Cognates are advantaged in early bilingual expressive vocabulary development

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

[1] Concordia University

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 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 infants 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 children produced a greater percentage of cognate words than non-cognate words on the MacArthur-Bates Communicative Development Inventories. Moreover, the magnitude of the cognate advantage increased with age. The findings suggest cognate learning is facilitated in early bilingual vocabulary development. Just as in monolingual infants, these results suggest that phonological overlap supports bilingual language acquisition.

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

Cognates are advantaged 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 appear 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" in English and *"banane"* 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), more 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 vocabulaires grow

(Legacy et al., 2017). By the age of 27 months, bilingual toddlers recognize target words more accurately when preceded by its translation equivalent (Floccia et al., 2020).

Some translation equivalents share form and thus phonological overlap between languages, typically due to a shared etymology; these are called cognates. Cognates range in their degree of phonological similarity: For example, English "banana" and French *"banane"* are identical except for their last phoneme, while English "pants" and French *"pantalon"* differ across multiple phonemes, and even have different number of syllables. Some typologically close languages even have form-identical cognates, such as the word *"si"* which means yes in both Spanish and Catalan.

Cognates appear to have a special status in bilingual language processing. Previous research has reported a cognate facilitation effect where bilinguals are better and quicker at identifying 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-language learners between the ages of eight and 13 years old, where the 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 4-8 year-old children (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 phonological neighborhood density effects, where monolingual 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), where 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 neighbours than those with few 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 phonological neighbours after learning new words facilitated the production of the new words. Therefore, learning a new word with close phonological neighbours seems to help learners maintain the new word in memory, therefore 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 18-36 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 report 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 *"si"* in both Spanish and Catalan) and form-similar (e.g., "hand" is *"mano"* in Spanish and "ma" in Catalan) cognates. Results indicated that 28% of the words produced by the bilingual infants was composed of 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 18-36 month-old sample studied by Gampe et al. (2021).

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 few to no form-identical cognates. Despite the presence of cognates, note that they belong to different language families: English is a Germanic language and French is a Romance language.

We collected monthly vocabulary data on English-French bilingual infants' word

production starting between the ages of 16–20 months and ending 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). From these two forms, we limited our analysis to 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 produced both "apple" and *"pomme"*, or both "banana" and *"banane"*, 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 where possible. .

We hypothesized that cognates would be more readily produced by English-French bilinguals than non-cognates. Thus, we predicted that English-French 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

Measures

Web-based MacArthur-Bates Communicative Development Inventory: Words and Sentences (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 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. 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. Although caregivers were asked to fill out the forms every month, it was possible for them to skip some months when necessary. 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 Zoom. This was repeated every five months to track any potential changes in the infant's language exposure.

Identification of Translation Equivalents and Cognates

Analytical Strategy

Results

Descriptive Measures of Number of Words Produced

Dependent Variable 1: Cognate Words Versus Non-Cognate Words

Complete List.

Matched List.

Dependent Variable 2: Cognate Pairs Versus Non-Cognate Pairs

Complete List.

Matched List.

Summary of Analyses

Taken together, our two sets of analyses revealed that overall bilingual infants produced more cognates than non-cognates. This pattern is modulated by age, where bilingual infants produced more cognates than non-cognates over time. This result was consistent not only across the two sets of analyses, but also across the complete and the matched list.

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 advantage for cognates in infancy and its magnitude was modulated by age. Although there are many more potential non-cognate than cognates available to French-English bilingual infants, our results showed that infants produced proportionally more cognates (e.g., English "banana"–French *"banane"*) than non-cognates (e.g., English "dog"–French *"chien"*) over time. Moreover, this pattern is unlikely to be due to confounding factors, since the same pattern of results was consistently found independent of whether one or both words in a translation equivalent pair are learned, as well as in a carefully matched list of translation equivalents that were matched for part of speech, typical age of acquisition, and word category when possible.

Together with previous findings, our results begin to paint a developmental picture of the effects of cognate status on early vocabulary productions. Both our results, as well as those of Bosch & Ramon-Casas (2014) indicate that vocabulary facilitation for form-similar cognates is either difficult to detect or absent at 18 months. Form-identical cognates may be already advantaged by 18 months (Bosch & Ramon-Casas, 2014), although this could not be examined in our study due to the paucity of form-identical cognates in French and English. From 18–27 months, our data indicate that bilingual children acquire form-similar cognates at a faster rate than non-cognates. This cognate facilitation effect can 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).

The advantage for cognates observed in bilingual children is likely to be due to an interconnected network between bilinguals' two languages. Previous studies have

documented that words from bilinguals' two languages are in fact linked and are processed in parallel. For example, cross-language priming studies on young bilingual children demonstrated that words in both languages were simultaneously activated when bilingual children were using related words in either language and this coactivation begins to emerge from 18 months (e.g., De Anda & Friend, 2020; Jardak & Byers-Heinlein, 2018; Singh, 2014). Moreover, phonology may play a role in this interconnected network, as similar-sounding words were shown to be coactivated upon hearing a phonologically-related word in one of the two languages (Von Holzen & Mani, 2012). In other words, words that are semantically- or phonologically-related are linked across bilinguals' two languages. It has been suggested that children more easily learn words that share associative cues and words acquired in one language facilitate the acquisition in bilingual children's other languages (Bilson et al., 2015). Cognates are possibly easier to learn than non-cognates because there are more associative cues. While both cognates and non-cognates share semantic overlap across languages, cognates share an additional phonological overlap. In other words, the phonological similarity in cognates facilitates vocabulary acquisition across languages.

The effect of phonological similarity would predict that the ease with which bilingual infants learn words across the two languages could be impacted by how similar words in their two languages are. For those who are learning close language pairs like Spanish and Catalan, their two languages share many words that are very similar-sounding to one another — sometimes share the identical forms (e.g, *"si"* meaning 'yes' in both languages). Form-identical cognates are likely to be more salient and frequent in the input and thus easier to acquire than form-similar cognates or non-cognates as only one form needs to be acquired for both languages (Bosch & Ramon-Casas, 2014). On the other hand, for those who are learning languages that share a lesser degree of phonological similarity like English and French, there would be fewer cognates available, and potentially very few form-identical cognates. Therefore, it is possible that when languages are very similar and

share many form-identical cognates, these types of cognates will be acquired preferentially by the infants, but when languages are somewhat less similar and share mostly form-similar cognates, these will be acquired preferentially instead. 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 acquisition of form-identical cognates, form-similar cognates, and non-cognates. This would further test if there is any possible difference in the nature of cognate advantage and if different language pairs may create a different schedule for vocabulary acquisition.

Despite the difference in the nature, 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.

However, an important consideration is that our study does not address whether the same cognate advantage would apply to vocabulary comprehension, although there is evidence that points in this direction. Young bilingual 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. There is an overall mixed evidence for whether the effect of cognate status is absent in comprehension (Schott & Byers-Heinlein, 2019) or is present in both comprehension and production (Kelley and Kohnert, 2012). However, it is possible that the

cognate advantage is modulated by the level of difficulty of the vocabulary item for both comprehension and production. It has been found that although the cognate advantage was found 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 sort of task, 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 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.

Conclusion

The present study evaluated whether an advantage for cognates is present in bilingual infants' early vocabulary productions. English-French bilingual infants' showed an advantage for cognates, with proportionally more cognates being produced than non-cognates, a pattern which magnified as the infants grew older and learned more vocabulary. In combination with previous results (Bosch & Ramon-Casa, 2014), these findings suggest a plausible developmental trajectory for the production of bilingual infants, starting with an advantage for the production of form-identical cognates followed by an advantage for form-similar cognates. Moreover, this cognate advantage may also be the first step in the development of bilingual infants' productive vocabularies, which later results in the cognate facilitation effect observed in childhood and adulthood. These findings may affect our expectations from bilingual infants in terms of bilingual language

productions such as the age at which certain translation equivalent words and/or pairs would be produced. Furthermore, they suggest that we may need to have different expectations regarding language development for bilingual infants learning different language pairs, such as the types of translation equivalent words/pairs that are produced and when they emerge. 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.