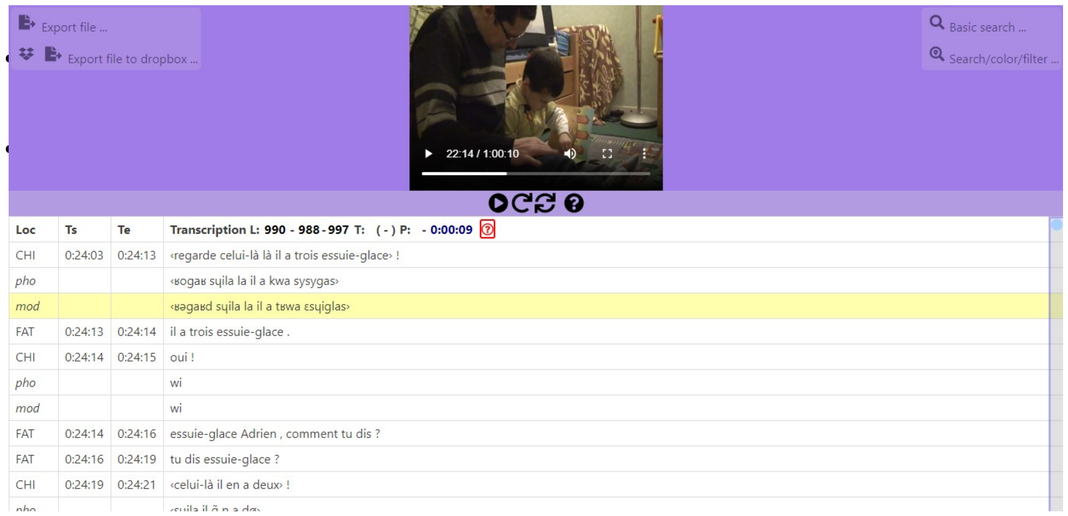
A computationally-based approach to the understanding of child's phonological development

Andrea Briglia, *STIH lab*
andrea.briglia@sorbonne-universite.fr

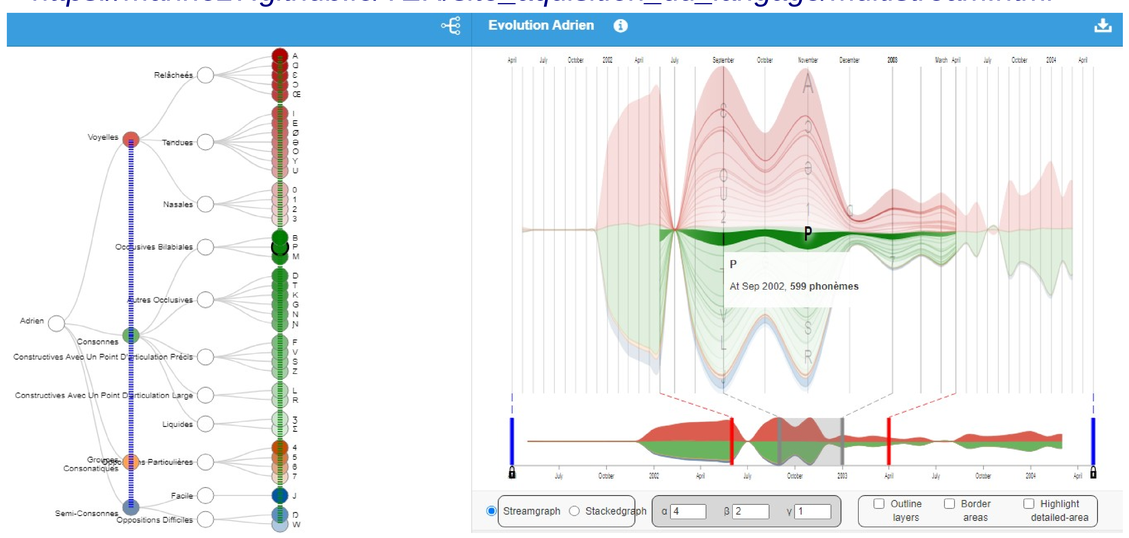
LETTRES SORBONNE UNIVERSITÉ

- **Data**. CoLaJE (1) is an open access French database made up of 7 child spoken language monthly *in vivo* videorecordings. There are around thirty longitudinal samples per child, each child counts approx 7000 sentences and 28'000 words. IPA transcriptions of *pho, mod* and orthograohic for CHI.



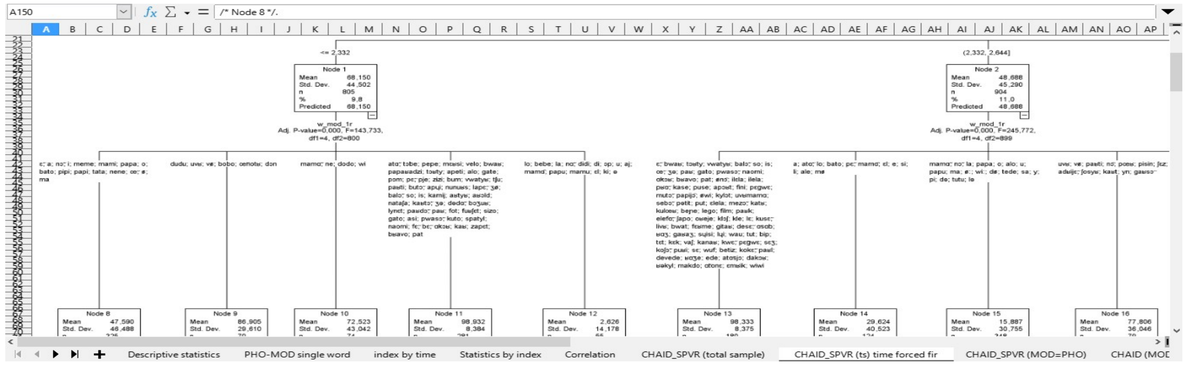
Loc	Ts	Te	Transcription L: 990 - 988 - 997 T: (-) P: - 0:00:09
CHI	0:24:03	0:24:13	‹regarde celui-là là il a trois essuie-glace› !
pho			‹ʁɔgaʁ sɥila la il a kwa sysygas›
mod			‹ʁəgaʁd sɥila la il a tʁwa esɥiglas›
FAT	0:24:13	0:24:14	il a trois essuie-glace .
CHI	0:24:14	0:24:15	oui !
pho			wi
mod			wi
FAT	0:24:14	0:24:16	essuie-glace Adrien , comment tu dis ?
FAT	0:24:16	0:24:19	tu dis essuie-glace ?
CHI	0:24:19	0:24:21	‹celui-là il en a deux› !
pho			‹ɥila il ɑ̃ n a dø›

- **METHOD/2** Multiresolution Streamgraph (3). To each French phoneme is assigned a color, here is represented *pho.*
https://marine27.github.io/TER/site_aquisition_du_langage/multistream.html



- **HP** Phonetic and phonological variations are not random, they rather follow and underlying logic made up of constraints and tendencies. Comparisons between similarly sampled children could give us a way to outline common preferential learning patterns

- **METHODS/1** We first use PHON to align *pho* and *mod* lines, then we set an algorithm able to calculate the SPVR (Sentence Phonetic Variation Rate). We use CHAID $\chi 2$ (2), a decision tree technique conceived to overcome in a non parametric way the problem of multiple comparisons. Three steps: merging, splitting and stopping.



- **CONCLUSIONS** Broadly speaking, FLA and phonological theories, such as the three pillars of Clements' « Theory of traits », can be quickly confirmed by using these two ways of representing massive longitudinal datasets. Despite so, particular learning paths for a given consonantal cluster are hard to study because of the high variability between children. Here an example of the paradigmatic case of plosive-liquids :
- /gR/ → /dR/ ; /dR/ → /gR/
- /tR/ → /kR/ ; /kR/ → /tR/
- « 52 Q : ben attends, on essaie de/ de l'touver, si on le trouve pas ze/ ben c'est pas grave hein, ça c'est un gros euh c'est bien.. si on le krouve pas, alors c'est pas grave donc. . » (4)

- **REFERENCES**

1.Morgenstern A., ; Parisse C. « The Paris Corpus ». Cambridge Univ Press. 2012

2.Kass G. « An exploratory technique fo investigating large quantities of categorical data ». Applied Statistics 29 (2), 1980

3.Cuenca et al. « Mutltiresolution Streamgraph approach to time series » IEEE, 2018

4. Sauvage J. « L'acquisition du langage. Un système complexe » Louvain, 2015.