Cognates are advantaged over non-cognates in early bilingual expressive vocabulary development

Lori Mitchell[1], Rachel Ka-Ying Tsui[1], & Krista Byers-Heinlein[1]

[1] Concordia University

Author Note

Correspondence concerning this article should be addressed to Rachel Ka-Ying Tsui, Department of Psychology, 7141 Sherbrooke St. West, Montreal, QC, Canada, H2T1V2. Rachel Ka-Ying Tsui is now at Laboratory for Language Development, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako-shi, Saitama, Japan, 351-0198. E-mail: rachelkytsui@gmail.com

Abstract

Bilingual infants grow up with the unique experience of needing to learn two words for most concepts. These words are called translation equivalents, and translation equivalents that also sound similar (e.g., banana—*banane*) are called cognates. Research has consistently shown that children and adults process and name cognates more easily than non-cognates. The present study explored if there is such an advantage for cognate production in bilinguals' early vocabulary development. Using longitudinal expressive vocabulary data collected from 47 English–French bilingual children starting at the ages of 16–20 months up to 27 months (a total of 219 monthly administrations in both English and French), results showed that overall children produced a greater proportion of cognate words than non-cognate words on the MacArthur-Bates Communicative Development Inventories. The findings suggest that cognate learning is facilitated in early bilingual vocabulary development. Just as phonological overlap supports monolingual infants in learning phonologically similar words in only one language, phonological overlap also supports bilinguals in learning phonologically similar words across their two languages.

Keywords: bilingualism, infants, cognates, translation equivalents, phonological similarity, expressive vocabulary development

Cognates are advantaged over non-cognates in early bilingual expressive vocabulary development

Infants understand some words during the first year of life, and begin to produce words around their first birthday (Fenson et al., 2007). To do so, infants must represent both the phonological and semantic aspects of a word, and associate the two. Intriguingly, the network of words that children already know appears to shape the words that they will learn: monolingual infants are more likely to learn words that are phonologically (Luce & Pisoni, 1998; Coady & Aslin, 2003) and semantically (Coady & Aslin, 2003) similar to those they already know. Bilingual infants provide a unique perspective into understanding children's developing lexical networks, as they must acquire translation equivalents, which are cross-language synonyms with complete or nearly complete semantic overlap (e.g., "apple" in English and *"pomme"* in French; De Houwer et al., 2006; Legacy et al., 2017; Pearson et al., 1995; see White et al. (2017) for a discussion of convergence in bilinguals' semantic representations). For infants acquiring typologically or historically related languages, some of these translation equivalents will be cognates, which are also phonologically similar (e.g., "banana" /bənænə/ in English and *"banane"* /bɑnɑn/ in French). This current study aimed to understand the impact of cognate status on the acquisition of words in young bilinguals by examining whether bilingual infants produce cognates more readily than non-cognates in early language development.

Translation Equivalents

Translation equivalents are an important part of early language development for bilingual children. While early researchers claimed that bilinguals avoid learning translation equivalents (Volterra & Taeschner, 1978), recent work shows that bilingual infants acquire translation equivalents from an early age (Legacy et al., 2017; Pearson et al., 1995). Bilingual infants begin to produce translation equivalents by 16 months, and produce more translation equivalents with age as their vocabularies grow (Legacy et al., 2017). The strong semantic overlap of a word in one language seems to facilitate the acquisition of its translation equivalent in the other language, at least at younger ages when bilingual infants have smaller vocabularies (Bilson et al., 2015; Tsui et al., in press). By the age of 27 months, bilingual toddlers recognize a target word more accurately when preceded by its translation equivalent (Floccia et al., 2020).

For bilingual infants, some translation equivalents sound very similar. Specifically, cognates are a type of translation equivalents that have significant phonological overlap, typically due to a shared etymology[1]. Cognates range in their degree of phonological similarity: For example, English "banana" /bənænə/ and French *"banane"* /bɑnɑn/ are highly phonetically similar, while English "pants" /pænts/ and French *"pantalon"* /pɑ̃ tɑlɔ̃/ are more different including a different number of syllables. Some typologically close languages even have form-identical cognates, such as the word *"sí"* /si/ which means "yes" in both Spanish and Catalan.

Cognates appear to have a special status in bilingual language processing and production. Previous research has reported a cognate facilitation effect where bilinguals are better and quicker at identifying and naming cognates than non-cognates when performing vocabulary tasks (e.g., Costa et al., 2000; Kelley & Kohnert, 2012; Sheng et al., 2016). This type of advantage for cognates has been reported in bilingual adults (e.g., Costa et al., 2000) as well as in school-aged children (e.g., Kelley & Kohnert, 2012; Sheng et al., 2016). For example, Kelley and Kohnert (2012) provide evidence for the cognate facilitation effect in Spanish-speaking English learners between the ages of 8 and 13 years old, where children identified and named more cognates than non-cognates in receptive and expressive vocabulary tasks. A similar cognate advantage has been found for picture naming and translation tasks for English–Spanish and English–German 4- to 8-year-old children where

[1] Cognates can also overlap in their orthography, but we do not address orthography in this paper as our participants were too young to read.

bilingual children were more accurate in naming cognates and faster at translating cognates than non-cognates (Schelletter, 2002; Sheng et al., 2016). Therefore, cognates seem to be advantaged in school-aged bilingual children's language processing and production.

Effects of Phonological Similarity on Early Word Learning

The advantage for cognates could be attributed to the phonological overlap between words, which may make them easier to learn. Existing literature on monolinguals has reported that children are more likely to produce words that sound similar to other words in their lexicons (e.g., "at" and "cat," "hat" and "cat"), especially at younger ages (e.g., Jones & Brandt, 2019). For instance, looking at 300 British English-speaking children aged 12 to 25 months, Jones and Brandt (2019) found that the strength of phonological similarity between words was an important predictor for word production (but not comprehension), whereby young children tended to produce words that follow similar phonological patterns. Similarly, using archival expressive vocabulary data from 1,800 16- to 30-month-old American infants, it was shown that infants produced more nouns with many phonological neighbours than those with few phonological neighbours (Storkel, 2009). It is possible that the high degree of phonological similarity aids word acquisition through sounds already established in the lexicons. For example, Demke et al. (2002) found that hearing real-word phonological neighbours facilitated the learning of new pseudowords. Another possibility is that the words that share a high degree of phonological similarity in the language input are learned first by infants, as supported by a recent study looking at the developing lexicons of young infants across 10 languages (Fourtassi et al., 2020). Overall, learning a new word with close phonological neighbours seems to help learners maintain the new word in memory, making similar-sounding words easier to acquire and produce (e.g., Coady & Aslin, 2003; Demke et al., 2002; Jones & Brandt, 2019).

Extending this notion to bilingual infants, some evidence suggests that phonological similarity facilitates vocabulary learning across languages as well. For example, Gampe et al. (2021) examined parent-reported vocabulary size of 18- to 36-month-old children learning Swiss German and another language. Children learning languages that were more phonologically similar to Swiss German (e.g., standard German, Dutch, English) produced more words than children learning languages that were more phonologically dissimilar (e.g., Turkish, French). Moreover, children learning more similar languages learned more cognate translation equivalents, while the number of non-cognate translation equivalents was similar across groups. These results are consistent with other studies reporting that language distance affects early bilingual language acquisition (e.g., Blom et al., 2019; Gampe et al., 2021; Havy et al., 2016; Sheng et al., 2016).

However, not all studies have reported a generalized advantage for cognates in vocabulary learning. In a study of younger children, Bosch and Ramon-Casas (2014) used parent reports to examine word production in 18-month-olds learning Spanish and Catalan, two strongly related languages that share many form-identical (e.g, "yes" is *"sí"* /si/ in both Spanish and Catalan) and form-similar (e.g., "hand" is *"mano"* /mɑno/ in Spanish and *"mà"* /mɑ/ in Catalan) cognates. Results indicated that 28% of the words produced by the bilingual infants were form-identical cognates, while less than 2% of words were form-similar cognates or non-cognate translation equivalents (Bosch & Ramon-Casas, 2014). One explanation for this finding is that for form-identical cognates, infants only need to learn a single form for a particular concept, which they can then transfer across their languages. Based on these results, bilingual infants may not benefit from cognates' phonological overlap unless that overlap is perfect. Indeed, there is some evidence that Spanish–Catalan infants are somewhat insensitive to phonological distinctions in form-similar cognates (Ramon-Casas et al., 2009; Ramon-Casas & Bosch, 2010), perhaps even representing them as form-identical. Another interpretation of this result is that the effect of cognates on bilingual vocabulary learning changes across development, which could explain the discrepant results of the 18-month-old sample studied by Bosch & Ramon-Casas (2014), and the 18- to 36-month-old sample studied by Gampe et al. (2021). Specifically, it is possible that an advantage for cognates is detectable first for form-identical cognates when they are present in the languages (as in Bosch & Ramon-Casas, 2014), and then later for form-similar cognates as children grow older and learn more words overall (as in Gampe et al., 2021). In other words, cognate status and age might interact.

Current study

To better understand the impact of phonological overlap on bilingual infants' vocabulary learning, we examined the production of cognate and non-cognate translation equivalents in French–English bilingual infants. English and French share many form-similar cognates due to historical language contact (Choi, 2019), although only a few form-identical cognates. Despite the presence of cognates, note that these two languages belong to different language families: English is a Germanic language and French is a Romance language. Previous work looked at learners of closely related languages with many form-identical cognates (Spanish and Catalan; Ramon-Casas & Bosch, 2010), or else a heterogeneous group of bilinguals learning many different language pairs (Gampe et al., 2021). Thus, our study provided an important test of the generalizability of these results in a new and homogeneous population of young bilinguals.

We collected monthly vocabulary data on French–English bilingual infants' word production starting when children were between the ages of 16–20 months and ending when were up to 27 months of age using the MacArthur-Bates Web-Communicative Developmental Inventory: Words and Sentences form in American English (Fenson et al., 2007) and Québec French (Trudeau et al., 1999). Uniquely, our dataset was longitudinal, allowing us to investigate potential developmental effects. We focused our analysis on translation equivalent pairs and then classified the pairs according to cognate status (cognate or non-cognate words). We counted children's production of both translation equivalent pairs (e.g., whether they produced both "apple" /æpəl/ and *"pomme"* /pɔm/, or both "banana" /bənænə/ and *"banane"* /bɑnɑn/), as well as individual words independent of whether children produced its translation equivalent. Since it is not possible to randomly assign our main variable of interest (cognate status), we analyzed both a complete list of cognate and non-cognate words, as well as a carefully selected subset of these cognate and non-cognate words which were matched on age of acquisition and on word category (e.g., words about food) where possible.

We hypothesized that French–English bilinguals would more readily produce cognates than non-cognates. Thus, we predicted that French–English bilingual infants would produce proportionally more translation equivalent words and pairs that were cognates than non-cognates. We likewise anticipated an interaction between cognate status and age, with a stronger effect of cognate status at older ages as the infants' vocabulary size (and the number of translation equivalent words and pairs produced) grew.

Method

The present research was approved by the Human Research Ethics Committee at Concordia University [certification #10000439]. Participation was on a voluntary basis and the families were free to withdraw at any time. The study design was pre-registered at https://osf.io/rh7av.

Participants

The current study comprised data from 50 French–English bilingual infants (26 females) collected from August 2020 to May 2021, as part of a larger ongoing longitudinal study which aims to collect data from 100 bilingual infants. Participating infants were aged between 16 and 20 months at the onset of participation (mean starting age = 17.98 months, SD = 1.15, range = 16.20 – 20.40), and were aged between 16 to 27 months at their final time of participation (M = 21.96 months, SD = 3.20, range = 16.30 – 27.14). Participants were recruited from Québec, Canada through government birth lists, social media, and participating families' referrals. Inclusion criteria were the following: full-term pregnancy (i.e., at least 37 weeks of gestation), normal birth weight (> 2500 grams), and no reported developmental delays or any hearing or vision problems. Bilingual infants were defined as those exposed to each of English and French for at least 10% and at most 90% of the time over the course of their lives since birth, with less than 10% of exposure to a third language[2]. To capture a wider range of bilingual experience, the language exposure range in this study was wider than some studies (e.g., Morin-Lessard & Byers-Heinlein, 2019; Sebastián-Gallés & Bosch, 2009) but similar to the range used in others (e.g., Hoff & Ribot, 2017; Place & Hoff, 2011).

In total, parents completed 230 English CDI administrations and 226 French CDI administrations. We retained only cases where both the English and French were completed at the same time point to be able to determine infants' translation equivalent knowledge. This left us with 219 completed administrations from 47 infants. Six infants contributed data at only one time point, and 41 infants contributed data at more than one time point, with participants contributing an average of 4.70 measurements for each language (SD = 2.51, range = 1 – 10). On average across the 219 administrations, participating infants were exposed to English 48.8% of the time (SD = 17.3, range = 11 – 84), to French 50.6% of the time (SD = 17.7, range = 16 – 88), and to a third language 0.6% of the time (SD = 1.5, range = 0 – 5). Of the 47 bilingual infants, 26 were English-dominant (M = 60.1% English exposure, SD = 10, range = 49 – 84), 20 were French-dominant (M = 66.4% French exposure, SD = 12.7, range = 51 – 88), and 1 reported equal exposure to both English and French. The average maternal education level was 17.32 years (SD = 2.29, range = 12 – 23), and 89.40% of the mothers had completed a university degree or higher.

[2] We have also run the analyses using a stricter 25%-75% inclusion criterion. The results are reported in the supplemental materials.

Measures

Web-based MacArthur-Bates Communicative Development Inventory: Words and Sentences (Web-CDI).

The number of words produced in English and French was obtained monthly via the web-based versions of the MacArthur-Bates Web-Communicative Development Inventories: Words and Sentences form (Web-CDI; https://webcdi.stanford.edu/), using the American English version (Fenson et al., 2007) and the Québec French adaptation ("Mots et Énoncés"; Trudeau et al., 1999). Our study focused on the vocabulary checklist component of the CDIs, with 680 words in the English version and 664 words in the Québec French version. We asked the caregiver most familiar with the infant's vocabulary in each language to complete the respective version, although following the instructions on the Web-CDI they could seek help from others who often speak the corresponding language with the infant. The English forms were completed by mothers (88%), fathers (7%), and both parents (5%), whereas the French forms were completed by mothers (84%), fathers (11%), and both parents (5%). Thus, most of the time, the same caregiver (usually the mother) filled out both forms. Generally, whichever caregiver completed forms in a particular language did so throughout the study, with the exception of 2 participants (4.3%) whose English forms were filled out by different caregivers for some administrations, and 3 participants (6.4%) whose French forms were filled out by different caregivers for some administrations. Infants' demographic information including age and sex was also collected at the start of the Web-CDI.

Language Exposure Questionnaire (LEQ) using the Multilingual Approach to Parent Language Estimates (MAPLE).

The infant's language exposure and background was measured with an adaptation of the Language Exposure Questionnaire (LEQ; Bosch & Sebastián-Gallés, 2001), using the Multilingual Approach to Parent Language Estimates (MAPLE; Byers-Heinlein et al., 2020). During a 15- to 20-minute structured interview, the primary caregiver(s) were asked questions about the infant's language exposure from birth until their current age. This provided a global estimate of the percentage of exposure that the infant had to each of their languages across all contexts.

Procedure

Data collection for this study began in August 2020 and ended in May 2021, although the start date of participation varied across participants. On the first day of each month, links to the English and French Web-CDI forms were sent to the caregivers by email. On the forms, the words that were checked off in previous months were automatically filled in the following months; thus, caregivers only needed to check off the new words that their child produced each month. This was intended to reduce the burden on participants, and increase the response rate. Parents were instructed to consider words produced even if the child's production were not yet adult-like (e.g., the child produced "raff" instead of "giraffe"). We asked that the Web-CDI forms be completed during the first week of each month. A reminder was sent on the 8th of the month, and an extra week was given for caregivers who had not yet completed the forms. If caregivers still did not complete the form, they were asked to resume their participation the following month. Once the forms were completed, caregivers received a brief report about their child's vocabulary knowledge at that time point, including the total number of words that their child produced as well as the breakdown of the categories (such as animals, food, furniture, etc.) for which their child produced words.

At the first data collection time point, caregivers also completed the LEQ questionnaire with a trained research assistant over the online video chat application Zoom. This was repeated every five months to track any potential changes in the infant's language exposure. This was particularly important as data collection overlapped with the COVID-19 pandemic, thus it was important to closely track language exposure changes due to lockdowns, return to daycare, etc.

Identification of Translation Equivalents and Cognates

A list of translation equivalents on the English and French forms of the CDI was created by three proficient English–French bilingual adults who carefully examined the English and French versions of the CDIs; a total of 611 translation equivalent pairs were identified (the full list is available at https://osf.io/7fz6c/; Gonzalez-Barrero et al., 2020). Next, bilingual research assistants identified 138 of the possible 611 translation equivalent pairs as cognates, with the remaining 473 words as non-cognates. Phonological similarity of the identified cognates were further confirmed by bilingual undergraduate students who were asked to evaluate the phonological overlap of recordings of those words. This method was preferred to other methods that focus on orthography (overlap in spelling), since infants acquire language through spoken words as opposed to reading. These steps were carried out in the Concordia Infant Research Laboratory for different projects, prior to the current study.

From the list of 611 translation equivalents, we further excluded any translation equivalent pairs that had complex relationships rather than one-to-one mappings. For example, "noodle" forms a translation equivalent pair with either the French word *"nouilles"* or *"pâtes"*, where both French words are listed together as one item on the French CDI form. These pairs were removed because we could not know which form (e.g., *"nouilles"* or *"pâtes"*) the infant produced, and we were not able to classify these pairs as either cognates or non-cognates.

Following this procedure identified a complete list of 537 translation equivalents (131 cognates[3] and 406 non-cognates), which were compared in a first set of analyses. However, note that the cognates and non-cognates in this list could vary systematically on correlated

[3] Among these 131 cognates on the complete list, we could identify 11 cognates that were potentially form-identical: choo choo, grr, meow, vroom, woof, Cheerios, Coke, pizza, muffin, toast, and jeans (only the last three were on the matched list). Note that each of these words was either a sound effect, a brand name, or a conventionalized borrowing. Even for these words, adult speakers often pronounce them slightly

factors including variations in parts of speech and differences in age of acquisition of certain words between languages. Within the full list of translation equivalents, we thus identified a matched subset of cognates and non-cognates that would be used in a second set of analyses.

Our procedure for identifying the matched list was pre-registered, and designed to minimize potential experimental biases. The matched list was first restricted to nouns as infants show a noun bias in language acquisition (Caselli et al., 1995), and doing so matched the cognates and non-cognates for part of speech. Next, the remaining 272 translation equivalents (cognates = 90, non-cognates = 182) were matched on age of acquisition and word category where possible (e.g., food, furniture, etc.). However, data on age of acquisition, which was obtained from the wordbankr package (Version 0.3.1; Frank et al., 2017), was not available for 41 translation equivalents which were therefore removed, leaving 231 possible items (cognates = 81, non-cognates = 150). Using the optmatch package (Version 0.9.14; Hansen & Klopfer, 2006) in the R statistical language (R Core Team, 2019), each cognate item was matched to a non-cognate item according to the typical age of acquisition in both English and French obtained from the wordbankr package (Version 0.3.1; Braginsky, 2018) with the closest match possible on word category. There were 52 pairs that matched exactly based on these criteria. For example, the cognate pair "chair"–*"chaise"* and the non-cognate pair "bed"–*"lit"* matched because they are typically acquired at age 21 months in English and French and are both in the furniture category (Frank et al., 2017). The remaining 29 pairs were matched on age of acquisition as well, allowing a possible one-month deviation in either English, French or both. For example, the cognate pair "mittens"–*"mitaine"* and the non-cognate pair "slipper"–*"pantoufle"* matched since the English words are acquired at 28 and 27 months respectively (one-month

differently in French and and English such that words align with each language's phonology, for example differences in the exact realization of particular phonemes, and stress pattern differences in bisyllabic words muffin and pizza.

deviation), both French words are acquired at 22 months of age (Frank et al., 2017), and both are clothing. Thus, the final items (81 cognates[4], 81 non-cognates) included in the matched list were as similar as possible in all respects except their cognate status.

Analytical Strategy

Analyses were run on two different dependent variables to examine whether bilingual infants would produce more cognates than non-cognates over their vocabulary development. The first dependent variable was the proportion of items on the word list that infants produced, where translation equivalents were counted as separate items. For example, the word "banana" would be counted as a produced cognate, whether or not its translation equivalent *"banane"* was produced. The second dependent variable was the proportion of translation equivalent pairs infants produced. Here, pairs were counted only if the infant produced both items in a pair. For example, the pair "banana"–*"banane"* was counted as a produced cognate pair if and only if the child could produce both words in the pair.

For each dependent variable we conducted analyses using (1) the complete list of cognates and non-cognates (537 translation equivalents pairs in total) and then restricted the analysis to (2) a matched list (nouns only and matched on age of acquisition; 162 translation equivalent pairs in total). Based on the two dependent variables and the two sets of words, we therefore ran a total of four models. Logistic mixed effects analyses were performed in the R statistical language (Version 4.0.2; R Core Team, 2019) using the lme4 package (Bates et al., 2015). Mixed effects models are appropriate for repeated measures data. This type of model also accounts for missing data and does not require each participant to contribute the same number of datapoints. A logistic model was appropriate as our dependent variable was a proportion. Regression weights reflected the total number of cognates and non-cognates to account for the different number of words between the

[4] Among these 81 cognates on the matched list, there are 3 form-identical cognates and the remaining 78 pairs are form-similar cognates.

cognate and non-cognate lists. The lmerTest package (Kuznetsova et al., 2017) was used to calculate p-values. Goodness-of-fit tests for the logistic regression models were estimated using the DHARMa package (Hartig, 2022). Analysis scripts and the data set used in the present study are available at [https://osf.io/rh7av/].

Results

Descriptive Measures of Number of Words Produced

Out of the complete list (a possible 537 translation equivalent pairs with $537 \times 2 = 1074$ words), bilingual infants on average produced a total of 157 words (SD = 158), with a range of 0 – 709 words, which constituted 14.6% of the words on the complete list. Moreover, bilingual infants produced an average of 39 complete translation equivalent pairs where both the English and French words were produced (SD = 50.61, range = 0 – 243), which constituted 7.3% of the translation equivalent pairs on the complete list.

Restricting to the matched list which contained 162 translation equivalent pairs with $162 \times 2 = 324$ words, bilingual infants produced an average of 51 words (SD = 59.71, range = 0 – 248), which constituted 15.7% of the words on the matched list. On average, bilingual infants produced a total of 12 complete translation equivalent pairs (SD = 20.77, range = 0 – 92), which constituted 7.6% of translation equivalent pairs on the matched list.

Dependent Variable 1: Cognate Words Versus Non-Cognate Words

In this analysis, the dependent variable was the total proportion of words infants produced on the relevant list. Proportion was used as opposed to raw number of words to provide a more comparable description of production of cognates versus non-cognates, since the number of cognate words and non-cognates words differed especially in the complete list. Our predictor variables were age (in days) and cognate status. Age was continuous and was centered at the mean age of 547.6 days (approximately 18 months) for ease of interpretation. Cognate status was categorical with two levels (cognates versus non-cognates) with non-cognates as the reference level. We ran separate logistic regression models for the complete and matched lists. The initial model specification included a random slope of age and cognate status by participants, which was pruned to a random intercept to achieve model convergence. The final model was:

proportion_word ~ age * cognate_status + (1|participant)

Complete List. Out of the complete list which contained 262 cognate words (i.e., adding the 131 English cognate words and 131 French cognate words) and 812 non-cognate words (i.e., adding the 406 English non-cognate words and 406 French non-cognate words), bilingual infants produced an average of 54 cognate words (SD = 45.56, range = 0 – 204) and 103 non-cognate words (SD = 113.09, range = 0 – 505). The proportion of cognate words produced was 0.21 (SD = 0.17, range = 0 – 0.78), whereas the proportion of non-cognate words produced was 0.13 (SD = 0.14, range = 0 – 0.62). A Q-Q plot visualization and goodness-of-fit tests on the model's residuals showed that our model had a good model fit, $D = 0.06$, $p = .125$. Table 1 shows the coefficient estimates for the model and Figure 1 Panel A visualizes the model. We observed significant main effects of age and cognate status, as well as a significant interaction. Overall, the pattern of results indicated that infants produced a greater proportion of cognates than non-cognates, with a slightly steeper learning curve for non-cognates than for cognates, although non-cognate production did not "catch up" to cognate production during the ages we observed.

Matched List. Out of the 162 cognate (i.e., adding the 81 English cognate words and 81 French cognate words) and 162 non-cognate words (i.e., adding the 81 English non-cognate words and 81 French non-cognate words) on the matched list, bilingual infants produced an average of 27 cognate words (SD = 31.52, range = 0 – 135) and 23 non-cognate words (SD = 28.4, range = 0 – 113). The overall mean proportion of cognate words produced was 0.17 of words (SD = 0.19, range = 0 – 0.83), whereas the proportion of non-cognate words produced was 0.14 (SD = 0.18, range = 0 – 0.7). A Q-Q plot visualization as well as goodness-of-fit tests on the model's residuals showed a good model fit, $D = 0.06$, $p = .095$. Table 1 also shows the coefficient estimates for the matched list model and Figure 1 Panel B visualizes the model. Similar to the patterns reported in the complete list model, there were significant effects of age and cognate status, once again showing that infants produced a greater proportion of cognates than non-cognates on the matched list. However, for the matched list there was no interaction between cognate status and age, indicating that the magnitude of the cognate advantage for this list was stable as infants grew older.

Table 1. Coefficient estimates from the mixed-effects logistic models predicting proportion of words produced.

	Complete list				Matched list			
	Estimate	SE	z	p	Estimate	SE	z	p
Intercept	-2.520	0.150	-16.800	<.001	-2.690	0.214	-12.500	<.001
cognate_status	0.740	0.015	49.500	<.001	0.234	0.026	8.860	<.001
age_days	0.010	0.000	83.200	<.001	0.012	0.000	45.200	<.001
cognate_status * age_days	-0.001	0.000	-5.120	<.001	0.000	0.000	0.933	0.351

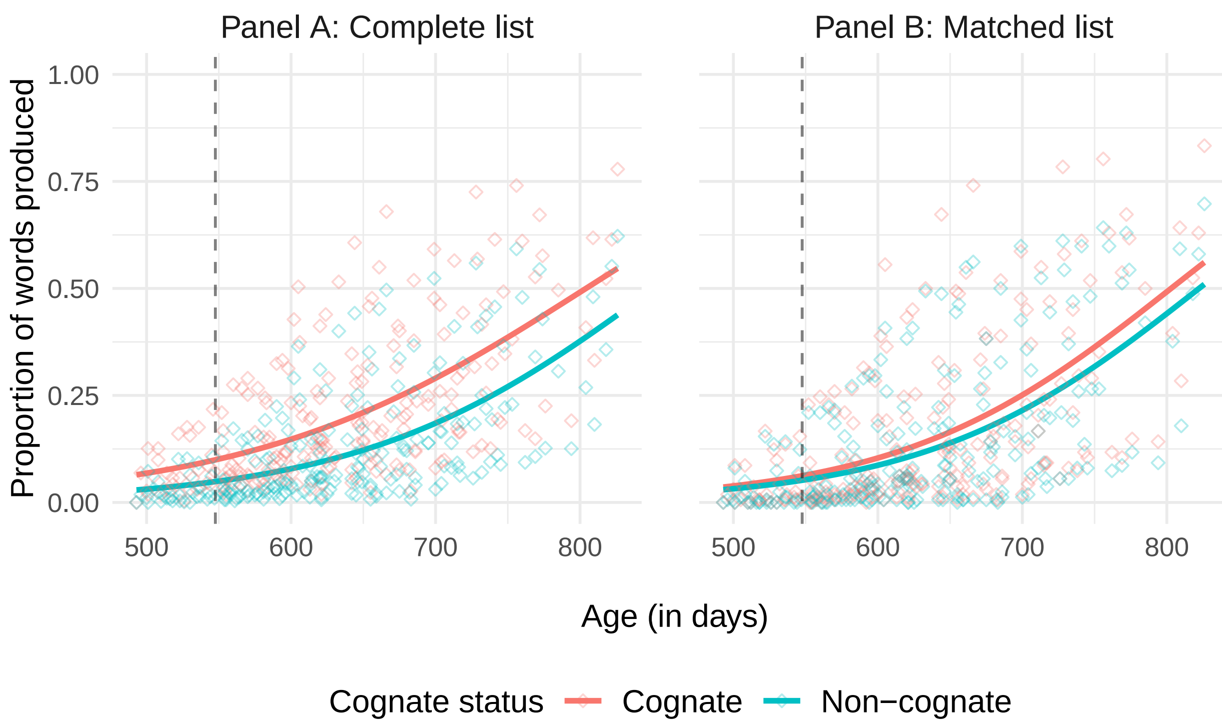

Figure 1. Proportion of words produced by age and cognate status, with Panel A representing the complete list and Panel B representing the matched list. Note that the black dashed line represents the mean age of 547.6 days which serves as the reference level for age in our models.

Dependent Variable 2: Cognate Pairs Versus Non-Cognate Pairs

In this analysis, the proportion of translation equivalent pairs produced was entered as the dependent variable. Age and cognate status were entered as our predictor variables, with non-cognates set as the reference level. Again, we ran separate logistic models for the complete and matched lists. The initial model specification, which included a random effect of age and cognate status by participants, had to be reduced for model convergence; therefore, the final model was:

proportion_pair ~ age * cognate_status + (1|participant)

Complete List. Out of the complete list which contained 537 translation equivalent pairs (131 cognates and 406 non-cognates), infants produced an average of 17 cognate pairs (SD = 18.1, range = 0 – 82) and 22 non-cognate pairs (SD = 32.93, range = 0 – 167). The proportion of cognate pairs produced was 0.13 (SD = 0.14, range = 0 – 0.63) whereas the proportion of non-cognate pairs produced was 0.05 (SD = 0.08, range = 0 – 0.41). A Q-Q plot visualization and goodness-of-fit tests on the model's residuals revealed that our model showed a good model fit, $D = 0.04$, $p = .398$. Table 2 shows the coefficient estimates for the model and Figure 2 Panel A visualizes the model. There were significant effects of age and cognate status, showing that overall infants produced a greater proportion of cognates than non-cognates. Similar to the pattern reported in the first set of analyses, the interaction between age and cognate status suggested a slightly steeper learning curve for non-cognates than cognates, although an advantage for cognates was still apparent even at 27 months.

Matched List. Out of the 162 translation equivalent pairs, bilingual infants produced an average of 7 cognate pairs (SD = 12.21, range = 0 – 58) and 5 non-cognate pairs (SD = 8.83, range = 0 – 42). The proportion of cognate pairs produced was 0.09 (SD = 0.15, range = 0 – 0.72) and the proportion of non-cognate pairs produced was 0.06 (SD = 0.11, range = 0 – 0.52). A Q-Q plot visualization and the goodness-of-fit test on the model's residuals ($D = 0.09$, $p = .002$) suggested that the logistic model did not fully capture the distribution of the data, but we nevertheless retained the model on theoretical grounds (the dependent variable was proportion) and to facilitate comparison to the previous models. The coefficient estimates for the matched list model is shown in Table 2, and Figure 2 Panel B visualizes the model. Similar to the results for the complete list, the main effects of age and cognate status were statistically significant, showing that infants produced a larger proportion of cognates than non-cognates. However, unlike the results for the complete list, the interaction between age and cognate status was not statistically significant, showing that the magnitude of the cognate difference was reasonably stable across age.

Table 2. Coefficient estimates from the mixed-effects logistic models predicting proportion of translation equivalent pairs produced.

	Complete list				Matched list			
	Estimate	SE	z	p	Estimate	SE	z	p
Intercept	-3.680	0.152	-24.200	<.001	-4.320	0.252	-17.100	<.001
cognate_status	1.210	0.029	41.200	<.001	0.660	0.061	10.800	<.001
age_days	0.011	0.000	46.400	<.001	0.014	0.001	25.800	<.001
cognate_status * age_days	-0.002	0.000	-6.100	<.001	-0.001	0.001	-0.858	0.391

Summary of Analyses

Overall, the result patterns were largely consistent across the two sets of analyses, whereby bilingual infants produced a greater proportion of cognates than non-cognates. Infants increased their production of both cognates and non-cognates across age, and for the complete list (although not the matched list) this increase was slightly steeper for non-cognates than cognates, although production of cognates remained proportionally greater than that of cognates at the oldest age we observed (27 months).

Discussion

This current study evaluated whether phonological similarity facilitates vocabulary learning in bilinguals, by examining whether cognates are advantaged in bilingual infants' early vocabulary production. Using monthly expressive vocabulary data, our longitudinal dataset revealed an overall advantage for cognates in infancy. Across ages, infants

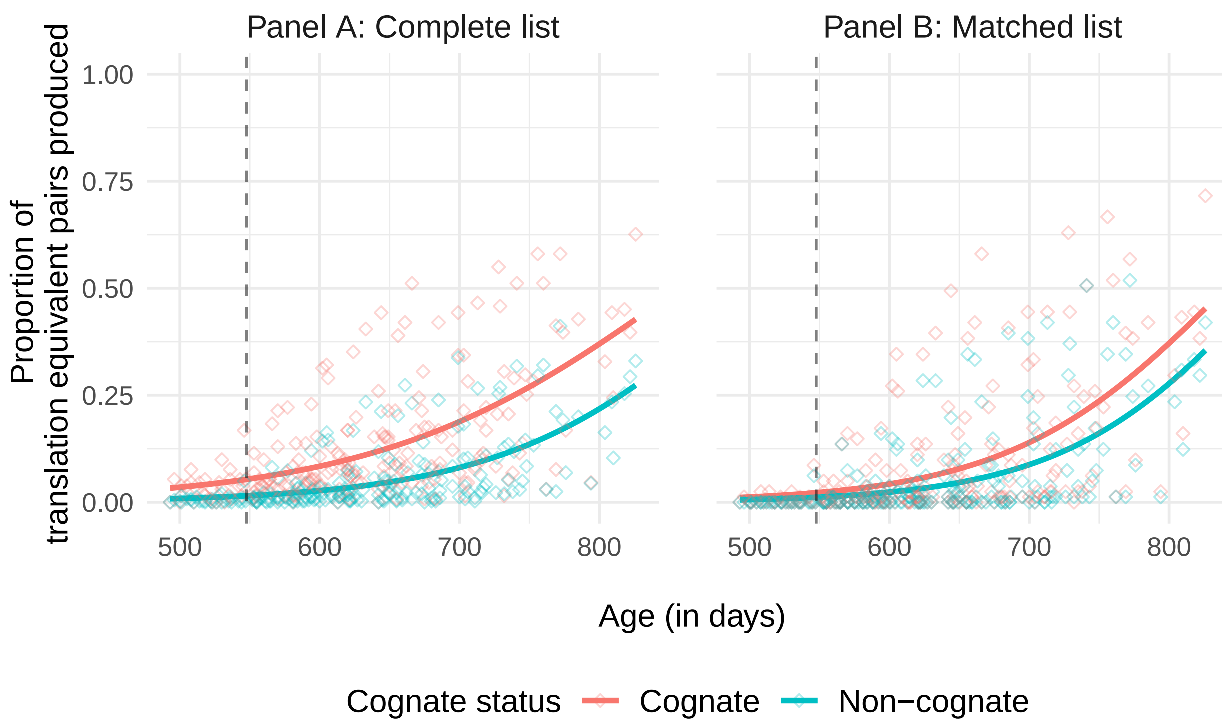

Figure 2. Proportion of translation equivalent pairs produced by age and cognate status, with Panel A representing the complete list and Panel B representing the matched list. Note that the black dashed line represents the mean age of 547.6 days which serves as the reference level for age in our models.

produced proportionally more cognates (e.g., English “banana” /bənænə/–French *”banane”* /bɑnɑn/) than non-cognates (e.g., English “apple” /æpəl/–French *”pomme”* /pɔm/), although note that in raw terms children still produced a greater number of non-cognates than cognates, due to the greater absolute frequency of non-cognate translation equivalents on the French–English CDI checklists.

Together with previous findings, our results begin to paint a developmental picture of the effects of cognate status on early vocabulary productions. Spanish and Catalan have both form-similar and form-identical cognates, and Bosch & Ramon-Casas (2014) reported a cognate advantage for form-identical but not form-similar cognates at 18 months. Cognates in French and English are almost exclusively form-similar, and we found an advantage for these form-similar in infants aged 16 to 27 months. Other studies have also reported an advantage for non-cognate translation equivalents (Bilson et al., 2015), which might vary with age (Tsui et al., 2022). Overall, translation equivalents with the largest phonological overlap appear to be the most advantaged in early production and thus their effect might be detectable from age 18 months, with potential advantages in children's production of form-similar and non-cognate translation equivalents strengthening across the second and third year of life.

The robust cognate advantage across different bilingual infant populations points to the possibility that the origin of the cognate facilitation effect observed in childhood and in adulthood emerges from infancy. Previous studies which examined the cognate facilitation effect in bilingual adults and school-aged children have reported that bilinguals are better at processing cognates; for example, they can identify and/or name cognates more easily and quickly in a vocabulary task (Costa et al., 2000; Kelley & Kohnert, 2012; Sheng et al., 2016). Thus, the cognate facilitation effect appears to be robust in vocabulary production across the lifespan, with the advantage for cognates in production emerging early on, as our study results suggested.

We interpret these results in light of theories that emphasize the interconnectedness of the two languages in the developing bilingual lexicon (DeAnda et al., 2016). Studies show that, even across languages, words that are semantically related are acquired sooner by bilingual children (Bilson et al., 2015) and are co-activated in language processing (e.g., DeAnda & Friend, 2020; Jardak & Byers-Heinlein, 2018; Singh, 2014). Moreover, young monolinguals find it easier to learn words that are phonologically similar to one another (Coady & Aslin, 2003; Demke et al., 2002; Jones & Brandt, 2019), and young bilinguals co-activate phonologically-related words both within and across languages (Von Holzen & Mani, 2012). These two sets of findings were confirmed in a study of monolingual children across 10 languages, who were more likely to acquire words with a high degree of semantic or phonological association (Fourtassi et al., 2020). Unique to bilinguals, cognates have a high degree of both semantic and phonological overlap, which our results show facilitate their acquisition.

There are several specific ways that cognates' phonological and semantic overlap might advantage their learning. One possibility is that, for cognates, bilingual children might only need to map one phonological form (or slightly varied phonological forms for the cases of form-similar cognates) to label the same referent across the two languages, whereas for non-cognate translation equivalents bilingual children have to memorize two completely different forms for the same referent. Indeed, bilingual children learning similar languages learn more cognate translation equivalents and have a larger vocabulary size in general (Gampe et al., 2021). Thus, transfer effects could explain the cognate advantage we observed in production, and would predict an early-emerging cognate advantage for word comprehension as well. Another possibility is that hearing a cognate word activates and strengthens phonological representations for both languages (e.g., hearing "banana" could activate and strengthen both "banana" and *"banane"*), thus accelerating cognate learning (Schott et al., 2022). Bilingual children have been found to identify and name cognates easier and faster than non-cognates, suggesting that the phonological overlap in cognates could support bilinguals' lexical decoding and processing (Kelley & Kohnert, 2012; Sheng et al., 2016). Finally, the closer the phonological form of cognates, the more similar they might be for children to articulate. Support for such a hypothesis comes from study which showed that bilingual children not only learned phonologically-similar words faster but produced phonologically-similar nouns more frequently and more evenly than form-dissimilar nouns across their two languages (Schelletter, 2002). Note that these three possible mechanisms are not mutually exclusive, and could each contribute to the cognate advantage we observed.

There are several other factors that could also contribute to children's faster learning of cognates than non-cognates. For example, Bosch & Ramon-Casas (2014) brought up several additional possibilities including frequency in the language input, reference to more complex concepts, or production difficulty due to changes in phonological forms, although they could not provide direct evidence due to the limited items on their vocabulary checklists. Our study attempted to account for several of these factors, by analyzing a subset of cognates and non-cognates that were carefully matched for part of speech, typical age of acquisition, and word category when possible. With this carefully controlled subset, we again found a production advantage for cognates. Thus, while such additional factors could potentially contribute to the cognate advantage, our results suggest that such third variable explanations are unlikely to underlie our results.

The cognate advantage can, at least in part, explain why bilingual children learning more similar languages show accelerated vocabulary development relative to bilinguals acquiring less similar languages (Blom et al., 2020; Gampe et al., 2021; Sheng et al., 2016). It has been shown that the more overlap shared across the two languages, the easier the words are learned by bilingual children (Bosma et al., 2019). Therefore, for those who are learning close language pairs that share a high degree of phonological overlap like Spanish and Catalan, their two languages share many cognates which sometimes are even form-identical, meaning that they are pronounced the same way in both languages (e.g, *"sí"* /si/ meaning "yes" in both languages). On the other hand, for those who are learning languages that share a lesser degree of phonological similarity like English and French, there are potentially very few form-identical cognates. It is possible that when languages are very similar and share many form-identical cognates, bilingual infants can benefit from these words from a very young age. On the other hand, when languages are somewhat less similar and share mostly form-similar cognates, children may need more time to detect and benefit from cognates. Overall, we suggest that there is a gradual timeline for the facilitative effect of cognates in infancy, which starts off with form-identical cognates then form-similar cognates (Bosma et al., 2019). Future studies could include additional language pairs which are less similar than Spanish and Catalan but more similar than English and French, such as Spanish and Italian (Schepens et al., 2013), to directly compare the timeline regarding the acquisition of form-identical cognates, form-similar cognates, and non-cognates. Moreover, while previous studies suggested that bilingual children learning more similar languages learned more cognate pairs than those learning less similar languages (Gampe et al., 2021), it would also be important for future studies to further examine whether the advantage for cognates is of the same nature across different language pairs. A final interesting direction would be to use a continuous metric to quantify the degree of phonological overlap in form-similar cognate pairs, to more precisely examine how phonological overlap contributes to word learning.

An important avenue for future research would be to examine whether the same cognate advantage would be observed in receptive vocabulary acquisition, and indeed some evidence points in this direction. Some work with Spanish-Catalan bilinguals has suggested that infants show less perceptual sensitivity to cross-language phonological distinctions in cognates due to their phonological similarity (Ramon-Casas et al., 2009; Ramon-Casas & Bosch, 2010), suggesting that cognates may hold a different status in early bilinguals' receptive lexicons compared to non-cognates. However, more recent research with French-English bilingual toddlers had an opposite finding, whereby cognates were represented in more phonetic detail than non-cognates (Schott et al., 2022). There is also evidence that the cognate advantage is modulated by the level of difficulty of the vocabulary item for both comprehension and production. One study found that although the cognate advantage was observed in easier items, the effect was even greater in vocabulary items that were considered to be medium or hard (Kelley & Kohnert, 2012). This may suggest that infants would have a cognate advantage in any vocabulary task—either receptive or expressive, especially for less-familiar words where they may use the cognate word they have already acquired for help (Kelley & Kohnert, 2012), which is the case when infants are acquiring new words and learning to pronounce them. Therefore, we could expect a cognate advantage in both comprehension and production, serving different purposes: in comprehension, a cognate advantage would help activate the representations for the words in both languages, whereas in production, cognates may also facilitate the acquisition of the word in the individuals' other language in terms of pronunciation, as was seen in our study. Future research could explore the difference between comprehension and production in bilingual infants' language acquisition while simultaneously looking at the cognate advantage. Moreover, future studies could also consider looking into the advantage for cognates and its impact on bilingual children's online production or pronunciation of cognate words.

Conclusion

The present study demonstrated that French–English bilingual infants show an advantage for cognates in vocabulary production, with proportionally more cognates being produced than non-cognates. This finding can, at least in part, explain why children learning typologically similar languages show faster vocabulary growth than those learning more distant languages (Blom et al., 2020; Gampe et al., 2021; Sheng et al., 2016). Altogether, our study provides a greater understanding of the effect of similar-sounding words on infants' language acquisition over time. Future studies with data from other populations of bilinguals will be important to more fully understand the effect of the cognate advantage in early bilingual vocabulary development.

Acknowledgements

The present research was approved by the Human Research Ethics Committee at Concordia University [certification #10000439]. We are grateful to all the families who participated in this research. This manuscript derives from a thesis submitted by Lori Mitchell in partial fulfilment for an honours degree in Psychology at Concordia University. This work was supported by grants to Byers-Heinlein from the Natural Sciences and Engineering council of Canada [2018-04390]) and the National Institutes of Health [1R01HD095912-01A1]. Byers-Heinlein holds the Concordia University Research Chair in Bilingualism and Open Science. This study has been presented as a poster at the 2021 Boston University Conference on Language Development, and the data has also been presented as a talk at the 2022 International Congress of Infant Studies. We thank the members of the Concordia Infant Research Lab for their comments on earlier versions of this paper.

References

Bates, M., D., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134. https://doi.org/10.1016/j.cognition.2015.03.013

Blom, B., E., & Timmermeister, M. (2019). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171. https://doi.org/10.1177/0142723719892794

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322. https://doi.org/10.1177/0165025414532559

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *21*(1), 29–49. https://doi.org/10.1207/S15327078IN0201_3

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2019). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism*, *22*(4), 371–385. https://doi.org/doi.org/10.1080/13670050.2016.1254152

Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jardak, A., … Tamayo, M. P. (2020). MAPLE: A multilingual approach to parent language estimates. *Bilingualism: Language and Cognition*, *23*(5), 951–957. https://doi.org/10.1017/S1366728919000282

Caselli, B., M. C., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159–199. https://doi.org/10.1016/0885-2014(95)90008-x

Choi, H. (2019). Strategy training for English-French cognate awareness: Contributions to Korean learners' L3 French competency. *Electronic Journal of Foreign Language*

Teaching, *16*(1), 68–79. Retrieved from https://e-flt.nus.edu.sg/wp-content/uploads/2020/09/choi.pdf

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*(2), 441–469. https://doi.org/10.1017/S0305000903005579

Costa, C., 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), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism*, *10*(3), 331–347. https://doi.org/10.1177/13670069060100030401

DeAnda, S., & Friend, M. (2020). Lexical-semantic development in bilingual toddlers at 18 and 24 months. *Frontiers in Psychology*, *11*, 3500. https://doi.org/10.3389/fpsyg.2020.508363

DeAnda, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). Lexical processing and organization in bilingual first language acquisition: Guiding future research. *Psychological Bulletin*, *142*(6), 655–667. https://doi.org/10.1037/bul0000042

Demke, G., T. L., & Siakaluk, P. D. (2002). The influence of exposure to phonological neighbours on preschoolers' novel word production. *Journal of Child Language*, *29*(2), 379–392. https://doi.org/10.1017/S0305000902005081

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories (CDIs)* (2nd ed.). Baltimore, MD: Brookes Publishing.

Floccia, C., Luche, C. D., 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.

https://doi.org/10.1016/j.jml.2019.104086

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. https://doi.org/10.1111/cogs.12847

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Gampe, Q., A., & Daum, M. M. (2021). Does linguistic similarity affect early simultaneous bilingual language acquisition? *Journal of Language Contact*, *13*(3), 482–500. https://doi.org/10.1163/19552629-13030001

Gonzalez-Barrero, A. M., Schott, E., & Byers-Heinlein, K. (2020). *Bilingual adjusted vocabulary: A developmentally-informed bilingual vocabulary measure*. Preprint. https://doi.org/10.31234/osf.io/x7s4u

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*(3), 609–627. https://doi.org/10.1198/106186006X137047

Hartig, F. (2022). *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. R package version 0.4.5. Retrieved from https://CRAN.R-project.org/package=DHARMa

Havy, B., M., & Nazzi, T. (2016). Phonetic processing when learning words: The case of bilingual infants. *International Journal of Behavioral Development*, *40*(1), 41–52. https://doi.org/10.1177/0165025415570646

Hoff, E., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics*, *190*, 241–245. https://doi.org/10.1016/j.jpeds.2017.06.071

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.

https://doi.org/10.1111/cdev.13050

Jones, S., & Brandt, S. (2019). Do children really acquire dense neighbourhoods? *Journal of Child Language*, *46*(6), 1260–1273. https://doi.org/10.1017/S0305000919000473

Kelley, A., & Kohnert, K. (2012). Is there a cognate advantage for typically developing Spanish-speaking English-language learners? *Language, Speech, and Hearing Services in Schools*, *43*(2), 191–204. https://doi.org/10.1044/0161-1461(2011/10-0022)

Kuznetsova A., P. B., Brockhoff, & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or chien? Translation equivalents in the receptive and expressive vocabularies of young French–English bilinguals. *Journal of Child Language*, *44*(4), 881–904. https://doi.org/10.1017/S0305000916000295

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1. https://doi.org/10.1097/00003446-199802000-00001

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. https://doi.org/10.1017/S0305000909990055

Morin-Lessard, E., & Byers-Heinlein, K. (2019). Uh and euh signal novelty for monolinguals and bilinguals: Evidence from children and adults. *Journal of Child Language*, *46*(3), 522–545. https://doi.org/10.1017/s0305000918000612

Pearson, B. Z., Fernández, S. C., & Oller, D. K. and. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language*, *22*(2), 345–368. https://doi.org/10.1017/s030500090000982x

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence

2-year-olds' bilingual proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ramon-Casas, M., & Bosch, L. (2010). Are non-cognate words phonologically better specified than cognates in the early lexicon of bilingual children? In *Proceedings of the 4th conference on laboratory approaches to spanish phonology* (pp. 31–36). Sommerville, MA: Cascadilla.

Ramon-Casas, S., M., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*(1), 96–121. https://doi.org/10.1016/j.cogpsych.2009.02.002

Rocha-Hidalgo, J., & Barr, R. (2022). Defining bilingualism in infancy and toddlerhood: A scoping review. *International Journal of Bilingualism*, *0*(0). https://doi.org/10.1177/13670069211069067

Schelletter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism: Language and Cognition*, *5*(2), 93–107. https://doi.org/10.1017/S1366728902000214

Schepens, J., Dijkstra, T., Grootjen, F., & Heuven, W. J. B. van. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLOS ONE*, *8*(5), e63006. https://doi.org/10.1371/journal.pone.0063006

Schott, E., Moore, C., & Byers-Heinlein, K. (2022). *Banana and banane: Cross-language phonological overlap supports bilingual toddlers' word representations*. Preprint. https://doi.org/10.31219/osf.io/hgdvq

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, *12*(6), 874–887.

https://doi.org/10.1111/j.1467-7687.2009.00829.x

Sheng, L., L., & Fulton, A. (2016). A robust demonstration of the cognate facilitation effect in first-language and second-language naming. *Journal of Experimental Child Psychology*, *141*, 229–238. https://doi.org/10.1016/j.jecp.2015.09.007

Singh, L. (2014). One world, two languages: Cross-language semantic priming in bilingual toddlers. *Child Development*, *85*(2), 755–766. https://doi.org/10.1111/cdev.12133

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, *36*(2), 291–321. https://doi.org/10.1017/S030500090800891X

Trudeau, N., Frank, I., & Poulin-Dubois, D. (1999). Une adaptation en français québecois du MacArthur Communicative Development Inventory [a Quebec French adaptation of the MacArthur Communicative Development Inventory]. *Revue d'orthophonie Et d'audiologie*, *23*, 31–73.

Tsui, R. K.-Y., Gonzalez-Barrero, A. M., Schott, E., & Byers-Heinlein, K. (2022). Are translation equivalents special? Evidence from simulations and empirical data from bilingual infants. *Cognition*, *225*, 105084. https://doi.org/10.1016/j.cognition.2022.105084

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, *5*, 311–326. https://doi.org/10.1017/S0305000900007492

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. https://doi.org/10.1016/j.jecp.2012.08.001

White, M., A., & Storms, G. (2017). Convergence in the bilingual lexicon: A pre-registered replication of previous studies. *Frontiers in Psychology*, *7*, 2081. https://doi.org/10.3389/fpsyg.2016.02081