



**HAL**  
open science

## The protolexicon at the end of the first year: Evidence from consonant vs. vowel processing in frequent words and part-words in French.

Camillia Bouchon, Sharon Peperkamp, Juan Manuel Toro, Thierry Nazzi

### ► To cite this version:

Camillia Bouchon, Sharon Peperkamp, Juan Manuel Toro, Thierry Nazzi. The protolexicon at the end of the first year: Evidence from consonant vs. vowel processing in frequent words and part-words in French.. International Congress of Infant Studies, Jul 2024, Glasgow, Ecosse, United Kingdom. hal-04780683

**HAL Id: hal-04780683**

**<https://univ-paris8.hal.science/hal-04780683v1>**

Submitted on 13 Nov 2024

**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.

# The protolexicon at the end of the first year: Evidence from consonant vs. vowel processing in frequent words and part-words in French

Bouchon C.<sup>a</sup>, Toro J.-M.<sup>b,c</sup>, Peperkamp S.<sup>d</sup> & Nazzi T.<sup>e</sup>

- a. Laboratoire CHArt–UPEC (Cognitions Humaine et Artificielle), Université Paris Est Créteil, 94380 Bonneuil sur Marne, France.
- b. Universitat Pompeu Fabra, 08002 Barcelona, Spain.
- c. Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.
- d. Laboratoire de Sciences Cognitives et Psycholinguistique (ENS-PSL, CNRS, EHESS), Paris, France.
- e. Integrative Neuroscience and Cognition Center, CNRS – Université de Paris, Paris, France.

At the end of the first year, infants’ mean receptive vocabulary approximates 50 words (CDI measures for French: Kern, 2007). However, these words might be “blended together” with part-words in a protolexicon, as suggested by French-learning-11-month-olds looking similarly at frequent words (*e.g. bonjour, lapin*; hello, rabbit) and frequent part-words (*e.g. korin*, from “encore un”: *one more*) (Ngon et al., 2011). As such, frequency, and not lexical status, would determine whether a portion of the speech input belongs to the protolexicon. Following this idea, the current study evaluated whether consonants and vowels are processed similarly in frequent part-words as they are in frequent words, for which recognition is more impacted by consonantal than vocalic alterations (Poltrock & Nazzi, 2015).

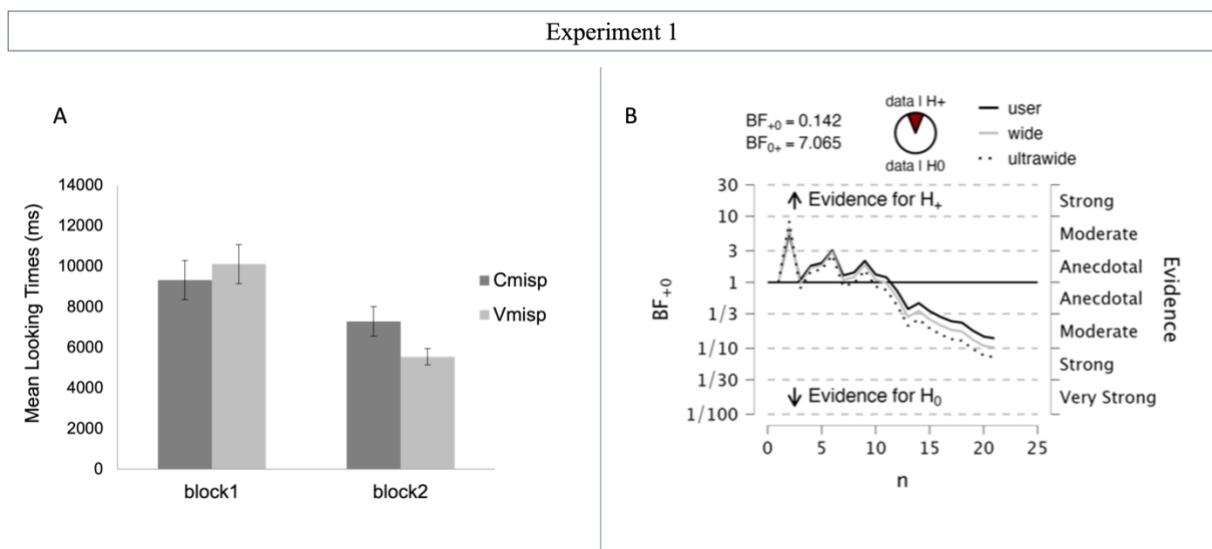
We ran two experiments with HPP. For both, items were selected from a toolbox for phonologizing infant-directed speech corpora automatically from CHILDES orthographic corpora transcription of speech input to French-learning infants under two years of age, considering four phonological rules, *i.e.* liaison, liquid deletion, enchaînement, and “je”-devoicing (Carbajal, Bouchon, Dupoux & Peperkamp, 2018). From the resulting phonological transcriptions, all consonant initial disyllabic words *and* part-words were ranked on their frequency of occurrence. Exp 1’s stimuli derived from the 10 most frequent *part-words* extracted from this ranking (*e.g. korin*) into 10 consonant- (*e.g. kotin*) and 10 vowel mispronunciations (*e.g. korai*). Exp.2’s stimuli derived from the 10 similarly frequent *words* extracted from this ranking (*e.g. bouton*, button) into 10 consonant- (*e.g. bousson*) and 10 vowel mispronunciations (*e.g. bouto*). Note that Exp. 1 and 2’s items were 1.8 times less frequent than in Poltrock & Nazzi (2015).

Exp. 1 compared 12-month-olds’ processing of consonant vs. vowel mispronunciations of the 10 frequent part-words. A repeated-measures ANOVA of looking times showed no effect of type of mispronunciation (**Fig. 1.A**) and a Bayesian paired-sample t-test offered moderate to strong support for H<sub>0</sub> (**Fig. 1.B**), suggesting that the absence of bias towards consonant or vowel

mispronunciations within part-words does not emerge from infants' randomly response. This result could be due to the lack of lexicality of part-words, and/or lower frequency than in Poltrock & Nazzi, (2015) which were 1.8 more frequent than our frequent part-words.

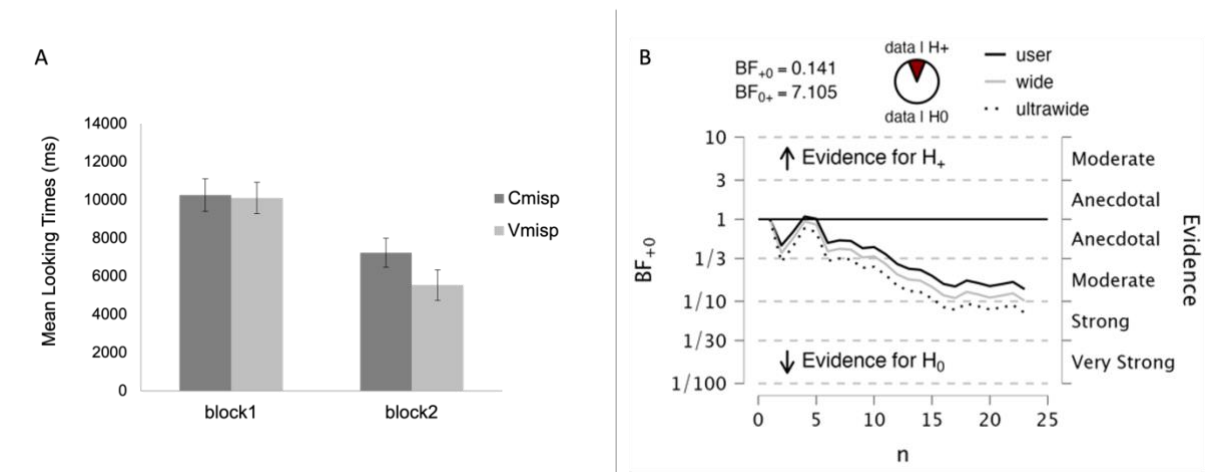
Exp. 2 compared infants' processing of consonant *vs.* vowel mispronunciations of the 10 frequent disyllabic words. These words were familiar to 26% of 12 month-olds on average (Kern S., personal communication), and 33% according to Wordbank. A consonant bias in Exp. 2 would attribute the absence of bias in Exp. 1 to lexical status and not frequency. A repeated-measures ANOVA showed no effect of type of mispronunciation (**Fig. 2.A**) and a Bayesian paired-sample t-test offered moderate to strong support for  $H_0$  (**Fig. 2.B**). Therefore, consonants and vowels are processed equally within less frequent words. Together, the two experiments suggest that small changes in frequency might prevent words from entering infant's protollexicon and/or the display of the consonant bias at this age. How lexical status might interact with frequency for the emergence of the consonant bias remains to be understood.

495 words



**Figure 1. A.** Experiment 1. Patterns of mean looking time in ms to consonantal (Cmisp) and vocalic mispronunciations (Vmisp) of part-words, by block. Bars indicate standard errors across individuals ( $n = 21$ ). **B.** Descriptive plot of the Bayesian paired-sample t-test analysis for Exp 1.  $H_+$  is the alternative hypothesis of a consonant bias (*i.e.* longer LTs for Vmisp, than Cmisp, *e.g.* *korin*: *korai* > *kotin*).

Experiment 2



**Figure 2. A.** Experiment 2. Patterns of mean looking time in ms to consonantal (Cmisp) and vocalic mispronunciations (Vmisp) of words, by block. Bars indicate standard errors across individuals ( $n = 23$ ). **B.** Descriptive plot of the Bayesian paired-sample t-test analysis.  $H_+$  is the alternative hypothesis of a consonant bias (*i.e.* longer LTs for Vmisp than Cmisp, *e.g.* *bouton*: *bouto* > *bousson*).