Brain signatures of native and non-native words in French-learning 24-month-olds: The effect of vocabulary skills

Oytun Aygün, Pia Rämä

▶ **To cite this version:**

Oytun Aygün, Pia Rämä. Brain signatures of native and non-native words in French-learning 24-month-olds: The effect of vocabulary skills. Journal of Neurolinguistics, 2022, 62, pp.101055. 10.1016/j.jneuroling.2021.101055 . hal-03662912

HAL Id: hal-03662912
https://u-paris.hal.science/hal-03662912

Submitted on 8 Jan 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.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0911604421000713
Manuscript_e83482a5b4aca3df26cad0e06409da91

Brain signatures of native and non-native words in French-learning 24-month-olds: the effect of vocabulary skills

Oytun, Aygün [a] & Pia, Rämä [a,b]

[a] Université de Paris, Paris, France, [b] Université de Paris, CNRS, Integrative Neuroscience and Cognition Center, F-75006 Paris, France

Corresponding author:

Rämä, Pia

45, rue des Saints-Pères

75006 Paris

France

Email address: pia.rama@parisdescartes.fr

Declarations of interest: none

© 2021 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract

The event-related potential (ERP) technique provides a temporally accurate measure to distinguish among different linguistic processes. Here, we measured ERPs in response to known words, pseudowords and nonwords in 24-month-old French-learning children to investigate how individual vocabulary skills contribute to the processing of native-like and non-native-like words during a listening task. The N200 was more pronounced for pseudowords than for nonwords while no difference was found between known words and pseudowords. The amplitude difference between known words and pseudowords was, however, correlated with the productive vocabulary. Toddlers with a higher vocabulary score exhibited a bigger difference than toddlers with a lower vocabulary score. Similarly for the frontally distributed late negativity, only those toddlers with higher vocabulary knowledge exhibited a gradient pattern of activity in response to three word types while children with lower vocabulary skills exhibited a similar responsiveness to each word type. Our results suggest that vocabulary skills contribute to the magnitudes of brain signals in response to native and non-native words in a non-referential listening task.

Keywords: lexical processing, vocabulary skills, phonotactics, event-related potentials, developing brain, toddlers

1. Introduction

Between 6 and 12 months of age, infants improve their ability to discriminate between native phonemes while discrimination of non-native phonemes declines (e.g., Werker & Tees, 1984; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; for a review, see Werker, 2018). Native and non-native phonetic discrimination ability at the age of 7.5 months has been shown to predict later vocabulary development: better native language perception predicted higher productive vocabulary score at 18- and 24-months while better non-native phonetic perception predicted lower vocabulary score (Kuhl, Conboy, Padden, Nelson & Pruitt, 2005; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008). Infants become also sensitive to language-specific phonotactic features of their native language before their first birthday (e.g., Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993). Phonotactic knowledge helps infants to segment words from continuous speech which then allows infants to recognize and understand frequently heard nouns (e.g., Tincoff & Jusczyk, 1999, 2012; Bergelson & Swingley, 2012; Parise & Csibra, 2012). Vocabulary growth has a slow pace in the beginning but children start to produce new words at a remarkable rate around the end of the second year of life (e.g., Reznick & Goldfield, 1992; Ganger & Brent, 2004). Consequently, by 24 months children produce (with an extensive individual variation) in average two to three hundred words (e.g., Fenson et al., 1993). At the same time, they also increase the speed and accuracy of native word recognition from speech and lexical representations become more stable (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Fernald, McRoberts, & Swingley, 2001). On the other hand, the ability to learn non-native words diminishes with age and increasing vocabulary knowledge during the second year of life: 14-month-old infants were

capable of associating novel visual items with non-native sounds, while older infants (18- to 20-months), especially with higher vocabulary skills, associated only native sounds with novel pictures (Graf Estes, Edwards, & Saffran, 2011; May & Werker, 2014; Hay, Graf Estes, Wang, & Saffran, 2015). Less is, however, known about brain signals associated with processing of phonotactically native and non-native words in the developing brain.

Neuroimaging and electrophysiological techniques, such as near-infrared spectroscopy (NIRS) and event-related potentials (ERPs), have been recently used to study language development in infant populations. These techniques capture linguistic processes even in the absence of an overt (motor or verbal) reaction or attention that are often required in behavioral tasks. While the NIRS detects neural activity with relatively high spatial accuracy, the ERPs measure electrical brain activity in a millisecond range, thus providing a temporally accurate measure of language processes. The ERP components correspond, at least partly, to distinct linguistic sub-processes (e.g., acoustic, phonological, or lexical-semantic) providing a possibility to investigate the timing and the sequence of these processes. Besides, the ERP technique is easily applicable to young children (for reviews see, e.g., Johnson et al., 2001; Thierry, 2005). Moreover, several studies have shown that even though the effects were not found behaviorally, the ERPs showed sensitivity to experimental or participant group manipulations (e.g., de Haan & Nelson, 1997; Rivera-Gaxiola, Csibra, Johnson, & Karmiolff-Smith, 2000a, b; Kozou et al., 2005; Sokka et al., 2016) providing further benefits to use the ERPs to map language development.

Increasing word knowledge during the second year of life has been associated with specialized electrophysiological activity. When infants were presented with known and unknown words, the amplitudes of two ERP components, the N200 and N375, were larger for known than

for unknown words (Mills, Coffey-Corina, & Neville, 1993, 1997; Conboy & Mills, 2006). In younger infants (13- to 17 months), the components were broadly distributed over the scalp (Mills, Coffey-Corina, & Neville, 1997) while in 20-month-olds they were observed mainly over the left temporal and parietal recording sites (Mills et al., 1993). Also vocabulary skills contributed to the findings: the N200 was more pronounced for known than for unknown words both in low- and high-producers while the difference in the amplitude of the N375 between word types was found only in the high-producers. In bilingual infants, the occurrence and the distribution of ERPs were different for dominant and non-dominant languages (Conboy & Mills, 2006). Another ERP study, conducted in 11-month-old English-learning infants, suggested that a more pronounced N2 component in response to familiar words compared to unfamiliar words is associated with an automatic capture of attention towards familiar words (Thierry, Vihman, & Roberts, 2003). These findings suggest that age, vocabulary skills, and language experience are associated with an increased neural specialization in processing of word meaning.

In a picture-word congruency task, a more pronounced frontally distributed negativity between 100 and 500 ms post word-onset was observed for congruous words, that is, for those words that correctly labeled the pictures when compared to the same words in incongruous picture contexts (Friedrich & Friederici, 2004, 2005). Also, a short-term familiarization to previously unknown word-forms augments the magnitude of the N200-500 component in 10- to 12-month-old infants during the test phase (e.g., Kooijman, Hagoort, & Cutler, 2005; Kooijman, Hagoort, & Cutler, 2009; Goyet, de Schonen, & Nazzi, 2010; Männel & Friederici, 2010). These familiarity effects between 200 and 500 ms after word onset were observed mainly over the frontal recording sites (Kooijman et al., 2009; Goyet et al., 2010). While the N200-500 component (or temporally similar early negativity) has been associated with word familiarization

and lexical congruence, the N400 component has been linked to more mature lexical-semantic processing, semantic integration, and lexical expectancy both in adults (for review, see Kutas & Federmeier, 2011) and in children (e.g., Friedrich & Friederici, 2004; Torkildsen et al., 2006; Rämä, Sirri, & Serres, 2013; Sirri & Rämä, 2015). The N400 effect (more negative response to incongruent *versus* congruent stimulus pairs) in adult participants typically appears over the central-posterior recording sites. In young children, the N400 effect has been observed over distributed scalp positions and also in variable time windows from 400 ms to 1200 ms. Some studies have also reported a late frontally distributed negativity, appearing 600 ms to 1200 ms after word-onset (e.g., Mills et al., 1993, 1997; Friedrich & Friederici, 2004; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007a). This late component has been suggested to reflect further lexical search (Junge, Kooijman, Hagoort, & Cutler, 2012), lexical-semantic processing (e.g., Friedrich & Friederici, 2004) or increased attentional processes during language tasks in young children (e.g., Friedrich & Friederici, 2004; Torkildsen et al., 2006, 2007a).

Some evidence shows that the infant brain is also differentially responsive to meaningless pseudowords (i.e., nonsense words that obey the phonetic, prosodic, and phonotactic rules of a given language) and to nonwords (i.e., nonsense words that violate phonotactic regularities of a language). The N200 and N375 components were found for pseudowords but not for backwardly presented words in 20-month-old, English-learning infants (Mills et al., 1993) while in younger infants (13- to 17-months) no significant difference between pseudowords and backwardly presented words was found (Mills et al., 1997). Differences between the two word types were found over the right hemisphere with a broad distribution from anterior to posterior sites. It should be noted that while backwardly presented words serve as a control for the auditory stimuli by having similar physical characteristics of speech stimuli, they violate not only phonotactic,

but also phonetic and prosodic features of speech signal. In another study, conducted in 18-month-olds, German-speaking infants were presented with monosyllabic pseudowords either with a phonotactically legal or illegal consonant onset structure (Steber & Rossi, 2020). The results demonstrated a larger negativity for legal than for illegal words over the anterior and posterior scalp positions in a time window of 500 ms to 1000 ms. In a lexical-semantic picture-word context, a (short-lived) larger negativity for pseudo- than for nonwords was obtained over the frontal scalp positions in 12-month-olds while only in the 19-month-olds, a larger broadly distributed negativity was found (Friedrich & Friederici, 2005). The broadly distributed negativity might reflect a more mature mechanism of semantic integration, in response to phonotactically legal, but not to illegal words (Friedrich & Friederici, 2005). Interestingly, only those 19-months-olds whose later language outcome (tested at 30-months) was age-adequate presented a significant difference between the processing of pseudo- and nonwords while children with lower vocabulary knowledge eleven months later did not show differential responses between the word types (Friedrich & Friederici, 2006). These effects regarding the retrospective language outcome were found over two temporally separate time-windows: between 250 ms and 500 ms, and between 600 ms and 800 ms. To sum up, previous behavioral evidence shows that increasing vocabulary knowledge diminishes the ability to associate novel visual items with phonotactically non-native labels or non-linguistic sounds (Graf Estes, Edwards, & Saffran, 2011; May & Werker, 2014). There are only a few previous ERP studies that investigated brain signals in response to phonotactically legal *versus* illegal words in 12 to 20 month old infants (Mills et al., 1997; Friedrich & Friederici, 2005; Steber & Rossi, 2020). However, in these ERP studies either stimulus properties or concurrent comprehensive and expressive language skills were not fully controlled or systematically analyzed. The current

study aims to clarify whether the magnitudes and the timings of the ERPs in response to pseudowords and nonwords are dependent on vocabulary knowledge to further understand the role of vocabulary in nonnative word processing.

In the present study, 24-month-old, French-learning toddlers were presented with known words, pseudowords and nonwords during the experiment in a non-referential context and their comprehensive and expressive vocabulary skills were evaluated by a parental questionnaire. The participants were presented with four different pseudowords (e.g., “meautai”), four phonotactically illegal nonwords (e.g., “zbaiton”), and four familiar words in French (e.g., “bateau”, boat). Our aim was to investigate how comprehensive and productive vocabulary skills contribute to the magnitudes and the timings of the ERPs in response to phonotactically legal and illegal words in French-learning toddlers. Based on earlier behavioral findings (Graf Estes, Edwards, & Saffran, 2011; May & Werker, 2014), we expected that 24 month old French-learning children with higher vocabulary skills would exhibit a larger difference in the amplitudes between the word types indicating a greater sensitivity to word nativeness and meaning also in a non-referential task. Second, we aimed to investigate whether the ERPs are differentially modulated in response to pseudowords and nonwords during the course of the experiment. A short exposure to previously unknown words has been shown to augment the magnitude of the N200-500 component in 10- to 12-month-olds (e.g., Kooijman et al., 2005; Goyet et al., 2010) while in older infants an increase in the amplitudes of the N200-500 component for novel words was observed only after a referential training task (Mills, Plunkett, Prat & Schafer, 2005). We expected modulations to occur in response to phonotactically legal pseudowords but not in response to phonotactically illegal nonwords.

2. Material and methods

2.1. Participants

The final participant sample included twenty-three (17 girls) 24-month-old children (min: 23 months 7 days, max: 25 months 5 days, mean: 24 months 9 days). All children came from monolingual French-speaking families in the Parisian region. They were recruited through a database of parents who previously volunteered to participate in child development studies. All children were born full-term and none of them suffered from any known hearing or language impairment. Also, there was no atypical development of language in the families. The mean number of years for maternal education was 16 (range from 11 to 19 years). Parents gave written informed consent before participation. Twenty-four other participants were tested but their data were rejected due to noisy (less than ten accepted trials in one experimental condition) electroencephalogram (EEG) recording (n = 14,), due to technical problems while recording (n = 1), or refusal to put on the EEG cap (n = 9).

The French translation of the MacArthur-Bates Communicative Development Inventory for Words and Sentences (CDI; Fenson et al., 1993) was used to measure the productive vocabulary size. The parents were asked to fill the inventory at home during the two weeks following the study. The mean number of produced words was 274 (SD = 190, range from 16 to 686 words) and the mean number of comprehended words was 496 (SD = 143, range from 262 to 751 words). The mean numbers and the ranges were similar to those reported in earlier studies testing French-learning children of the same age (e.g., Rämä et al., 2013; Helo, Azaiez, & Rämä,

2017; Helo, van Ommen, Pannasch, Danteny-Dordoigne & Rämä, 2017). Based on the median split (467 words for comprehension) of the comprehensive vocabulary inventory scores, the participants were divided into comprehensive language groups: normal-to-high comprehenders (9 girls) and normal-to-low comprehenders (8 girls). Independently, based on the median split (243 words for production) of the productive vocabulary inventory scores, the participants were divided into productive language groups: normal-to-high producers (11 girls) and normal-to-low producers (6 girls). Children in the low comprehender group comprehended on average 378 words (SD = 66, range 262 – 449 words) and in the low producer group produced on average 120 words (SD = 77, range 16 – 231 words) words while children in the high comprehender group comprehended on average 605 words (SD = 99, range 467 – 751 words) and in the high producer group produced on average 415 words (SD = 145, range 243 – 686 words). The study was conducted in conformity with the declaration of Helsinki and approved by the Ethics Committee of the University of Paris Descartes.

2.2. Stimuli

The audiostimuli consisted of three types of CVCV or CCVCV words: (a) familiar French words (ballon / balɔ̃ / (ball), cheveux / ʃəvø / (hair), maison / mɛzɔ̃ / (house), bateau / bato / (boat)), (b) pseudowords (lechon / ləʃɔ̃, vabeux / vabø, meautai / motɛ, bonsa / bɔ̃za, and (c) nonwords (vchonbe / vʃɔ̃bə, chbeula / ʃbøla, bmeausai / bmozɛ, zbaiton / zbɛtɔ̃ (Table I). Words were selected from three databases: Cross Linguistic Lexical Norms (cdi-clex.org), Stanford Word bank (wordbank.standford.edu) and from our own BabyLab database. Before the experiment, the parents were sent a questionnaire and asked to evaluate if their child understood

and/or produced the (familiar) words used in our experiment (among nine other words). They were also asked to express their confidence about the judgement for each word with a 5-point scale (*I am sure my child knows the word, I am somewhat confident that my child knows the word, I don't know, I am somewhat confident that my child does not know the word, I am sure my child does not know word*). Three parents were not sure whether their child knew one of the words. These children were presented with a visual display of four images, and were asked to point to the correct image after naming. Two of them were able to point to the correct image and were included, one could not, and he was excluded.

Pseudowords and nonwords were constructed by swapping phoneme positions between two word pairs: ballon/balɔ̃ and cheveux/ʃəvø, and maison/ mɛzɔ̃ and bateau/bato). As an example, pseudoword /lechon/ was constructed from the words /ballon/ and /cheveux/ by combining the third phoneme of ballon/balɔ̃, the second phoneme of cheveux/ʃəvø, the first phoneme of cheveux/ʃəvø and the fourth phoneme of ballon/balɔ̃. Three other pseudowords and also CCVCV nonwords were constructed in a similar manner but an additional consonant was added to CCVCC words to create an illegal onset consonant cluster in French (Dell, 1995). There was no meaning in any of the pseudo- or nonwords. Bi-phoneme (token) frequencies were retrieved from the database Lexique (New, Pallier, Brysbaert, & Ferrand, 2004). Average phoneme frequencies and word durations (ms) are shown in Table I. The duration of words varied between 487 ms and 690 ms (mean duration 578 ms, SD = 69 ms). Mean durations of familiar, pseudowords, and nonwords were 561 ms, 526 ms, and 646 ms, respectively.

The words were recorded with TASCAM Linear PCM Recorder (DR-07MKII). The speaker was a female native French phonetician, and she pronounced the words written in the

phonetic alphabet. She was asked to pronounce the words slowly, clearly and as neutral as possible. The stimuli were normalised with Praat software (version 6.0.28).

2.3. Procedure

Before the testing, parents received a picture book that explained in a child-directed manner what happens during their laboratory visit (e.g., EEG capping, testing session). This book was meant to help both parents and children to be prepared for the testing session. After arriving at the laboratory, the children were allowed to play in the laboratory environment, and they were also familiarized with the EEG cap. During the experiment the child sat on the lap of a parent in a dimly lit room at 70 cm from a computer screen and two loudspeakers from which the sound was delivered. Parents were instructed not to communicate with the child during the experiment. Children were allowed to watch a silent cartoon on the screen throughout the experiment or to play with small toys positioned on the table in front of them during the experiment. Each word (12 words) was presented 60 times during the experiment. Each word type (familiar words, pseudowords and nonwords) included four examples, thus, each child heard each word type 240 times and the total number of stimuli was 720. The presentation order was pseudo-randomised (with consecutive sets of 12 trials) to ensure an even distribution of each word type during the experiment. The same word was never repeated more than twice. The mean stimulus onset asynchrony (SOA) was 1000 ms with a jitter ranging from 950 to 1050 ms. The experiment lasted about 12 minutes.

2.4. EEG recording and pre-processing

Continuous electroencephalogram (EEG) was recorded (band-pass was 0.1 – 100 Hz, sampling rate was 250 Hz) from 128 electrodes using a Geodesic Sensor Net (GSN, NetStation EGIS V2.0). Impedances were kept below 50 kΩ. EEG was filtered (0.3–30 Hz), segmented (900 ms, beginning 100 ms before target word onset), and ocular artefacts were removed with an ocular artefact removal (OAR) algorithm (Gratton, Coles, & Donchin, 1983). The 100-ms pre-stimulus period determined the baseline for amplitude measures. The epochs including eye-movement artefacts, using the electrodes above and below the eyes, as well as motion artefacts, exceeding ± 180 lV in any channel were excluded. The epochs with more than 30% contaminated channels were also rejected. Individual bad channels were replaced using spherical spline interpolation.

The data was divided into three equal bins according to the number of repetitions of each word type. Each bin consisted of twenty repetitions of each word, thus 80 repetitions of each word type, and lasted about four minutes. The ERPs were analysed for early, middle and late bins. To obtain ERPs, the epochs were averaged separately for each subject, word type (known word, pseudoword, nonword), and phase (early, middle, late). The averaged waveforms were re-referenced to the average of mastoids and baseline corrected. The average waveforms were grand-averaged across all participants. The mean number of accepted EEG epochs was 46 (range: 23–73; SD = 15), 46 (range: 20–73; SD = 15), and 45 (range: 20–72; SD = 16) for the known words, native pseudowords, and non-native pseudowords, respectively. The mean number of accepted EEG epochs was 48 (range: 25–72; SD = 16), 46 (range: 15–72; SD = 17), and 43 (range: 16–73; SD = 18) for the early, middle, and late bin, respectively. The mean number of accepted trials was not significantly different among the word types (one-way ANOVA: $F(2, 68) = 0.004$, $p = 0.996$) or bins (one-way ANOVA: $F(2, 206) = 0.485$, $p = 0.618$).

2.5. Data and statistics analyses

Similarly to previous literature (e.g., Mills et al., 1993, 1997; Friedrich & Friederici, 2005; Kooijman et al., 2005; Bosseler, Clarke, Tavabi, Larson, Hippe, Taulu, & Kuhl, 2020), the following three measurement windows were identified from our data: early positivity (time window of 100-250 ms), N200 (time window of 250-500 ms) and late negativity (time window of 500-800 ms). Mean amplitudes were measured over the time-windows post stimulus, relative to a -100 to 0 ms prestimulus baseline period.

Repeated measures analysis of variance (ANOVA) included Word Type (known word, pseudoword, nonword), Phase (early, middle, late), Recording Area (frontal, parietal) and Hemisphere (left, right) as within-subject factors and Productive or Comprehensive Vocabulary Group (separately) as between-subject factors. The mean amplitudes extracted from 16 electrodes over the frontal and the parietal recording areas were included in the statistical analyses. The frontal and the parietal recording sites included the following electrode positions: 12, 13 (FC1), 18, 19 (F1), 20, 24 (F3), 28 (FC5), and 29 (FC3) (frontal left), 5, 112 (FC2), 10, 4 (F2), 118, 124 (F4), 117 (FC6), and 111 (FC4), (frontal right), 52 (P5), 53, 59 (P7), 60 (P3), 61 (P1), 65 (PO7), 66, and 67 (PO3) (left parietal), and 92 (P6), 86, 91 (P8), 85 (P4), 78 (P2), 90 (PO8), 84, and 77 (PO4) (right parietal). The statistical analyses were conducted with SPSS (IBM SPP statistics, version 24) and the Greenhouse-Geisser correction was applied for non-sphericity when appropriate.

3. Results

3.1. Early positivity (P100)

The amplitudes of the P100 were slightly more pronounced for nonwords (mean = 2.52 μV, SD = 3.87 μV) than for known words (mean = 2.10 μV, SD = 3.65 μV) and pseudowords (mean = 2.12 μV, SD = 3.61 μV). There was, however, no main effect of Word Type ($F(2, 44) = 0.52$, $p = 0.588$, $\eta^2 = 0.023$) or Hemisphere ($F(1, 22) = 1.48$, $p = 0.236$, $\eta^2 = 0.063$) on the amplitudes of positivity, but the interaction between Word Type and Hemisphere reached the significance ($F(2, 44) = 3.48$, $p = 0.041$, $\eta^2 = 0.137$). In further analyses, however, significant main effect of Word Type were not observed either over the left ($F(2, 44) = 0.92$, $p = 0.402$, $\eta^2 = 0.040$) or the right ($F(2, 44) = 0.52$, $p = 0.590$, $\eta^2 = 0.023$) hemisphere, suggesting that the P100 is not modulated by word type (Fig. 1).

When we re-examined the positivity with Productive and Comprehensive Vocabulary Groups as between subject variables, we observed, besides the interaction between Word Type and Hemisphere (Comprehension: $F(2, 42) = 3.40$, $p = 0.045$, $\eta^2 = 0.139$; Production: $F(2, 42) = 3.37$, $p = 0.046$, $\eta^2 = 0.138$), an interaction between Word Type and Comprehensive Vocabulary Group ($F(2, 42) = 3.85$, $p = 0.037$, $\eta2 = 0.155$). The amplitudes were more positive for nonwords than for pseudowords in the normal-to-low comprehenders ($t(10) = -2.31$, $p = 0.044$, $d' = 0.773$) while in the normal-to-high comprehenders, there was a tendency for more positive

amplitudes for pseudowords than for nonwords (t(11) = -2.09, p = 0.061, d' = 0.624). Also, an interaction between Hemisphere and Productive Vocabulary Group (F(1, 21) = 5.96, p = 0.024, η2 = 0.221) was found. However, even though the amplitudes were larger over the right hemisphere in the high-vocabulary group and larger over the left hemisphere in the low-vocabulary group, significant differences between the left and right hemispheres in either vocabulary group were not found (all p > 0.05).

Figure 1 here

3.2. N200

The amplitudes of the N200 were more negative for known words (mean = -0.98 μV, SD = 4.61 μV) and pseudowords (mean = -1.22 μV, SD = 4.80 μV) than for nonwords (mean = 1.02 μV, SD = 4.58 μV). The main effect of Word Type (F(2, 44) = 9.95, p = 0.000, η^2 = 0.311) and the interaction between Word Type and Recording Area was significant (F(2, 44) = 12.35, p < 0.001, η^2 = 0.359) on the amplitudes of the N200. The main effect of Word Type was observed over the frontal (F(2, 44) = 17.98, p < 0.001, η^2 = 0.450) but not over the parietal (F(2, 44) = 2.85, p = 0.072, η^2 = 0.115) recording sites (Figs. 1 & 2). The amplitudes were significantly more negative for known words than for nonwords (t(22) = -5.03, p < 0.001, d' = 0.973) over the frontal recording sites (Fig. 3 A). The difference between pseudowords and nonwords was also significant (t(22) = -4.63, p < 0.001, d' = 0.907) while no difference was found between known words and pseudowords (t(22) = -0.97, p = 0.341, d' = 0.199). No other main effects (Phase or Hemisphere) or interactions were observed.

When we re-examined the N200 component with Productive and Comprehensive Vocabulary Groups as between subject variables over the frontal recording sites, we observed, besides the main effect of Word Type (Comprehension: $F(2, 42) = 19.18$, $p < 0.001$, $\eta^2 = 0.477$; Production: $F(2, 42) = 18.14$, $p < 0.001$, $\eta^2 = 0.463$), the main effect of Comprehensive Vocabulary Group ($F(1, 21) = 9.18$, $p = 0.006$, $\eta 2 = 0.304$). The amplitudes were significantly more negative in the normal-to-low than in the normal-to-high comprehenders ($t(21) = 3.03$, $p = 0.006$, $d' = 0.814$, Fig. 3 C). Even though the interaction between Word Type and Vocabulary Group was not significant (($F(2, 42) = 2.20$, $p = 0.132$, $\eta 2 = 0.095$), we calculated correlations between vocabulary scores and differences in amplitudes between two word types (known words *minus* pseudowords, known words *minus* nonwords, and pseudowords *minus* nonwords). The results showed that the number of produced words correlated with the difference in amplitudes between known words and pseudowords. The bigger the difference was in the amplitudes (more negativity for known words than for pseudowords), the bigger the productive vocabulary size was ($r = -0.597$, $p = 0.003$) indicating that children with higher vocabulary skills were more sensitive to word meaning. No other correlations were found.

Figure 2 here

3.3. Late negativity

The amplitudes of the late negativity were more pronounced for known words (mean = -3.45 μV, SD = 5.50 μV) and for pseudowords (mean = -3.25 μV, SD = 5.03 μV) than for nonwords (mean = -2.36 μV, SD = 4.84 μV). The main effect of Word Type ($F(2, 44) = 2.51$, $p = 0.104$, $\eta^2 = 0.102$) was not significant on the amplitudes of the late negativity, but the interaction between Word Type and Recording Area reached significance ($F(2, 44) = 8.95$, $p = 0.001$, $\eta^2 = 0.289$). Similarly to the N200, the main effect of Word Type was observed over the frontal ($F(2, 44) = 5.58$, $p = 0.014$, $\eta^2 = 0.202$) but not over the parietal ($F(2, 44) = 2.01$, $p = 0.147$, $\eta^2 = 0.084$) recording sites (Fig. 1 & 2). The amplitudes were more negative for known words than for nonwords ($t(22) = -2.67$, $p = 0.014$, $d' = 0.480$) over the frontal recording sites (Fig. 3 B). The difference between pseudowords and nonwords was almost significant ($t(22) = -2.07$, $p = 0.050$, $d' = 0.397$) while no difference was found between known words and pseudowords ($t(22) = -1.88$, $p = 0.074$, $d' = 0.374$). No other main effects (Phase or Hemisphere) or interactions were observed.

When we re-examined the late negativity with Productive and Comprehensive Vocabulary Groups as between subject variables over the frontal recording sites, we observed, besides the main effect of Word Type (Comprehension: $F(2, 42) = 5.28$, $p = 0.017$, $\eta^2 = 0.201$; Production: $F(2, 42) = 6.11$, $p = 0.008$, $\eta^2 = 0.225$), a significant interaction between Word Type and Productive Vocabulary Group ($F(2, 42) = 5.10$, $p = 0.016$, $\eta2 = 0.196$). There was a significant difference between amplitudes for known words and nonwords ($t(11) = -3.51$, $p = 0.005$, $d' = 0.949$), and pseudowords and nonwords ($t(11) = -2.73$, $p = 0.020$, $d' = 0.782$) in normal-to-high word producers (Fig. 3 D). Also, the difference between known words and pseudowords almost reached the significance ($t(11) = -2.11$, $p = 0.059$, $d' = 0.652$). No significant differences were found in the normal-to-low word producers (all $p > 0.05$). The

results indicate that the late negativity was modulated as a function of word type only in the group of normal-to-high producers. Moreover, there was a tendency for correlation between productive vocabulary skills and the difference in amplitudes between words and pseudowords. The bigger the difference was in the amplitudes, the bigger the productive vocabulary size was. However, the correlation did not quite reach statistical significance ($r = -0.402$, $p = 0.057$). No other correlations were found.

Figure 3 here

4. Discussion

In the present study, we investigated the effects of comprehensive and productive vocabulary skills on brain activity in response to known words, pseudowords and nonwords during the course of an ERP experiment in French-learning toddlers. The N200 was more pronounced for pseudowords than for nonwords while no difference was found between known words and pseudowords. The amplitude difference between known words and pseudowords was, however, correlated with the productive vocabulary. Toddlers with a higher vocabulary score exhibited a bigger difference than toddlers with a lower vocabulary score. Also, the amplitude of the frontally distributed late negativity, measured in the time window of 500 ms to 800 ms, was modulated by an interaction between word type and vocabulary group: children with higher (productive) vocabulary knowledge exhibited a gradient pattern of activity in response to different word types while children with lower vocabulary skills responded similarly to each

word type. There are only a few earlier ERP studies investigating brain signals in response to phonotactically illegal nonwords in an explicit (Friedrich & Friederici, 2005) or an implicit (Mills et al., 1993; Steber & Rossi, 2020) experimental context in a similar age group. The effects of concurrent vocabulary skills were not systematically evaluated in these studies. In accordance with earlier findings, our results showed that the N200 and late negativity were larger for pseudowords than for nonwords, suggesting that phonotactically legal pseudowords activate lexical search more efficiently than illegal nonwords also in 24-month-old toddlers. In addition, our results provided further evidence about the influence of concurrent vocabulary skills on processing of pseudowords and nonwords.

A negative component around 200 to 500 ms after word onset has been earlier associated with word recognition, word meaning, familiarity, and learning (Mills et al., 1993, 1997; Friedrich & Friederici, 2005; Kooijman et al., 2005; Kooijman et al, 2009; Goyet et al., 2010; Männel & Friederici, 2010). Our results showed that the vocabulary group had the main effect on the N200 amplitudes: the amplitudes were more negative for normal-to-low than for normal-to-high comprehenders, regardless of the word type. This finding is interesting in the light of previous evidence showing that the N200 in response to lexical priming (higher negativity for congruent *versus* incongruent lexical items) was shown to be more pronounced in infants at a familial risk of dyslexia (Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007b) and in infants with lower language skills (Friedrich & Friederici, 2006). The enhanced lexical priming effect in children with lower vocabulary skills was suggested to be due to a slower or more effortful lexical-phonological processing (Friedrich & Friederici, 2006). Our results further suggest that more effortful processing in children with lower vocabulary skills might explain the increased N200 also in a non-referential experimental condition. The N200-500 is typically

observed over the frontal sites, and it is larger for familiar or known words compared to unknown words. In the current experiment, the negative response between 250 and 500 ms was larger for phonotactically familiar compared to phonotactically unfamiliar words but equally strong for known words and for pseudowords. A similar finding regarding known words and unfamiliar words has been earlier reported in 17- to 21-month-old English learning infants (Mills et al., 2005). The authors suggested that the use of infant directed speech in their study may have boosted attention towards the unfamiliar words thereby masking effects of semantic processing of familiar words. In our study, the words were not pronounced in infant directed speech so participants' word knowledge is a more likely explanation than a boosted attention towards pseudowords.. The amplitude difference between known words and pseudowords was shown to be correlated with productive vocabulary skills, demonstrating that the higher the vocabulary score was, the bigger the difference was in amplitudes between known and pseudowords. In toddlers with higher vocabulary scores, the amplitude was more negative for words than for pseudowords while in toddlers with lower vocabulary scores, the amplitudes for known and pseudowords were similar than for pseudowords in high producers. Even though all the participants, independent of their vocabulary score, knew the experimental words, the strength of word representations as well as the age of their acquisition might have differed in the two groups, that could explain the correlation effect.

The amplitude of the late negativity, measured in the time window of 500 to 800 ms, was modulated by an interaction between lexical status and vocabulary group: children with higher (productive) vocabulary knowledge exhibited a gradient pattern of activity in response to different word types while children with lower vocabulary skills exhibited a similar responsiveness to each word type. Earlier research has shown that vocabulary skills modulate the

occurrence (e.g., Torkildsen, Hansen, Svangstu, Smith, Simonsen, Moen & Lindgren, 2009; Rämä et al., 2013), distribution (e.g., Mills et al., 1993, 2005), and timing (e.g., Junge, Kooijman, et al., 2012) of the language-related ERPs. For instance, 20-month-old toddlers with high productive vocabularies presented an N400 incongruity effect to violations of newly trained associations between novel words and pictures, whereas 20-month-olds with low productive vocabularies did not (Torkildsen et al., 2009). Also, the N400 priming effect for spoken words was obtained in 24-month-olds and a subgroup of 18-month children with normal-to-high productive skills, which suggests that more advanced productive vocabulary skills contribute to the taxonomic organization of the developing lexical-semantic system (Rämä et al., 2013). Children with lower vocabularies have been shown to present a bilateral or a more scattered distribution of ERPs in response to known and unknown words than children with higher vocabularies, suggesting that neural networks become more specialized with increasing vocabulary skills (Mills et al., 1993; Mills et al., 2005). Moreover, it has been shown that, even though both high and low vocabulary groups displayed a familiarity effect at the time window of 200 ms to 500 ms, only infants with greater vocabularies continued to exhibit the effect at a later time window from 500 ms to 650 ms which led the authors to suggest that the extended response might reflect an additional processing stage related to lexical search for the word (Junge, Kooijman, et al., 2012). Our results demonstrated the amplitude difference between pseudowords and nonwords both at the earlier and the later time windows, suggesting that both word (phonotactic) familiarity and lexical search processes were affected by vocabulary knowledge. However, even though the ERP technique provides a temporally accurate measure to distinguish among different (linguistic) processes and track their time courses (Kutas & Federmeier, 20110), several cortical processes and components contribute to the scalp distribution of the ERP

responses (for review, see Näätänen & Picton, 1987). Accordingly, our results do not suggest that there are two distinct neural processes and further investigations are needed to distinguish between different cognitive and neural events underlying the waveforms.

In accordance with earlier findings in 11- (Thierry et al., 2003) and 20- (Mills et al., 1993) month-olds, the P100 was insensitive to word type also in our study. Previously, the P100 was shown to be more pronounced over the left hemisphere in 20-month-olds (Mills et al., 1993) and in 13- to 17-month-olds with higher comprehension skills (Mills et al., 1997) while in 13- to 17-month-olds with lower comprehension skills, the P100 was bilaterally distributed (Mills et al., 1997), suggesting that age and vocabulary skills contribute to a more focalized, left-hemispheric, distribution of activity. In our study, the distribution of the P100 was not asymmetrically distributed. Also, none of the components were modulated during the course of the experiment. Earlier studies have demonstrated that only a few repetitions of novel words cause an increase in negativity both in explicit (Friedrich & Friederici, 2011; Junge, Cutler & Hagoort, 2012) and implicit (e.g., Kooijman et al., 2005; Kooijman et al., 2009) tasks, at least in younger infants. In older infants, an increase in N200-500 amplitude was observed only when the novel words were paired with new objects, not when repeated in isolation (Mills et al., 2005). Similar to our finding, a greater negativity in response to pseudowords than to nonwords (when repeated in isolation) was not modulated during the course of the experiment in 18-month-olds (Steber & Rossi, 2020). There are at least two possibilities to explain the lack of modulation in our study: it is possible that the modulation in toddlers occurs only in the presence of a referential learning, or modulation occurs already after presentation of a few items and thus not detectable in our study.

To conclude, our results demonstrated that frontally distributed N200 and late negativity were modulated by word type and vocabulary skills. The responses were more pronounced for pseudowords than for nonwords while no difference was found between known words and pseudowords. However, vocabulary knowledge affected toddlers' reactivity to different word types: only toddlers with larger vocabularies exhibited a gradient pattern of signal magnitude in response to different word types. It has been earlier suggested that children with less vocabulary knowledge might have less detailed phoneme perception (Graf Estes et al., 2011) or weaker memory representations for native phoneme clusters (Friedrich & Friederici, 2006). Also, accumulation of vocabulary and phonotactic knowledge may limit the pertinent lexical candidates to be learnt and children with larger vocabulary are more likely to reject phonotactically illegal labels in a word learning task (Graf Estes et al., 2011). Based on our findings, we further propose that vocabulary knowledge contributes to the processing of native-like and non-native-like words in a non-referential task both at the level of word recognition and lexical search. Our results also support the use of ERP technique to investigate temporal sequences of linguistic processes in developmental populations.

Acknowledgements: We thank all families for their participation and contribution to this research. This work is supported by PME DIM Cerveau et Pensée 2013 and a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083 – LabEx EFL). It contributes to the IdEx Université de Paris – ANR-18-IDEX-0001.

References

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences,* 109(9), 3253-3258.

Bosseler, A. N., Clarke, M., Tavabi, K., Larson, E. D., Hippe, D. S., Taulu, S., & Kuhl, P. K. (2020). Using magnetoencephalography to examine word recognition, lateralization, and future language skills in 14-month-old infants. *Developmental Cognitive Neuroscience*, 47, 100901.

Conboy, B. T., & Mills, D. L. (2006). Two languages, one developing brain: Event□related potentials to words in bilingual toddlers. *Developmental Science*, 9(1), F1-F12.

de Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six□month□old infants: A neurobehavioral study. *Child Development*, 68(2), 187-210.

Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua*, 95(1-3), 5-26.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i-185.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., ... & Reilly, J. S. (1993). MacArthur Communicative Inventories: User's guide and technical manual. San Diego.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228-231.

Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In Approaches to Bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition. Volume 1 (pp. 97-123). Benjamins.

Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287-295.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16(8), 1465-1477.

Friedrich, M., & Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17(11), 1785-1802.

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, 43(1), 1-12.

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40(4), 621.

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, 1332, 75-89.

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy*, 16(2), 180-197.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4), 468-484.

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child development*, 86(1), 10-22.

Helo, A., Azaiez, N., & Rämä, P. (2017). Word Processing in Scene Context: An Event-Related Potential Study in Young Children. *Developmental Neuropsychology*, 42(7-8), 482-494.

Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior and Development*, 49, 248-266.

Johnson, M. H., de Haan, M., Oliver, A., Smith, W., Hatzakis, H., Tucker, L. A., & Csibra, G. (2001). Recording and analyzing high-density event-related potentials with infants using the Geodesic Sensor Net. *Developmental Neuropsychology*, 19(3), 295-323.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, 50(14), 3702-3712.

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, 15(4), 463-473.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402-420.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24(1), 109-116.

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten□month□olds. *Infancy*, 14(6), 591-612.

Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, 199(1-2), 31-39.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the" critical period". *Language Learning and Development*, *1*(3-4), 237-264.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13-F21.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.

Männel, C., & Friederici, A. D. (2010). Prosody is the key: ERP studies on word segmentation in 6-and 12-month-old children. In Cognitive Neuroscience Society (CNS) 2010 Annual Meeting (pp. 261-261).

May, L., & Werker, J. F. (2014). Can a click be a word?: Infants' learning of non□native words. *Infancy*, 19(3), 281-300.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5(3), 317-334.

Mills, D. L., Coffey□Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, 13(3), 397-445.

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, 20(1), 19-31.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 516-524. doi.org/10.3758/bf03195598

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, 23(7), 728-733.

Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18-and 24-month old children. *Brain and Language*, 125(1), 1-10.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28(3), 406.

Rivera-Gaxiola, M., Csibra, G., Johnson, M.H., & Karmiolff-Smith, A. (2000a). Electrophysiological correlates of cross- linguistic speech perception in Native English speakers. *Behavioral Brain Research*, **111**, 13–23.

Rivera-Gaxiola, M., Csibra, G., Johnson, M.H., & Karmiolff-Smith, A. (2000b). Electrophysiological correlates of category goodness. *Behavioral Brain Research*, **112**, 1–11.

Sirri, L., & Rämä, P. (2015). Cognitive and neural mechanisms underlying semantic priming during language acquisition. *Journal of Neurolinguistics*, *35*, 1-12.

Sokka, L., Leinikka, M., Korpela, J., Henelius, A., Ahonen, L., Alain, C., ... & Huotilainen, M. (2016). Job burnout is associated with dysfunctions in brain mechanisms of voluntary and involuntary attention. *Biological Psychology*, 117, 56-66.

Steber, S., & Rossi, S. (2020). So young, yet so mature? Electrophysiological and vascular correlates of phonotactic processing in 18-month-olds. *Developmental Cognitive Neuroscience*, 43, 100784.

Thierry, G. (2005). The use of event□related potentials in the study of early cognitive development. *Infant and Child Development*, 14(1), 85-94.

Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport*, 14(18), 2307-2310.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2), 172-175.

Tincoff, R., & Jusczyk, P. W. (2012). Six□month□olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432-444.

Torkildsen, J. v. K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., ... & Lindgren, M. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, 19(6), 431-454.

Torkildsen, J. v. K, Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007a). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, 20(4), 332-351

Torkildsen, J. v. K, Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007b). Brain responses to lexical-semantic priming in children at-risk for dyslexia. *Brain and Language*, 102(3), 243-261.

Torkildsen, J. v. K, Hansen, H. F., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, 108(2), 73-88.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*(4), 703-728.

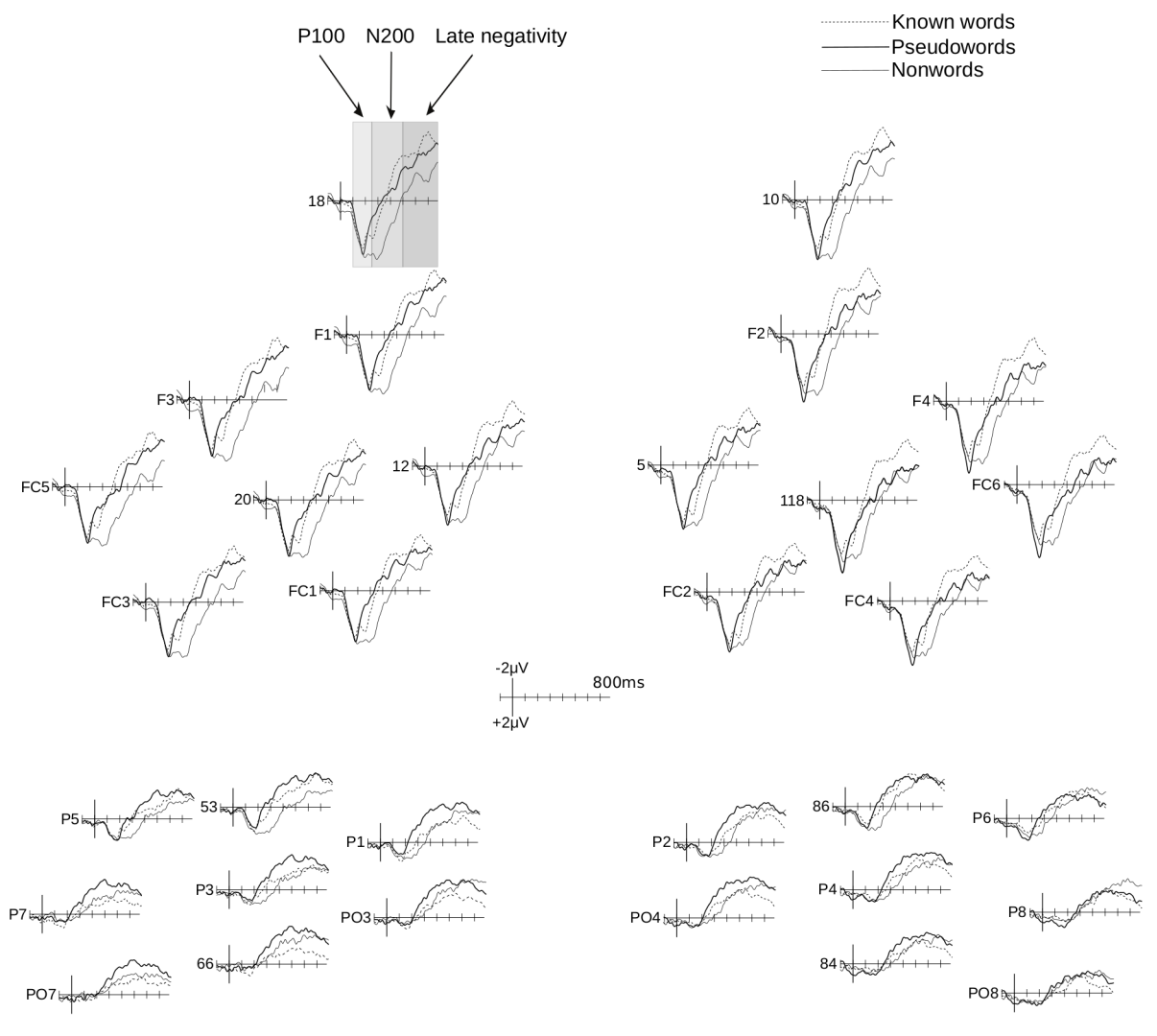
Figure Legends

Fig. 1. The grand-average ERPs to known words (dashed lines), pseudowords (thick lines), and nonwords (thin lines) over the frontal and parietal recording sites. Vertical lines indicate word onset. The light, medium, and dark grey boxes indicate the time windows for Positivity (100–250 ms), N200 (250–500 ms), and late negativity (500–800 ms) amplitude analyses, respectively. The x-axis of the scale is from -100 ms to 800 ms with respect to word onset.

Fig. 2. Isopotential maps showing the across-participants calculated scalp distributions of the difference between the ERPs to (known words – pseudowords), (known words – nonwords), (pseudowords – nonwords), separately for the N200 (250-500 ms after word onset) and the Late negativity (500-800 ms after word onset). The head viewed from above. L = left, R = right.

Fig. 3. A. Mean amplitudes for the N200 over the frontal recording sites for known words, pseudowords and nonwords in all participants. **B.** Mean amplitudes for the late negativity over

the frontal recording sites for known words, pseudowords and nonwords in all participants. **C.** Mean amplitudes for the N200 over the frontal recording sites for known words, pseudowords and nonwords in the normal-to-low (light grey bars) and normal-to-high (dark grey bars) word comprehenders. **D.** Mean amplitudes for the late negativity over the frontal recording sites for known words, pseudowords and nonwords in the normal-to-low (light grey bars) and normal-to-high (dark grey bars) word producers. Vertical bars indicate SEMs * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$


P100
N200
Late negativity
Known words
Pseudowords
Nonwords
18
10
F1
F2
F3
F4
FC5
20
12
5
118
FC6
FC3
FC1
FC2
FC4
-2μV
800ms
+2μV
P5
53
P1
P2
86
P6
P7
P3
PO3
PO4
P4
P8
66
84
PO7
PO8

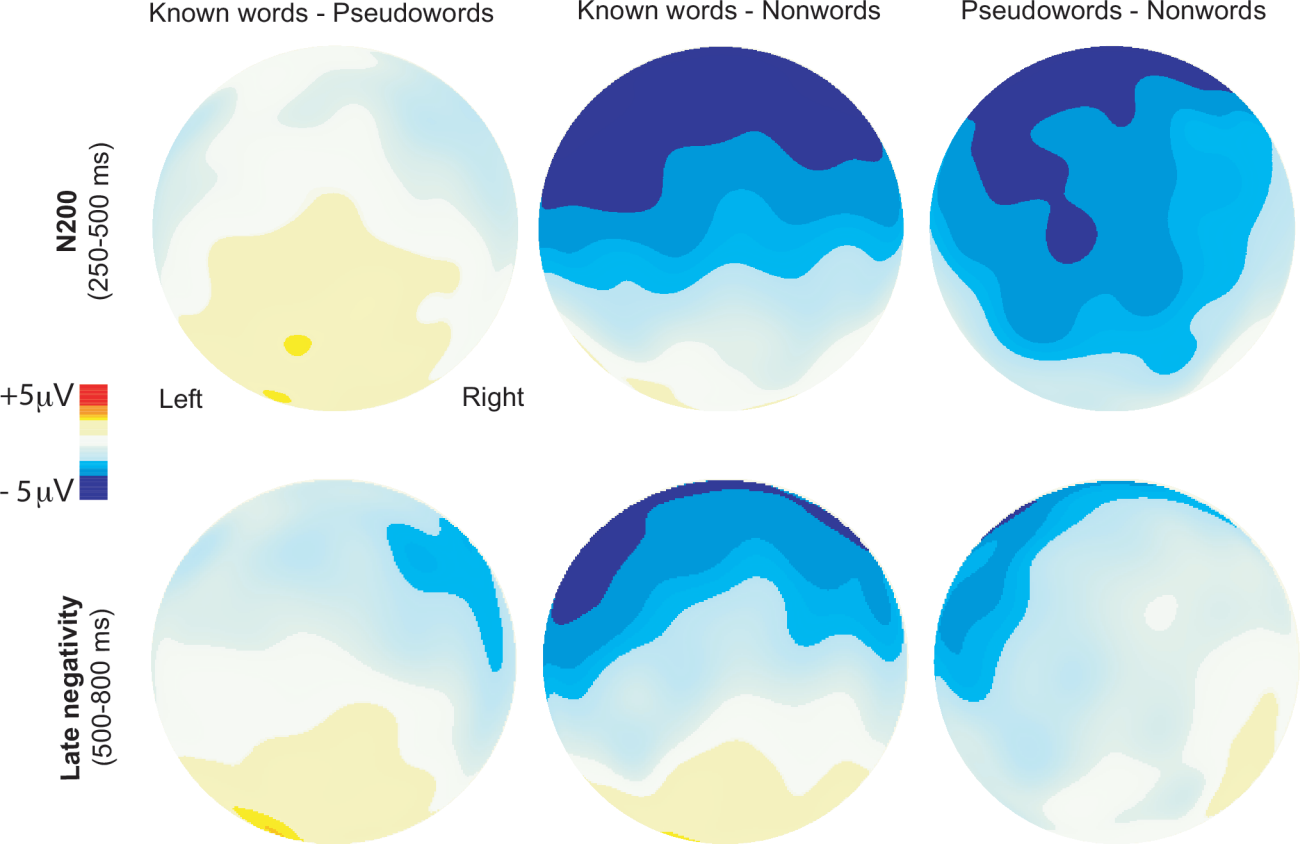

Known words - Pseudowords
Known words - Nonwords
Pseudowords - Nonwords
N200
(250-500 ms)
+5μV
- 5μV
Left
Right
Late negativity
(500-800 ms)

Table 1 : Average Biphoneme Frequency, Word Frequency and Duration of the Stimuli

	AvbiF	WordF	Duration (ms)
Known Words			
ballon /balɔ̃/	6469	27	506
maison /mɛzɔ̃/	6371	570	496
bateau /bato/	4209	107	585
cheveux /ʃəvø/	2907	116	657
means	**4989**	**205**	**561**
Native Pseudowords			
lechon /ləʃɔ̃/	786	0	487
vabeux /vabø/	3600	0	533
meautai /motɛ/	5449	0	523
bonsa /bɔ̃za/	1285	0	561
means	**2780**	**0**	**526**
Non-native Pseudowords			
vchonbe /vʃɔ̃bə/	397	0	628
chbeula /ʃbøla/	5208	0	627
bmeausai /bmozɛ/	1703	0	690
zbaiton /zbɛtɔ̃/	4059	0	637
means	**2842**	**0**	**646**
Total means			**578**
SD			69

Table 1: Biphoneme Frequencies of Stimuli…
last one from : **freqfilms**: la fréquence du mot selon le corpus de sous-titres (par million d'occurrences);
biphoneme frequencies from: