challenging type of speech [2].
- In this study, we focused on **speech rhythm measurement** through syllabic prominence of polysyllabic words

corpus Ortolang

pipeline GitLab

## Corpus

✔ The CLES is a state certificate established by the French Ministry of Higher Education and Research, designed for university-level language proficiency assessment. [1]

✔ The collection of spontaneous L2 speech recordings comprises 2 types of role-play:

— CLES B2 —
Argumentative discussions
(2 or 3 candidates)
Mean dur.: 9'35"

— CLES B1 —
Vocal messages
(monologues)
Mean dur.: 3'20"

✔ Each recording is provided with high-quality certification-level proficiency ratings.

| Type | Nb. of speakers | Duration |
|---|---|---|
| 3-speaker | 15 | 1h03'44'' |
| 2-speaker | 232 | 18h16'50'' |
| 1-speaker | 13 | 39'28'' |
| **Total** | **260** | **20h00'02''** |

✔ Each candidate assumed a specific given role, either advocating for or against the subject.

✔ The objective is to negociate and work towards a compromise.

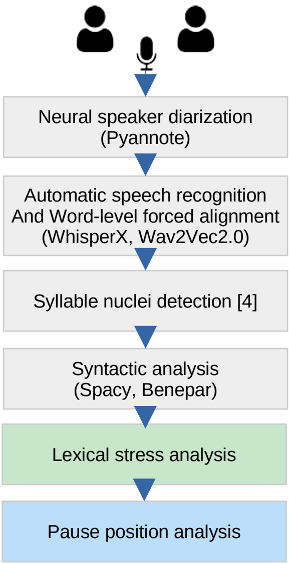
**Public portion:**

- 128 speakers
- French as L1: 93%
- 48% F, 52% M
- 62 groups
- Total duration: 10 h. (mean: 9'35", min 5'12", max 14'30")

| Proficiency | Nb. of speakers | Proportion |
|---|---|---|
| B2 | 151 | 58 % |
| B1 | 75 | 29 % |
| Non-validated | 34 | 13 % |
| **Total** | **260** | |

**Note:** a similar corpus is being made involving native speakers and Japanese-L1 speakers of English [3].

## Processing Pipeline



Neural speaker diarization (Pyannote)

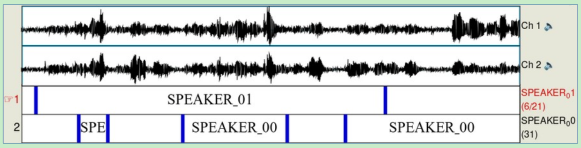
Automatic speech recognition And Word-level forced alignment (WhisperX, Wav2Vec2.0)

Syllable nuclei detection [4]

Syntactic analysis (Spacy, Benepar)

Lexical stress analysis

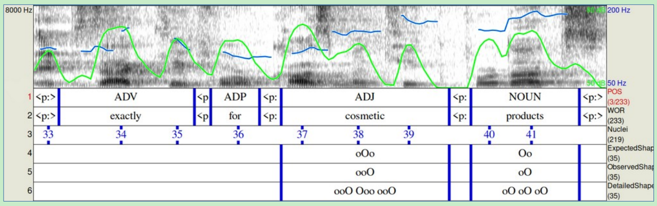
Pause position analysis



◄ Speaker diarization output in TextGrid format

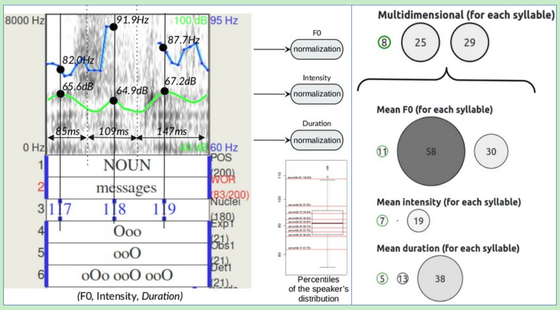
Lexical stress analysis annotation example ►

Only polysyllabic words with adequat number of syllable nuclei detected are annotated, in order to filter bad word alignments.



## Lexical stress analysis:

- Lexical stress is estimated from prosodic measures on syllable nuclei, based on F0, intensity and syllable duration.
- Each prosodic measure is converted in speaker percentile, (50 means median prominence level for any speaker, 0 is minimum and 100 is maximum)
- Comparison with the prescripted stress pattern (CMU dictionary).



## Pipeline Evaluation & Limitations:

- As the pipeline combines several modules, errors can occur at different levels, often leading to incorrect annotations.

▲ Syllable detection and word alignment often mismatches, leading to a limited nb. of target words (only **41%** of polysyllabic words in the study below were **target words**).

▲ Manual evaluation of random 100 target words showed that **17%** were miss-recognized or miss-aligned, potentially leading to wrong judgments that can be problematic in a real assessment context.

▲ **Intrinsinc vowel length** and **word ending lengthening** need to be considered in order to improve stress estimation.

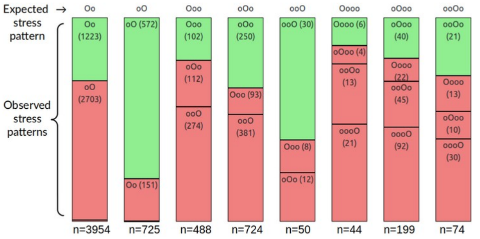
▲ Some cases of **vowel devoicing** also impacted F0 measures (tackled with linear interpolation for now)

## Lexical stress position accuracy and degree of prosodic contrast

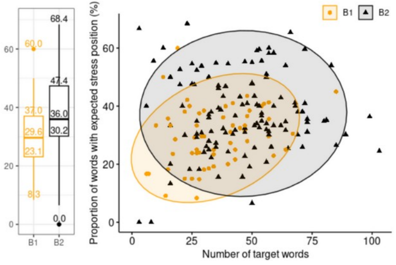
- **Sub-corpus:**
  - French-L1 speakers with either B1 or B2 speaking proficiency (n=176, 11 hours)
  - Speaking B1 level: 34%, B2 level: 66%
  - 6350 target words.
- **Hypothesis:**
  - Position accuracy B2>B1.
  - Shift to last syllable.
  - Stress mainly by duration change.
  - F0 and intensity used mainly by high proficiency speakers.

- **Main observations:**
  - ► Mean stress position accuracy varies greatly among speakers (0~68.4%, mean: 35.4%).
  - ► B2 speakers perform better than B1 in terms of stress position accuracy (36% vs. 29.6%) and prosodic contrast.
  - ► Syllabic prominence is often detected on the last syllable of words, which might be caused by L1 influence.
  - ► The better the speaker mean stress position accuracy, the higher pitch and intensity of expected stressed syllable.
  - ► Duration parameter is the most discriminant.

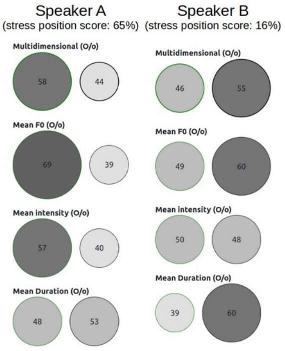
**Number of observed stress patterns for each expected pattern**



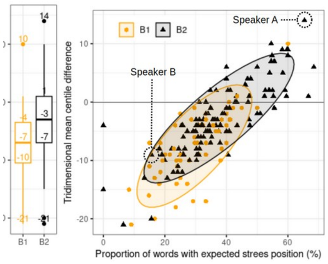
**Proportion of target words with expected stress position per speaker**



**Prosodic contrast**



**Prosodic contrast per speaker**



SpeakerA

SpeakerB

**References:**

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

[2] Coulange S, Kato T, Rossato S, Masperi M. (2024). Enhancing Language Learners' Comprehensibility through Automated Analysis of Pause Positions and Syllable Prominence. Languages 9(3):78

[3] Coulange, S., Konishi, T., Kato, T., Sugahara, M., Rossato, R., Masperi, M. (2024). A corpus of spontaneous dialogues in L2 English by French and Japanese L1 speakers for automated assessment of fluency. 6th International Symposium on Learner Corpus Studies in Asia and the World (LCSAW6), Feb. 2024, Kobe, Japan.

[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.