The contribution of early language exposure to the cortical tracking of speech

Jose Pérez-Navarro[1], Anastasia Klimovich-Gray[2], Mikel Lizarazu[1], Giorgio Piazza[1], Nicola Molinaro[1,3]*, & Marie Lallier[1]*

[1] Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastian, Spain

[2] University of Aberdeen, Aberdeen, United Kingdom

[3] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

* The last two authors made equal contributions to this manuscript

Abstract

Despite the relevance of the cortical tracking of speech for the development of speech perception skills, no study to date has explored whether and how it is shaped by language experience and exposure. In this study we expand the current theories of cortex-speech tracking by showing that the cortical tracking of speech in unbalanced bilingual children is modulated by language exposure. We recorded electroencephalography in 35 Basque-Spanish bilingual children (6 y.o.) with a markedly unbalanced bilingual profile (>70 % exposure to Basque; <30 % to Spanish) while they listened to continuous speech in each of their languages. We assessed their cortical tracking of speech at the acoustic temporal (speech envelope), lexical (lexical frequency), and semantic (sentence-level semantic distance) levels. Phase alignment between the speech envelope and the EEG activity (speech-brain coherence) in the delta frequency band (0.5 - 1.5 Hz) was significant and similar in both languages. Interestingly, mTRF modeling showed that the cortical encoding of the speech envelope in the language with the least exposure was the most robust. Moreover, mTRF modeling showed that children had a more sensitive cortical tracking of semantic information in their most experienced language. Lastly, only for the language with dominant exposure, the cortex-speech tracking of the envelope was linked to the phonological abilities, while their cortical tracking of lexico-semantic information predicted their vocabulary knowledge. Our findings inform the developmental theories by showing that the cortical tracking of acoustic temporal and lexico-semantic speech features depends on the accumulated experience within a language during the early years of language acquisition.

Introduction

Continuous exposure to spoken language in a wide variety of contexts makes its acquisition appear spontaneous and effortless. Despite this apparent ease, developmental evidence shows that complex brain systems and processes supporting language comprehension emerge between perinatal stages and around three years of age, and become language-selective between birth and the start of primary school (for reviews see Gervain, 2015; Kuhl, 2004; Skeide & Friederici, 2016). However, little is known about how these neurocognitive processes evolve and mature during childhood as a function of language exposure or experience, two essential contributors to language development (e.g., Carbajal & Peperkamp, 2020; Gámez et al., 2019; Gathercole & Thomas, 2005; Paradis et al., 2016; Pearson et al., 1997; Thordardottir, 2011).

Cortical tracking of speech and language acquisition

Accurate multidimensional models of language development require understanding how different levels of language exposure and proficiency modulate fundamental brain mechanisms underlying language acquisition, including speech comprehension. We focus on the cortical tracking of speech (CTS), a neurocognitive process that is critical for understanding continuous speech (Giraud & Poeppel, 2012). CTS is typically referred to as the dynamic alignment of brain oscillatory activity to the temporal modulations of the speech envelope (Obleser & Kayser, 2019). In this study we refer to such speech envelope tracking as *acoustic-temporal* CTS, which takes place within delta (aligning with prosodic phrasing) and theta (syllable timing) frequency bands and has been shown to support speech comprehension in adults (e.g., Ding & Simon, 2012; Gross et al., 2013; Luo & Poeppel, 2007; Molinaro & Lizarazu, 2018; Peelle et al., 2013). In addition, a growing number of studies are investigating CTS beyond the speech envelope at higher order lexical and semantic levels (Brodbeck et al., 2022; Broderick et al., 2018, 2021). Here, we term this level of linguistic tracking as *lexico-semantic* CTS. We hypothesize that early

language experience shapes the maturation of CTS at both acoustic-temporal and lexico-semantic levels. In fact, behavioral evidence shows that language experience supports increasingly efficient processing and comprehension skills for continuous speech. For example, in monolingual environments, the amount of child-directed input is positively linked to word processing speed, the ability to encode lexical information from continuous speech input (Hurtado et al., 2008; Weisleder & Fernald, 2013). In addition, in bilingual contexts, the proportional exposure to both languages correlates with language-specific word processing speed (Hurtado et al., 2014).

Here, we hypothesize that language experience contributes to CTS at two different levels. First, accumulated input in a language should be crucial for building up knowledge about the temporal statistics embedded in the speech envelope (hereafter acoustic-temporal tracking), which contributes to the efficient dynamic alignment of cortical oscillatory activity to relevant phonological units in the speech signal (Giraud & Poeppel, 2012; Luo & Poeppel, 2007). Accordingly, greater experience within a given language should be associated with more precise acoustic-temporal CTS (Lizarazu et al., 2021). Second, richer language experience should support the development of a neurocognitive model guiding the extraction of higher order linguistic information via lexico-semantic CTS (Broderick et al., 2021; ten Oever & Martin, 2021). Therefore, linguistic experience should be positively associated with the cortical tracking of lexico-semantic information.

CTS and phonological development

Language experience supports the development of language knowledge across linguistic domains (see also Gathercole & Thomas, 2005; Oller & Eilers, 2002; Paradis & Jia, 2016; Pérez-Navarro & Lallier, 2023). In particular, accumulated language experience enhances phonological acquisition (Nittrouer, 1996; Nittrouer & Burton, 2005; for a review, see Nittrouer, 2002) which,

in turn, supports the comprehension of continuous speech. For example, experience with a specific language benefits the performance of bilinguals when repeating nonwords matching the phonological features of that language (Messer et al., 2010; Parra et al., 2011). Relatedly, research in children with hearing difficulties shows that both a reduced quality and quantity of the speech input delays the development of both phonological and broader speech comprehension skills (Briscoe et al., 2001; Nittrouer & Burton, 2001, 2005). Of particular relevance for the purpose of our study, Goswami (2011, 2017) proposes that the cortical tracking of speech amplitude modulations within the speech envelope supports phonological development during infancy and childhood. Within this framework, the accurate processing of the speech envelope is thought to guide and shape the emergence of phonological representations (Goswami & Leong, 2013; Leong & Goswami, 2014). Accordingly, a growing number of studies have now shown that acoustic-temporal CTS is at play as early as 4 months of age and developing throughout childhood (Attaheri et al., 2022; Menn, Ward, et al., 2022; Ríos-López et al., 2020).

Studies of atypical language development also suggest that CTS mediates the contribution of language experience to the acquisition of adequate linguistic abilities. When compared to their typically developing peers, children who are at risk of developing dyslexia receive less hyperarticulated speech from their mothers (Kalashnikova et al., 2020), which could have an impact when accessing phonological information from the speech signal. In addition, several electrophysiological studies have shown that phonological deficits in children with dyslexia are tightly linked to atypical CTS within the delta and theta bands (Destoky et al., 2020; Di Liberto et al., 2018; Granados Barbero et al., 2022; Molinaro et al., 2016; Power et al., 2016) in comparison to both chronological-age-matched and reading-age-matched younger peers who are also matched on how much they have been exposed to written inputs (Di Liberto et al., 2018; Power et al., 2016). This suggests that impaired CTS could be causally implicated in atypical phonological development, and be somehow independent from linguistic experience (specially

for written language). However, Destoky et al. (2020) reported poorer CTS in children with dyslexia only when compared to chronological and not to reading age-matched controls. Given these somewhat divergent findings, the extent to which accumulated language experience contributes to the maturation of CTS during phonological development remains unclear. Our study tackles this question by exploiting the largely different experiences that unbalanced bilingual children have within their two languages. Importantly, this group provides a unique opportunity to quantify the effect of accumulated language experience on acoustic envelope-level CTS and phonological development within the same participants. To gain a more comprehensive understanding of the relevance of CTS for the development of language skills, we go a step forward by exploring whether CTS is also linked to lexico-semantic knowledge through exposure and experience with a particular language.

CTS and lexico-semantic development

Both lexico-semantic and syntactic knowledge, whose development is tightly related to accumulated language experience (Carbajal & Peperkamp, 2020; Oller et al., 2007; Paradis & Genesee, 1996; Pearson et al., 1997), have been linked to the efficiency of cortical oscillatory mechanisms for speech processing. In monolingual children, Panda et al. (2020) showed that lexico-semantic (vocabulary) knowledge was linked to the synchronization between language-related brain areas through cortical oscillatory activity during continuous speech listening, although they did not assess CTS. With respect to CTS, it has been shown that in adults it is modulated by both context-driven word predictability (Broderick et al., 2021; Klimovich-Gray et al., 2021; Koskinen et al., 2020; Molinaro et al., 2021) and the syntactic structuring of speech (Kaufeld et al., 2020; Meyer et al., 2017; Meyer & Gumbert, 2018). Of particular relevance for our hypotheses, these studies converge on the finding that the predictability of lexico-semantic and syntactic information shapes CTS in adults, possibly as they have developed efficient

neurocognitive language models as a function of accumulated language experience. However, the contribution of language experience to the development of lexico-semantic CTS during childhood is an open question that, to our knowledge, no study has directly addressed previously. By addressing such a question, we will be able to add evidence to the body of neurocognitive models for tracking linguistic information from a developmental perspective.

The current study

Overall, the evidence reviewed above points to the relevance of acquired language knowledge through accumulated linguistic experience for the tuning of CTS at both the acoustic-temporal level (linked to phonology) and higher order linguistic levels (associated with lexico-semantics and syntax). Here, we study the relationship between input and CTS at both levels by capitalizing on unbalanced bilingual children, whose accumulated experience within their two languages varies considerably. By following a within-participant approach, we can limit interindividual variability, and boost the robustness of CTS contrasts. Therefore, we are able to explore within a single group of children the role of greatly different language experiences in each of their two languages on the cortical tracking of acoustic-temporal speech information (i.e., the envelope), and on the much less investigated mechanism of cortical tracking of lexico-semantic information in continuous speech. In addition, we aim to assess whether the cortical tracking of these two types of speech information is related to phonological and lexico-semantic skills measured behaviorally.

Results

Language measures

In line with their significantly higher language experience (i.e., percentage of exposure) within Exp(+) than within Exp(-), $t(63) = 35.13$, $p < .001$ ($\beta = 72.298$, $SE = 2.058$), participants

showed significantly higher vocabulary knowledge in Basque than in Spanish, $t(32) = 13.18$, $p < .001$ ($\beta = 0.077$, $SE = .006$; see Figure 2). However, there was no between-language difference in phonological abilities, $t(30.1) = -0.84$, $p > .05$ ($\beta = -0.014$, $SE = .017$), nor in the comprehension of the stories that participants listened to during the EEG session, $t(28) = 0.12$, $p > .05$, ($\beta = 0.005$, $SE = .039$). Indeed, Bayesian t-tests yielded moderate evidence for between-language similarities in phonological abilities, $BF_{10} = .285$, *error* = .034 %, *median difference* = -0.151, *CI* [-0.492 .184], and story comprehension respectively, $BF_{10} = .22$, *error* = .024 %, *median difference* = -0.04, *CI* [-0.417 .334].

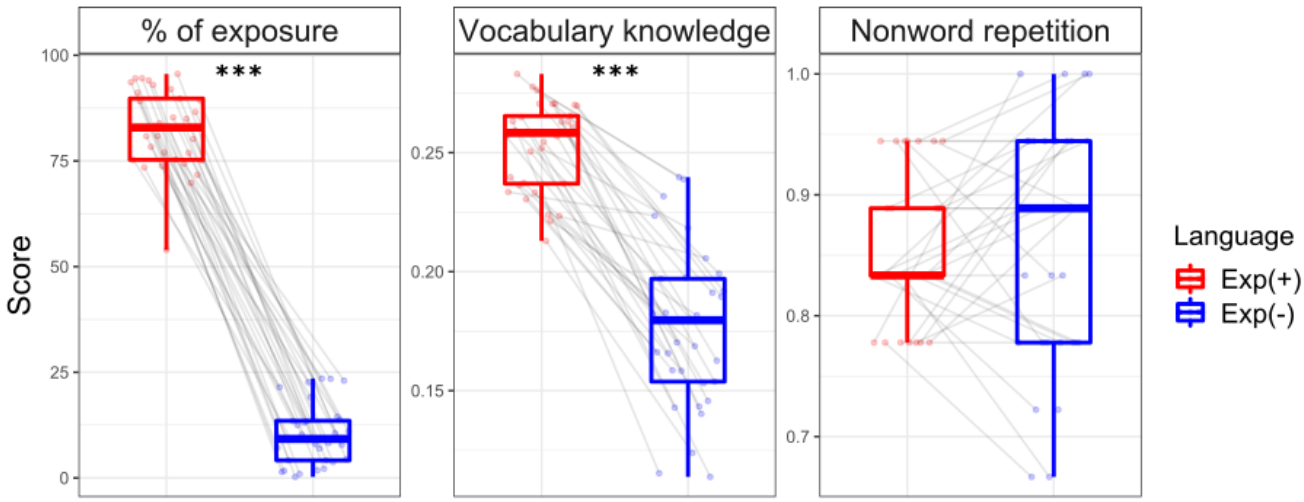


Figure 2. Amount of linguistic exposure and performance on vocabulary (picture naming) and phonology (nonword repetition) tasks in each language. Points represent each individual score in the different measures and languages. Boxplots represent group estimates, with horizontal lines within each box marking the median score. Upper and lower hinges mark the first and third quartile, and whiskers show 1.5 * interquartile range. Lines connect the scores of each participant across both languages. Asterisks indicate a significant difference between languages (***p < 0.001).

Acoustic-temporal CTS

Coherence

Within the delta frequency band, speech-brain coherence was significant between 0.5 and 1.5 Hz in Exp(+), *cluster statistic* = 271.73, $p < .001$, *SD* = .001, and Exp(-), *cluster statistic* = 278.17, $p = .001$, *SD* = .001 (Figure 3, A). However, there were no between-language difference. Coherence in this delta range had considerably overlapping topographies in both languages (Figure 3, B).

In the theta range (4 – 7 Hz), we did not find significant speech-brain coherence in any of the languages (all cluster-corrected *p-values* > .05). We also did not observe significant coherence in the specific 5.5 Hz bin that aligned closely to the syllable rate of both languages ($p > .05$).

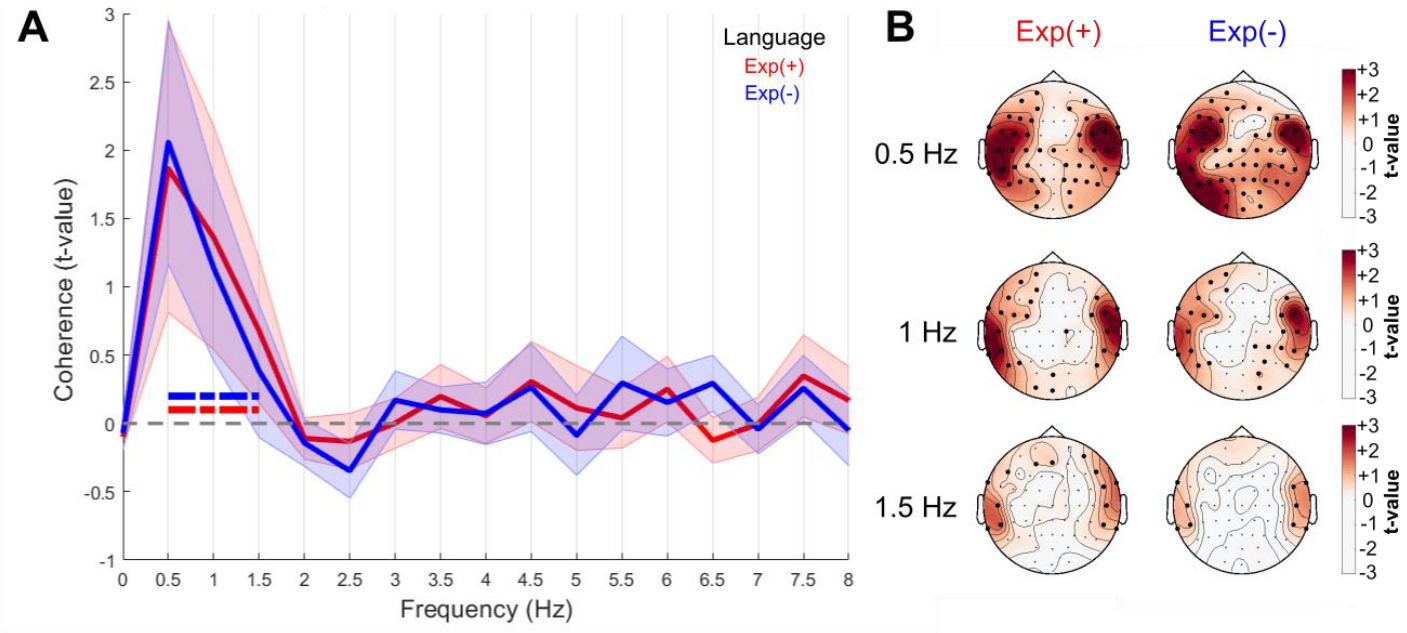


Figure 3. A) Speech-brain coherence across frequency bands for Exp(+) (red), Exp(-) (blue) against their random surrogate (flipped version). The discontinuous horizontal lines on the bottom mark the frequency range in which there was significant coherence for Exp(+) (red), and Exp(-) (blue). Red and blue shaded areas represent the standard error. B) Topography of speech-brain coherence in the 3 significant frequencies in Exp(+) (left) and Exp(-) (right). The colormap marks the size of the difference in coherence (normalized) between genuine speech and its flipped version (i.e., red, higher coherence; white, lower). Bigger black circles within each topographic map signal electrodes with significant speech-brain coherence from CBPT.

mTRF

Our control tests, one-sample t-tests (one-tailed) against no different prediction correlation from 0, showed that our three regressors of interest contributed above chance level to the mTRF model (speech envelope: $t(65) = 12.78$, $p < .001$, $d = 1.57$; lexical frequency: $t(65) = 6.50$, $p < .001$, $d = .8$; semantic distance: $t(65) = 5.27$, $p < .001$, $d = .65$; Supplemental Figure 3), which enabled us to conduct between-languages comparisons of their fit.

Regarding acoustic-temporal CTS, CBPT yielded significantly higher prediction correlation coefficients (Pearson's r) between envelope and EEG mTRF in the less exposed language (Exp(-)) than in the more exposed one (Exp(+)), *cluster statistic* = -27.65, $p = .01$, $SD = .004$ (Figure 4, B). Regarding the temporal weights in response to the speech envelope, CBPT did not show significant differences between languages (*cluster-corrected p-values* > .05; Figure 4, A). This suggests differences in how strongly children rely on envelope tracking between their two languages (i.e., stronger reliance on envelope tracking in their less experienced language), but no between-language differences in the temporal characteristics of the response to the envelope.

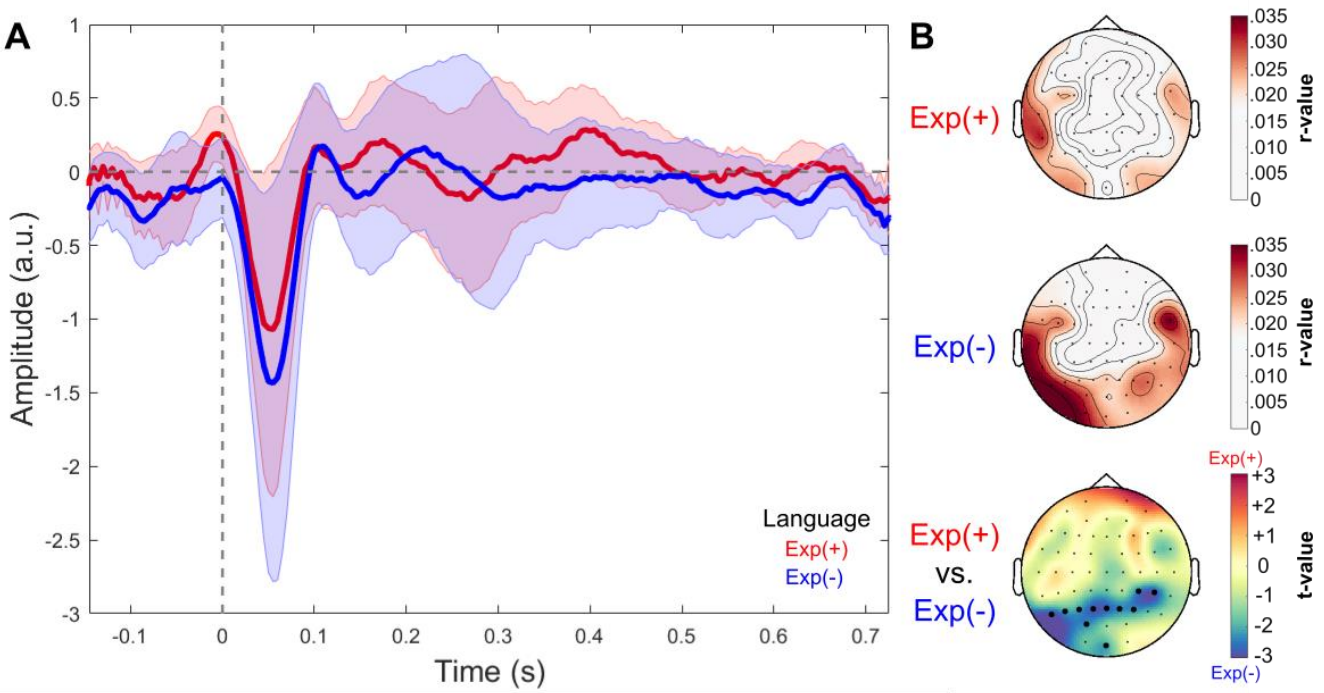


Figure 4. A) Temporal weights of the EEG signal TRF to the speech envelope in Exp(+) (red) and Exp(-) (blue). Red and blue shaded areas represent the standard error. B) Topography of the prediction correlation coefficients (r-values) between speech envelope and EEG TRF for Exp(+) (top), Exp(-) (middle), and their difference (bottom). The colormap for the two top topographies marks the degree of correlation (r-values); while the bottom topography is color mapped according to the size of the difference (t-value) between Exp(+) (relatively higher correlation coefficients in red) and Exp(-) (relatively higher correlation coefficients in blue). Bigger black circles signal electrodes in which CBPT yielded a significantly bigger prediction correlation at envelope level in Exp(-).

Lexico-semantic CTS: lexical frequency and semantic distance

CBPT did not yield significant between-language differences in the prediction correlation coefficients (model-fit r values derived from mTRF) between lexical frequency and EEG activity (*cluster-corrected p-values* > .05). The temporal weights in response to lexical frequency showed a similar latency to lexical-level TRFs in adults (e.g., Broderick et al., 2021) and did not show significant between-language differences (*cluster-corrected p-values* > .05; Figure 5, A).

Regarding EEG responses to semantic distance, we found relevant differences between languages. While mTRF prediction correlation coefficients (r-values) for semantic distance did not differ between languages (*cluster-corrected p-values* > .05); CBPT yielded a significant difference in the temporal response (i.e., weights) to semantic distance between the Exp(+) and Exp(-) languages (*cluster statistic* = 2827.4, p = .009, *SD* = .003). Thus, within this group of participants, EEG signal showed significantly higher sensitivity to semantic information in Exp(+) than in Exp(-) between 70 and 230 milliseconds after word onset (Figure 5, B). This suggests earlier tracking of the relevant contextual-semantic information for the language to which bilingual children were exposed the most.

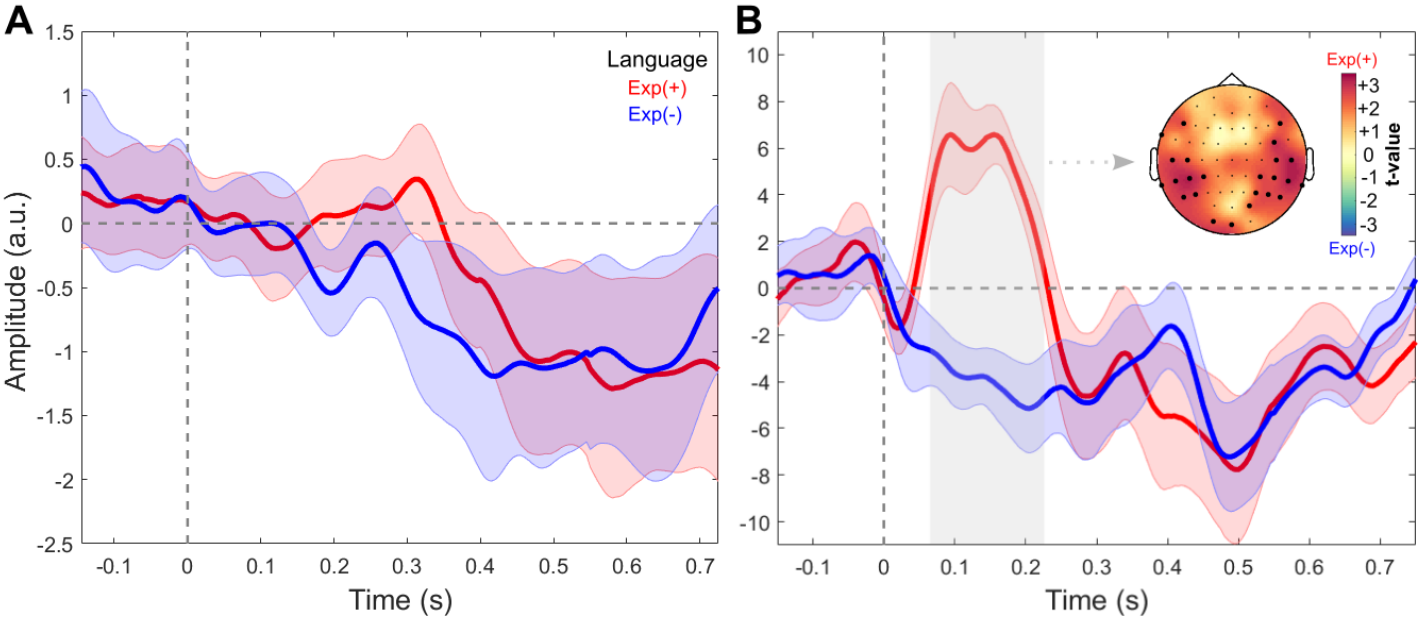


Figure 5. A) Temporal weights of the EEG signal TRF to lexical frequency in Exp(+) (red) and Exp(-) (blue). Red and blue shaded areas represent the standard deviation of the temporal weights. B) Temporal weights of the EEG signal TRF to semantic distance in Exp(+) (red) and Exp(-) (blue). Red and blue shaded areas represent the standard error of the temporal weights. The gray rectangle marks the latency of significant between-languages differences. The distribution over the scalp of such a significant difference in semantic distance (Exp(+) > Exp(-)) in this time window is represented by bigger dots in the top right topographical map.

Relationship between CTS and behavioral language measures

Acoustic-temporal CTS and lexico-semantic CTS composite scores were computed by aggregating EEG metrics for a more robust estimate of CTS-behavior relationships, while limiting the possibility of inflating type I error due to multiple tests. More specifically, an acoustic-temporal CTS composite was composed by averaging the normalized values of: 1) speech-brain coherence at delta (0.5-1.5 Hz), 2) speech-brain coherence at theta (4-7 Hz), and 3) prediction correlation values for envelope TRF. The lexico-semantic CTS composite consisted of: 1) prediction correlation values for lexical frequency TRF, and 2) prediction correlation values for semantic distance TRF. These composite scores were validated by the strong positive correlations found between (i) coherence to the speech envelope at delta and theta, and the

envelope mTRF (*average Pearson's r coefficient* = .59; Exp(+): .6; Exp(-): .57; all *p-values* < .01) and (ii) lexical frequency and semantic distance mTRFs (r = .79; Exp(+): .81; Exp(-): .78; all *p-values* < .01).

Regarding the relationship between acoustic-temporal CTS and behavioral language measures, there was a significant interaction between phonological abilities and language as predictors of acoustic-temporal CTS [$F(2, 46.52) = 4.69$, $p < .014$]. This interaction was driven by the significant and positive relationship between acoustic-temporal CTS and phonological abilities (nonword repetition) in Exp(+) [$t(45.69) = 3.05$, $p = .004$ ($\beta = .62$, $SE = .21$)], but not in Exp(-) [$t(48.93) = .62$, $p > .05$ ($\beta = .09$, $SE = .14$)] (Figure 6). There was no significant relationship between envelope-CTS and vocabulary knowledge in either of the two languages (all *p-values* > .05).

In parallel to the relationship between phonological abilities and envelope-level CTS, we found that vocabulary knowledge was positively related to lexico-semantic CTS. In this case the interaction between vocabulary knowledge and language was only marginally significant [$F(2, 57) = 2.81$, $p = .068$]. Nonetheless, this interaction was also driven by an underlying significant and positive relationship between vocabulary knowledge and lexico-semantic CTS only in Exp(+), $t(57) = 2.37$, $p = .021$ ($\beta = .41$, $SE = .17$), and not in Exp(-), $t(57) = -0.006$, $p > .05$ ($\beta = -0.001$, $SE = .14$) (Figure 6). There was no significant relationship between lexico-semantic CTS and phonological abilities in either of the two languages (all *p-values* > .05).

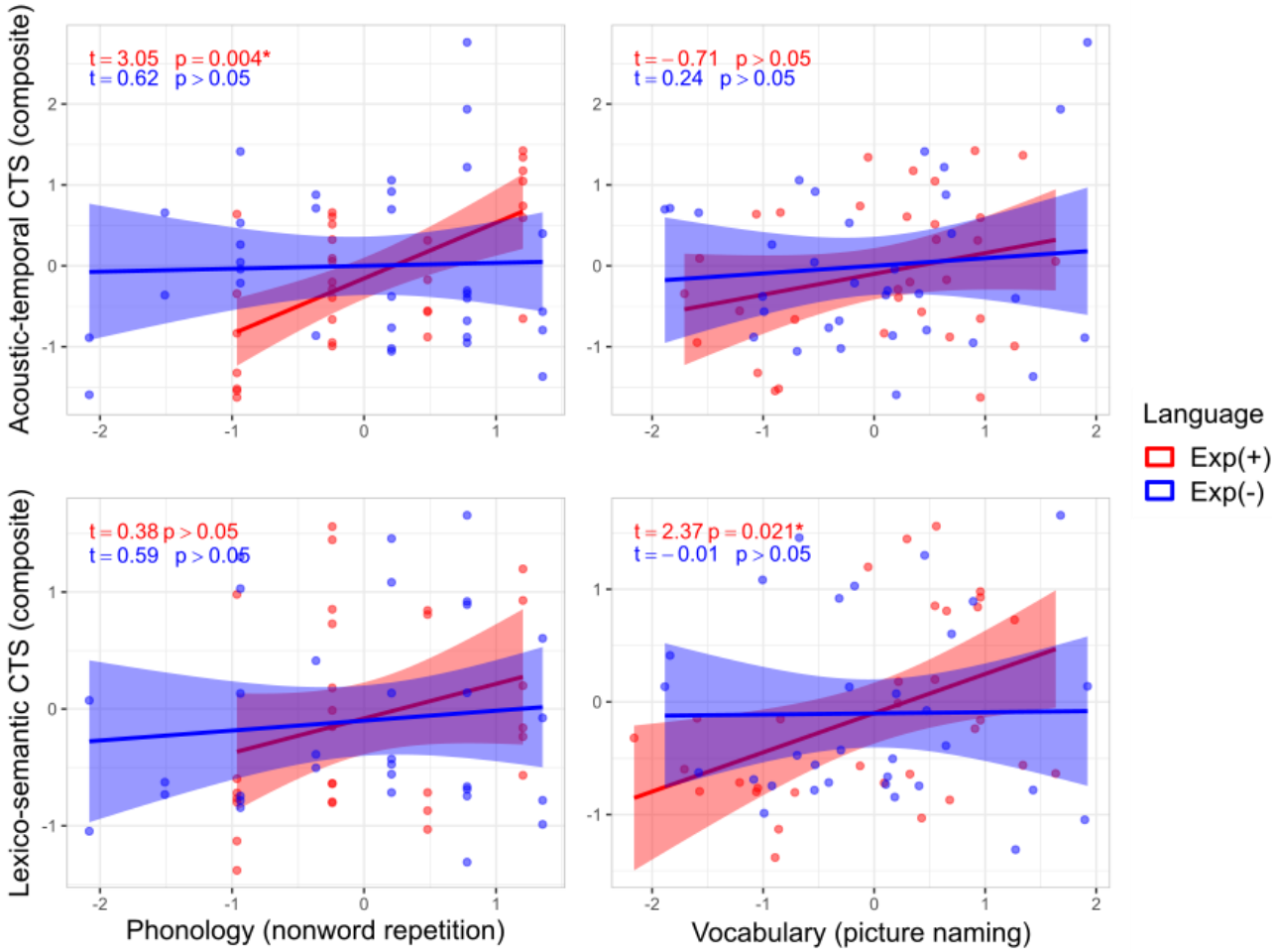


Figure 6. Relationship between CTS (acoustic-temporal and lexico-semantic) and language knowledge (phonology and vocabulary). Lines represent the slopes of linear regressions between CTS and behavioral metrics in each language respectively, with shaded areas around the lines marking 95% confidence intervals. Estimates in top left corners are the t- and p-values of linear mixed effect models for each relationship; asterisks indicate a significant effect.

Discussion

The main goal of the present study was to estimate the impact of linguistic exposure on the cortical tracking of speech, by assessing the CTS of 6-year-old unbalanced bilingual children in both their languages. We found relevant differences between the more experienced language (Basque) and the less experienced language (Spanish) in the cortical tracking of acoustic-temporal (speech envelope) and lexico-semantic speech features. Such differences can be

understood from a developmental perspective in which linguistic input plays a relevant role in the maturation of the cortical tracking of different types of linguistic information encoded in the speech signal. Additionally, and given that CTS was linked to our measures of language knowledge (in vocabulary and phonology), our results inform the relevance of CTS for indexing the development of specific language abilities. Moreover, the present study is, to our knowledge, the first to assess the cortical tracking of two different languages (and at different levels of acoustic-linguistic complexity) within the same group of bilingual children. The implications of these findings for understanding speech processing and language development are discussed below.

Linguistic experience and cortical tracking of speech in language development

Regarding the relevance of the cortical tracking of the speech envelope for efficient speech processing, our results contribute to the growing evidence that this neural mechanism plays an important role from early in life (see Menn, Michel, et al., 2022 for evidence in infants, and Ríos-López et al., 2020 for evidence in early childhood). We report a robust cortical tracking of the speech envelope in both of the languages of bilinguals through our two different metrics, namely speech-brain coherence and envelope-level mTRF models.

Regarding speech-brain coherence, significant CTS was found within the 0.5 to 1.5 Hz delta range, the timescale of prosodic stress occurrences, whose tracking has been shown to be relevant for efficient lexical segmentation (Kooijman et al., 2009). In line with previous research in children (Ríos-López et al., 2020), we did not find significant theta-band speech-brain coherence, which has been related to the tracking of syllabic speech units in continuous speech and might be more relevant at later stages of development, such as adulthood (Doelling et al., 2014; Peelle et al., 2013). It is possible that the developing language system of children is more sensitive to the slower delta frequency timescales for tracking continuous speech, and only starts

exploiting faster syllabic amplitude modulations once their phonological representations are more developed, possibly through sufficient experience and proficiency. Indeed, a recent cross-linguistic study by Peter et al. (2022) showed that acquired language-specific knowledge shaped the cortical tracking of syllabic information in adults. They demonstrated that adult native speakers of French (a syllable-timed language) showed enhanced cortical tracking of the syllable rate (~5 Hz) compared with English and Japanese listeners (stress- and mora-time languages respectively), for whom syllable-timing might be less relevant for extracting phonological information. However, in another recent study, Menn et al. (2022) did find significant speech-brain coherence at the syllable rate in infants as young as 9 months of age. Task requirement differences could explain these divergent findings between Menn et al. (2022) and our study at the theta frequency band. In Menn and collaborators' study, infants' CTS was measured while they were listening to their mothers in relatively short (40 second) live interactions (audiovisual stimuli), whereas in our study, CTS metrics were extracted from considerably longer 15-minute continuous stories with static images (very limited visual information). It is therefore possible that the length of the tasks and the use (or lack) of additional visual cues modulated CTS differently between studies, in line with the attested supportive effect of visual information for theta-band cortical tracking of speech (Lakatos et al., 2008; Power et al., 2013).

Regarding the spatial distribution of speech-brain coherence over the scalp, we observed a clear overlap between the two languages of these bilingual children, showing no differences in the location of significant electrode clusters, and the presence of a wide bilateral distribution over the scalp (Figure 3, B). Because estimating the cortical source of sensor-level EEG effects can only be done coarsely, we did not have a specific prediction beyond this bilateral temporal distribution that is typically shown in response to auditory stimulation. Nonetheless, a widely distributed speech-brain coherence topography was to an extent predicted by Sánchez-Alonso and Aslin (2022), who suggested that the use of naturalistic stimuli in developmental research

(the case in our study) would result in more widespread and less left-lateralized brain responses to speech than is usually shown in more constrained settings (see also Hamilton et al., 2018; Huth et al., 2016). Such bilateral distribution of speech-brain coherence is also in line with previous meta-analytic evidence that, while monolinguals and bilinguals who acquired their second language after 6 years of age show a clear left lateralization of language processing, early bilingual experience (as in our study) is associated with a more bilateral hemispheric involvement (Hull & Vaid, 2006, 2007). In the same vein, a recent study shows that, as proficiency in a second language increases, the recruitment of language processing brain regions more closely resembles that of native listeners (Zhang et al., 2023). It should be noted that the finding that the same delta frequencies drove significant speech-brain coherence in both languages, accompanied by greatly overlapping spatial distributions, could be related to the fact that Basque and Spanish have fairly similar phonologies at many levels, including prosody (Ezeizabarrena & Alegria, 2015; Hualde & Urbina, 2011), which could result in similar neurocognitive mechanisms at delta frequency for processing the prosody of each language.

The lack of significant between-language differences in speech-brain coherence analyses shows that the young children in our study robustly tracked the speech envelope in both their more and less experienced languages. Nonetheless, when it comes to envelope-level TRFs, we found that linguistic experience did shape children's cortical tracking of speech envelope information. Namely, the continuous EEG activity of this group of children tracked the speech envelope of their less experienced language (Spanish) more faithfully than that of their more experienced one (Basque). This was observed despite the fact that, at the time of testing, children had developed similar receptive language abilities in both their languages, illustrated by the lack of relevant between-language differences in phonological abilities and in story comprehension.

These findings help contextualize the role of cortical tracking of the speech envelope for language development. During language acquisition, envelope tracking is tightly linked to phonological development and broader language outcomes (Goswami et al., 2010; Kalashnikova et al., 2019; Vanvooren et al., 2017). Relatedly, a recent study showed that the envelope CTS of adult second-language learners increased with second language proficiency (Lizarazu et al., 2021). Therefore, the fact that the children encoded envelope-level information more faithfully in their less experienced language suggests they relied to a larger extent on low-level acoustic information to achieve a similar level of comprehension as in their more exposed language. We propose that envelope-level CTS could thus serve as a route for speech comprehension during language learning, possibly by compensating while increasing higher-order linguistic knowledge is being learned. In line with this interpretation, studies show that adults rely more on envelope-level CTS under noisy and effortful conditions than under normal conditions (Hauswald et al., 2022; Schmidt et al., 2023). Thus, it is likely that as individuals can exploit an increasingly proficient language model for CTS, they gradually shift from relying on low-level temporal information towards tracking higher order linguistic units and structures.

Beyond the speech envelope, we report initial evidence on the cortical tracking of abstract linguistic features from continuous speech in children. Such evidence provides a developmental perspective to the role of language knowledge in the cortical tracking of higher order linguistic information, which is being increasingly studied in adults (e.g., Broderick et al., 2021; Kaufeld et al., 2020; Reetzke et al., 2021). Specifically, we found that children showed an early increased temporal sensitivity to semantic-level information only in their more experienced language (Basque). In particular, the temporal response to semantic distance showed a peak between 70 and 230 ms in their more experienced language (Basque), but not in their less experienced one (Spanish). We did not expect the cortical response to the semantic context to show a peak at such early latency. Indeed, this temporal response was considerably earlier than that found in adults

in response to a regressor containing the same sentence-level semantic distance information in English, which ranged between 200 and 500 ms (Broderick et al., 2021). However, it falls within a similar latency range (~50-200 ms) to several adult studies that found early contextual (typically semantic) influence on the cortical tracking of continuous speech (Brodbeck et al., 2018, 2023; Broderick et al., 2019; Donhauser & Baillet, 2020).

A plausible interpretation of our finding is that when children have more established language knowledge (only in the more experienced language here), they can capitalize on the semantic context to predict and efficiently track the phonological and/or lexical information from upcoming speech. Two interrelated factors suggest that, in the case of our study, such difference between languages in the cortical exploitation of semantic information might be driven by lexical, but not phonological, facilitation. First, Basque and Spanish overlap to a great extent in their phonological systems, but not in their vocabularies (Ezeizabarrena & Alegria, 2015). Second, the children showed similarly developed phonological abilities in both languages, but significantly greater lexical knowledge in their more experienced language. Therefore, it is plausible that their increased cortical sensitivity to contextual semantic information in their more experienced language is related to a more efficient lexical (rather than phonological) processing of the upcoming word in continuous speech. This interpretation is consistent with the positive relationship found between the cortical tracking of lexico-semantic information and vocabulary knowledge.

Cortical tracking of speech as an index of language abilities

The relevance of acoustic-temporal and lexico-semantic CTS for explaining behavior was evaluated by assessing their link to individual differences in language knowledge within both phonology and vocabulary domains. There was a specific relationship between acoustic-temporal (envelope) CTS and phonological abilities, only in the more experienced language

(Basque). The cortical tracking of the speech envelope was positively linked to nonword repetition performance, but not to vocabulary knowledge in this language nor language abilities in the less experienced language. This finding is in line with Molinaro et al., (2016) and Di Liberto et al. (2018), who respectively reported that the cortical tracking of the envelope and phonetic features were related to phonological abilities in dyslexic children. Moreover, our finding is coherent with the hypothesis that sensitivity to prosodic and syllabic speech AMs (assessed directly here through speech-brain coherence) subserves the temporal sampling of phonological units within these timescales (phrases, words, and syllables) during phonological development (Goswami, 2011, 2017; Goswami & Leong, 2013; Leong & Goswami, 2014). Accordingly, bilingual participants who exhibited higher CTS to such speech AMs were those who showed more developed phonological processing abilities. Thus, our results support that phonological abilities and neurocognitive abilities for tracking speech AMs develop hand in hand (Boets et al., 2007; Di Liberto et al., 2018; Lundberg et al., 1988; Power et al., 2016; Vanvooren et al., 2017). In the same vein, Ríos-López et al. (2021) found that, in pre-readers, delta envelope CTS predicted later reading outcomes.

It is relevant to note that, while tracking speech AMs contributes to the development of phonological skills (Leong & Goswami, 2014, 2017; Molinaro et al., 2016), studies in adults highlight that CTS might be more reliant on speech rate than on speech envelope modulations, at least in normal and effortless listening environments (Hauswald et al., 2022; Schmidt et al., 2023). Therefore, the role of acoustic-temporal CTS for language development might decrease as soon as individuals can rely on sufficient lexico-semantic knowledge to track continuous speech (Kaufeld et al., 2020; Molinaro et al., 2021). A complementary possibility is that, along with language development, delta oscillatory activity shifts gradually from acoustic-temporal CTS (at envelope-level) towards semantic and syntactic CTS. This possibility aligns with a suggested developmental shift when exploiting prosodic information, namely from recognizing

word boundaries early in life towards extracting semantic and syntactic information later in childhood (Friederici & Männel, 2014; Guasti, 2002; Männel et al., 2013). Especially relevant to our findings is the study by Kaufeld et al. (2020), which aimed to disentangle the role of prosody from the role of linguistic structure and meaning for delta-band CTS in adults. They found that linguistic content modulated delta neural oscillatory activity beyond stimulus-driven prosodic timing. It is therefore possible that, during language development, delta CTS helps map the acoustically driven prosodic timing during language learning, while concurrently allowing the accumulation of knowledge about the linguistic units and structures present at similar acoustic timescales, such as syntactic information. Interestingly, Menn, Ward, et al. (2022) reported that 10-month-old infants could track stressed syllables within the delta frequency band and that these skills predicted vocabulary knowledge at 24 months. It is therefore possible that, in our study, children relied more on extracting envelope-level phonological information in their less familiar language, hence the significantly stronger prediction correlation of envelope-level TRFs in that language; and utilized the same delta tracking to extract increasingly more abstract linguistic information in their more exposed language. Future longitudinal studies assessing the different reliance on acoustic-temporal or linguistic features for tracking the speech signal as a function of age and language knowledge, will help clarify whether and when this potential developmental shift takes place.

Parallelly to the acoustic-temporal CTS-phonology relationship, we also observed a significant positive relationship between lexical abilities (vocabulary) and the cortical tracking of lexico-semantic features conveyed by the speech signal, only in the most proficient language (Basque). These findings are in line with Hurtado et al. (2014)'s behavioral results showing that, in infants growing up in bilingual environments, the relative amount of exposure to their languages was positively correlated with speech processing efficiency and vocabulary knowledge in each language. Our results add to this finding by showing that vocabulary

knowledge is related to the cortical temporal encoding of lexico-semantic information in continuous speech, but only when a given language is acquired early and with sufficient exposure (Basque in our study). Indeed, these results provide initial evidence for the behavioral relevance of the cortical tracking of lexico-semantic information during language development. Interestingly, CTS was linked to both phonological and vocabulary performance only in the more experienced language, suggesting that the more stable language representations and processes in this language might map more consistently and faithfully onto individual differences in CTS. In contrast, weaker knowledge in the less experienced language might not map as readily onto the neural resources involved in CTS. Accordingly, the language knowledge of children was overall more variable in their less experienced language (Figure 2). There is also the possibility that, with a larger sample, we could have detected significant behavioral-CTS relationships in the less exposed language, by accounting for the higher intra- and inter-individual variability.

Overall, the present study informs theoretical, empirical, and computational accounts of the relevant role that linguistic experience plays in continuously shaping cortical oscillatory activity mechanisms that extract linguistic information from speech. Given our and previous findings, we propose that acoustic-temporal (envelope) CTS serves as phonological route for speech processing while a language is being learnt. Later on, and as a function of maturational factors such as age and linguistic experience, expert neurocognitive language models start exploiting lexical, semantic, and syntactic information to organize linguistic meaning (Martin, 2020; Martin & Doumas, 2017), group words within or between phrases (Meyer et al., 2017), and predict upcoming linguistic information (Molinaro et al., 2021; ten Oever & Martin, 2021).

In summary, our findings shed new light on the role that linguistic experience plays in the cortical tracking of speech, a neurocognitive mechanism supporting speech comprehension. For the first time, we show that, during language development, the maturation of the cortical tracking

of the temporal speech envelope and lexico-semantic speech features is tightly linked to the amount of exposure to a language and the subsequent knowledge attained in that language. We also highlight that individual differences in specific language abilities are linked to the cortical tracking of related linguistic features conveyed by continuous speech.

Method

Participants

Participants were 35 (18 females) Basque-Spanish bilingual children between 6 and 7 years of age (*Mean age* = 6.92; *SD* = .11). We selected them from a broader longitudinal study (Pérez-Navarro & Lallier, under review) based on their amount of exposure to each language when they were 6.4 years old (the last stage of that longitudinal study). Namely, they were the group of children who had the most experience within their most exposed language (> 70% of their waking hours exposed to Basque, referred to as *Exp(+)* henceforth) and the least within their least exposed language (< 30% of their waking hours exposed to Spanish, *Exp(-)* henceforth). This selection criterion enabled us to investigate the influence of differential linguistic experience on CTS of two different languages within the same participants, thus limiting inter-subject variability for our language comparisons. Importantly, since participants had started formal reading instruction only a few months prior to participation in the study, the amount of written input was not sufficient to be considered a main source of language input at the time of testing. Thus, the potential influence of written language exposure on phonological abilities (Dehaene et al., 2015; Goswami & Bryant, 1990; Morais et al., 1979) and CTS in our study was considered to be minimal.

Participants had normal hearing, no history of neurological disorders, nor familial risk of developmental language disorder or any other cognitive-related genetic pathology. The study was approved by the BCBL Ethics Committee and complied with the Declaration of Helsinki.

Behavioral session

The relative amount of linguistic experience within each language was assessed through a comprehensive linguistic questionnaire completed by the children's parents. From this questionnaire, we extracted a composite index of the proportional amount of experience within each language (Supplemental Formulas 1 to 3). In addition to the amount of language experience, we assessed children's phonological and lexico-semantic language knowledge.

Phonological abilities were evaluated with a nonword repetition task consisting of 18 items for each language, ranging from 3- to 6-syllable nonwords (6 nonwords per syllable length and language). In this task, participants were randomly presented with the 36 auditory nonwords, each constructed to follow either the stress and syllabic phonotactic constraints of Basque or those of Spanish. Upon participant's oral repetition of each nonword, an experimenter coded each repetition as either correct or incorrect. Mean accuracy for each language was taken as a measure of language-specific phonological abilities.

Lexico-semantic knowledge was assessed through a picture-naming task consisting of 45 items for each language. The items were different in each language version, although matched in terms of word frequency. In each language version of the task, participants were asked to name the pictures that appeared on screen in a random order, without any time constraints. The experimenter coded each trial as either correct or incorrect upon participant's response. The order of presentation of the two languages was counterbalanced across participants. In order to harmonize participants' performance in the picture-naming task across both languages, we weighted performance on the different items (1, correct; 0, incorrect) by the inverse of their Zipf

lexical frequency (van Heuven et al., 2014). For Basque words, lexical frequency was extracted from EHME database (Acha et al., 2014); and for Spanish words from EsPal database (Duchon et al., 2013). This way, more frequent (and presumably easier) words had a smaller weight on participants' overall performance than less frequent ones that were less likely to be known by children. Thus, our measure of performance in the picture-naming task was Zipf-frequency-weighted accuracy.

Importantly, our phonological and lexico-semantic measures have previously proven sensitive to differential amounts of language exposure in bilinguals (Gámez et al., 2019; Paradis et al., 2016; Thordardottir et al., 2006).

Electroencephalography (EEG) session

EEG task: speech listening

Participants listened to continuous streams of natural speech in the form of storytelling. We used two stories that were adaptations of two short books targeted at 6-year-old children, and followed a very similar narrative structure. Importantly, this literary style targeted at children offers a great number of different words and, consequently, considerable variability in their lexical frequencies (Nation et al., 2022; see Supplemental Figure 1), which was relevant to explore the effect of language experience and knowledge on the cortical tracking of lexico-semantic information.

The stories were about the history of outer space exploration and the evolution of life on Earth. For each story, both a Exp(+) (Basque) and a Exp(-) (Spanish) version was narrated by the same female native speaker of both languages in a child-directed speech register. Each story lasted about 15 minutes, because this duration has proven sufficient to robustly estimate cortical tracking of speech with EEG (Destoky et al., 2019). Half of the participants listened to the 'outer

space' story in Basque, and the 'life on Earth' story in Spanish, and vice versa for the other half of participants. The order of language presentation was counterbalanced across participants. Participants were asked to listen attentively to the stories that were presented to them over speakers. They were sitting in a comfortable upright position and asked to look at static images depicting the story narrative, presented on the center of a computer screen positioned at 80 cm from their eyes. Every 5 minutes approximately, participants were asked three simple yes/no questions (9 per story in total) to check whether they were paying attention to, and comprehending, the stories.

EEG preprocessing

EEG data was recorded using a 64 Ag-AgCl electrodes standard setting (actiCAP, Brain Products GmbH, Germany). One electrode was placed over the outer canthus of each eye, and one below the left eye to monitor eye movements and blinks. Electrode impedance was always below 15 kΩ and remained below 10 kΩ in the vast majority of electrodes across participants. During data collection, raw EEG signal was amplified (BrainAmp DC, Brain Products GmbH, Germany), online high-pass filtered at 0.05 Hz, digitized using a sampling rate of 1000 Hz, and referenced to the midline central electrode (Cz).

To obtain the best possible temporal alignment between acoustic stimuli and EEG signal, we digitized the speakers' acoustic signal, with a sampling rate of 1000 Hz (Polybox, Brain Products GmbH, Germany). This allowed us to compensate for varying lags between the digitized trigger and the actual presentation of acoustic stimuli. We then cross-correlated the amplitude values of the speakers' acoustic signal and its corresponding audio template to ensure an optimal alignment. Thus, before EEG data preprocessing, triggers that marked the onset of each speech fragment were realigned to the time of maximum correlation with the actual presentation of the acoustic signal.

All EEG data preprocessing steps, and later data transformations were conducted at sensor level in MATLAB (version R2014B, MathWorks, 2014), using both custom code and functions from FieldTrip toolbox (version 20180604, Oostenveld et al., 2011). First, we downsampled EEG and audio signals to 200 Hz. Second, we bandpass filtered the signal between 0.2 and 40 Hz with a zero-phase fourth-order finite impulse response filter, using the default transition bandwidth in the FieldTrip toolbox for bandpass filtering. Third, we detected physiological artifacts through independent component analysis (ICA, runica method) on the filtered signal. After visual inspection of ICA, we subtracted independent components related to eye movements and blinks from the EEG signal (*mean number of rejected components per participant* = 2.06, *SD* = .5). Interpolation of bad channels was achieved using the weighted average of their neighbors. Our EEG data exclusion criteria was to not further analyze participant datasets for which more than 30% of the data was rejected. Although no participant exceeded such a threshold, two participants were excluded, as one of them did not want to remain seated while listening to the stories and the other fell asleep during the recording. Thus, we ended up with a sample of 33 analyzable EEG and behavioral datasets in each language.

For speech-brain coherence analyses, we divided the continuous EEG signal from each storytelling and language into 2000-ms epochs (the inverse of our lowest frequency of interest as well as frequency resolution, 0.5 Hz) with a temporal overlap of 1000 milliseconds. We discarded epochs and channels in which overall voltage departed more than 3 z-values from the average of all epochs and channels respectively (*mean percentage of epochs removed per participant* = 1.14 %, *SD* = .66; *mean number of channels removed per participant* = 1 out of 64, *SD* = 1.93). For mTRF analyses, we proceeded with the continuous EEG signal (without epoching).

CTS indexes

We used two types of CTS metrics (Figure 1). The first one was coherence (Halliday et al., 1995), which measures the phase correlation between two signals (here the brain and the speech signals, thus termed speech-brain coherence hereafter), which was used to extract the brain tracking of acoustic temporal speech information (i.e., the speech envelope). The second set of CTS metrics were extracted from the multivariate temporal response function (mTRF, Crosse et al., 2016) of continuous brain oscillatory activity to acoustic temporal and linguistic (i.e., lexico-semantic) information. mTRFs consist of the linear mapping of the values of continuous stimuli (acting as regressors) on continuous brain activity (acting as response, EEG in our case).

Speech-brain coherence

We computed speech-brain coherence as the phase correlation between EEG and speech envelope (i.e., the Hilbert-transformed audio signal). To achieve this, we used custom functions in MATLAB following the specifications described in Molinaro and Lizarazu (2018). We circumscribed our speech-brain coherence analysis to the 0.5 – 10 Hz frequency range, which spans over the delta (0.5 – 4 Hz) and theta (4 – 7 Hz) frequency bands, which align to the timescales of prosodic phrasing (~1000-2000 ms) and syllables (~150-200 ms) respectively. In order to test whether speech-brain coherence within theta band was specifically related to syllable tracking, we estimated the syllable rate of the stories (based on an automatic algorithm, (de Jong & Wempe, 2009; in Praat, Boersma & Weenink, 2021). The overall average syllable rate was 5.63 Hz (*SD* = .33) and was highly similar in Basque (5.7 Hz, *SD* = .28) and Spanish (5.56 Hz, *SD* = .36). Given that our speech-brain coherence analysis had a frequency resolution of 0.5 Hz, the frequency bin that aligned most closely to the syllable rate was 5.5 Hz.

Coherence values vary between 0 (no linear phase relation) and 1 (total linear phase relation). To find the moment of maximum speech-brain synchronicity, we computed coherence between

both signals at 6 different time lags (ranging from 40 to 140 ms in steps of 20 ms) in two arrays of sensors that have previously shown speech-brain coherence effects, located symmetrically within the left (i.e., T7, C3, TP7, and CP3) and right (i.e., T8, C4, TP8, CP4) temporal hemispheres. This a priori selection of sensors was only used to determine the time of maximum coherence, and not for localizing statistical coherence effects in our analyses (which were located through cluster-based permutation tests). A 60 ms lag of the EEG with respect to the speech signal was the time point of maximum coherence across participants and conditions, to which we circumscribed our speech-brain coherence estimates for statistical analyses.

mTRFs

We computed one mTRF encoding model per participant and language, which included four regressors: word onsets, the speech envelope, lexical frequency, and sentence-level semantic distance. Word onsets were employed as a control regressor, to account for the presumably high variability in the mTRF models associated with the mere responses to word onsets. Speech envelope served as a regressor for acoustic-temporal CTS, and lexical frequency and semantic distance for lexico-semantic CTS. Importantly, computing envelope-level mTRF also served to verify that it was related to the envelope tracking measured with coherence and to establish a positive control for mTRF analyses of lexico-semantic features, which are less salient in the signal. The speech envelope regressors have the same operationalization (i.e., the Hilbert-transformed audio signal) as in the speech-brain coherence analyses (also following the specifications in Molinaro & Lizarazu 2018).

Lexical and semantic regressors consisted of continuous vectors (one per language and story) containing bursts at the onset of every content word in each story. The amplitude of such bursts corresponded to latent variables that were used as proxies for lexical and semantic information. For the lexical regressor, we used lexical frequency: the amplitude of each burst was the inverse

of the Zipf lexical frequency of its corresponding word (van Heuven et al., 2014; see Supplemental Figure 1). For the semantic regressor, we computed sentence-level semantic distance vectors following a similar approach to Broderick et al. (2021; see Supplemental Figure 2). We first obtained the semantic representation of each story word in a 300-dimensional space through the fasttext Python package for text representations (Joulin et al., 2016). Then, the amplitude of each burst was computed as 1 minus the Pearson correlation of the semantic dimensions of a word with the average of all its preceding words within a sentence. This way, lexical items with a bigger semantic correlation with the preceding words within a sentence had lower amplitude bursts, and less semantically related words had higher amplitude bursts relative to their preceding context.

After obtaining all the regressors of interest, we used the mTRF toolbox (Crosse et al., 2016) in MATLAB to fit one mTRF encoding model with the EEG response with all features (envelope, lexical frequency, and sentence-level semantic distance), and for each language and participant. Our mTRF analysis time window was 900 ms, spanning from -150 ms pre- to 750 ms post-stimulus. In order to train and test the encoding model, we split our continuous EEG signal and each corresponding feature vector (~ 16 minutes) into 8 folds of equal length (~ 2 minutes). We trained the encoding model in 7 folds and tested its accuracy on the remaining one, repeating this process for 30 iterations per fold. We then extracted two mTRF metrics for between-languages comparisons of CTS. First, we assessed whether and to what extent each individual regressor contributed significantly to the model fit (through *prediction correlation*, r-values). Thus, within 30 iterations per fold for each regressor of interest (envelope, lexical frequency, and semantic distance), we contrasted the full model fit (all regressors) versus the fit of the model where one regressor was permuted, and averaged such r-values across iterations. The resulting fold-average r-values, the above-chance prediction correlation coefficient between each feature, and the mTRF (at electrode level) were used in further statistical analyses as proxies for the extent to

which each feature was linear mapped by EEG activity and to confirm that this feature contributed to the model-fit above the chance level across all subjects (see mTRF group analysis section). For the main analysis, the prediction correlation coefficient (the difference between the full model r-value and the specific feature-permuted model r-value) served as an electrode-level estimate of how faithfully EEG activity mapped a given speech feature (e.g., lexical information).

The second mTRF metric was the *temporal weights* of the EEG response to each regressor. Such weights form an ERP-like response that indexes EEG temporal sensitivity to changes in a given regressor (see Crosse et al., 2021). Thus, the temporal weights allowed us to test whether there were relevant between-languages differences in the amplitude or the latency of the temporal response to the envelope, lexical, or semantic information.

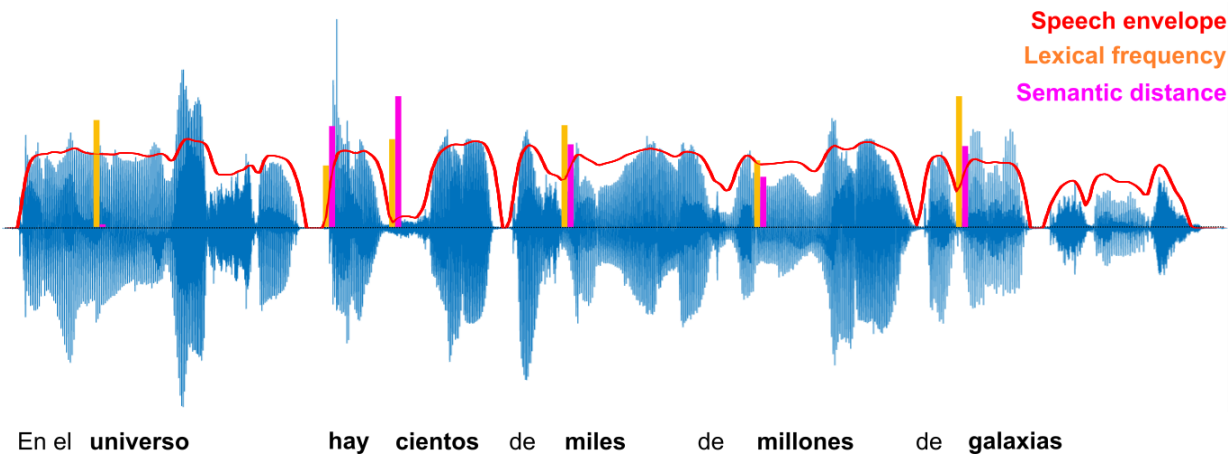


Figure 1. Graphical summary of the CTS analyses. In blue, the speech waveform of "En el universo, hay cientos de miles de millones de galaxias" ("In the universe, there are hundreds of billions of galaxies"). Speech-brain coherence and envelope-level mTRF models are based on the relationship between the speech envelope (in red) and EEG activity. Lexical frequency and semantic distance mTRFs are obtained from the EEG response to bursts of different amplitude (lexical frequency, orange; sentence-level semantic distance, pink) at the onset of each content word.

Statistical analysis

Behavioral language measures

In order to assess between-languages differences in language performance, we fitted linear mixed effect (LME) models with language (Exp(+), Exp(-)) as predictor for each behavioral language measure (language experience, vocabulary knowledge, and phonological ability), and participants as random intercepts, to account for between-participant differences. LME models were fitted using the lmer function from lme4 package (Bates et al., 2015) in R (version 4.2.1, R Core Team, 2022). In addition, we used Bayesian t-tests in JASP (version 0.16.4, JASP team, 2022) to test the strength of the evidence for between-languages similarities in our behavioral measures in the case of non-significant between-languages differences.

Speech-brain coherence

We used cluster-based permutation tests (CBPTs) to statistically estimate speech-brain coherence. CBPT is an efficient way of estimating the presence of a statistical effect in a high dimensional space, as it is able to account for the spatial adjacency of electrodes and test for significant effects that are shared across a group of electrodes (a cluster) (Maris & Oostenveld, 2007). In our case, we ran dependent-samples two-tailed CBPTs in FieldTrip with 1000 permutations. Our CBPTs required at least 2 electrodes to form a cluster as a way to limit the possibility of single-electrode false-alarm effects, and we corrected for multiple comparisons based on the number of a-priori significant clusters. We first assessed whether there was above-chance speech-brain coherence in each language by contrasting through CBPT the phase alignment of EEG and genuine speech envelope versus the phase alignment of EEG and a surrogate version of the envelope that did not follow the original speech order (i.e., flipped speech envelope surrogate). Then, we contrasted through CBPT the speech-brain coherence

values of Exp(+) and Exp(-) to analyze whether there was a significant between-languages effect on this CTS metric.

mTRFs

Before our analyses of interest, we conducted a control test to verify that, at the group level, there were significant prediction correlations between our three regressors of interest and the mTRF model. Namely, we conducted a one-sample t-test (one-tailed) against the null hypothesis (no different prediction correlation from 0) for each of the regressors. After such control, and similarly to when contrasting speech-brain coherence between languages, we compared Exp(+) and Exp(-) through 6 CBPTs, one per mTRF metric (i.e., prediction correlation and temporal weights) and regressor (i.e., envelope, lexical frequency, and semantic distance). Both prediction correlation and temporal weights were analyzed at electrode level.

Relationship between CTS and behavioral language measures

In order to assess the correspondence between CTS and behavioral language performance in the phonological and lexico-semantic domains, we fitted two LME models, one with acoustic-temporal CTS and one with lexico-semantic CTS as outcome measure. In the first LME model, phonological abilities and vocabulary knowledge, as well as their interaction with language, served as predictors of acoustic-temporal CTS, a composite of speech-brain coherence and envelope-level TRFs. In the second LME model, the same set of measures were employed as predictors of lexico-semantic CTS: a composite score of lexical- and semantic-level TRFs. We included participants as random intercepts in both LME models, which allowed us to account for baseline individual differences in the different languages and language domains. As for our behavioral language analyses, we used the lmer function from lme4 package in R. In order to test for omnibus main effects and interactions of the predictors, we used the anova function in base R.

Acknowledgements

The data, analytic code, and materials necessary to reproduce the analyses presented here will be made publicly accessible at the Open Science Framework repository of the project upon publication. The analyses presented here were not preregistered. We would like to thank the participants and their families, as well as the BCBL lab department, for their contribution and commitment over the years of this longitudinal study. Special thanks to Araitz Garnika, Leire Eizagirre, Ioanna Taouki, for their support with stimuli preparation and EEG data collection, to Giovanni Di Liberto for his support with mTRF modeling, and to Caroline Handley for her proofreading and suggestions on this manuscript. This research was supported by the FPI grant BES-2016-078125 (to ML and JPN) and by the PID2022-136989OB-I00 (to ML) by Ministerio Español de Economía, Industria y Competitividad MINECO and Fondo Social Europeo; through project RTI2018-096242-B-I00(MCIU/AEI/FEDER, UE) funded by Ministerio de Ciencia, Innovación y Universidades (MCIU), the Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER); by the Basque Government through the BERC 2018-2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490.

References

Acha, J., Laka, I., Landa, J., & Salaburu, P. (2014). EHME: A New Word Database for Research in Basque Language. *The Spanish Journal of Psychology*, *17*. https://doi.org/10.1017/sjp.2014.79

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung

speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.46, retrieved 15 May 2021 from http://www.praat.org/) [Computer software].

Boets, B., Wouters, J., van Wieringen, A., & Ghesquière, P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory. *Neuropsychologia*, *45*(8), 1608–1620. https://doi.org/10.1016/j.neuropsychologia.2007.01.009

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological Processing, Language, and Literacy: A Comparison of Children with Mild-to-moderate Sensorineural Hearing Loss and Those with Specific Language Impairment. *Journal of Child Psychology and Psychiatry*, *42*(3), 329–340. https://doi.org/10.1111/1469-7610.00726

Brodbeck, C., Bhattasali, S., Cruz Heredia, A. A., Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *eLife*, *11*, e72056. https://doi.org/10.7554/eLife.72056

Brodbeck, C., Kandylaki, K. D., & Scharenborg, O. (2023). *Neural representations of non-native speech reflect proficiency and interference from native language knowledge* (p. 2023.04.15.537014). bioRxiv. https://doi.org/10.1101/2023.04.15.537014

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, *172*, 162–174. https://doi.org/10.1016/j.neuroimage.2018.01.042

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current Biology*, *28*(5), 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *Journal of Neuroscience*, *39*(38), 7564–7575. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Broderick, M. P., Di Liberto, G. M., Anderson, A. J., Rofes, A., & Lalor, E. C. (2021). Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84597-9

Carbajal, M. J., & Peperkamp, S. (2020). Dual language input and the impact of language separation on early lexical development. *Infancy*, *25*(1), 22–45. https://doi.org/10.1111/infa.12315

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*, 705621. https://doi.org/10.3389/fnins.2021.705621

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), Article 4. https://doi.org/10.1038/nrn3924

Destoky, F., Bertels, J., Niesen, M., Wens, V., Ghinst, M. V., Leybaert, J., Lallier, M., Ince, R. A. A., Gross, J., Tiège, X. D., & Bourguignon, M. (2020). Cortical tracking of speech in noise accounts for reading strategies in children. *PLOS Biology*, *18*(8), e3000840. https://doi.org/10.1371/journal.pbio.3000840

Destoky, F., Philippe, M., Bertels, J., Verhasselt, M., Coquelet, N., Vander Ghinst, M., Wens, V., De Tiège, X., & Bourguignon, M. (2019). Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. *NeuroImage*, *184*, 201–213. https://doi.org/10.1016/j.neuroimage.2018.09.006

Di Liberto, G. M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., & Lalor, E. C. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, *175*, 70–79. https://doi.org/10.1016/j.neuroimage.2018.03.072

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85 Pt 2*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Donhauser, P. W., & Baillet, S. (2020). Two Distinct Neural Timescales for Predictive Speech Processing. *Neuron*, *0*(0). https://doi.org/10.1016/j.neuron.2019.10.019

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Ezeizabarrena, M.-J., & Alegria, A. (2015). Early coda production in bilingual Spanish and Basque. In *The Acquisition of Spanish in Understudied Language Pairings. Edited by Tiffany Judy & Silvia Perpiñán.* John Benjamins Publishing Company.

Friederici, A. D., & Männel, C. (2014). Neural correlates of the development of speech perception and comprehension. In *The Oxford Handbook of Cognitive Neuroscience* (pp. 171–192). Oxford University Press. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2028117

Gámez, P. B., Griskell, H. L., Sobrevilla, Y. N., & Vazquez, M. (2019). Dual Language and English-Only Learners' Expressive and Receptive Language Skills and Exposure to Peers' Language. *Child Development*, *90*(2), 471–479. https://doi.org/10.1111/cdev.13197

Gathercole, V. C. M., & Thomas, E. M. (2005). Minority language survival: Input factors influencing the acquisition of Welsh. *ISB4*. Proceedings of the 4th International Symposium on Bilingualism, Somerville, MA.

Gervain, J. (2015). Plasticity in early language acquisition: The effects of prenatal and early childhood experience. *Current Opinion in Neurobiology*, *35*, 13–20. https://doi.org/10.1016/j.conb.2015.05.004

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Goswami, U. (2017). A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Current Directions in Psychological Science*, 56–63. https://doi.org/10.1177/0963721417727520

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read* (pp. viii, 166). Lawrence Erlbaum Associates, Inc.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, *4*(1). https://doi.org/10.1515/lp-2013-0004

Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2010). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

Granados Barbero, R., Ghesquière, P., & Wouters, J. (2022). Development of Atypical Reading at Ages 5 to 9 Years and Processing of Speech Envelope Modulations in the Brain. *Frontiers in Computational Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fncom.2022.894578

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLOS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Guasti, M. T. (2002). *Language acquisition: The growth of grammar* (pp. xiv, 474). The MIT Press.

Halliday, D. M., Rosenberg, J. R., Amjad, A. M., Breeze, P., Conway, B. A., & Farmer, S. F. (1995). A framework for the analysis of mixed time series/point process data—Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Progress in Biophysics and Molecular Biology*, *64*(2–3), 237–278. https://doi.org/10.1016/s0079-6107(96)00009-0

Hamilton, L. S., Edwards, E., & Chang, E. F. (2018). A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Current Biology*, *28*(12), 1860-1871.e4. https://doi.org/10.1016/j.cub.2018.04.033

Hauswald, A., Keitel, A., Chen, Y.-P., Rösch, S., & Weisz, N. (2022). Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *European Journal of Neuroscience*, *55*(11–12), 3288–3302. https://doi.org/10.1111/ejn.14912

Hualde, J. I., & Urbina, J. O. de. (2011). *A Grammar of Basque*. Walter de Gruyter.

Hull, R., & Vaid, J. (2006). Laterality and language experience. *Laterality: Asymmetries of Body, Brain and Cognition*, *11*(5), 436–464. https://doi.org/10.1080/13576500600691162

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia*, *45*(9), 1987–2008. https://doi.org/10.1016/j.neuropsychologia.2007.03.002

Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers*. *Bilingualism: Language and Cognition*, *17*(1), 189–202. https://doi.org/10.1017/S136672891300014X

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), Article 7600. https://doi.org/10.1038/nature17637

JASP team. (2022). *JASP* (0.16.4) [Computer software].

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). *FastText.zip: Compressing text classification models* (arXiv:1612.03651). arXiv. https://doi.org/10.48550/arXiv.1612.03651

Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at three years: A significant relationship. *Developmental Science*, *0*(ja), e12836. https://doi.org/10.1111/desc.12836

Kalashnikova, M., Goswami, U., & Burnham, D. (2020). Infant-directed speech to infants at risk for dyslexia: A novel cross-dyad design. *Infancy*, *25*(3), 286–303. https://doi.org/10.1111/infa.12329

Kaufeld, G., Bosker, H. R., Oever, S. ten, Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0302-20.2020

Klimovich-Gray, A., Barrena, A., Agirre, E., & Molinaro, N. (2021). One Way or Another: Cortical Language Areas Flexibly Adapt Processing Strategies to Perceptual And Contextual Properties of Speech. *Cerebral Cortex*, *bhab071*. https://doi.org/10.1093/cercor/bhab071

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. https://doi.org/10.1080/15250000903263957

Koskinen, M., Kurimo, M., Gross, J., Hyvärinen, A., & Hari, R. (2020). Brain activity reflects the predictability of word sequences in listened continuous speech. *NeuroImage*, *219*, 116936. https://doi.org/10.1016/j.neuroimage.2020.116936

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, *308*, 141–161. https://doi.org/10.1016/j.heares.2013.07.015

Leong, V., & Goswami, U. (2017). Difficulties in auditory organization as a cause of reading backwardness? An auditory neuroscience perspective. *Developmental Science*, *20*(6), e12457. https://doi.org/10.1111/desc.12457

Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., & Molinaro, N. (2021). Language Proficiency Entails Tuning Cortical Activity to Second Language Speech. *Cerebral Cortex*, *bhab051*. https://doi.org/10.1093/cercor/bhab051

Lundberg, I., Frost, J., & Petersen, O.-P. (1988). Effects of an Extensive Program for Stimulating Phonological Awareness in Preschool Children. *Reading Research Quarterly*, *23*(3), 263–284. JSTOR.

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86–94. https://doi.org/10.1016/j.dcn.2013.01.003

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, A. E. (2020). A Compositional Neural Architecture for Language. *Journal of Cognitive Neuroscience*, *32*(8), 1407–1427. https://doi.org/10.1162/jocn_a_01552

Martin, A. E., & Doumas, L. A. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biology*, *15*(3), e2000663. https://doi.org/10.1371/journal.pbio.2000663

MathWorks. (2014). *MATLAB* (version 8.4.0.150421 (R2014b)) [Computer software]. The MathWorks Inc.

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Menn, K. H., Ward, E. K., Braukmann, R., van den Boomen, C., Buitelaar, J., Hunnius, S., & Snijders, T. M. (2022). Neural Tracking in Infancy Predicts Language Development in Children With and Without Family History of Autism. *Neurobiology of Language*, *3*(3), 495–514. https://doi.org/10.1162/nol_a_00074

Meyer, L., & Gumbert, M. (2018). Synchronization of Electrophysiological Responses with Speech Benefits Syntactic Information Processing. *Journal of Cognitive Neuroscience*, *30*(8), 1066–1074. https://doi.org/10.1162/jocn_a_01236

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(9), 4293–4302. https://doi.org/10.1093/cercor/bhw228

Molinaro, N., & Lizarazu, M. (2018). Delta(but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, *48*(7), 2642–2650. https://doi.org/10.1111/ejn.13811

Molinaro, N., Lizarazu, M., Baldin, V., Pérez-Navarro, J., Lallier, M., & Ríos-López, P. (2021). Speech-brain phase coupling is enhanced in low contextual semantic predictability

conditions. *Neuropsychologia*, 107830. https://doi.org/10.1016/j.neuropsychologia.2021.107830

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *37*(8), 2767–2783. https://doi.org/10.1002/hbm.23206

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323–331. https://doi.org/10.1016/0010-0277(79)90020-9

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book Language and Its Implications for Children's Language, Literacy, and Development. *Current Directions in Psychological Science*, *31*(4), 375–380. https://doi.org/10.1177/09637214221103264

Nittrouer, S. (1996). The Relation Between Speech Perception and Phonemic Awareness. *Journal of Speech, Language, and Hearing Research*, *39*(5), 1059–1070. https://doi.org/10.1044/jshr.3905.1059

Nittrouer, S. (2002). From Ear to Cortex. *Language, Speech, and Hearing Services in Schools*, *33*(4), 237–252. https://doi.org/10.1044/0161-1461(2002/020)

Nittrouer, S., & Burton, L. T. (2001). The Role of Early Language Experience in the Development of Speech Perception and Language Processing Abilities in Children with Hearing Loss. *Volta Review*, *103*(1), 5–37.

Nittrouer, S., & Burton, L. T. (2005). The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status. *Journal of Communication Disorders*, *38*(1), 29–63. https://doi.org/10.1016/j.jcomdis.2004.03.006

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.08.004

Official Gazette of the Basque Country, BOPV 25-04-2008 (2008).

Oller, D. K., & Eilers, R. E. (2002). *Language and Literacy in Bilingual Children*. Multilingual Matters. https://digitalcommons.library.umaine.edu/fac_monographs/188

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*(2), 191–230. https://doi.org/10.1017/S0142716407070117

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Intell. Neuroscience*, *2011*, 1:1-1:9. https://doi.org/10.1155/2011/156869

Panda, E. J., Emami, Z., Valiante, T. A., & Pang, E. W. (2020). EEG phase synchronization during semantic unification relates to individual differences in children's vocabulary skill. *Developmental Science*, *n/a*(n/a), e12984. https://doi.org/10.1111/desc.12984

Paradis, J., & Genesee, F. (1996). Syntactic Acquisition in Bilingual Children: Autonomous or Interdependent? *Studies in Second Language Acquisition*, *18*(1), 1–25. https://doi.org/10.1017/S0272263100014662

Paradis, J., & Jia, R. (2016). Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, *20*(1), e12433. https://doi.org/10.1111/desc.12433

Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: Individual variation in long-term outcomes. *Journal of Child Language*, *43*(3), 553–580. https://doi.org/10.1017/S0305000915000562

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58. https://doi.org/10.1017/S0142716400009863

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pérez-Navarro, J., & Lallier, M. (2023). *The contribution of the amount of linguistic exposure to bilingual language development: Longitudinal evidence from preschool years*. PsyArXiv. https://doi.org/10.31234/osf.io/jwkt5

Peter, V., van Ommen, S., Kalashnikova, M., Mazuka, R., Nazzi, T., & Burnham, D. (2022). Language specificity in cortical tracking of speech rhythm at the mora, syllable, and foot levels. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-17401-x

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1–10. https://doi.org/10.1016/j.bandl.2016.06.006

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00777

R Core Team. (2022). *R: A language and environment for statistical computing* [Manual]. https://www.R-project.org/

Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain and Language*, *213*, 104891. https://doi.org/10.1016/j.bandl.2020.104891

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2020). Development of neural oscillatory activity in response to speech in children from 4 to 6 years old. *Developmental Science*, e12947. https://doi.org/10.1111/desc.12947

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2021). Right-hemisphere coherence to speech at pre-reading stages predicts reading performance one year later.

Journal of Cognitive Psychology, *0*(0), 1–15. https://doi.org/10.1080/20445911.2021.1986514

Sanchez-Alonso, S., & Aslin, R. N. (2022). Towards a model of language neurobiology in early development. *Brain and Language*, *224*, 105047. https://doi.org/10.1016/j.bandl.2021.105047

Schmidt, F., Chen, Y.-P., Keitel, A., Rösch, S., Hannemann, R., Serman, M., Hauswald, A., & Weisz, N. (2023). Neural speech tracking shifts from the syllabic to the modulation rate of speech as intelligibility decreases. *Psychophysiology*, e14362. https://doi.org/10.1111/psyp.14362

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), Article 5. https://doi.org/10.1038/nrn.2016.23

ten Oever, S., & Martin, A. E. (2021). An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. *eLife*, *10*, e68066. https://doi.org/10.7554/eLife.68066

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1), 1–21. https://doi.org/10.1080/14769670500215647

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English: *Quarterly Journal of Experimental Psychology*. https://journals.sagepub.com/doi/10.1080/17470218.2013.850521

Vanvooren, S., Poelmans, H., De Vos, A., Ghesquière, P., & Wouters, J. (2017). Do prereaders' auditory processing and speech perception predict later literacy? *Research in Developmental Disabilities*, *70*, 138–151. https://doi.org/10.1016/j.ridd.2017.09.005

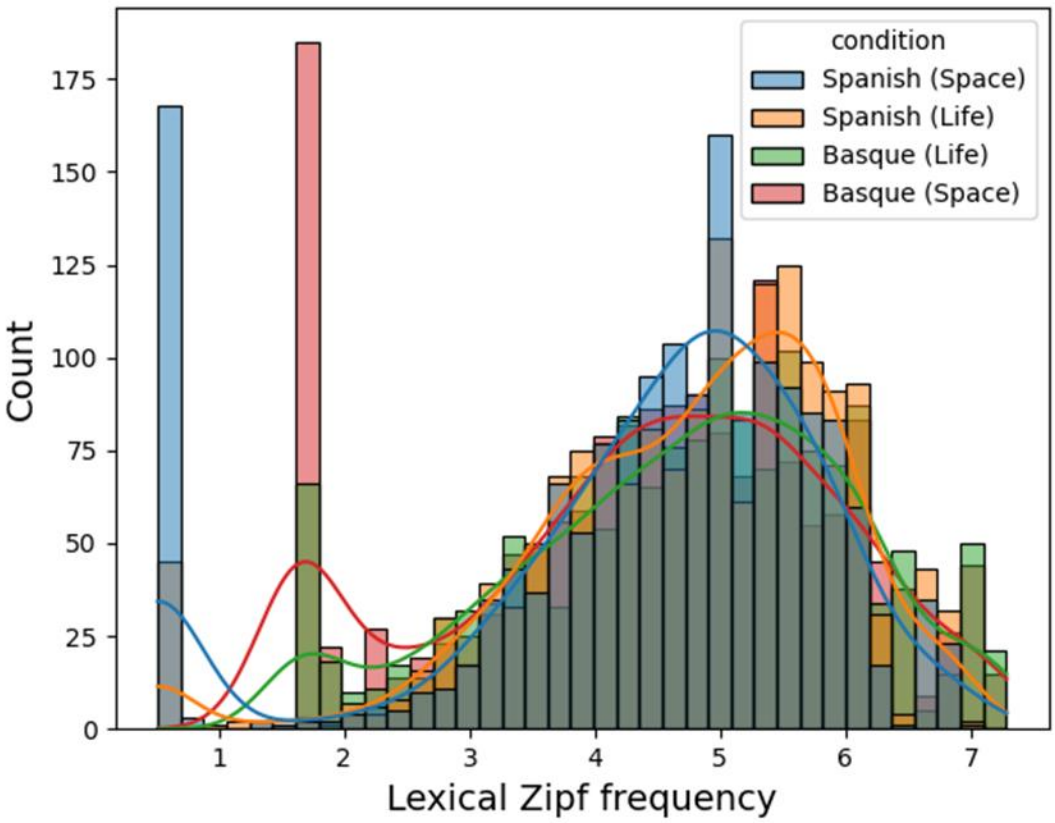
Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Zhang, R., Wang, J., Lin, H., Turk-Browne, N. B., & Cai, Q. (2023). Neural signatures of second language proficiency in narrative processing. *Cerebral Cortex*, *33*(13), 8477–8484. https://doi.org/10.1093/cercor/bhad133

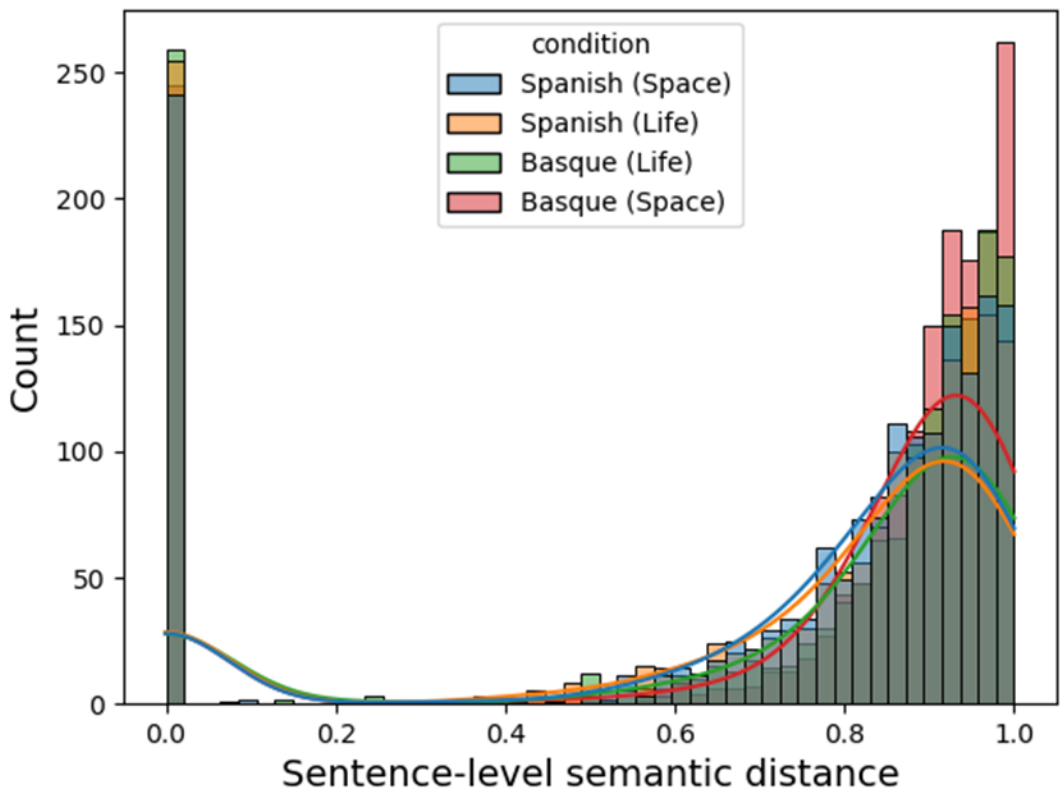
CRediT authorship contribution statement

Jose Pérez-Navarro: Conceptualization, Methodology, Formal analysis, Writing – Original draft, Writing – Review & Editing Investigation, Data curation, Visualization, Funding acquisition; **Anastasia Klimovich-Gray**: Methodology, Formal analysis, Writing – Review & Editing, Data Curation, Resources, Visualization; **Mikel Lizarazu**: Methodology, Formal analysis, Writing – Review & Editing, Data Curation, Resources, Visualization; **Giorgio Piazza**: Data collection, Data curation, Writing – Review & Editing; **Nicola Molinaro**: Conceptualization, Writing – Original draft , Writing – Review & Editing, Supervision; **Marie Lallier**: Conceptualization, Writing – Original draft, Writing – Review & Editing, Supervision.

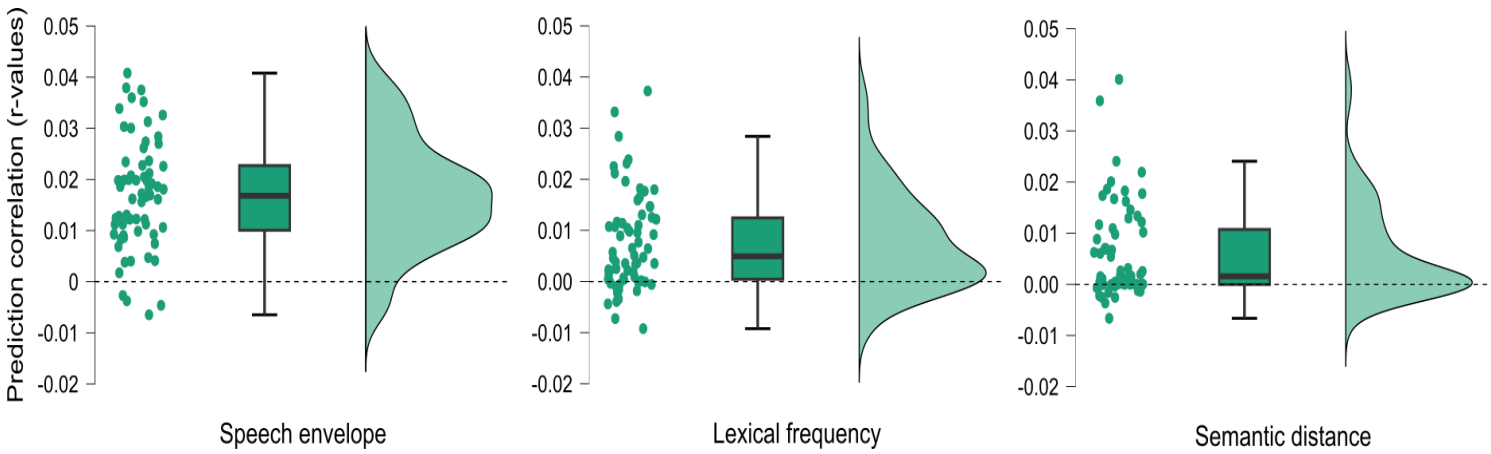
Supplemental material



Supplemental Figure 1. Zipf frequency distributions of content words in Spanish (1st, blue; 2nd, yellow), and Basque (1st, green; 2nd, red) stories. Bars represent content word count per Zipf frequency bin and smooth lines mark the density distribution.



Supplemental Figure 2. Sentence-level semantic distance distributions of content words in Spanish (1st, blue; 2nd, yellow), and Basque (1st, green; 2nd, red) stories. Bars represent content word count per bin of sentence-level semantic distance and smooth lines mark the density distribution.



Supplemental Figure 3. Distributions of average prediction correlations between each regressor and the mTRF model. Each participant is represented by two points (one per language). In the boxplots, horizontal lines within each box mark the median score. Upper and lower hinges mark the first and third quartile, and whiskers show 1.5 * interquartile range. Density plots mark the group-level distribution of prediction correlations for each regressor.

Supplemental Formula 1. Formula and rationale of the amount of exposure (AoE) to a language composite index (AoE_L). The $CurrentHomeAoE_L$ and $CurrentSchoolAoE_L$ values are themselves composite indexes (*SF2* and *SF3* respectively):

$$AoE_L = \frac{(Age - Age of Acquisition_L) \times BirthToAgeAoE_L + \left((CurrentHomeAoE_L \times .75) + (CurrentSchoolAoE_L \times .25)\right)}{2}$$

The rationale behind the multipliers of the different indexes that build the composite measure is weighting the contexts in which participants are exposed to a given language by the relative number of waking hours that they represent in an everyday life situation. We took into consideration the relative amount of waking hours that children spend at home (~3147 h; 75% of the total of waking hours) and at school (~1050 h; 25% of waking hours), as well as the relative amount of school hours that children spend in classes (~875 h; 83%) and leisure time (~175 h; 17%) according to their schooling system (Official Gazette of the Basque Country, 2008).

Supplemental Formula 2. Formula of the *CurrentHomeAoE*$_L$ composite index:

$$CurrentHomeAoE_L = (Mother_L \times .5) + (Father_L \times .35) + (NuclearFamily_L \times .08) + (MaternalExtendedFamily_L \times .035) + (PaternalExtendedFamily_L \times .035)$$

In Supplemental Formula 2, we took into consideration the proportion of waking hours that children spend in contact with the different members of their familial environment (Leaper et al., 1998; Pancsofar & Vernon-Feagans, 2006). In the case of one-parent families, we adapted the above formula to *CurrentHomeAoE*$_L$*, which aggregates the input from both parents into a single value (*Parent*$_L$), as well as the ones from the maternal and paternal extended family (*Extended Family*$_L$), as follows:

Supplemental Formula 3. Formula of the *CurrentSchoolAoE*$_L$ composite index:

$$CurrentSchoolAoE_L = (SchoolLearningContent_L \times .83) + (SchoolLeisureTime_L \times .17)$$