

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:

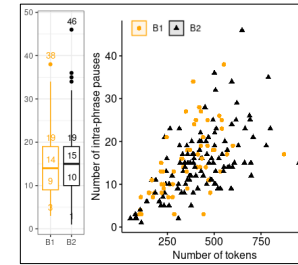
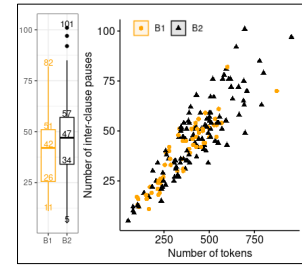
- Speech detection and neural speaker diarization (Pyannote)
- ASR & Word-level Forced Alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)
- Constituency analysis (Benepar)



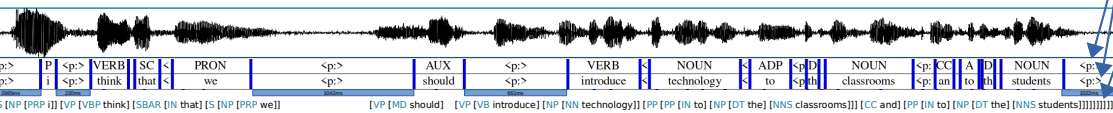
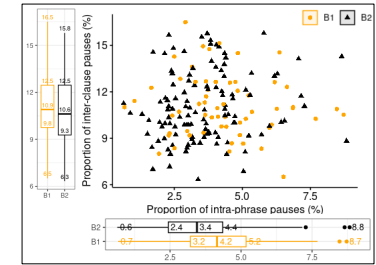
Structural analysis:

- 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)



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

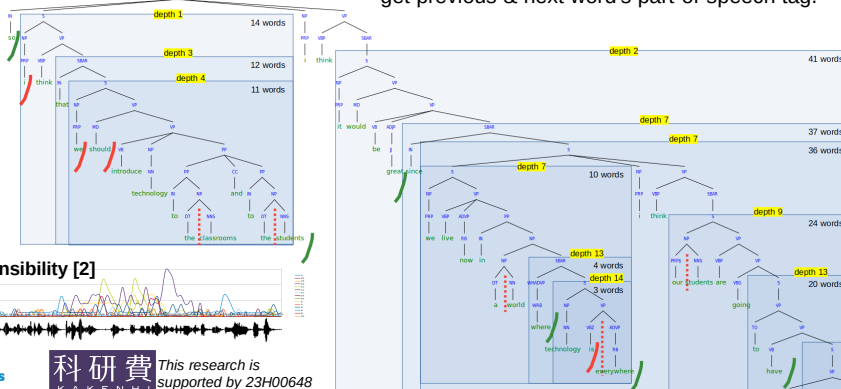
POS	Context	duration	wordLeft	wordRight	wordLeftDep	wordRightDep	wordLeftTag	wordRightTag	wordLeftCat	wordRightCat
2	ADV	0.9620201	ad	1	IN	1	I	IN	1	3
4	PRON	2.9847168	sp	1	3	PRP	think	VP	14	3
6	VERB	0.040073126	think	1	3	PRP	that	SBAR	12	4
8	SCONJ	0.00146557	that	1	4	IN	we	NP	11	1
10	PRON	1.5430228	we	1	5	PRP	should	VP	10	5
12	AUX	0.00120061	should	1	6	MD	introduce	VP	9	2
14	VERB	0.00109610	introduce	1	7	VB	technology	NP	1	1
16	NOUN	0.00102691	technology	1	8	NN	to	PP	7	1
18	ADP	0.00130210	to	1	9	IN	the	NP	2	1
20	DET	0.040073126	the	1	10	DT	classrooms	NP	2	10
22	NOUN	0.040073126	classrooms	1	11	NNS	and	CC	1	10
24	CCONJ	0.040073126	and	1	12	CC	to	PP	3	9
26	ADP	0.020036578	to	1	13	IN	use	VP	2	10
28	DET	0.040073126	use	1	14	DT	students	NP	1	10
30	NOUN	0.040073126	students	1	15	NNS				
32	PRON	0.00067341	they	1	16	PRP	will	VP	42	3
34	VERB	0.020036578	think	1	17	VP	it	SBAR	41	3
36	PRON	0.00020274	it	1	18	PRP	would	VP	40	3
38	AUX	0.00067341	would	1	19	MD	be	VP	39	6
40	AUX	0.00067341	be	1	20	VP				

Feature extraction:

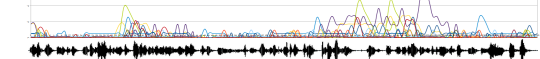
- Pauses as inter-word <p> tag of duration 180ms-2s.
- Total: 72,594 <p> tags, including 21,942 pauses.
- For each <p> tag:
 - get starting and ending biggest constituent label, number of words and syntactic depth;
 - get previous & next word's part-of-speech tag.

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.

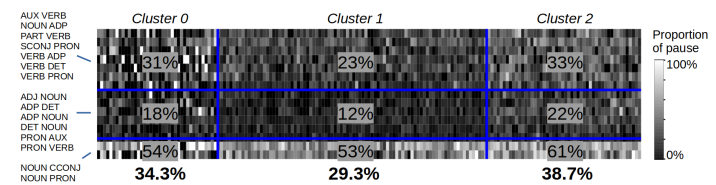


Dynamic rating of comprehensibility [2]

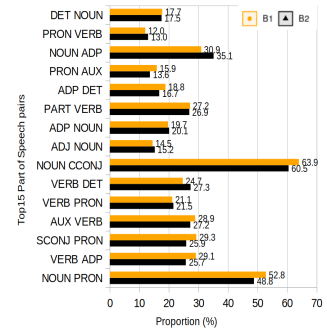


Lexical analysis:

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



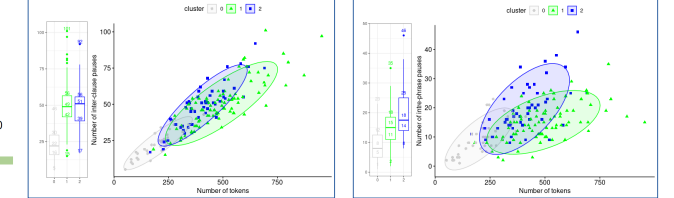
Proportion of pause for the 15 most frequent part-of-speech contexts



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-phrasal pause frequency.

Absolute number of inter-clause and intra-phrasal 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]