

# A new interactive graphical approach to visualise children phonemes' acquisition over time

Andrea Briglia, Jérémie Sauvage

## ► To cite this version:

Andrea Briglia, Jérémie Sauvage. A new interactive graphical approach to visualise children phonemes' acquisition over time. PaPE 2021 (Phonetics and Phonology in Europe), Universitat Pompeu Fabra Barcelona, Universitat de Barcelone. Association of Laboratory Phonology, Jun 2021, Barcelona, Spain. <https://pape2021.upf.edu/program-4/>. hal-03318570

HAL Id: hal-03318570

<https://hal.archives-ouvertes.fr/hal-03318570>

Submitted on 10 Aug 2021

**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.

Public Domain

## A new interactive graphical approach to visualise children phonemes' acquisition over time

Andrea Briglia    J r mi Sauvage

CoLaJE is a French open access database (available in CHILDES) containing longitudinal recordings of *in vivo* child spoken language from seven children that have been regularly filmed one hour every month from their first year of life until 5. Each record has been transcribed in CHAT, for two children an additional transcription in orthographic norm and IPA is provided. Data is coded in two tier : "*pho*" represents what the child says and "*mod*" represents what the child should have said according to the adult's norm. This sampling scheme meets child language representativeness standards (Stahl & Tomasello, 2004). Data is turned into a machine-readable format by converting it into *.txt* to make it suitable for Python.

We wrote a set of scripts to extract every occurrence of each of the 36 phonemes specific to French. It would have been possible to code Child-Directed Speech too, as everything adults say has been transcribed in these *corpora*, but we decided to focus on production only, even considering that the data amount was already huge. We ordered the 36 phonemes into a hierarchical list according to their articulatory effort (Sauvage, 2015). Aim was to obtain a global overview of the phonemes' acquisition over time and, at the same time, having the possibility to focus on the specific development of any given phoneme in relation to every other one: to do so, we chose to adopt the Multiresolution Streamgraph (Cuenca et al. 2018). This graph has been conceived to overcome readability and scalability issues in the visualisation of multiple time series: it allows a navigation through the different hierarchical levels providing a way to aggregate and disaggregate its structure, as well as to filter and highlight a specific element.

Through this visualisation it is possible to know how many occurrences of any phoneme has been pronounced by any of the seven children in any of the sessions. Results are coherent to current literature on consonants acquisition (McLeod et al., 2018).

A comparison has been made on the children of this dataset, allowing us to establish the amount of differences and similarities between them by overlapping graphs in a synoptic way.

We then tested whether the occurrences in this graph fit with Clements' « Theory of phonological features » : according to the « markedness avoidance » principle, we found that marked phonemes occur most of the times less than unmarked ones (e.g « p » occurs more times than « b », « t » occurs more times than « d », « k » more times than « g » and so on). By focusing on the same group of consonants specified by the voicing contrast, we found how the « feature economy » principle - according to which features present in one segment tend to be used to define others - is coherent to what it is possible to observe from the graphs: as this feature specifies more phonemes than any other one in French, those features appear more frequently and - *a fortiori* - children prioritize their acquisition.

These preliminary results are part of a Phd project: we are currently working on the analyses of the same *corpora* through features instead of phonemes. By visualising the hierarchy of features in the Multistream, we suppose to improve current knowledge on the path of phonological variations (*i.e* the sequence of temporary achieved stages toward the adult norm). To give an example , « tʁaktœʁ » is never pronounced « dʁaktœʁ » but instead « kʁaktœʁ » as « k » is voiceless as « t ». In this way we would try to improve knowledge about the relation between feature hierarchy and acquisition.

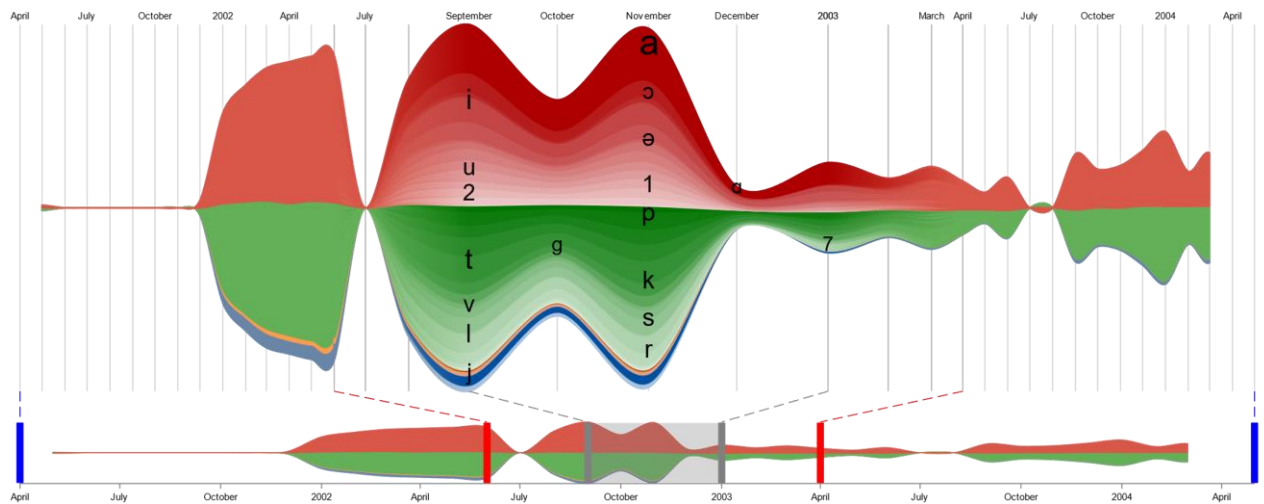


Figure 1. *Multiresolution Streamgraph* obtained from *Madeleine's dataset*  
[https://erickedu85.github.io/presentations/ecuenca\\_multistream\\_vis\\_2018.pdf](https://erickedu85.github.io/presentations/ecuenca_multistream_vis_2018.pdf)

- [1] Morgenstern A. ; Parisse C. (2012), « *The Paris Corpus* ». French language studies 22. 7-12. Cambridge University Press. Special Issue.
- [2] Tomasello, M. and Stahl, D. (2004). « *Sampling children's spontaneous speech: How much is enough?* » . Journal of Child Language, 31:101–121
- [3] Sauvage J. (2015). « *L'acquisition du langage. Un système complexe*. L'Harmattan-Louvain-la-neuve.
- [4] Cuenca E., Sallaberry A., Wang Y., Poncet P. (2018). « *MultiStream: A Multiresolution Streamgraph Approach to explore Hierarchical Time Series* ». IEEE Transactions on visualization and computer graphics, vol.24, no. 12
- [5] McLeod S.; Crowe K. (2018). “*Children's Consonant Acquisition in 27 Languages: A Cross-linguistic Review*”. American Journal of Speech-Language Pathology. P 1-26
- [6] Clements, G. N. & R. Ridouane (2011). “ *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*”. John Benjamins. Amsterdam