Developmental trajectories of bilingual word recognition

Gonzalo Garcia-Castro

Serene Siow

Kim Plunkett

Nuria Sebastian-Galles

# Introduction

Building a mental lexicon is a major achievement in the development of an infant: by storing representations of how familiar words sound and what they mean, an infant is able to make sense of their linguistic input. The foundations of an initial lexicon are in place before their first year of life ([Bergelson & Swingley, 2012](#ref-bergelson20126), [2015](#ref-bergelson2015early); [Hallé & Boysson-Bardies, 1994](#ref-halle1994emergence); [Parise & Csibra, 2012](#ref-parise2012electrophysiological); [Tincoff & Jusczyk, 1999](#ref-tincoff1999some); [Vihman, 2004](#ref-vihman2004cross)). This initial lexicon consists of only a few items; mainly words for people, interjections, body parts, and food ([Tardif et al., 2008](#ref-tardif2008baby); [Tincoff & Jusczyk, 2012](#ref-tincoff2012six)), but it undergoes rapid growth during the second year of life ([Bloom, 2002](#ref-bloom2002children); [Ganger & Brent, 2004](#ref-ganger2004reexamining); [Goldfield & Reznick, 1990](#ref-goldfield1990early); [McMurray, 2007](#ref-mcmurray2007defusing)). According to parental reports, the average 15-month-old infant already understands more than 100 words, and by two years, they understand more than 400 words ([Frank et al., 2021](#ref-frank2021variability)). This accelerated lexical developmental is reflected in infants’ trajectories of word recognition: infants recognise familiar words faster and more efficiently as they approach their second birthday ([Fernald et al., 1998](#ref-fernald1998rapid), [2001](#ref-fernald2001half); [Hurtado et al., 2007](#ref-hurtado2007spoken)). Despite being exposed to a more complex linguistic input, bilinguals show equivalent trajectories of word acquisition and word recognition to their monolingual peers’ ([Byers-Heinlein et al., 2023](#ref-byers-heinlein2023sometimes); [De Houwer et al., 2014](#ref-de2014bilingual); [Hoff et al., 2012](#ref-hoff2012dual); [Legacy et al., 2018](#ref-legacy2018vocabulary); [Pearson & Fernández, 1994](#ref-pearson1994patterns); [Vihman et al., 2007](#ref-vihman2007onset)). This is a remarkable deed for two reasons. First, bilingual infants receive a relative impoverished linguistic input in each of their languages compared to monolinguals ([Costa & Sebastián-Gallés, 2014](#ref-costa2014does)). Second, they face a more complex referential context, since they often learn two labels for each referent, one in each language. The mechanisms that allow bilingual’ trajectories of lexical developmental to keep up with monolinguals’ are still unclear.

Previous studies have pointed to the interplay between bilingual’ languages as a candidate facilitative mechanism behind bilingual lexical acquisition. In particular, cross-language similarity at the lexical level has been associated with larger vocabulary sizes. ([**floccia2018introduction?**](#ref-floccia2018introduction)) collected vocabulary data from a sample of 367 bilingual children living in the United Kingdom, who were learning English and an additional language (out of a diverse pool of 13 languages). Using an edit distance-based metric of phonological similarity, the authors computed the average lexical similarity between the translation pairs of the English vocabulary checklist, and of each of the other languages. Children’s vocabulary sizes in the additional language was positively associated with the amount of language similarity between their two languages. For instance, children learning English and Dutch (which share high lexical similarity) showed larger vocabulary sizes than those learning English and Mandarin (which share low lexical similarity). The authors suggested that the acquisition of words in the additional language might be facilitated by their cognate status (i.e., being phonologically similar to their translation equivalent). Larger vocabulary sizes might then be expected in bilinguals learning two languages sharing a high proportion of cognates. Recent studies have reported an earlier acquisition of cognate words ([Mitchell et al., 2022](#ref-mitchell2022cognates); [**garcia-castro2023cognate?**](#ref-garcia-castro2023cognate)), providing further evidence that cross-language similarity at the lexical level might boost vocabulary growth in bilinguals.

The facilitative effect of cross-language similarity is in line with the language non-selective account of bilingual lexical access. There is robust evidence suggesting that bilingual adults activate both languages in parallel—even during monolingual situations—which influences their performance in word recognition and production tasks ([Ana I. Schwartz & Diaz, 2007](#ref-schwartz2007reading), [2007](#ref-schwartz2007reading); [Dijkstra et al., 1999](#ref-dijkstra1999recognition), [2010](#ref-dijkstra2010cross); [Dufour & Kroll, 1995](#ref-dufour1995matching); [Groot, 1992](#ref-groot1992determinants); [**marian1999?**](#ref-marian1999)). For instance Costa et al. ([2000](#ref-costa2000cognate)) presented Catalan-Spanish bilinguals with a series of pictures of familiar objects. For each object, participants were asked to name it in Catalan (their dominant language). Unbeknownst to participants, the authors manipulated the cognate status pictures’ labels in Catalan and Spanish. In half of the trials, the labels associated to the pictures were cognates (i.e., phonologically similar translation equivalents, e.g., *gat*-*gat* [*cat*]) in Catalan and Spanish, whereas in the other half of the trials the labels were non-cognates (e.g., *taula*-*mesa* [*table*]). Participants named pictures faster in cognate trials than in non-cognate trials. Critically, Spanish monolinguals who completed the same task, naming the pictures in Spanish, showed equivalent naming times in both conditions. These results revealed that bilinguals activated their Spanish phonology, despite performing the naming task exclusively in Catalan: the visual recognition of the presented pictures led to the activation of its associated phonological forms in both languages in parallel, which influenced the subsequent the dynamics of word production.

Parallel activation has also been reported in he developing lexicon ([Bosma & Nota, 2020](#ref-bosma2020cognate); [Floccia et al., 2020](#ref-floccia2020translation); [Jardak & Byers-Heinlein, 2019](#ref-jardak2019labels); [Poarch & Van Hell, 2012](#ref-poarch2012cross); [Singh, 2014](#ref-singh2014one); [Von Holzen et al., 2019](#ref-von2019impact)). For instance, Von Holzen & Mani ([2012](#ref-von2012language)) found evidence of cross-language phonological priming in twenty 21- to 42-months-old children learning German and English. In their experimental task, each trials begun with the auditory presentation of an English prime word embedded in a carrier phrase. Then, the target label was auditorily presented in German. Finally, the target and a distractor pictures were presented side-by-side. The authors created three conditions. In the phonological priming condition, the auditorily presented prime and target labels were phonologically related (e.g., *slide*-*Klide* [*dress*]). In the priming through translation condition, both labels were phonologically related through translation: the auditorily presented prime label did *not* overlap with the target label (e.g., *leg*-*Stein* [*stone*]), but its translation in German did [*Bein*]. In the unrelated condition, prime and target labels were phonologically unrelated in both languages. Interestingly, participants showed a weaker target preference in priming through translation trials, compared to unrelated trials. This was interpreted evidence of the lexical representations of the prime labels being activated in both English and German, leading to interference effects in later target recognition.

Previous studies had reported similar cross-language priming effects. The conclusion that such effects occurred at the lexical level, as opposed to being the result of mere acoustic similarity, was nonetheless unwarranted. In studies where the auditorily presented prime label was phonologically overlapped with the target label (e.g., [Marian & Spivey, 1999](#ref-marian1999activation)), priming effects might be explained by the activation of the phonological segments of the prime label, which facilitated the recognition of the target label, which shared such phonological segments. This might have been the phonological priming trials in Von Holzen & Mani ([2012](#ref-von2012language)), in which prime and target labels shared overt phonological overlap. In contrast, the priming through translation results suggest that cross-language activation in the developing lexicon occurs at the lexical level. Since participants’ target recognition was modulated by the non-presented prime translation, such an inhibitory effect must stem at the lexical level. The fact that phonological priming facilitated the recognition of the target words (both labels belong to different languages), and that phonological priming through translation interfered target recognition (translation and target belong to the same language) points to cross-language connections playing an excitatory role, and to within-language connections playing an inhibitory role.

Despite extending the available knowledge on the developing bilingual lexicon, the study by Von Holzen & Mani ([2012](#ref-von2012language)) has some methodological challenges that prevent strong conclusions about language non-selective access. First, participants’ received their English input at school, where they heard German and English (between three and four hours each) from a German native speaker. Except for four children who also listened to English at home, participants’ only exposure to English was 3-4 hours per day, for the time they attended daycare (an average of 12.35 months). Assuming that infants are exposed to a daily average of 12 total hours of speech, and they attended daycare only during weekdays, their estimated exposure to English throughout the week represented between 10.50% and 14.00% for the 21 month-olds (the youngest in the sample), and between 5% and 7% for the 43-month-olds. Conventionally, children are considered bilingual if exposed to 25% or more of the time to a second language ([Byers-Heinlein et al., 2021](#ref-byers-heinlein2021multilab); [Rocha-Hidalgo & Barr, 2023](#ref-rocha-hidalgo2023defining)). Participants in Von Holzen & Mani ([2012](#ref-von2012language)) might therefore not be generalisable to other populations of bilinguals who received a more balanced dual language input. Second, the presence of cognates in the stimuli lists lead to some primes in the phonological priming condition sharing substantial cross-language phonological overlap thought translation with the target words like *fire* //–*Feuer* // and *Eier* //, or *Moon* //–*Mond* // and *Huhn* //. Conversely, some trials in the phonological priming through translation shared within-language overlap with the target, such as *Soup* //–*Suppe* //—*Puppe* // and *Puppe* //. Given that the number of total trials in both conditions amounted to six trials, a considerable proportion of the trials presented this methodological limitation. Finally, the observed pattern of results was dependent on the inclusion of a single item, *Buch* [*book*, in German]. When this item was included, monolinguals showed a similar interference effect to bilinguals in the priming through translation condition, in which only bilinguals with knowledge of the translation should show an effect. Additionally, given that monolinguals and bilinguals were not directly, it is not possible to conclude that the differences in statistical significance between monolingual and bilingual target looking, compared to chance levels, does not necessarily provide a significant difference between the two groups. As a consequence, the conclusion that the interference effects observed in the priming through translation condition is exclusive to bilinguals (which would suggest that the priming effect were not due to the intrinsic properties of the stimuli) in unwarranted.

Additionally, although the priming through translation paradigm provides a convenient framework to investigate parallel activation in the bilingual lexicon, its design prevents strong interpretations about the extent to which parallel activation is the result of the experimental manipulations of interest. The presentation of primes and targets in both languages in the same task—or even within the same trial—presents participants with a bilingual context that might lead to a higher baseline activation of lexical representations of both languages, contrary to a context in which parallel activation results from genuine cross-language activation ([Grosjean, 1997](#ref-grosjean1997bilingual)). Another limitation of priming through translation studies is that, when prime words are auditorily presented, the task constrains the naming context to one of the languages. The consequence is a potential asymmetry between the activation dynamics of the lexical representation of both labels: while the participant has heard one label, the activation of its translation must spread across semantic links. This might lead to a lower strength, and higher latency of the activation of the translation. In summary, current evidence on the cross-language interplay in the initial lexicon should be considered carefully.

In the present study, we put the hypothesis that bilingual infants activate both languages in parallel to a more severe test. We used an extension of the implicit naming task by Mani & Plunkett ([2010](#ref-mani2010infant)), in which participants were primed with a picture presented in silence. Previous studies using this paradigm have successfully provided evidence of language non-selectivity. In particular, Von Holzen & Mani ([2014](#ref-von2014bilinguals)) found that bilingual adults generate implicit labels in both labels for visually fixated pictures. The authors presented German-English bilinguals with 120 prime-target pairs. Primes were presented as familiar pictures in silence. After prime picture offset, target words were presented auditorily. Participants’ N400 ERP components were recorded from target word onset. The authors manipulated the phonological relationship between the prime and target word-forms within participants’ L1 and across L1 and L2. When prime and target were phonologically identical (*Affe*-*Affe*, German for *monkey*), or similar (*Fahne*-*Sahne*, German for *flag* and *ice-cream*, respectively) within L1, participants’ showed a reduced N400 amplitude, compared to when prime and target were phonologically unrelated (*Messer*-*Seil*, German for *knife* and *rope*). Critically, a similar effect was found when prime and target were phonologically related through translation (*Rustsche*-*Kleid*, German for *slide* and *dress*). This suggests that participants activated prime labels in both languages in parallel, and that both labels impacted the dynamics of target words recognition.

This paradigm provides participants with an exclusively monolingual context in which they perform the task in only one of their languages. If differences in task performance are found associated with covert manipulation of cross-linguistic properties of the stimuli, such differences may be more conclusively linked to the parallel activation of lexical representations in participants’ lexicons. By using an implicit naming paradigm, the activation of the lexical representation of the prime may occur in both languages as the result of lexical retrieval after visual recognition of the prime picture, since participants are not presented with any auditory prime label. This contrasts with priming through-translation tasks, in which the prime label in one of the languages is directly activated by auditory word recognition, while the label of its translation is expected to be activated as the result of cascaded activation spreading through their semantic association. This may lead to an asymmetry in the time course of the activation of both labels, which might in turn affect the dynamics of phonological priming in subsequent word recognition.

In the present investigation, we presented 20- to 32-months-old toddlers with a cross-linguistic phonological priming paradigm. We manipulated both the phonological overlap between the prime word and target word, and the cognate status of the prime word, to test the hypothesis that bilingual toddlers implicitly name familiar pictures in both languages. If so, phonological priming effect should be larger when the prime label in both languages shares phonological overlap with the target words (*Related/Cognate* condition), compared to when only the label of the prime in one language overlaps phonologically with the target (*Related/Non-cognate*). We also also included a control *Unrelated* condition in which regardless of the cognate status of the prime word, neither of the prime labels in any language overlapped phonologically with the target word. To test cross-language priming effects, we registered participants’ time course of target picture looking in the three conditions. We predicted within-language priming effects, revealed as differences in target looking between *Unrelated* and *Related/Non-cognate* trials, as previously reported Mani & Plunkett ([2011](#ref-mani2011phonological)). We further predicted cross-language priming effects, revealed as differences in target looking between *Related/Non-cognate* and *Related/Cognate* trials. This effect should be stronger in bilinguals, if not completely absent in monolinguals, who should be indifferent to the phonological properties of the stimuli in the other language.

Following previous studies associating the emergence of phonological links between lexical items to the size of children’s vocabulary ([Mani & Plunkett, 2011](#ref-mani2011phonological)), we also compared the explanatory power of two indices of receptive vocabulary size (dominant-language vocabulary size and total vocabulary size) against participants’ age. We used parental reports of vocabulary checklist to estimate participants’ vocabulary size. We computed vocabulary size in the dominant language (L1 vocabulary) as the proportion of words in the dominant language reported by their caregivers to be understood by the child in a vocabulary checklist, and the total vocabulary size as the proportion of words reported as understood by the child in *both* languages. If within-language or cross-language phonological associations in the lexicon emerge as a function of vocabulary growth, the size of the differences between conditions should larger in participants with larger vocabulary sizes. A model including participants’ L1 or total vocabulary size as a predictor should then show better performance than a model that instead includes participants age. In the case of bilinguals, if the emergence of cross-language priming effects is associated with the growth of their vocabulary is *both* languages, a model including total vocabulary size as a predictor should perform better than a model that includes L1 vocabulary size. The opposite pattern would indicate that cross-language priming effects emerge as a function of the growth of L1 vocabulary.

Testing bilingual populations at early ages presents the limitation that the vocabulary size of participants’ non-dominant language is relatively small, which substantially constrains the number of items that can be used to as stimuli in the task, as well as the number of trials that can be constructed to generate stimuli lists. In order to circumvent the problem of limited vocabulary knowledge in the non-dominant language, tested participants only in their dominant language, therefore directing our research to the question of how possessing a second language affects processing of the dominant one in bilingual toddlers ([Costa & Sebastián-Gallés, 2014](#ref-costa2014does) provide a fuller description of the advantages of such approach).

# Study 1

## Methods

All materials, data, and reproducible code can be found at the OSF ([https://osf.io/hy984/](https://osf.io/ckydb/)) and GitHub (<https://github.com/gongcastro/cognate-priming>) repositories. This study was conducted according to guidelines laid down in the Declaration of Helsinki, and was approved by the Drug Research Ethical Committee (CEIm) of the IMIM Parc de Salut Mar, reference 2020/9080/I. Before every testing session, caregivers were asked to read and sign an informed consent form, and were given a token of appreciation at the end of it.

### Participants

We collected data from 162 children living in the Metropolitan Area of Barcelona (Spain), tested at the Laboratori de Recerca en Infància at the Universitat Pompeu Fabra. Families were recruited from maternity rooms in private hospitals and social media, and contacted via phone when the child’s age spanned between 20 and 32 months. From the 162 children that participated, 81 participated once, 55 participated twice, and 26 participated three times. Recurrent participants were tested with at least 2.06 months of difference. We gathered a total of 269 testing sessions. Participants were divided into monolinguals and bilinguals based on their relative degree of exposure to Catalan and Spanish, estimated using the Language Exposure Questionnaire (LEQ, [Bosch & Sebastián-Gallés, 2001](#ref-bosch2001evidence)). We categorised participants as monolingual if exposed to more than 80% or more of the time to their dominant language, and as bilingual otherwise. Eighty-three of the participants were categorised as monolinguals (49 female, 34 male) and 80 as Catalan/Spanish bilinguals (34 female, 48 male) (see **?@tbl-participants** for a detailed summary of participants’ age and language profile). Participants’ vision was normal, none used glasses or any other type of vision corrector.

|  |  |  |  |  | *Degree of Exposure (%)* | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *Age (months)* | | *Spanish* | *Catalan* | *English* |
| *N sessions* | *N participants* | *M (SD)* | *Range* | *M (SD)* | *M (SD)* | *M (SD)* |
| Monolingual | | | | | | | |
| Catalan | 95 | 53 | 25.7 (3.8) | 20.0 (32.3) | 4.3 (5.8) | 95.3 (6.0) | 0.3 (2.1) |
| Spanish | 53 | 30 | 25.4 (3.8) | 20.0 (32.0) | 91.2 (6.5) | 8.4 (6.6) | 0.3 (0.9) |
| Bilingual | | | | | | | |
| Catalan | 72 | 51 | 25.3 (3.8) | 19.4 (31.7) | 37.5 (10.4) | 61.8 (10.7) | 0.2 (0.8) |
| Spanish | 49 | 33 | 25.1 (3.4) | 20.2 (31.3) | 61.0 (11.1) | 38.7 (10.6) | 0.3 (1.4) |

**?(caption)**

We collected vocabulary data using parental responses to the Barcelona Vocabulary Inventory (BVQ, [Garcia-Castro et al., 2023](#ref-garcia-castro2023bvq)), an online vocabulary checklist inspired in several adaptations of the the Communicative Developmental Inventory (CDI, [Fenson et al., 1994](#ref-fenson1994variability)) developed to assess the vocabulary size of Catalan-Spanish bilingual toddlers. Families received a link to the BVQ immediately after each experimental session, and were given two weeks to fill it. We calculated two measures of receptive vocabulary size for each participant: *L1 vocabulary size* (proportion of words reported as acquired in the checklist of the dominant language), and *total vocabulary size* (proportion of the words in both checklists reported as acquired). 132 (49%) Families failed to provide a complete response to the BVQ within the two-week time limit, or did not provide a successful response to the questionnaire. For missing questionnaire responses, we imputed the vocabulary size of the participant using single imputation, using the vocabulary size scores of a pool of 586 additional participants for which a successful response for the questionnaire had been gathered. We used participants age in months and their language profile (monolingual/bilingual) as predictors. We used the mice R package ([Van Buuren & Groothuis-Oudshoorn, 2011](#ref-van2011mice)) to perform imputation using the Bayesian linear regression method.

Overall, there was a substantial increase in both total and L1 vocabulary sizes associated to age (see [Figure 1](#fig-vocabulary)). Monolinguals showed overall higher L1 vocabulary sizes than monolinguals: they were reported to understand an average of 56.9% (*SD* = 17.74%) of the words at around 21 months, 75.9% (*SD* = 14.79%) at 25 months, and 82.6% (*SD* = 14.95%) at 30 months, while bilinguals were reported to understand an average of 49.1% (*SD* = 22.51%) of the words at around 21 months, 68.0% (*SD* = 20.33%) at 25 months, and 79.0% (*SD* = 13.80%) at 30 months. Total vocabulary sizes showed the opposite pattern, with monolinguals infants being reported to understand 40.50% (*SD* = 15.989%) of the words at around 21 months, 59.83% (*SD* = 19.619%) at 25 months, and 63.30% (*SD* = 21.128%) at 30 months, and bilinguals being reported to understand 47.93% (*SD* = 23.261%) of the words at around 21 months, 64.05% (*SD* = 17.150%) at 25 months, and 75.83% (*SD* = 17.170%) at 30 months.

|  |
| --- |
| Figure 1: Participant vocabulary sizes across ages and language profiles. Vocabulary size scores are presented using two measures: L1 vocabulary size (proportion of words marked as *Understands* in the vocabulary checklist of the dominant language), and total vocabulary size (proportion of words marked as *Understands* in vocabulary checklists of both dominant and non-dominant languages). Grey lines connect vocabulary size scores from the recurrent participants. For visualisation purposes, participants were group into categorical age groups at 21, 25, 30 months, the ages around which most participants were tested. Only participants included in the final data set are shown. |

### Stimuli

We used 62 distinct words included in the BVQ to create the stimuli lists. We created six stimuli lists: three in Catalan, and three in Spanish. Each list contained 32 trials, each involving a prime-target-distractor group. Each word played a role as either prime, *or* as target and distractor across the three lists in their corresponding language. For instance, the Catalan word *cadira* appeared as *prime* in the three lists, but never as *target* or *distractor*; the Catalan word *bici* appeared as *target* and *distractor* across the three lists, but never as a prime. Target-distractor pairings were held constant across the three lists in each language. For instance, in all Catalan lists the word *bici* was paired with the word *porta*. Target-distractor pairings were also yoked, so that each member of the same target-distractor pair appeared once as target and once as distractor in each list. For instance, the *bici*-*porta* paired appeared twice in each of the three Catalan lists: once with *bici* as target and *porta* as distractor, and once with *porta* as target and *bici* as distractor. This counterbalancing avoided participants encountering looking at the target word guided solely by that word having being named in a previous trial. Finally, prime words appeared only once in each list: each target-distractor pair was associated with a different prime word in both appearances. In each list, the same prime word was presented alongside a different target-distractor pair. For instance, the Catalan prime word *barret* was presented with the *bici*-*porta* target-distractor pair in one list, with the *bici*-*porta* pair in another list, and with *berenar*-*amanida* in the remaining list. To avoid competition between target and distractor pictures, semantically related target-distractor pairs were perceptually distinct ([Arias-Trejo & Plunkett, 2009](#ref-arias2009lexical); [Floccia et al., 2020](#ref-floccia2020translation)). The order of the trials was randomised across experimental session, so that each time a participant was tested, the order in which the prime-target-distractor was presented was randomised. Each participant was randomly assigned to one of the three lists in the corresponding (dominant) language, and always the same list across their experimental sessions in the case of a recurrent participant.

In 16 of the 32 trials of the same list (henceforth *related* trials), the prime and the target words were phonologically related, sharing phonological onset (at least first phoneme). In the other 16 trials (*unrelated* trials), prime and target did not share phonological onset. 8 of the 16 *related* trials included a cognate prime (*cognate* trials), and the other 8 included a non-cognate trials (*non-cognate* trials). A prime word was considered cognate if its Catalan and Spanish translation shared phonological onset. Especial attention was paid to avoiding semantic or taxonomic relationships between prime and target words, and between prime and distractor words. Target and distractor word pairs were phonologically unrelated (did not share phonological onset). Some of them shared semantic features or a taxonomic relationship. This is the case of words associated with especially salient referents such as animals or food. To avoid infants guiding their gaze to these objects based on their saliency, we paired animals and food items together. The position of the target and distractor pictures (right or left) for each target-distractor pair was alternated, so that in one list the target would appear on the left, in another list it would appear on the right, and so on.

We examined the overall equivalence of the three trial types by comparing them across three variables relating to the target word: lexical frequency, word prevalence, animacy. **?@tbl-stimuli** shows a detailed summary of the stimuli properties, broken down by trial type and testing language. Lexical frequencies were extracted from the Catalan and Spanish corpora of the CHILDES database ([MacWhinney, 2000](#ref-macwhinney2000childes); [Sanchez et al., 2019](#ref-sanchez2019childes)) as counts per million words, and transformed into Zipf scores for easier cross-language comparison ([Van Heuven et al., 2014](#ref-van2014subtlex); [Zipf, 1945](#ref-zipf1945meaning)). We defined word prevalence as the proportion of same-aged infants who were reported to understand the word in the BVQ database.

The auditory stimuli were natural exemplars of the selected target words, spoken by Central Catalan-Spanish proficient bilingual female speaker who was instructed to pronounce each word in a toddler-directed manner. Recordings were made with an Audio-Tecnica 328 microphone (AT2050) at a sampling rate of 44100 Hz, in a soundproof room at the *Laboratori de Recerca en Infancia* at University Pompeu Fabra. We used the Audacity and Praat ([Boersma & Van Heuven, 2001](#ref-boersma2001speak)) software packages to record and edit the audio files. The speaker was presented with a list of words in Catalan. The order of the words was pseudo-randomised, and each word was produced three times in a row before moving to the next word in the list. After going through all the words in the list, the speaker went through the word list again generating three tokens for each word, now in an inverse order (from bottom of the list to the top). We then repeated the same procedure for the list of Spanish words. The resulting audios were manually chunked into individual word-forms. For each of the six tokens produced for each word, the most adequate was selected for further processing. The audios were then transformed to stereo by duplicating them into two channels, denoised, and finally normalised. The mean duration of the final audios was 1.23 (*SD* = 0.17) and 1.08 (*SD* = 0.14) seconds for the Catalan and Spanish lists.

To make the pronunciation of the words as familiar as possible to each infant, we generated additional pronunciation variants for some words in Catalan and Spanish. Catalan words involving the // phoneme in their Central Catalan variant (e.g., /) were also recorded with such phoneme replaced by /j/ (e.g., /), a phonological process common in the Metropolitan Area of Barcelona. Spanish words involving the // phoneme were also generated replacing such phoneme with // to better accommodate Latin variants of Spanish. Before every experimental session, caregivers were asked to utter three written words involving the // phoneme (in the case of participants tested in Catalan) or the // phoneme (in the case of participants tested in Spanish). Each token contained the critical phoneme at onset, inter-vocalic position, and coda. The experimenter assigned the participant to the Catalan or Spanish stimuli list involving the closest variant to that of caregivers’.

Each word was depicted by a realistic phonographic representation of a typical exemplar. Image backgrounds were removed from the original pictures using the GNU Image Manipulation Program (GIMP), resized to a rectangle of a maximum of 400 pixels hight or wide, and finally placed in the centre of a 50% grey rectangle square of 500$$13.23 cm (11.613º visual angle from participants’ perspective).

|  | Prevalence (%) | Frequency | Animacy (%) | Duration (s) |
| --- | --- | --- | --- | --- |
| Catalan | | | | |
| Cognate | 36.8 (17.6) | 3.9 (0.6) | 22.9 | 1.2 (0.2) |
| Non-cognate | 37.9 (15.9) | 3.8 (0.5) | 25.0 | 1.3 (0.2) |
| Unrelated | 35.4 (14.8) | 3.9 (0.5) | 21.9 | 1.2 (0.2) |
| Spanish | | | | |
| Cognate | 30.0 (13.4) | 4.2 (0.5) | 12.5 | 1.1 (0.1) |
| Non-cognate | 31.4 (14.0) | 4.3 (0.5) | 14.6 | 1.1 (0.1) |
| Unrelated | 28.9 (14.8) | 4.2 (0.5) | 16.7 | 1.1 (0.1) |

**?(caption)**

### Procedure

Testing took place in a sound-proof room. Participants sat on their caregivers’ lap in a dimly lit testing booth while the experimenter conducted the experiment from outside. Caregivers were instructed to keep their eyes shut (to avoid recording their gaze, instead of the participant’s), to be still, and to avoid interacting with the participant verbally or non-verbally. Participants sat at approximately 65 cm from the eye-tracker and a XX-in screen of screen resolution. We used a custom Matlab XXXX script using the PsychToolbox-3 extension ([Brainard & Vision, 1997](#ref-brainard1997psychophysics); [Kleiner et al., 2007](#ref-kleiner2007s); [Pelli & Vision, 1997](#ref-pelli1997videotoolbox)) to present the stimuli, and the Tobii Analytics SDK 3.0 to interact with the eye-tracking while the experiment was running. Sampling rate was set at 120 Hz. A 5-point calibration was performed before every experimental session, in which the picture of a colourful beach ball was presented. We set a 55% grey background for the calibration and stimuli presentation. Auditory stimuli were presented through two loudspeakers located behind the screen, one to each side. The experimenter monitored the experimental from outside the room using a centrally located video camera place above the screen. After a successful calibration the experimenter triggered the onset of the first trial. Trials were presented uninterruptedly and without intervention of the experimenter until the 32 trials were presented, or the experimental session had to be stopped because of the participant’s behaviour.

|  |
| --- |
| Experimental task design with examples in Catalan. In each trial, the prime image is presented in silence for 1,500 ms. The the auditory target label is presented, and finally the target and distractor pictures are presented side-by-side for 2,000 ms. In cognate trials (*n* = 8), Catalan *and* Spanish prime labels shared phonological onset with the target label. In non-cognate trials (*n* = 8), only the Catalan prime label shared phonological onset with the target label. In unrelated trials (*n* = 32), none of the prime labels shared phonological onset with the target label. |

Each trial started with the presentation of an attention getter for 3,000 milliseconds. Then, the prime picture was presented in silence in the centre of the screen for 1,500 milliseconds. Fifty milliseconds after the offset of the prime image, an auditory label was played from the loudspeakers and, 700 milliseconds after the onset of the auditory label, the target and distractor pictures were presented side-by-side during 1,000 milliseconds until the end of the trial. After this, the attention getter of the next trial was immediately presented. Each experimental session took approximately 10 minutes.

### Data analysis

#### Data processing

We defined a time windows of interest from 200 ms after target and distractor pictures onset until the end of the trial at 2,000 ms when both pictures disappeared from screen. To avoid modelling fixations driven by processes other than auditory word recognition ([Fernald et al., 1998](#ref-fernald1998rapid), [2001](#ref-fernald2001half)). Missing eye-tracker samples were interpolated using the last-observation-carried-forward (see [Zettersten et al., 2022](#ref-zettersten2022peekbank) for a similar approach), with a maximum of 20 maximum consecutive missing samples being interpolated (an equivalent of 166.67 ms).

We gathered data from 8,608 trials from 269 testing sessions, generated from 162 distinct participants. We excluded trials in which participants failed to provide 50% valid eye-tracking samples (equivalent to 750 ms) during the prime phase (*n* = 1,815) or 50% valid samples (equivalent to 1,000 ms) during the target-distractor phase (*n* = 1,262). We also excluded trials in which participants did not provide at least 10% of valid samples (equivalent to 100 ms) of for the target *and* the distractor each (*n* = 2,461) (see [Floccia et al., 2020](#ref-floccia2020translation); [**manidurrantfloccia?**](#ref-manidurrantfloccia) for a similar approach).

After trials that matched any of the aforementioned exclusion criteria from the dataset, we excluded participants who did not provide at least two valid trials in the *Cognate* condition, the *Non-cognate* condition, or the *Unrelated* condition (*n* = 32), and participants with a L1 vocabulary size lower than 10% (*n* = 3). The final dataset included 5,019 trials from 237 testing sessions, generated by 150 distinct participants. Of those participants, 82 provided data from one experimental session, 49 provided data from two experimental sessions, and 19 provided data from three experimental sessions. **?@tbl-attrition-trials** shows a detailed description of the trial attrition.

|  |  | Trials | | |
| --- | --- | --- | --- | --- |
| *Participants* | *Cognate* | *Non-cognate* | *Unrelated* |
| Bilingual | 104 (17) | 596 (372) | 555 (413) | 1,140 (796) |
| Monolingual | 133 (15) | 741 (443) | 747 (437) | 1,506 (862) |
| *N* | 237 | 1,337 | 1,302 | 2,646 |

**?(caption)**

#### Modelling approach

We used Bayesian Hierarchical General Additive Models (HGAMs) to model our data and test our hypotheses ([Pedersen et al., 2019](#ref-pedersen2019hierarchical)). First, we defined our response variable as the empirical logit of target looking using [Equation 1](#eq-elogit) across all eye-tracking samples for each each participant in each condition ([Agresti, 2012](#ref-agresti2012categorical); [Barr, 2008](#ref-barr2008analyzing)). We used a Gaussian distribution to model our response variable. To test our hypotheses, we then included *Condition*, *Group*, and either *Age* (), *L1 vocabulary* (), or *Total vocabulary* () as fixed effects in the model. We also added all two-way and three-way interactions between the predictors. We set two *a priori* contrasts for the *Condition* predictor: one comparing *Unrelated* and *Related/Non-cognate* trials (sum-coded as -0.5 and +0.5, with *Related/Cognate* trials coded as 0), and another comparing *Related/Non-cognate* and *Related/Cognate* trials (sum-coded as -0.5 and +0.5, with *Unrelated* trials coded as 0). We also set two *a priori* contrasts for the *Group* predictor: one comparing *Monolingual (ENG)* with *Monolingual (CAT/SPA)* participants (sum-coded as -0.5 and +0.5, with *Bilingual (CAT-SPA)* participants coded as 0), and another comparing *Monolingual (ENG)* with *Bilingual (CAT-SPA)* participants (sum-coded as -0.5 and +0.5, with *Monolingual (CAT/SPA)* participants coded as 0). Before entering the model, the numeric predictors *Age*, *L1 vocabulary*, and *Total vocabulary* were standardised by subtracting from each observation the mean of the predictor, and dividing the result from the standard deviation of the predictor.

We included *Session*—which indexes individual testing sessions that may belong to the same participant—as grouping variable, nested within *Participant* grouping variable—which indexes a distinct participant. This nested random effects structure incorporates the longitudinal design of data collection, in which multiple participants were tested more than once at different ages. We added by by-session intercepts and *Condition* slopes, and by-participant intercepts and *Age* slopes. To model the time course of target looking across time bins, we included B-splines for the main effect of *Time*, and one for an adjustment of the *Time* cubic spline by *Group* and *Relatedness* ([Wood, 2017](#ref-wood2017generalized)). For both splines, we specified basis functions or *knots*. [Equation 2](#eq-model) shows a formal implementation of the model.

The comparison between the three models including *Age*, *L1 vocabulary*, *Total vocabulary* as predictors was done using leave-one-out cross-validation (LOO-CV) as a benchmark of model performance, using Pareto-smoothed importance sampling (PSIS) to approximate it ([Vehtari et al., 2017](#ref-vehtari2017practical)). We implemented the model using brms ([Bürkner, 2017](#ref-burkner2017brms)), an R interface to the Stan probabilistic language (2.33.0) ([Carpenter et al., 2017](#ref-carpenter2017stan)). We ran four iteration chains using the by-default No U-Turn Sampler algorithm with 500 iterations each and an additional 500 warm-up iterations per chain.

#### Statistical inference

We assessed the practical relevance of the estimated regression coefficients of the model following Kruschke & Liddell ([2018](#ref-kruschke2018bayesian)). First, we specified a region of practical equivalence (ROPE) from NA to +NA, in the probability scale. This region indicates the range of values that we considered equivalent to zero. We then summarised the posterior distribution of each regression coefficient with the 95% highest density interval (HDI). This interval contains the true value of this coefficient with 95% probability, given the data. Finally, we calculated the proportion of posterior samples in the 95% HDI that fell into the ROPE, noted as , which indicates the probability that the true value of the regression coefficient falls into the ROPE (and therefore should be considered equivalent to zero). For example, indicates that, given our data, there is a 80% probability that the true value of the coefficient falls within the ROPE, and can therefore be considered equivalent to zero.

We followed a similar procedure to compare the effect of the predictors of interest on the time course of target looking. First, we defined a ROPE from NA to +NA in the logit scale. We then computed the difference between of the regression coefficient of the continuous predictor *Age*, the difference coefficients *Group*, *Condition*, and *Age* at each time bin, which indicates the marginal effect of the predictor on target looking at that specific

## Results

### Model comparisons

|  | *ELPD* | *SE ELPD* | *ELPD (diff.)* | *SE ELPD (diff)* |
| --- | --- | --- | --- | --- |
| Base models | | | | |
| Model 2 | -20,592.69 | 135.90 | --- | --- |
| Model 3 | -20,593.64 | 135.82 | -0.95 | 0.96 |
| Model 1 | -20,786.74 | 133.47 | -194.05 | 31.91 |
| Model 0 | -20,787.48 | 133.68 | -194.79 | 31.93 |
| Age and vocabulary size models | | | | |
| Model 5 | -20,593.51 | 135.84 | --- | --- |
| Model 3 | -20,593.64 | 135.82 | -0.13 | 2.12 |
| Model 6 | -20,593.73 | 135.86 | -0.22 | 2.24 |
| Model 4 | -20,594.95 | 135.83 | -1.44 | 2.15 |
| \*ELPD\*: sum of expected log pointwise predictive density for a new dataset. \*SE \(ELPD\)\*: standard error of the \*ELPD\*, which indictes the uncertainty about the predictive performance for unknown future data. \*ELPD (diff)\*: pairwise difference in \*ELPD\* for two models. The difference is computed relative to the model with highest \*ELPD\*. \*SE ELPD (diff.)\*: standard error of component-wise differences of \*ELPD\* between two models. | | | | |

**?(caption)**

**?@tbl-loos** summarises the outputs of the LOO-CV model comparisons. Models and , showed substantially better predicted performance than models and , which indicates that adding the *Condition* predictor improves significantly the fit of the models. Both and models showed equivalent performance, suggesting that adding the two-way interaction between *Group*, and *Condition*, did not improve the the fit of the model substantially. We now report a description of the posterior distribution of the regression coefficients of model . We now report the median and 95% highest density interval of each fixed regression coefficient in (see **?@tbl-coefs**).

### Linear effects

**?(caption)**

Overall, the average participants’ looking time exceeded chance levels, as indicated by the fact that the 95% HDI of the intercept term excluded zero ( = 0.217, 95% HDI = [0.172, 0.256]), and all of its posterior samples were outside of the ROPE. The coefficient of *Age* had a positive sign, but its 95% HDI overlapped completely with the ROPE ( = , 95% HDI = [-0.017, 0.057]), indicating that participants from all ages showed equivalent overall target word recognition. The 95% HDI of the coefficient of *Group* also included zero ( = -0.011, 95% HDI = [-0.090, 0.075]) and completely overlapped with the ROPE, indicating an equivalent overall target preference in monolinguals and bilinguals. The 95% HDI of the first contrast of the *Condition* predictor—comparing *Unrelated* and *Related/Non-cognate* trials—included zero ( = 0.055, 95% HDI = [-0.027, 0.148]), and 72.75% of its posterior samples overlapped with the ROPE. The 95% HDI of the second contrast, comparing *Related/Non-cognate* and *Related/Cognate* trials, also included zero ( = -0.016, 95% HDI = [-0.123, 0.083]), and 88.69% of its posterior samples overlapped with the ROPE. The overall target preference was equivalent across both pairwise condition comparisons. The interaction term between the first *Condition* contrast contained zero ( = -0.016, 95% HDI = [-0.183, 0.156]), with 58.99% of its posterior samples overlapping with the ROPE. The interaction term between the second *Condition* contrast also contained zero ( = 0.091, 95% HDI = [-0.107, 0.301]), and 72.75% of its posterior samples fell within the ROPE. The outcomes of this model provide strong to moderate evidence against differences in overall target looking time across conditions and across monolinguals and bilinguals.

A comparison between models including *Age* (), *L1 vocabulary* (), and *Total vocabulary* ()—and their three-way and two-way interactions with *Group* and *Condition*—against model , which only included *Age* as a covariate is shown in **?@tbl-loos**. Overall, all models performed equivalently, with the simplest model showing slightly better performance. This suggests that participants’ target looking during the test phase can be predicted with relative accuracy without taking into account *L1 vocabulary*, or *Total vocabulary* sizes.

Both younger and older participants showed equivalent looking time across conditions, as suggested by the fact that the 95% HDI of the two-way interaction between *Age* and the first *Condition* contrast included zero ( = , 95% HDI = ), as well as the interaction between *Age* and the second *Condition* contrast ( = , 95% HDI = ). This pattern of results was found in both monolinguals and bilinguals, as the 95% HDI of the three-way interaction between *Age*, *Group*, and *Condition* included zero both the first ( = , 95% HDI = ) and second ( = , 95% HDI = ) contrast of *Condition*.

### Time course of target looking

An analysis of the time course of target looking revealed a similar pattern of results (see [Figure 2](#fig-epreds)). Posterior mean prediction for the three conditions overlap across the full time course of the trial in both language groups, and also in both extrema of the age range of participants (20 and 32 months).

|  |
| --- |
| Figure 2: Margin posterior mean predictions. A) Posterior mean predictions of the time course of target fixation in the test phase. B) Posterior mean prediction of the time course of the differences in target looking time between conditions. Lines and intervals indicate the mean and 95% credible interval of the posterior predictions. |

In an exploratory analysis, we compared the model-estimated time points at which participants’ gaze behaviour showed evidence of target looking above chance level (50%). We defined the recognition time of the trial as the earliest time point at which the lower bound of the 95% HDI posterior mean prediction of the target looking excluded zero. As expected, the recognition times were on average shorter for older infants than for younger infants. Differences between conditions were larger in bilinguals than in monolinguals, especially in the older group. The recognition time for monolingual 20-month-olds was 687.88 ms in the *Unrelated* condition, 639.39 ms in the *Related/Non-cognate* condition, and 639.39 ms in the *Related/Cognate* condition. In monolingual 30 month-olds, the recognition point was 639.39 ms in the *Unrelated* condition, 574.75 ms in the *Related/Non-cognate* condition, and 607.07 ms in the *Related/Cognate* condition. In bilingual 20 month-olds, the recognition point was 687.88 ms in the *Unrelated* condition, 639.39 ms in the *Related/Non-cognate* condition, and 639.39 ms in the *Related/Cognate* condition. In bilingual 30 month-olds, the recognition point was 687.88 ms in the *Unrelated* condition, 639.39 ms in the *Related/Non-cognate* condition, and 639.39 ms in the *Related/Cognate* condition.

|  |
| --- |
| Figure 3: Model-predicted recognition times for monlingual and bilinguals at 20 and 30 months of age. |

## Discussion

In Study 1, we adapted an implicit naming paradigm to investigate the potential role of within- and cross-language phonological priming in 21- to 32-month-old children. Both the overall target looking and its course time across the trial suggested strong word recognition, regardless of priming condition, participant language profile and age. Overall, participants’ performance was equivalent across all priming conditions. An exploratory analyses on word recognition times suggested a complex pattern of results, showing an age effect—in which older participants looked at the target above chance earlier than younger infants—and a weak interaction between priming conditions, age, and language group. These results conflict with previous findings in several ways.

First, previous investigations using implicit naming paradigm had provided strong evidence in favour of infants and toddlers generating internal labels when presented with pictures of familiar objects. Mani & Plunkett ([2010](#ref-mani2010infant)) and Mani & Plunkett ([2011](#ref-mani2011phonological)) reported priming effects in 18 and 24-month-old infants. In primed trials, participants’ target looking preference was influenced by the previous presentation of a phonologically related referent in silence. In control trials—in which prime and target labels were unrelated—no priming effect was found. This suggested that infants implicitly named the prime label, and that the resulting internal phonological form was detailed enough to interact with the subsequent auditory recognition of the target word. In the present study, we did not find evidence in favour of differences in target looking preference between related and unrelated trials, regardless of the age of participants or their language profile.

Both Mani & Plunkett ([2010](#ref-mani2010infant)) and Mani & Plunkett ([2011](#ref-mani2011phonological)), and the present study share a similar design, but sampled different children populations. Participants in Mani and Plunkett’s studies were raised in English speaking households in Oxford, while participants in the present study were learning Catalan or Spanish, if not a combination of both. We identify two ways in which this might have influenced the results. One is that the stimuli lists in Catalan and Spanish might have differed in some critical dimension, which interfered infants’ ability to generate implicit labels for prime pictures. For example, words in the present study were predominantly bi-syllabic, while those from Mani and Plunkett’s studies were mostly monosyllabic, and contained less phonemes on average. Longer word-forms might have been more difficult to retrieve in time for a prime effect to take place in target recognition. Another possibility is that exposure to a second language might have delayed the developmental trajectories of word production in bilinguals, in such way implicit naming emerges later ages. Given that 51.02% of monolinguals in Barcelona were also exposed to a second language, they might as well been affected by such delay in implicit naming trajectories.

The results from the Study 1 also conflict with previous reports of cross-language priming in toddlers. Particularly, Von Holzen & Mani ([2012](#ref-von2012language)) and Floccia et al. ([2020](#ref-floccia2020translation)) had provided evidence of phonological priming through translation. One critical different between both studies and the present one is the silent or auditory presentation of the prime labels. In priming through translation studies, the prime is auditorily presented to participants, and cross-language priming effects take place through the subsequent activation of the prime translation. In the present study, the prime label was not named, and participants were expected to generate both the prime label and its translation implicitly. It may be the case that infants in the present study faced a more challenging task, having to internally retrieve phonological forms without any auditory prompt like the one provided by priming through translation studies.

Finally, another possibility is that participants in the present study did in fact general prime labels, but such labels did not interfere over target recognition. This could be because they were not phonologically specified enough. This is unlikely, as previous studies have shown strong evidence of fine phonological representations in the early lexicon REFERENCE. Instead, it could be the case that the prime labels were activated, but phonological links between lexical representations are still weak or not present, so that priming effects are not possible until later in age. To our knowledge, this study is the first one to test bilingual children at this age on implicit naming paradigms, so evidence of priming is not present at the moment.

We found no differences between models including participants’ age or any of the two indices of vocabulary size (L1 and total vocabulary size) as predictors. Previous studies suggest that same-language vocabulary size predicts better bilingual infants’ performance in word recognition tasks, as compared to cross-language vocabulary size ([Marchman et al., 2010](#ref-marchman2010vocabulary)).

In summary, the possibly increased difficulty of the task induced by the properties of the stimuli, together with differences in task design and populations might have contributed to the null pattern of results shown in Study 1. Given the complexity of the picture drawn by these factors, we conducted a second study with the aim of replicating the original findings by Mani & Plunkett ([2010](#ref-mani2010infant)) and Mani & Plunkett ([2011](#ref-mani2011phonological)). If phonological priming effects are observed in a same-aged population using the adaptation of the task from Study 1, this would provide evidence in favour of cross-language differences in the developmental trajectories of implicit naming. If, on the contrary, a similar patter of null results is found, this would suggest either the existence of finer differences in task implementation compared to the original studies, or point to the effect of interest being less robust than suggested by the literature.

# Study 2

In this study, we conducted a conceptual replication of Mani & Plunkett ([2010](#ref-mani2010infant)) and Mani & Plunkett ([2011](#ref-mani2011phonological)). Using the same task from Study 1, we tested a same-aged group of monolinguals being exclusively raised in British English speaking homes, the same population tested in the original studies. As in Study 1, we manipulated the phonological relationship between the prime and target labels at word onset, as well as the cognate status of the prime label. Evidence from implicit naming of prime images would be revealed by differences in target looking or its time course across the trials between *Unrelated* trials, and *Related/Non-cognate* and *Related/Cognate* trials. Given that participants in this study are monolinguals—and contrary to the monolinguals in Barcelona, have received no exposure to any other language—we would not expect the cognate status of the prime labels to exert any effect on subsequent target recognition.

## Methods

### Participants

We collected data from 112 children (41 female, 68 male, 3 sex not reported, age: *Mean* = 26.36 months, *SD* = 4.01, *Range* = [20.03-32.5]), living in the Oxfordshire area (United Kingdom), tested at the Oxford BabyLab at the University of Oxford. Families were recruited from maternity rooms in private hospitals and social media, and contacted via phone when the child’s age spanned between 20 and 32 months. From the 112 children that participated, 97 participated once, and 15 participated twice. Recurrent participants were tested with at least XXX months of difference. We gathered a total of 127 testing sessions. All participants were being raised in exclusively British English monolingual homes.

### Vocabulary size

We collected vocabulary data using parental responses to the Oxford Communicative Development Inventory (OCDI) ([Hamilton et al., 2000](#ref-hamilton2000infant)). Families were sent the questionnaire immediately after each experimental session, and were given two weeks to fill it. Since all vocabulary sizes of English monolingual infants were only assessed in English, both *L1 vocabulary* and *total vocabulary* sizes were identical (see [Figure 4](#fig-vocabulary-oxf)).

|  |
| --- |
| Figure 4: Participant vocabulary sizes across ages and language profiles. Vocabulary size scores are presented using two measures: L1 vocabulary size (proportion of words marked as *Understands* in the vocabulary checklist of the dominant language), and total vocabulary size (proportion of words marked as *Understands* in vocabulary checklists of both dominant and non-dominant languages). Vocabulary size scores from the recurrent participants are shown linked. |

### Stimuli

New stimuli lists were generated to adapt the experimental procedure to English-learning infants. Stimuli lists were created following the same constraints as in Study 1. Auditory stimuli were recorded following the same procedure as in Study 1, produced by Southern British English native female speaker in a toddler-directed manner. Audio and picture processing were performed following the same workflow as in Study 1.

### Procedure and apparatus

Same as in Study 1. The study was run using a custom Matlab script, PresentMate, based on the PsychToolbox-3 extension ([Brainard & Vision, 1997](#ref-brainard1997psychophysics); [Kleiner et al., 2007](#ref-kleiner2007s); [Pelli & Vision, 1997](#ref-pelli1997videotoolbox)). Visual fixations were recorded using a Tobii TX 300 eye-tracker and a 23-in screen of 1920 1080 resolution. The Tobii Analytics SDK 3.0 was used to interact with the eye-tracking while the experiment was running. Sampling rate was set at 120 Hz. A 9-point calibration was performed before every experimental session, in which the picture of a colourful beach ball was presented.

### Data analysis

Same as in Study 1, but removing the *Group* predictor from the models, as only monolingual infants were tested.

We defined a time windows of interest from 200 ms after target and distractor pictures onset until the end of the trial at 2,000 ms when both pictures disappeared from screen. To avoid modelling fixations driven by processes other than auditory word recognition ([Fernald et al., 1998](#ref-fernald1998rapid), [2001](#ref-fernald2001half)). Missing eye-tracker samples were interpolated using the last-observation-carried-forward (see [Zettersten et al., 2022](#ref-zettersten2022peekbank) for a similar approach), with a maximum of 20 maximum consecutive missing samples being interpolated (an equivalent of 166.67 ms).

We gathered data from 3,484 trials from 110 testing sessions, generated from 97 distinct participants. We excluded trials in which participants failed to provide 50% valid eye-tracking samples (equivalent to 750 ms) during the prime phase (*n* = 829) or 50% valid samples (equivalent to 1,000 ms) during the target-distractor phase (*n* = 650). We also excluded trials in which participants did not provide at least 5% of valid samples (equivalent to 100 ms) of target or distractor looking in the test phase (*n* = 1,003) (see [Floccia et al., 2020](#ref-floccia2020translation) for a similar approach).

After trials that matched any of the aforementioned exclusion criteria from the dataset, we excluded participants who did not provide at least two valid trials in the *Cognate* condition, the *Non-cognate* condition, or the *Unrelated* condition (*n* = NA), and participants with a L1 vocabulary size lower than 10% (*n* = 2). The final dataset included 1,861 trials from NA testing sessions, generated by 80 distinct participants. Of those participants, 69 provided data from one experimental session, 10 provided data from two experimental sessions, and NA provided data from three experimental sessions. From the trials included in the final dataset dataset, 915 were *Unrelated* trials (502 previously excluded), were 468 *Related/Non-cognate* trials(241 previously excluded), and 478 were *Related/Cognate* trials (229 previously excluded).

## Results

|  | *ELPD* | *SE ELPD* | *ELPD (diff.)* | *SE ELPD (diff)* |
| --- | --- | --- | --- | --- |
| Base models | | | | |
| Model 1 | -8,431.76 | 73.79 | --- | --- |
| Model 0 | -8,492.39 | 72.48 | -60.63 | 15.68 |
| Age and vocabulary size models | | | | |
| Model 3 | -8,431.72 | 73.85 | --- | --- |
| Model 1 | -8,431.76 | 73.79 | -0.04 | 2.36 |
| Model 2 | -8,432.66 | 73.73 | -0.94 | 1.91 |
| \*ELPD\*: sum of expected log pointwise predictive density for a new dataset. \*SE \(ELPD\)\*: standard error of the \*ELPD\*, which indictes the uncertainty about the predictive performance for unknown future data. \*ELPD (diff)\*: pairwise difference in \*ELPD\* for two models. The difference is computed relative to the model with highest \*ELPD\*. \*SE ELPD (diff.)\*: standard error of component-wise differences of \*ELPD\* between two models. | | | | |

**?(caption)**

In line with the results in *Study 1*, including the *Condition* predictor increased the predictive performance of the model substantially, compared to a model including only the main effect of *Age* (see **?@tbl-loos-oxf**). We now report the median and 95% highest density interval of each fixed regression coefficient in . Overall, the average participant’ looking time exceeded chance levels, as indicated by the fact that the 95% HDI of the intercept term excluded zero ( = 0.217, 95% HDI = [0.172, 0.256]) and all of its posterior samples fell outside of the ROPE. The 95% HDI of the coefficient of *Age* had a positive sign, but did not exclude zero ( = 0.022, 95% HDI = [-0.017, 0.057]), and overlapped completely with the ROPE, indicating that participants from all ages showed equivalent overall target word recognition. The 95% HDI of the first contrast of the *Condition* predictor—comparing *Unrelated* and *Related/Non-cognate* trials—included zero ( = 0.055, 95% HDI = [-0.027, 0.148]), and 72.75% of its posterior samples fell within the ROPE. The 95% HDI of the second contrast—comparing *Related/Non-cognate* and *Related/Cognate* trials—also included zero ( = -0.016, 95% HDI = [-0.123, 0.083]), and 88.69% of its posterior samples fell within the ROPE. The overall target preference was equivalent across both pairwise condition comparisons.

Margin posterior mean predictions. A) Posterior mean predictions of the time course of target fixation in the test phase. B) Posterior mean prediction of the time course of the differences in target looking time between conditions. Lines and intervals indicate the mean and 95% credible interval of the posterior predictions.

**?(caption)**

**?@tbl-loos-oxf** shows a comparison between models including *Age* (), and *L1 vocabulary* ()—and their two-way and two-way interaction with *Condition*—against model , which only included *Age* as a covariate suggested that all models performed equivalently, with model showing a slightly better performance. Overall, this suggests that participants’ target looking during the test phase can be predicted with relative accuracy without taking into account *L1 vocabulary*, or *Total vocabulary* sizes.

An analysis of the time course of target looking revealed a similar pattern of results (see [Figure 2](#fig-epreds)). Posterior mean prediction for the three conditions overlap across the full time course of the trial in both language groups, and also in both extrema of the age range of participants (20 and 32 months).

|  |
| --- |
| Figure 5: A) Posterior mean predictions of the time course of target fixation in the test phase. Thin lines represent 100 individual posterior predictions. Thick lines indicate the mean of the posterior predictions. B) Posterior mean prediction of the time course of the differences in target looking time between conditions. Thin lines represent 100 individual posterior predictions. Black lines indicate the mean of the posterior predictions. |

In an exploratory analysis, we compared the model-estimated time points at which participants’ gaze behaviour showed evidence of target looking above chance level (50%). We defined the recognition time of the trial as the earliest time point at which the lower bound of the 95% HDI posterior mean prediction of the target looking excluded zero. As expected, the recognition times were on average shorter for older infants than for younger infants. Differences between conditions were larger in bilinguals than in monolinguals, especially in the older group. The recognition time for monolingual 20-month-olds was ms in the *Unrelated* condition, ms in the *Related/Non-cognate* condition, and ms in the *Related/Cognate* condition. In monolingual 30 month-olds, the recognition point was ms in the *Unrelated* condition, ms in the *Related/Non-cognate* condition, and ms in the *Related/Cognate* condition. In bilingual 20 month-olds, the recognition point was ms in the *Unrelated* condition, ms in the *Related/Non-cognate* condition, and ms in the *Related/Cognate* condition. In bilingual 30 month-olds, the recognition point was ms in the *Unrelated* condition, ms in the *Related/Non-cognate* condition, and ms in the *Related/Cognate* condition.

|  |
| --- |
| Figure 6: Model-predicted recognition times for monlingual and bilinguals at 20 and 30 months of age. |

## Discussion

Study 2 was aimed at investigating whether the phonological priming effects previously reported Mani & Plunkett ([2010](#ref-mani2010infant)) and Mani & Plunkett ([2011](#ref-mani2011phonological)) could be found using the adaptation of the implicit naming task form Study 1 in a sample with equivalent demographic characteristics to the original studies. Results from Study 2 did not provide any evidence in favour of phonological priming. English monolingual participants from all ages showed an equivalent pattern of target looking across both related and unrelated trials.

The absence of a phonological priming effect suggest that either English monolinguals did not generate implicit labels for the prime pictures presented in silence, or that if generated, such labels did not interact with the subsequent recognition of the target word. Both candidate explanations conflict with both Mani and Plunkett’s studies, and also with previous studies suggesting that infants 12-months and older already generate internal labels when presented with pictures of familiar objects.

One of the main differences between the implicit naming paradigm used by Mani & Plunkett ([2010](#ref-mani2010infant)) and the one in Studies 1 and 2 is the fact that target labels were auditorily presented before target and distractor pictures were presented on screen. Given that priming effects are very sensitive to the timing of the activation of the lexical presentation ([**REFERENCES?**](#ref-REFERENCES)), we decided to shorten the time span between prime image naming and target label presentation by removing the pre-naming interval of the test phase, in which both target and distractor pictures are presented form a baseline time interval before the target label is auditorily presented. By presenting the auditory target label immediately after the offset of the prime picture, the internally generated prime phonological form would be more likely to affect the activation of the target lexical representation. It is possible that this difference in task design might have led to the contrasting outcomes of Mani and Plunkett and the present study. Given that participants were not presented with the target and distractor pictures until after the target label was auditorily presented, one possibility is that participants faced the additional task of exploring both pictures to find the target referent, which infants in Mani & Plunkett ([2010](#ref-mani2010infant)) did not have to do, because of the pre-naming presentation of target and distractor in the test phase. This visual exploration of the screen might have increased the cognitive load of the task, rendering the potential effect of phonological priming to a weaker, undetectable effect.

Overall, the results from Study 2 support a more methodological explanation of the null pattern of results in Study 1, in which the lack of a phonological priming effect—both within and across languages—might be explained by differences in task design or replicability concerns.

# General discussion

In this paper, we investigated the developmental trajectories of cross-language co-activation in the developing lexicon. We tested a longitudinal an cross-sectional sample of 20-to-32 month olds in an implicit naming paradigm, in which we designed three trial types to manipulate the phonological overlap between the prime and target words within and across languages. In *Unrelated* trials, prime and target were phonologically unrelated in both languages. In *Related/Non-cognate* trials, prime and target labels shared phonological onset only in the dominant language of participants, in which they were tested. In *Related/Cognate* trials, the prime label was a cognate: prime and target labels shared phonological onset in both languages. In Study 1, we tested a cohort oh monolingual and bilingual infants learning Catalan, Spanish, or both, and found no phonological priming effect in monolinguals or bilinguals. In Study 2, we tested a cohort of English monolinguals, and found similar results. To our knowledge this is the first study investigating cross-language priming using an implicit naming paradigm. Such paradigm presents a critical asset for studying bilingual lexical processing. Since the prime pictures are presented in silence, the task is exclusively complete in the dominant language of participants. This provides a monolingual situation in which differences in performance across experimental conditions that involve cross-linguistic interactions can be more easily associated to the experimental manipulation in a causal way ([Grosjean, 1997](#ref-grosjean1997bilingual)). This contrast to other experimental paradigms in which participants are presented with—need to produce—items from both languages. Such covert manipulation of the similarity between stimuli from the two languages has allowed previous studies to provide strong support for the language-non selective account of lexical access e.g., ([Costa et al., 2000](#ref-costa2000cognate); [Thierry & Wu, 2007](#ref-thierry2007brain); [Von Holzen & Mani, 2014](#ref-von2014bilinguals)).

Previous studies have provided evidence of English monolingual infants 18-month-old and older being sensitive to phonological priming effects during auditory word recognition. Mani & Plunkett ([2010](#ref-mani2010infant)) had shown a facilitation effect in word recognition when 18-month-olds were previously presented with a familiar picture whose label shared phonological onset with the target word. Mani & Plunkett ([2011](#ref-mani2011phonological)) extended these results to 21-month-olds, showing an effect in the opposite direction: phonologically related primes interfered with target word recognition. The shift from facilitation to inhibition was interpreted as a shift from sub-lexical facilitation at the phonological level—in which target word recognition is facilitated by the prior activation of an overlapping set of phonological segments—to a lexical interference effect in which the activation of the lexical representation of the phonologically similar prime led to competition dynamics with the subsequently activated representation of the target word. The last conclusion was supported by the fact that the size of the interference effect was associated with the increase in participants’ vocabulary size, and also with the cohort size of the prime label—primes belonging to a larger neighbourhood of phonologically related words were subject to stronger interference effects during the task

Even earlier evidence of implicit naming was provided through ERP registering at 14 months of age. Duta et al. ([2012](#ref-duta2012erp)) presented name-known pictures in silence for one second. After one second, the picture’s label was presented auditorily, and the infants’ N400 component was registered. In some trials the auditory label corresponded to a canonical realisation of the word-form. In other trials, the word was mispronounced. Infants ERP signal differed when presented with correct pronunciations or mispronunciations, indicating that infants had generated expectations at the phonological level about the auditory labels of the picture before they were presented with its auditory label. A later study by Styles et al. ([2015](#ref-styles2015infant)) re-analysed the first 1,000 ms of the trials in the previous study (while the picture was being presented before the auditory label was presented), and found that the ERPs of name-known pictures (according to parental reports of receptive vocabulary) differed from those of name-unknown pictures (which infants would not be able to lexicalise). This provides further support to the account that at around 14 months of age, infants already name familiar pictures, even if presented in silence. In bilinguals, only one study had used an implicit naming paradigm to study cross-language priming. Von Holzen & Mani ([2014](#ref-von2014bilinguals)) adapted Mani & Plunkett ([2010](#ref-mani2010infant)) procedure to test German-English bilinguals. The authors manipulated the phonological overlap between the prime label in English (presented in silence) and a German target (auditorily presented) in both German and English. Participants’ ERPs revealed differences between trials in which the German prime label overlapped phonologically with the English target label, compared to trials in which none of the labels overlapped. These results suggest that adult bilinguals implicitly named prime pictures, and that the resulting label interacted with subsequent word recognition.

The results in the present investigation are in strong contrast with the aforementioned studies. Both studies 1 and 2 conflict with implicit naming.

Study 1 conflicts with cross-language studies. singh2014,

Kuipers & Thierry ([2015](#ref-kuipers2015bilingualism)) compared ERP responses to picture-word pairs in English monolinguals and Welsh-English bilinguals where the words were either semantically related or unrelated to a pictured object. Although both studies tested 30-month-olds, the conclusions the authors reached do not fully agree. Singh ([2014](#ref-singh2014one)) concluded that there is a discontinuity between adults and young children in the way the two lexicons interact, while Kuipers & Thierry ([2010](#ref-kuipers2010event)) and Kuipers & Thierry ([2015](#ref-kuipers2015bilingualism)) concluded that there are no differences in adults’ and toddlers’ lexical-semantic processing.

The present study presents methodological contributions. First, the longitudinal design of the data collection provided a more suitable design for drawing conclusions about developmental change in word recognition trajectories, compared to studies that tested participants from a single age group ([Floccia et al., 2020](#ref-floccia2020translation)), or studies that with an exclusively cross-sectional design ([Von Holzen & Mani, 2012](#ref-von2012language)). Second, the modelling approach in the present study incorporates the double source of repeated measures simultaneously into the random effects structure of the multilevel model. The models in Studies 1 and 2 accounts for the time course of target looking across the trials and its associated autocorrelation using Hierarchical General Additive Models ([Barr, 2008](#ref-barr2008analyzing); [Pedersen et al., 2019](#ref-pedersen2019hierarchical)). The models also incorporate the observations from the same testing session as a nested grouping variable inside participants, so that the test session-level parameters in each model are estimated taking into account the estimations of other testing sessions from the same participants. Overall, this approach allowed us to simultaneously model the participant-level variability, the testing session-level variability, and the time series consisting of participants’ gaze behaviour during the task. Finally, the adoption of a Bayesian approach to implement and estimate the model allowed the incorporation of prior knowledge about the distribution of the parameters to generate stable estimates, despite the complexity of the model.

Why. Auditory presentation, implicit label generation but no priming

rhyme vs onset

word length

Another possibility is that the current study lacked the appropriate statistical power to detect the effect sizes of interest. Reported differences in looking time associated to priming paradigms in the available literature in children are generally small. Despite the relatively larger sample size achieved in the present investigation, participant-level trajectories of target looking are highly variable, which might have increased the uncertainty of the models’ estimates.

The findings in the present investigation have important implications for current models of bilingual lexical development. For instance, recent accounts of bilingual words acquisition pointed to a facilitative role of phonological similarity between the two languages. A monograph by Floccia et al. ([2018](#ref-floccia2018vocabulary)) provided evidence of a language distance effect, where bilinguals learning two languages sharing higher lexical similarity knew more words in their less dominant language than those learning two language with lower lexical similarity. Later work by ([**mitchell2022cognate?**](#ref-mitchell2022cognate)), Siow et al. ([2022](#ref-siow2022effect)), and ([**garcia-castro2023cognate?**](#ref-garcia-castro2023cognate)) provided evidence for a more direct facilitation effect of cognateness. Cognates are form-similar translation equivalents. Cognateness is suggested to facilitate word acquisition through parallel activation: the co-activation of translation equivalents in speech might lead to an faster accumulation of learning instances with words, which ultimately results in an earlier age of acquisition. In the present study, we failed to observed cross-language effects during word recognition. These outcomes are not in line with parallel activation taking place during speech exposure in bilinguals. It is possible that the mechanisms involved in the cognate facilitation in bilingual word acquisition are not captured by the task used in the present study. Implicit naming might be tapping into lower-level functions, more related to acoustic and phonological processing, while word acquisition and learning might rely on more abstract sources of information that do recruit parallel activation.

recent work using parental reports of word acquisition have provided insights. For instance, previous work on lexical acquisition has provided some insight into how found that English-Spanish bilinguals’ vocabulary in English was influenced by the words they had already acquired in Spanish. This is consistent with previous accounts of lexical acquisition in which infants acquire new words through preferential acquisition Fourtassi et al. ([2020](#ref-fourtassi2020growth)).

One of the limitations of this study lies in the use of cognates as prime words.

Conclusions

# References

Agresti, A. (2012). *Categorical data analysis* (Vol. 792). John Wiley & Sons.

Ana I. Schwartz, J. F. K., & Diaz, M. (2007). Reading words in spanish and english: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, *22*(1), 106–129. <https://doi.org/10.1080/01690960500463920>

Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3633–3647.

Barr, D. J. (2008). Analyzing ‘visual world’eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

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

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*(4), 369–380.

Bloom, P. (2002). *How children learn the meanings of words*. MIT press.

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341–347.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49.

Bosma, E., & Nota, N. (2020). Cognate facilitation in frisian–dutch bilingual children’s sentence reading: An eye-tracking study. *Journal of Experimental Child Psychology*, *189*, 104699.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

Byers-Heinlein, K., Gonzalez-Barrero, A. M., Schott, E., & Killam, H. (2023). Sometimes larger, sometimes smaller: Measuring vocabulary in monolingual and bilingual infants and toddlers. *First Language*, *0*(0), 01427237231204167. <https://doi.org/10.1177/01427237231204167>

Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., & Wermelinger. (2021). A multilab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*, *4*(1).

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, *15*(5), 336–345.

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children’s vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35*(6), 1189–1211.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518. https://doi.org/<https://doi.org/10.1006/jmla.1999.2654>

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*(2), 166–180.

Duta, M., Styles, S., & Plunkett, K. (2012). ERP correlates of unexpected word forms in a picture–word study of infants and adults. *Developmental Cognitive Neuroscience*, *2*(2), 223–234.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.

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

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*(4), 1003–1015.

Floccia, C., Delle Luche, C., Lepadatu, I., Chow, J., Ratnage, P., & Plunkett, K. (2020). Translation equivalent and cross-language semantic priming in bilingual toddlers. *Journal of Memory and Language*, *112*, 104086.

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., et al. (2018). Vocabulary of 2-year-olds learning english and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 7–29.

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children’s semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.

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

Garcia-Castro, G., Ávila-Varela, D. S., & Sebastian-Galles, N. (2023). *Bvq: Barcelona vocabulary questionnaire database and helper functions*. <https://gongcastro.github.io/bvq>

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183.

Groot, A. M. de. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5).

Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*(1-2), 163–187.

Hallé, P. A., & Boysson-Bardies, B. de. (1994). Emergence of an early receptive lexicon: Infants’ recognition of words. *Infant Behavior and Development*, *17*(2), 119–129.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a british communicative development inventory. *Journal of Child Language*, *27*(3), 689–705.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.

Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by latino children learning spanish as their first language. *Journal of Child Language*, *34*(2), 227–249.

Jardak, A., & Byers-Heinlein, K. (2019). Labels or concepts? The development of semantic networks in bilingual two-year-olds. *Child Development*, *90*(2), e212–e229.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What’s new in psychtoolbox-3?*

Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206.

Kuipers, J. R., & Thierry, G. (2015). Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain and Language*, *149*, 27–32.

Kuipers, J.-R., & Thierry, G. (2010). Event-related brain potentials reveal the time-course of language change detection in early bilinguals. *Neuroimage*, *50*(4), 1633–1638.

Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2018). Vocabulary size and speed of word recognition in very young french–english bilinguals: A longitudinal study. *Bilingualism: Language and Cognition*, *21*(1), 137–149.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Mani, N., & Plunkett, K. (2010). In the infant’s mind’s ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*(7), 908–913.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206.

Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young spanish–english bilinguals. *Journal of Child Language*, *37*(4), 817–840.

Marian, V., & Spivey, M. (1999). Activation of russian and english cohorts during bilingual spoken word recognition. *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, 349–354.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631–631.

Mitchell, L., Tsui, R. K., & Byers-Heinlein, K. (2022). *Cognates are advantaged in early bilingual expressive vocabulary development*.

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

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, *44*(4), 617–653.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876.

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Poarch, G. J., & Van Hell, J. G. (2012). Cross-language activation in children’s speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *111*(3), 419–438.

Rocha-Hidalgo, J., & Barr, R. (2023). Defining bilingualism in infancy and toddlerhood: A scoping review. *International Journal of Bilingualism*, *27*(3), 253–274. <https://doi.org/10.1177/13670069211069067>

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). Childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*, 1928–1941.

Singh, L. (2014). One world, two languages: Cross-language semantic priming in bilingual toddlers. *Child Development*, *85*(2), 755–766.

Siow, S., Gillen, N. A., Lepadatu, I., Avila-Varela, D. S., Garcia-Castro, G., Sebastian-Galles, N., & Plunkett, K. (2022). *The effect of cognates on bilingual infant vocabulary trajectories: A study using bilingual CDIs of english and one additional language*.

Styles, S. J., Plunkett, K., & Duta, M. D. (2015). Infant VEPs reveal neural correlates of implicit naming: Lateralized differences between lexicalized versus name-unknown pictures. *Neuropsychologia*, *77*, 177–184.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby’s first 10 words. *Developmental Psychology*, *44*(4), 929.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535.

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

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

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*, 1–67.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for british english. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Vihman, M. (2004). Cross-linguistic experiments in word-form recognition. *The Journal of the Acoustical Society of America*, *115*(5\_Supplement), 2502–2502.

Vihman, M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in english, welsh, and english–welsh bilingual infants. *Applied Psycholinguistics*, *28*(3), 475–493.

Von Holzen, K., Fennell, C. T., & Mani, N. (2019). The impact of cross-language phonological overlap on bilingual and monolingual toddlers’ word recognition. *Bilingualism: Language and Cognition*, *22*(3), 476–499.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586.

Von Holzen, K., & Mani, N. (2014). Bilinguals implicitly name objects in both their languages: An ERP study. *Frontiers in Psychology*, *5*, 1415.

Wood, S. N. (2017). *Generalized additive models: An introduction with r*. CRC press.

Zettersten, M., Yurovsky, D., Xu, T. L., Uner, S., Tsui, A. S. M., Schneider, R. M., Saleh, A. N., Meylan, S. C., Marchman, V. A., Mankewitz, J., et al. (2022). Peekbank: An open, large-scale repository for developmental eye-tracking data of children’s word recognition. *Behavior Research Methods*, 1–16.

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, *33*(2), 251–256.