



Pause position analysis in spontaneous speech for L2 English fluency assessment

Sylvain Coulange, Tsuneo Kato

► To cite this version:

Sylvain Coulange, Tsuneo Kato. Pause position analysis in spontaneous speech for L2 English fluency assessment. 2023 Autumn Meeting of the Acoustic Society of Japan, Sep 2023, Nagoya, Japon, Japan. 150 2023 Autumn Meeting of the Acoustic Society of Japan, 2023. hal-04257261

HAL Id: hal-04257261

<https://hal.science/hal-04257261>

Submitted on 25 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Pause position analysis in spontaneous speech for L2 English fluency assessment

Sylvain COULANGE^{1,2}, Tsuneo KATO²

¹LIDILEM/LIG, Univ. Grenoble Alpes; ²SLPL, Doshisha Univ.

Context:

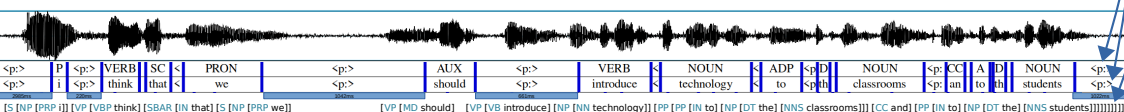
- Fluency assessment often deals with frequency and length of pauses,
- but a pause itself is **not necessarily a disfluency**. Pauses **may help** listeners to process the speech if they are strategically used, or they **can disturb** them if appearing at unexpected positions. [1]

→ We present a pipeline for pause position analysis in spontaneous L2 English.

- Pauses are defined here as silent or filled speech interruptions of duration ranging from 180ms to 2s.
- This study analyses:
 - pauses between clauses (considered as structurant)
 - pauses within phrases (considered as disfluent)
 - pauses' lexical environment (part-of-speech analysis, POS)

Processing pipeline:

- Speech detection and neural speaker diarization (Pyannote)
- ASR & Word-level Forced Alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)
- Constituency analysis (Benepar)



Corpus:

- L2 English spontaneous speech from 176 French learners recorded during CLES certification speaking session.
- Situation: 2 or 3 candidates discussing a polemical topic (role play) during 10min.

- Total 11 hours of continuous speech (per speaker: mean 3'44", min 32", max 6'51)
- Speaking B1 level: 34%, B2 level: 66%
- Speech duration: B1=B2, Nb tokens: B1<B2, Nb pauses: B1<B2, Silence proportion: B1=B2

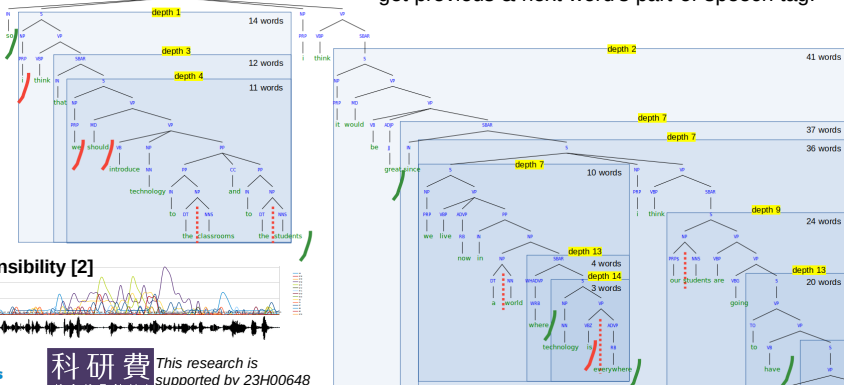
POS context	POS context duration	wordLeft	wordLeftEnd	wordLeftStart	wordLeftEnd	wordRight	wordRightStart	wordRightEnd	wordRightStart
1 ADV	0.00000000	start	1 IN	1 S	14 3 PRP				
2 ADV	0.00000000	start	1 IN	1 S	14 3 PRP				
3 ADV	0.00000000	start	1 IN	1 S	14 3 PRP				
4 ADV	0.00000000	start	1 IN	1 S	14 3 PRP				
5 ADV	0.00000000	start	1 IN	1 S	14 3 PRP				
6 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
7 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
8 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
9 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
10 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
11 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
12 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
13 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
14 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
15 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
16 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
17 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
18 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
19 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
20 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
21 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
22 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
23 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
24 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
25 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
26 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
27 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
28 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
29 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
30 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
31 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
32 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
33 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
34 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
35 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
36 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
37 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
38 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
39 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
40 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
41 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
42 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
43 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
44 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
45 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
46 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
47 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
48 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
49 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				
50 VERB	0.00000000	start	1 IN	1 S	14 3 PRP				

Feature extraction:

- Pauses as inter-word <p> tag of duration 180ms-2s.
- Total: 72,594 <p> tags, including 21,942 pauses.
- For each <p> tag:
 - get starting and ending biggest constituent label, number of words and syntactic depth;
 - get previous & next word's part-of-speech tag.

References:

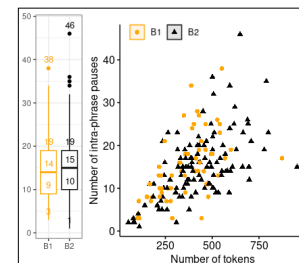
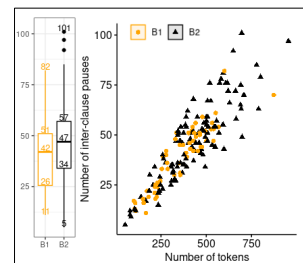
- [1] Isaacs, T., Trofimovich, P. and Foote, J. (2018). Developing a user-oriented second language comprehensibility scale for english-medium universities. *Language Testing* 35(2), 193-216.
- [2] Nagle, C., Trofimovich, P. and Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition* 41(4), 647-672.



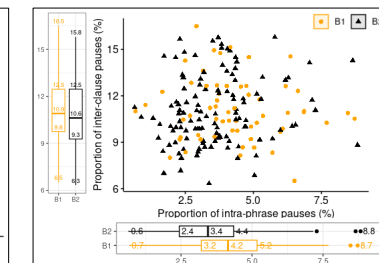
Structural analysis:

- Great variation of number of intra-phrasal pauses, less with inter-clausal pauses;
- No correlation between intra-phrasal and inter-clausal pause proportions;
- B2 speakers make less intra-phrasal pauses than B1 speakers;
- ...but difference between B1 and B2 is small.

Absolute number of inter-clause and intra-phrase pauses per speaker



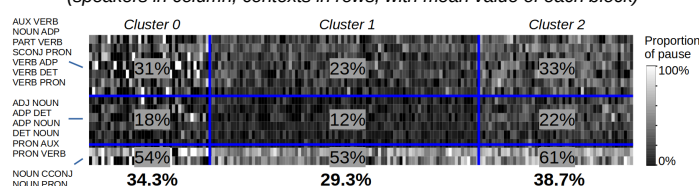
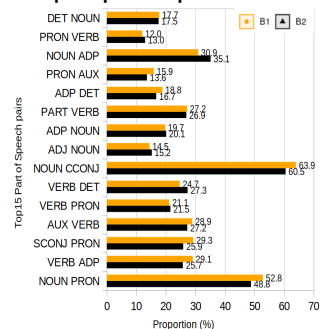
Proportion of inter-clause and intra-phrase pauses per speaker (nb pauses / nb tokens)



Lexical analysis:

- Pauses in same proportions for B1 and B2 for top15 most frequent POS contexts;
- B2 speakers make generally less pauses in these contexts.

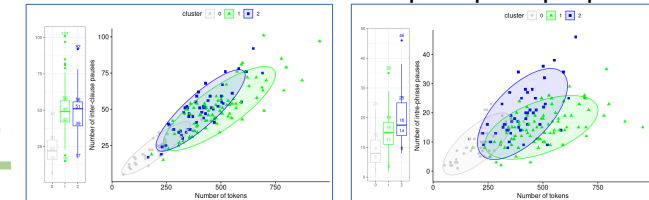
Proportion of pause for the 15 most frequent part-of-speech contexts



Grouping speakers according to pausing patterns:

- Ultimate segmentation is 3x3 groups;
- B1 and B2 speakers are mixed together in each group;
- Clusters 1 and 2 differentiate by overall frequency of pauses, Cluster 0 contains speakers with extreme values (too few occurrences) and intra-phrase pause frequency.

Absolute number of inter-clause and intra-phrase pauses per speaker



Discussion:

- Limited contrast between B1 and B2 speakers;
- Instead, large inter-speaker diversity in pausing pattern, especially within phrases;
- Need for investigating intra-speaker variability;
- Do more intra-phrasal pauses, and less inter-clausal pauses mean worse comprehensibility?
 - To investigate it, we plan to confront these results with dynamic ratings of comprehensibility [2]