

Dual language input and the impact of language separation on early lexical development

Maria Julia Carbajal  | Sharon Peperkamp 

Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives, ENS, EHESS, CNRS, PSL University, Paris, France

Correspondence

Maria Julia Carbajal, Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France.
Email: carbajal.mjulia@gmail.com

Funding information

Ecole des Neurosciences de Paris; Agence Nationale de la Recherche, Grant/Award Number: ANR-17-CE28-0007-01 and ANR-17-EURE-0017

Abstract

We examined properties of the input and the environment that characterize bilingual exposure in 11-month-old infants with a regular exposure to French and an additional language, and their possible effects on receptive vocabulary size. Using a diary method, we found that a majority of the families roughly followed a one-parent–one-language approach. Yet, the two languages co-occurred to various extents within the same half-hour both within and across speakers. We used exploratory correlation analyses to examine potential effects of the dual input on the size of infants' vocabularies. The results revealed some evidence for an impact of language separation by speakers.

1 | INTRODUCTION

Quantity and quality of input, as well as environmental factors, have long been argued to play an important role in language acquisition (Hoff, 2006). For instance, word frequency and syntactic complexity in the speech that children hear, socioeconomic status, and maternal responsiveness have all been found to influence language skills in monolingual children (Fernald, Marchman, & Weisleder, 2013; Hoff & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Tamis-LeMonda, Bornstein, & Baumwell, 2001; Weisleder & Fernald, 2013). For infants growing up in a bilingual environment, language exposure may differ in many more aspects. Bilingual children might vary in the contexts in which each language is used, in the characteristics of language use of their communicative partners (such as language choice, or linguistic proficiency), and in the degree of similarity between the target languages, among many other properties. The large number of potential sources of variability makes describing the bilingual experience a difficult task. However, in a world where an estimated half of the population is bilingual (Grosjean, 2010), characterizing the bilingual input and its impact on language acquisition is of great importance.

Thus far, the most investigated aspect of bilingual input has been the amount of exposure to each language (e.g., Cattani et al., 2014; David & Wei, 2008; Garcia-Sierra et al., 2011; Hoff et al., 2012; Hoff, Welsh, Place, & Ribot, 2014; Marchman, Martínez, Hurtado, Grüter, & Fernald, 2017; Pearson, Fernandez, Lewedeg, & Oller, 1997; Place & Hoff, 2011, 2016; Poulin-Dubois, Bialystok, Blaye, Polonia, & Yott, 2013; Thordardottir, 2011). Unless speakers in bilingual families are more talkative than in monolingual families, bilingual children will receive less input in each language than their monolingual peers in their unique language. In a study of bilingual toddlers learning English and an additional language, Cattani et al. (2014) investigated the amount of exposure to English that is necessary to perform like monolingual English toddlers in several language tasks. Their findings showed that English would need to comprise at least 60% of the bilingual's total input. Furthermore, the relative amount of exposure to a given language has been found to correlate with phonological development (Garcia-Sierra et al., 2011), vocabulary size (Cattani et al., 2014; David & Wei, 2008; Hoff et al., 2012; Pearson et al., 1997; Place & Hoff, 2011, 2016; Poulin-Dubois et al., 2013; Thordardottir, 2011), and grammatical skills (Gathercole, 2002a, 2002b, 2002c; Hoff et al., 2012; Place & Hoff, 2016) in that language.

Most studies investigating the amount of exposure in the bilingual's input have estimated the global percentage of each language throughout the child's life by means of parental questionnaires (such as the Language Exposure Questionnaire of Bosch & Sebastián-Gallés, 1997, or the Language Exposure Assessment Tool of DeAnda, Bosch, Poulin-Dubois, Zesiger, & Friend, 2016). However, a bilingual's language ratio may vary in time, due to life events such as moving from maternal care to attending a daycare center, or spending time in a country with a different language. In a longitudinal study of 13 French–English bilingual children from 1 to 3 years old, David and Wei (2008) found that their relative productive vocabulary (e.g., the proportion of French words in their total lexicon) adapted to changes of language proportions in their input throughout the months. Thus, linguistic abilities at a given time point in the child's life may be affected by recent changes in their input.

Other properties of the bilingual experience have been less explored, but a growing body of research suggests that many other factors beyond the relative amount of exposure can have an effect on language outcomes (De Houwer, 2018; Gathercole, 2014; Hoff & Core, 2013). One particular aspect that has been discussed since the beginnings of research on bilingualism is the impact of language choice of the primary caregivers. French linguist Maurice Grammont, as cited by Ronjat (1913), was perhaps the first to suggest that language separation by speaker—that is, following a one-person-one-language (OPOL) approach—was necessary to guarantee successful bilingual development. For decades, this has been a common advice given to parents raising bilingual children, but most research was based on case studies, often by linguists raising their own children (see reviews by Barron-Hauwaert, 2004, and Yamamoto, 2001).

Over the past 20 years, several studies have investigated the role of parental language separation with larger sample sizes and more systematic methods (Byers-Heinlein, 2013; De Houwer, 2007; Lyon, 1996; Place & Hoff, 2016; Yamamoto, 2001). In a large study with nearly 2,000 bilingual families of school-aged children in a Dutch-dominant region of Belgium, De Houwer (2007) found that an OPOL approach was neither necessary nor sufficient to guarantee that children would become actively bilingual. Indeed, the percentage of families where children only spoke the community's majority language was similar in families that followed an OPOL approach (26%) and those where both parents spoke both the majority language and a minority language (21%). Unfortunately, the conclusions of this study are limited, as the data consisted of a short parental questionnaire, asking only which language(s) each parent used at home, and whether the child spoke only the majority language or the minority language, or both. Furthermore, the study only examined children aged 6–10 years old, all of whom were attending Dutch-speaking schools. It is not uncommon for bilingual children to prefer using the language spoken at school (see, e.g., Wong Fillmore, 1991); the role of language separation

by speaker may thus be different at earlier developmental stages, when children spend substantially more time with their parents.

Two recent studies have looked at the effect of parental language separation on vocabulary development in toddlers. First, Byers-Heinlein (2013) examined a heterogeneous group of bilingual toddlers in Vancouver, Canada, who were exposed to English and one of a variety of other languages. Using a questionnaire to measure the frequency of language mixing (defined as code switching or word borrowing) in the input, she found a negative correlation of such mixing with receptive vocabulary in 1.5-year-olds, and a trend in the same direction with productive vocabulary in 2-year-olds. Second, Place and Hoff (2016) used the same questionnaire in a group of Spanish–English bilingual 2.5-year-olds in South Florida, USA, and found no clear evidence of a negative impact of mixing. Taken together, the results on language separation by individual speakers could mean that it affects language development only during the earliest stages, but more research is needed to reach conclusions on the impact of this variable. Specifically, while code switching is potentially the most challenging condition for bilingual language development, there are other ways in which a bilingual parent can fail to keep their two languages separate (e.g., they can switch from one language to another in between sentences or conversations), which could also be taken into account.

Regardless of whether parents stick to the OPOL principle or not, bilingual children may differ in how often both languages co-occur in time. Place and Hoff investigated this issue as well, using a language diary method (originally designed by De Houwer & Bornstein, 2003) with their population of Spanish–English bilingual families. Parents were asked to report every half-hour who spoke to the child and which languages were used. Time blocks were categorized as English-only, Spanish-only, or mixed (this last category indicated that both languages were used, regardless of speaker). Seven diaries were collected for each toddler, each on a different day of the week, over the course of 7 weeks. A first study (Place & Hoff, 2011), with a sample size of 29 two-year-olds, revealed that the number of time blocks in which only one language was used was correlated with productive vocabulary size in that language. Yet, the complementary measure of number of mixed blocks did not correlate with either language's vocabulary. In a follow-up with 90 2.5-year-olds (Place & Hoff, 2016), mixed blocks were further categorized as English- or Spanish-dominant. A correlation was found between number of English-dominant blocks and several English language outcome measures, including productive vocabulary, while Spanish-dominant blocks were unrelated to Spanish language skills. As details about who was speaking each language during mixed blocks were not collected in either study, it is difficult to interpret which aspects of language separation may or may not influence development.

Some other properties of the bilingual environment that have been found to influence language development are the number of speakers of each language (Gollan, Starr, & Ferreira, 2015; Place & Hoff, 2011), the presence of siblings (Bridges & Hoff, 2014; Silven, Voeten, Kouvo, & Lunden, 2014), parental strategies and attitudes toward bilingualism (Juan-Garau & Perez-Vidal, 2001; Nakamura, 2016), and the use and status of each language in the community (Gathercole & Thomas, 2009). Because of the large amount of variability within and across bilingual populations, it is hard to draw general conclusions about the impact of specific factors on bilingual language development from any specific study. What is true for a given population may not be true for another one. For instance, as argued by Place and Hoff (2016), their observed lack of an effect of language mixing may be due to the fact that mixing is a common behavior in the Spanish–English bilingual community in South Florida where the study was conducted. In the heterogeneous population studied by Byers-Heinlein (2013) in Vancouver, however, language mixing may be less common, as more than half of the families did not belong to a large bilingual community, and hence in these families, the native speaker(s) of a

language other than English were more likely to interact with relatively few other speakers of their language.

Overall, despite great progress made so far in identifying potential factors that could impact bilingual development, much work remains to be done. The goal of the present study is to further the characterization of bilingual exposure, by exploring different properties of the dual input in a group of 11-month-old infants. This is an age at which very little is known about bilingual input, particularly regarding language separation; yet, the first year of life is crucial for language development, and parents' behavior with respect to language separation might be different before and after their infant begins to produce words. As the measures of language outcome are limited at 11 months, we will focus on input description and offer an exploratory analysis of the potential impact of certain input factors on receptive vocabulary only.

We examine a heterogeneous population, consisting of infants who live in Paris and who are regularly exposed to French and an additional language (AL). In order to capture different quantitative and qualitative aspects of their input, we designed a modified version of the language diary previously used by Place and Hoff (2011, 2016) and asked parents to fill it out on two different days, a weekday and a weekend day. As in previous diary studies, parents reported every half-hour all the people who spoke to the infant. Additionally, however, they specified, for each speaker and each time block, the language(s) used to talk to the infant and the one(s) used to talk to other people. This modification allows us to disentangle effects of co-occurrence of the two languages in time from within-speaker effects of dual language use. These language diaries were complemented with a custom-made Language Environment Questionnaire. Using various measures of bilingual input derived from the diaries, we explore the sources of variability that characterize the dual exposure of this population. Finally, we investigate how these sources may influence early lexical development, as measured through a short MacArthur CDI.

2 | METHODS

2.1 | Participants

Fifty-eight families with bilingual 11-month-old infants living in or near Paris (France) participated (25 girls, 33 boys; mean age: 338 days, range: 319–356 days). An additional twenty-five families were excluded from analysis: 19 because they did not send one or both diaries back, two because they completed both diaries on a weekday or weekend day, three because of incomplete diary data, and one because of missing vocabulary data. All infants heard both French and an additional language (AL) on a regular basis. The additional languages that the infants were exposed to were Spanish ($n = 13$), English ($n = 10$), Italian ($n = 9$), German ($n = 7$), Polish ($n = 3$), Arabic, Catalan, Portuguese, Romanian, Russian ($n = 2$ each), Bulgarian, Greek, Hungarian, Japanese, Mandarin, and Swedish ($n = 1$ each). Out of the 58 participants, 38 heard their AL mainly from their mothers, 11 from their fathers, six from both parents, and three from a nanny. According to parental report, the mean amount of exposure to AL at home was 52% (range: 15–98%). Outside of home, however, most infants were reported to hear a majority of French, as it is the dominant language in the region they were being raised in, with an average of 84% (range: 40–100%).

This study was conducted in accordance with the ethical standards of the American Psychological Association, and it was approved by the ethical review board of the Université Paris Descartes (CERES). Parental written consent was obtained prior to data collection.

2.2 | Materials and procedures

The families were invited to come to our babylab in Paris prior to their participation in the language diary study. During this visit, the infant participated in a behavioral experiment not reported on here, and each family completed a custom-made Language Environment Questionnaire to collect information regarding the general language background of the infant, as well as a vocabulary questionnaire in French and, when available, the adaptation of this questionnaire in the infant's additional language.¹ Then, the primary caregiver was given the materials and instructions for the completion of the diaries. We describe each of these assessment tools below.

2.2.1 | Language Environment Questionnaire

A detailed Language Environment Questionnaire was designed to collect information regarding the infant's general language background. We will refer to this questionnaire as LEQ, but it should not be confused with the Language Exposure Questionnaire of Bosch and Sebastián-Gallés (1997), which is also known as LEQ.

Our questionnaire contained questions regarding the family composition and the languages used by parents, siblings, and other caregivers. Additionally, we collected information about the four adults who most regularly interacted with the infant. This included a measure of their proficiency in each of the two languages (noted in a six-point scale ranging from *doesn't speak the language* to *native speaker*), estimations of the hours per day spent with the infant during weekdays and weekend days, and measures of their language use when talking to the infant and when talking to other people, using the same scale as the one provided in the diaries (see description below). Finally, we asked for an estimation of the proportions of exposure since birth to French and AL that their infant had heard at home and out of home.

2.2.2 | Language Diaries

Each diary was constructed as a booklet containing one page per half-hour slot, beginning at 7:00 in the morning and finishing at 20:30 in the evening (a sample page is shown in the Appendix). Each page contained five rows, which served to annotate each of the speakers that interacted with the infant and their language use. If more than five people were present at a given time, the four people who interacted the most with the infant would be noted in the first four rows, and a summary of the remaining people would be noted in the fifth row (number of people and average language use). This was done to simplify the task of the annotator, who would otherwise have been required to keep track of the language use of a large number of people, possibly leading to inaccuracies in their report.

In each row, speakers were identified by their roles in the infant's life, that is, as “mother” or “aunt #1.” For each speaker, two columns were used to annotate their language use, the first one corresponding to the language(s) used to speak to the infant, and the second one to the language(s) spoken to other people in the presence of the infant. A five-point scale was given as options of language use; to avoid any potential ambiguities, the booklets were adapted to each language pair by writing the name of the AL explicitly, for example, *only French*, *mostly French*, *both equally*, *mostly Spanish*, and *only Spanish*. Two additional options were given to cover alternative scenarios, *none* (i.e., no language was spoken) and *other (specify)*. Finally, two boxes at the bottom of each page were provided to indicate

¹If some or all of the vocabulary data could not be provided during the laboratory visit (for instance, if the person present did not speak French), the questionnaire(s) would be taken home to be completed and sent back with the diaries.

the location and activity (e.g., “In the kitchen, having breakfast”), as well as any additional comments. The diaries were written in French, and translations were made available in English and Spanish.

The parents and/or regular caregivers were asked to complete two diaries, one on a weekday and one on a weekend day of their choice, within the month following their visit. The chosen days should be as typical as possible, in order to capture the infant's daily routine. Given that infants often spend time with different people throughout the day, families were told that different people could complete different parts of the diaries, such that for each part of the day, the person who was most present with the child filled in the corresponding time slots. Detailed instructions for completing the diaries were included in the first page of the booklet.

2.2.3 | Vocabulary questionnaire

To assess the receptive vocabulary of the infants, we used a modified version² of the European French short-form adaptation of the MacArthur CDI (Kern, Langue, Zesiger, & Bovet, 2010). Parents were instructed to indicate whether their infant understands each of the words in the list. In the case of English, Spanish, and Portuguese bilinguals, families were also given the corresponding short-form adaptations³ in those languages (English: Fenson et al., 2000; Spanish: Jackson-Maldonado, Marchman, & Fernald, 2013; Portuguese: Frota et al., 2015). Each form was filled in by the main caregiver who spoke the respective language to the infant.

2.2.4 | Coding and pre-processing

Diaries were coded with one entry per time block and per speaker. Given the big inconsistencies in the way the number of daycare workers present during daycare hours was reported, we grouped them all into one single row per time block. Note that this does not affect calculations of language proportions during those hours, as French was the only language spoken at the daycare centers that the infants attended. It does, however, affect calculations of number of speakers.

Speakers' language use was translated into French and AL percentages. While the exact proportions of each language used in a given time block are not known, we estimated them as in (1):

$$\begin{aligned}
 & \textit{only French: } 100\% \textit{ French—}0\% \textit{ AL} \\
 & \textit{mostly French: } 75\% \textit{ French—}25\% \textit{ AL} \\
 & \textit{both equally: } 50\% \textit{ French—}50\% \textit{ AL} \\
 & \textit{mostly AL: } 25\% \textit{ French—}75\% \textit{ AL} \\
 & \textit{only AL: } 0\% \textit{ French—}100\% \textit{ AL}
 \end{aligned} \tag{1}$$

Additionally, for each speaker in each time block, we defined a measure of *within-speaker language purity* (WSLP) as the percentage of the most spoken language. For instance, if a person spoke 75% French and 25% AL (or 25% French and 75% AL), then their WSLP value would be 75%. Thus, WSLP ranges from 50% to 100%. This measure intends to capture how often speakers use both

²This modified version included all original 81 words plus 10 additional words that were being used in the behavioral experiment the infants participated in.

³These forms were also extended to include the translation of the French words used in the behavioral experiment. The number of words in each language (and, in parenthesis, the original number of words in each form) was as follows: 93 (90) in English, 92 (90) in Portuguese, and 113 (104) in Spanish.

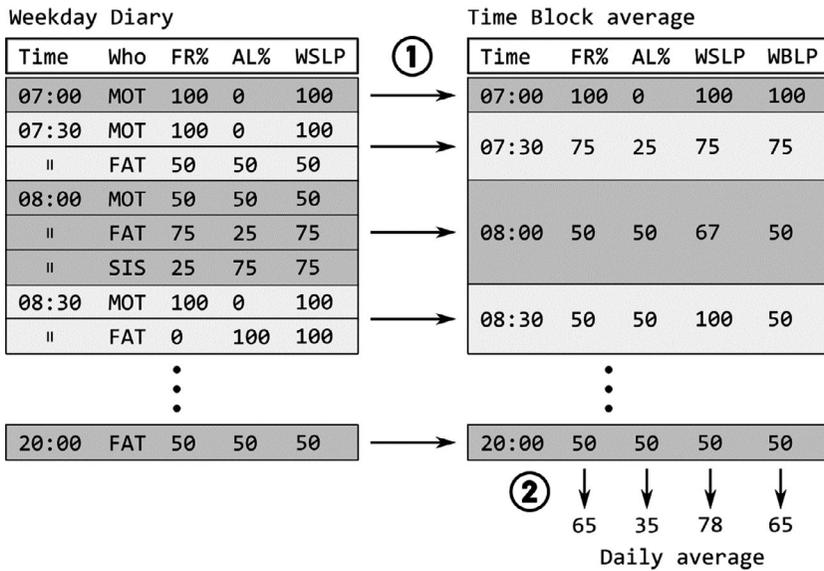


FIGURE 1 Example of computation of daily averages from direct speech during a weekday. (1) Mean percentage of French (here noted as FR), AL, within-speaker language purity (WSLP), and within-block language purity (WBLP) are computed for each time block (i.e., half-hour period) by averaging over all speakers. (2) Daily averages of percentage of French, AL, WSLP, and WBLP are obtained by averaging over all time blocks

languages within half an hour, regardless of whether code mixing took place or not. Finally, time blocks were classified as at home or out of home, based on the description of the place and the activity provided in the diaries.

Before processing the data, we filtered the diaries by keeping only close relatives and caregivers of the infant, that is, parents, siblings, grandparents, aunts/uncles, nannies, and daycare staff. From each diary, daily averages were computed as follows: First, an average of language use (% French, % AL, and WSLP) over all speakers was obtained for each half-hour. Here, we have made the simplifying assumption that the amount of speech is divided equally between all speakers within a given period of time, thus keeping the amount of speech per half-hour constant. Additionally, for each block we defined a measure of *within-block language purity* (WBLP) as the percentage of the most spoken language, regardless of speaker. For instance, if during a half-hour block the input was composed of 60% French and 40% AL (or 40% French and 60% AL), then WBLP would be 60%. Thus, WBLP ranges from 50% to 100%. Next, daily averages were calculated by averaging over all time blocks. These calculations were done separately for direct and indirect speech. An example of this process is shown in Figure 1.

Finally, weekly averages of all four measures were estimated by weighting the daily averages of weekdays $\times 5$ and of weekend days $\times 2$. Here, we have made the assumption that infants' routine reported on the weekday is likely to be repeated throughout the 5 days of the working week (and analogously for the weekend). Alternatively, if no assumptions regarding the routine were made, both diaries (weekdays and weekend days) could be given equal weights. A correlation analysis of the average exposure computed with and without applying different weights shows that the two measures are highly correlated ($r(114) = .95, p < .0001$), indicating perhaps that for many children the proportions of each language during weekdays and weekend days were similar. Thus, the non-weighted measure would likely yield comparable results in subsequent analyses.

Additionally, separate averages were obtained by speaker (i.e., by computing the mean percentages of French, AL, and WSLP over all rows of a given speaker over the two days, using the same weights as for the weekly averages) and by at-home or out-of-home location. The latter was done by first computing the time-block averages as explained in the first step of the daily averages, and then calculating separate averages over the blocks where the infant was at home and out of home, respectively. For the remainder of the paper, we will use the terms WBLP and WSLP to refer to the global weekly averages per infant, and ILP (individual language purity) to refer to the average WSLP calculated for each specific speaker.

3 | RESULTS AND DISCUSSION

We will first describe properties of the bilingual environment based on the Language Environment Questionnaire, followed by a characterization of the infants' input based on the language diaries. Finally, we will describe the vocabulary scores and examine potential correlations with infants' exposure (derived both from the LEQ and from the language diaries) in each language.

3.1 | Language Environment Questionnaire

The LEQ provides a first overview of the language background of the bilingual infants. Almost all infants lived alone with their parents and siblings (if any). Only one family did not follow this pattern: The infant lived with her mother and grandparents, while seeing her father on the weekends. Out of the 58 infants, 42 were first-borns, 11 had one older sibling, and the rest had two or more older siblings.

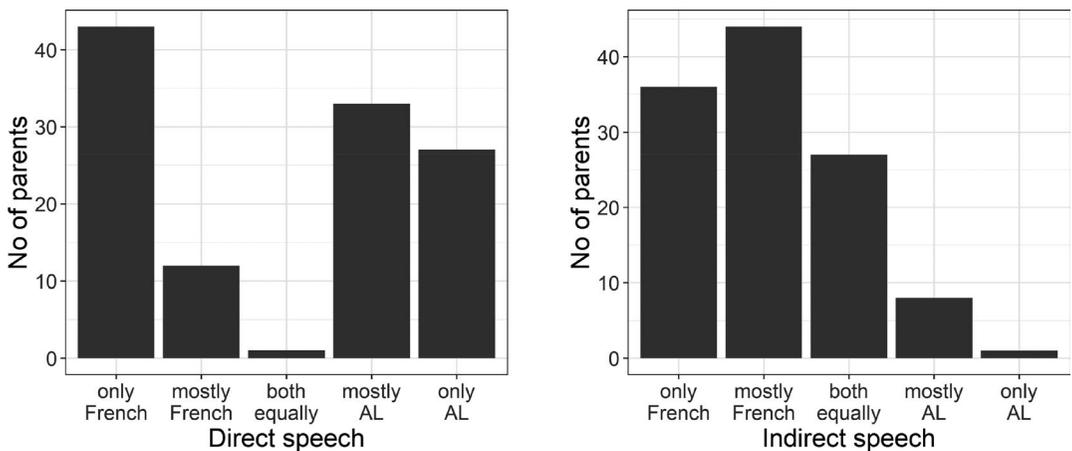
As mentioned above, the percentage of exposure to each language at home covered a wide range, from 85% French—15% AL, to 2% French—98% AL. For the great majority of the infants (49 out of 58), the parents used roughly the OPOL approach; that is, one parent spoke mostly or only French, and the other spoke mostly or only AL. Out of the 19 remaining cases, eight correspond to families in which both parents spoke the same language (three French-speaking families whose infant learnt AL from a nanny, and five AL-speaking families whose infants learnt French from a nanny or at daycare), and one corresponds to a family in which one parent spoke both languages equally often, while the other spoke only AL. In general, siblings spoke more French than AL, with 62% speaking mostly or only French, and the rest speaking both languages equally. No siblings were reported to speak mostly or only AL.

Table 1 shows the reported language behavior of the parents when talking to the infant, depending on their proficiency in each language. Parents were coded as speakers of a given language if they self-reported being a native or native-like speaker. If they reported at least native-like proficiency in both languages, they were coded as bilinguals; otherwise, they were coded as monolinguals. As can be seen in the table, bilingual parents generally used one language with the infant, most often the language not spoken by the other parent. In all cases in which both parents were bilingual, each one used a different language.

We then examined differences in the way parents used the languages with the infant and with other people. Figure 2 shows histograms of the languages used by fathers and mothers when addressing their infant and when talking to other people in the presence of their infant, as reported in the LEQ. Note that in spite of generally adhering to the OPOL approach, a fair amount of parents reported also speaking some amount of the other language when talking to their infant. This behavior was more common in AL speakers than in French speakers, which may be due partly to the influence of the community language, and partly to the fact that many AL-speaking parents were actually bilinguals

TABLE 1 Summary of parents' reported language use

Parents' languages	N	Main language used with the infant	
		Mother	Father
Both French monolinguals	3	French	French
French mother, AL father	3	French	AL
French mother, bilingual father	5	French	AL
Bilingual mother, AL father	5	French ($n = 3$), AL ($n = 2$)	AL
Bilingual mother, French father	20	AL	French
AL mother, French father	14	AL	French
AL mother, bilingual father	1	AL	Both equally
Both bilinguals	4	AL ($n = 3$), French ($n = 1$)	French ($n = 3$), AL ($n = 1$)
Both AL monolinguals	3	AL	AL

**FIGURE 2** Histograms of language use by parents in direct and indirect speech as reported in the LEQ (AL, additional language)

(52%, compared to 13% of French-speaking parents). Moreover, when talking to other people in the presence of their infant, parents used both languages more often, as can be seen by a prominent increase in *both-equally* responses, and a decrease in *only French* and *only AL* responses. Finally, when communicating to other people, French was used much more often than AL. This is not surprising given that French is the community language.

Finally, we examined secondary caregivers (nannies and daycare staff members). Twenty-two of the infants attended a daycare center regularly and had no nanny; 21 had a nanny but did not attend daycare; eight attended daycare and also had a nanny; and seven did not have any secondary caregivers. French was the only language spoken by daycare staff. On the other hand, out of 29 nannies, 22 spoke French to the infant, six spoke AL, and one spoke both languages. Thus, with the exception of these seven infants with AL or bilingual nannies, most infants heard their AL primarily from their close family.

As a preliminary conclusion, the data provided in the LEQ suggest that while our population is heterogeneous regarding the specific language pair, some aspects of their environment are common to

most infants. In particular, in this sample of a bilingual population in Paris, AL exposure came mainly from their close family, French had a strong presence outside of home, and most parents chose, to a certain extent, to keep languages separated. The information provided by the LEQ will serve as a complement to the information collected by the language diaries, which we describe next.

3.2 | Language diaries

First, in order to investigate the validity of the diaries as a way of estimating the infant's language input, we compared the exposure to AL both at home and out of home as calculated from the diary data, against the percentages reported by the parents in the LEQ. Since in the questionnaire we asked parents to consider all the speech the infant may have heard since birth when estimating the exposure to each language, we computed the diary estimates by pooling direct and indirect input.

For at-home exposure, we found a correlation of $r(56) = .70$ ($p < .0001$) between the parental and the diary estimations, indicating good agreement between both methods. For out-of-home input, a weak correlation was found ($r(55) = .25$, $p = .064$). However, this is likely due to the fact that most infants had a majority of French exposure outside their homes, according to both parental estimations and diary averages (for 71% of the infants, French comprised 75% or more of their exposure out of home according to both measures); the range of possible values is thus quite narrow. Furthermore, when estimating how much French and AL infants hear out of home, parents probably took into consideration a great amount of indirect French input from speakers who do not regularly interact with the infant (such as people on the street and in shops), while we only kept speech from regular speakers for our estimation. Thus, it is both possible that parents overestimated the weight of indirect French input and/or that we underestimated it. However, since all infants live in the same community, they are likely to have similar amounts of indirect French input from strangers out of home, and our estimations will therefore all be affected by a similar offset. Finally, since we will analyze direct and indirect input separately in the remainder of the study, this difference should be without consequences. We conclude that the diary estimates are overall in good agreement with the LEQ and are thus a reliable method to explore bilingual infant's exposure.

Next, we computed the average exposure to each language, separating direct from indirect input (regardless of whether it occurred at home or out of home). The distribution of exposure percentages for each language in direct and indirect speech is shown in Figure 3.

For direct speech, while the overall range of language exposure covers a wide spectrum (French min = 10%, max = 88%), on average infants heard more French than AL (mean French: 60%, mean AL: 39%). Three infants also heard a third language (OL), with a maximum of 18% of their direct input. For indirect speech, infants had a large majority of French exposure (mean French: 78%, mean AL: 20%). Eleven infants also heard a third or even a fourth language (OL), with a maximum of 23% of their indirect input. For 36 infants, the percentage of French in their indirect input was higher than in their direct input by at least five percentage points; for 14 others, there was a smaller increase; and for eight infants, there was a decrease in the amount of French (two of which were only caused by the presence of a third language, and not by an increase in AL). In total, only six infants had a higher amount of AL in indirect compared to direct input. However, infants' direct and indirect input showed a moderate correlation ($r(56) = .49$, $p < .0001$), indicating that the highest values of indirect AL input correspond to infants who also had relatively high amounts of direct AL input.

Infants' language exposure is not only characterized by how much of each language they hear, but also by how often the languages co-occur throughout the day. We thus examined the frequency of co-occurrence of both languages throughout the 2 days. By way of example, Figure 4 shows the percentages of direct French and AL for each half-hour block throughout weekdays and weekend days for

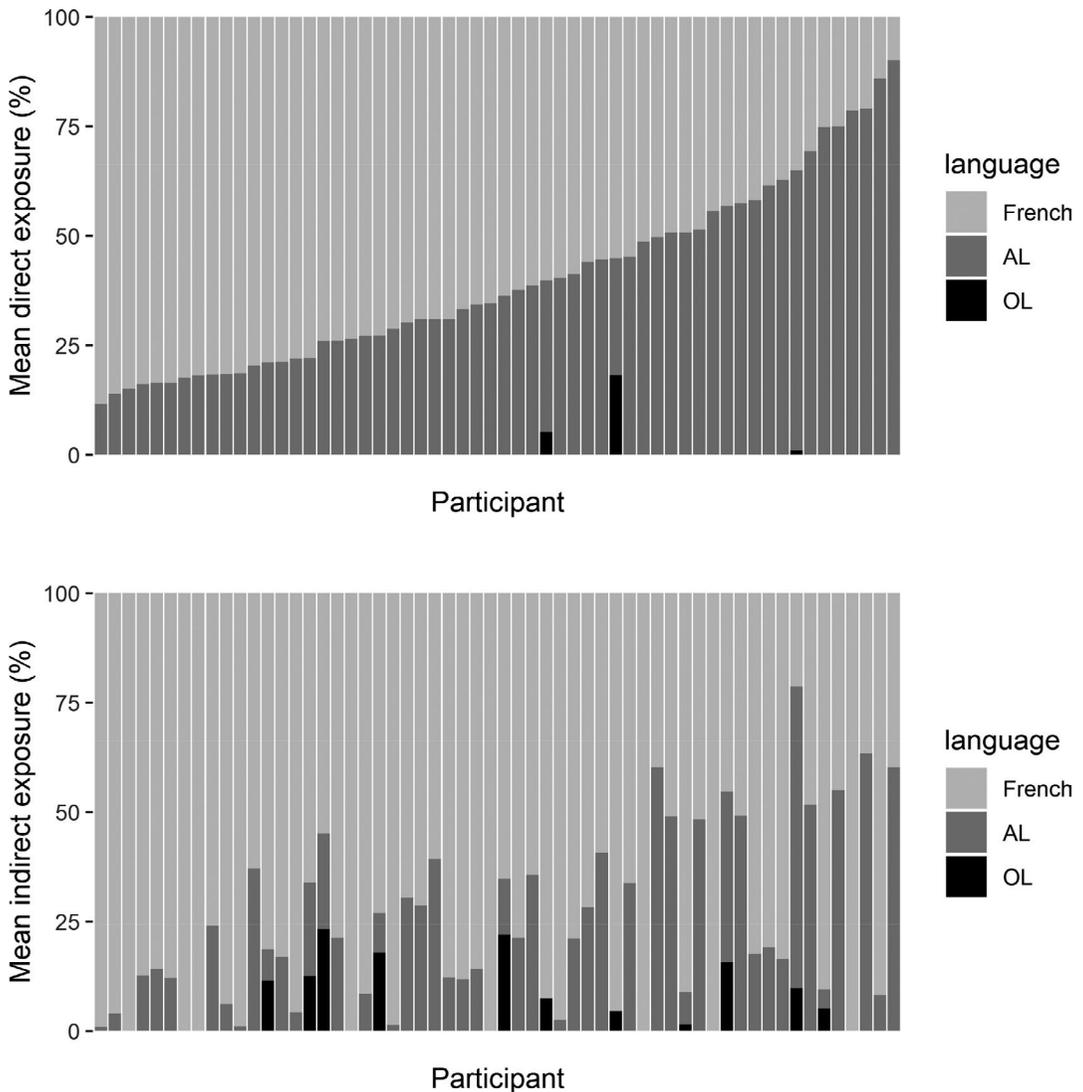


FIGURE 3 Estimates of infants' weekly average percentages of French, AL, and OL (other languages) in direct (top) and indirect (bottom) input based on their diary data. Participant order is the same in both panels

two infants with very similar average exposure to each language (infant *BB016*: mean French = 43%, mean AL = 57%; and infant *BB026*: mean French = 49%, mean AL = 51%).

Note that in spite of the similarity of their amounts of exposure to French and AL, these infants have drastically different experiences. In the case of *BB016*, the two languages are well separated in time: At home, AL is the only language spoken, while at daycare (from 9 a.m. to 6 p.m. on weekdays), only French is used. On the other hand, *BB026* often hears both languages within the same half-hour. This difference is quantified by the weekly average of within-block language purity (WBLP) of direct input, which was 97% for infant *BB016* and 77% for infant *BB026*. Overall, infants' average WBLP ranged from 64% to 99% in their direct input (mean = 84%, median = 85%, $SD = 6\%$), and from 63% to 100% in their indirect input (mean = 89%, median = 90%, $SD = 11\%$). This indicates that, for the great majority of infants, hearing both languages within the same half-hour was not uncommon.

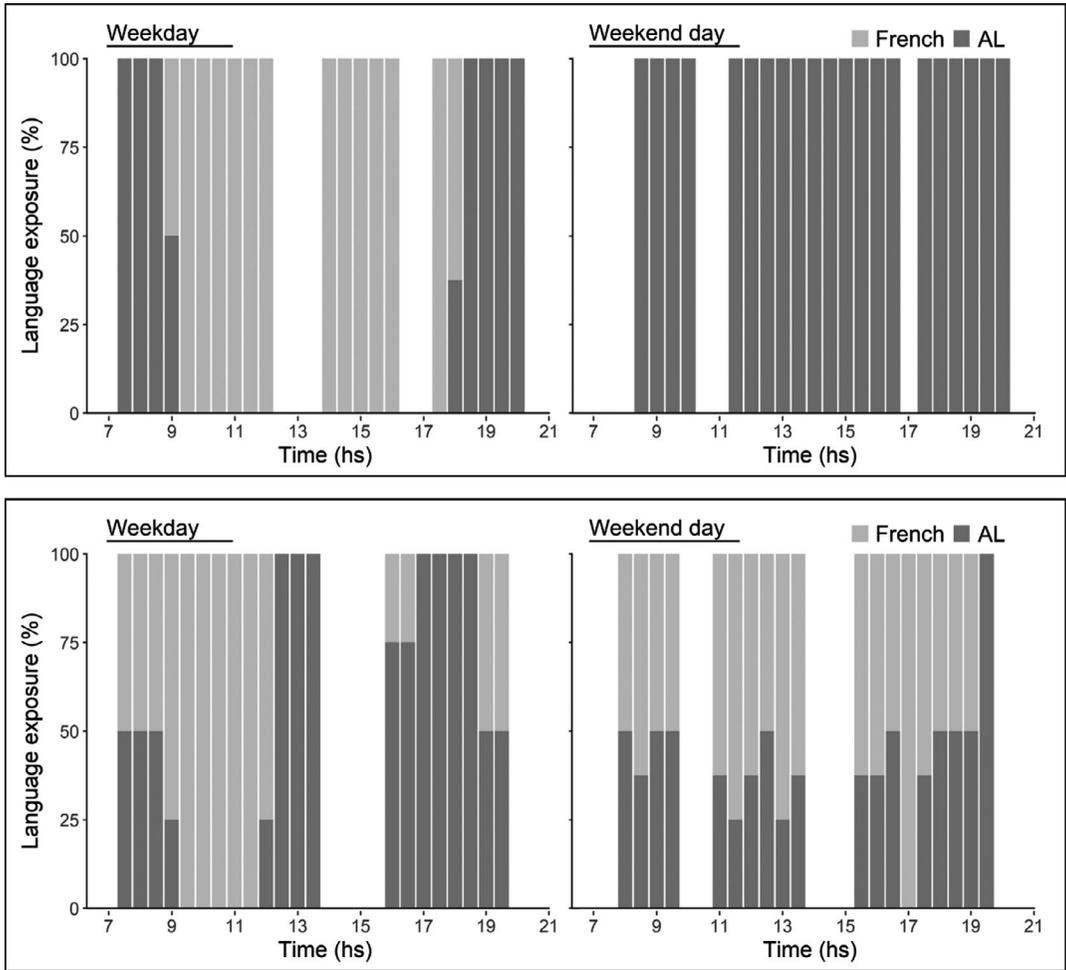


FIGURE 4 Proportions of direct French and AL exposure throughout the weekday and weekend day for infants *BB016* (top) and *BB026* (bottom). Empty time blocks represent time during which the infant did not receive any input (e.g., during naps)

The co-occurrence of both languages within the same half-hour block may be due to either the presence of bilingual speakers or the presence of monolingual speakers of two languages. To separate these cases, we computed the average within-speaker language purity (WSLP)⁴ for each child. To illustrate the difference between within-block language purity and within-speaker language purity, let us compare two infants who have similar averages of WBLP: *BB082* (77%) and *BB002* (79%). In the environment of infant *BB082*, speakers rarely used both languages when talking to the infant. The individual language purity (ILP) of the three most frequent speakers when talking to the infant was 100% from his father and daycare, and 99.3% from his mother. Overall, across two days and averaging all frequent speakers, the mean WSLP for this infant was 99.6%. On the other hand, in the environment of infant *BB002*, speakers sometimes used both languages. While this infant's father never used both

⁴Within-block language purity and within-speaker language purity are moderately correlated in direct speech ($r(56) = .35$, $p = .008$) and highly correlated in indirect speech ($r(56) = .90$, $p < .0001$). This is unsurprising, as infants whose caregivers frequently use both languages will inevitably encounter both languages co-occurring more often.

languages in the same half-hour block (ILP of 100%), his mother and sister often did, with an average purity of 84% and 73%, respectively. Across all frequent speakers throughout the two days, the mean WSLP for this infant was 85%. Overall, the average WSLP for all infants ranged from 81% to 100% in their direct input (mean = 95%, median = 98%, $SD = 5\%$), and from 65% to 100% in their indirect input (mean = 91%, median = 94%, $SD = 9\%$). The high median value of WSLP in direct input reflects a strong adherence to the OPOL principle by half of the families, while the other half allowed more or less language overlap by speaker.

It should be noted that the WSLP measure could be related, to some extent, to the balance between speakers' proficiency in both languages. Indeed, if a given speaker has very low proficiency in one of the two languages, the person will be less likely to use said language compared to someone who has equal proficiency in both. Thus, speakers showing an imbalance between their French and AL skills may be expected to have a higher individual language purity, and consequently contribute to a higher WSLP. To evaluate the influence of speakers' language balance in their amount of mixing, we built a linear mixed-effects model with the *packagelme4* (Bates, Maechler, Bolker, & Walker, 2015) in the *R* environment (R Core Team, 2017), including only data from parents for which we knew their proficiency level in each language based on the LEQ. For each speaker, we defined a "language balance" score as $5 - Abs(\text{French proficiency} - \text{AL proficiency})$. Thus, a score of 5 represents equal proficiency in both languages, while a score of 0 represents a lack of proficiency in one of the two languages and native proficiency in the other. We used each parent's ILP scores as dependent variable and included language balance (0 to 5), speech type (direct or indirect), speaker (father or mother), and the two-way interactions of language balance with speech type and with speaker as predictors.⁵ Subject (i.e., the infant) was used as a random variable.

The results revealed a main effect of language balance ($\beta = 2.30, p < .0001$), as well as a significant interaction of language balance by speech type ($\beta = -1.27, p = .005$). No other significant main effects or interactions were found. To investigate the interaction between language balance and speech type, we re-coded speech type using *treatment* coding and changing the baseline level, thus obtaining an estimate of the language balance effect for direct and indirect speech separately. This post hoc analysis revealed a significant effect of language balance only for indirect speech (direct: $\beta = 1.03, p > .1$; indirect: $\beta = 3.58, p < .0001$). Finally, using the *MuMIn* package in *R* (Barton, 2019), we calculated the percentage of the variance explained by the fixed effects of our model, showing that only 19% of the variance is captured by these variables. Taken together, these analyses show that while language balance plays a role in speakers' language separation, other contextual factors and parental attitudes (such as the decision to follow an OPOL approach) may play a bigger role in how well languages are separated by speakers in the child's environment.

Finally, we computed the number of speakers of each language that infants encountered throughout two days, that is, speakers who spoke on average over 50% of a given language when talking directly to the child. Despite a potential underestimation of the number of French speakers due to our data filtering process described earlier, infants encountered on average more French than AL speakers (French: $M = 2.4$, range = 1–5; AL: $M = 1.4$, range = 1–7; $t(57) = 4.24, p < .0001$). This difference is unsurprising given that French is the majority language in the region. Furthermore, we found moderate correlations between the number of frequent speakers of a given language and the percentage of exposure to that language (French: $r(56) = .45, p = .0004$; AL: $r(56) = .26, p = .047$).

⁵Variables *speech type* and *speaker* were "sum" coded. The estimates for these variables are thus to be interpreted as "main" effects.

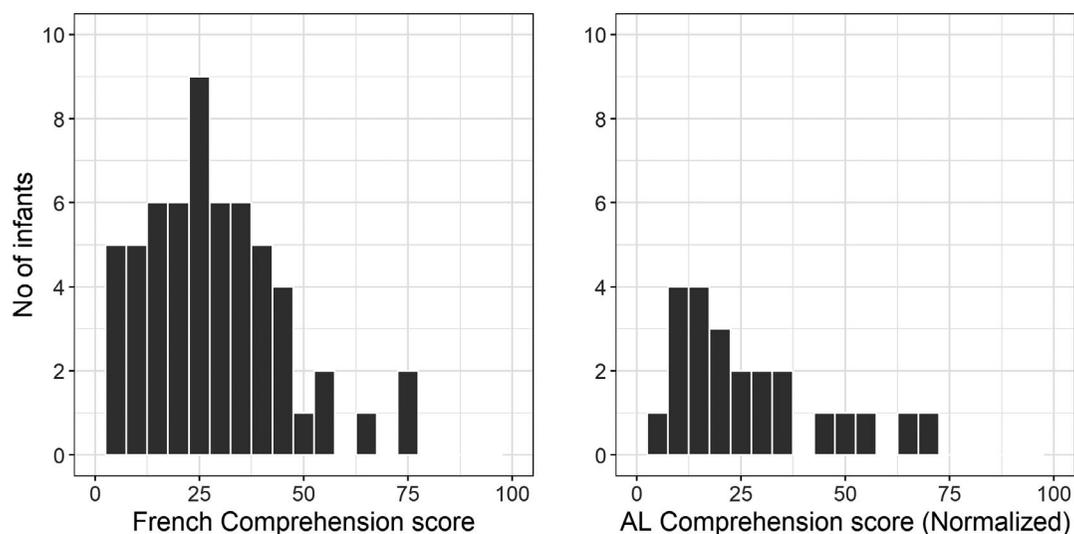


FIGURE 5 Histograms of bilinguals' vocabulary comprehension scores for French (left) and AL (right)

3.3 | Receptive vocabulary

Based on parental report in the short CDI, we computed for each infant a French comprehension score as the sum of all words the infant understands. On average, infants were reported to understand 28 out of 91 words (range: 4–75). For comparison, we looked at the comprehension scores of 64 11-month-old French monolinguals (mean age: 337 days) who participated in a laboratory study in our babylab during the same period. Their mean comprehension scores ($M = 28$ words, range: 7–83) were not significantly different from the scores of our bilingual infants ($t(119.5) = 0.18, p > .1$). This comparison suggests that, overall, our bilinguals are within normal ranges of receptive vocabulary for their age, at least in one of their two languages.

For 24 of the 58 infants, we also obtained the short CDI in their AL (namely in Spanish, English, and Portuguese). As these CDIs differ in their total number of words, for purposes of comparison we normalized the comprehension scores, dividing each score by the total number of words in the corresponding AL CDI, and multiplying it by 91 (the number of words in the French CDI). On average, infants were reported to understand 26 AL words (range: 0–71). A t test revealed a trend toward a smaller receptive vocabulary in AL compared to French for this subset of infants ($t(23) = -1.96, p = .06$). However, we found a high correlation of their comprehension scores across the two languages ($r(22) = .81, p < .0001$). The histograms of vocabulary comprehension scores for French and AL are shown in Figure 5.

3.4 | Effects of language exposure on receptive vocabulary

To explore possible effects of bilingual language exposure on language development, we computed first-order correlations between infants' French and AL comprehension scores on the one hand, and the different measures of bilingual exposure that we have defined on the other hand, that is, total percentage of exposure to French and AL, average within-block language purity, average within-speaker language purity, number of frequent speakers of each language, and maternal speaking behavior. Additionally, we include a measure of number of blocks with only French or only AL input, as it was found to have a significant effect on vocabulary size in Place and Hoff (2011, 2016). The density

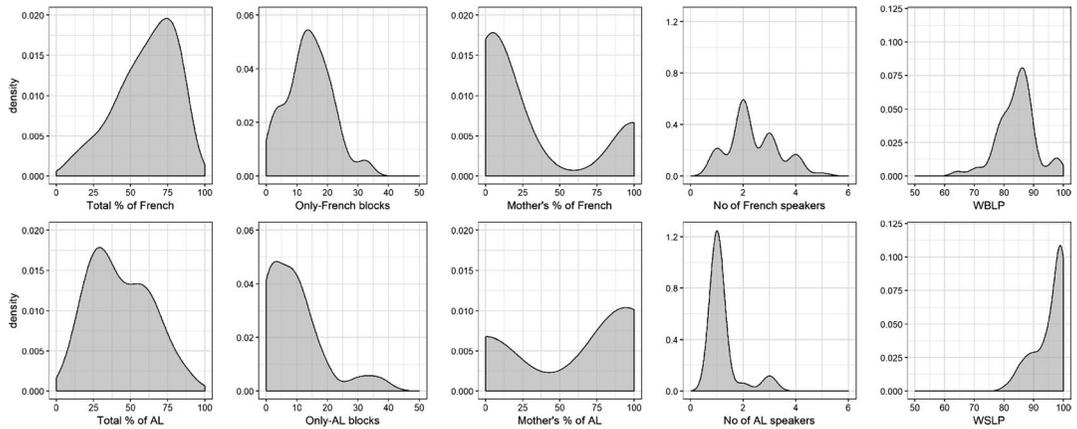


FIGURE 6 Density distributions of all input measures used in the correlation analyses

distribution of all these measures is shown in Figure 6. For all these analyses, we took into account direct input only and excluded the data from two infants whose French comprehension scores are outliers (Tukey's criterion).

Tables 2 and 3 show the results of the correlations for French and AL, respectively. It should be noted that these correlations are intended as an exploratory analysis, and any observed effects should be confirmed in future studies.

The correlation analyses showed overall few significant results, especially on French comprehension scores. Most surprisingly, the size of the receptive vocabulary in a given language did not seem to be associated with the total percentage of direct input in that language. However, we found a positive correlation between receptive vocabulary size and the percentage of each language spoken by the mother: The percentage of AL in maternal input correlated with AL comprehension scores. Furthermore, the number of blocks where only AL was spoken revealed a positive, although not significant ($p = .05$), correlation with AL comprehension scores.

Interestingly, we found a positive effect of within-speaker language purity, but not of within-block language purity, showing a significant association with French comprehension scores. This effect suggests that the frequency with which speakers use both languages in the same time block may affect infants' receptive vocabulary, with higher language overlap resulting in lower vocabulary scores. This observation is in line with the results of Byers-Heinlein (2013) showing that parental language mixing had a negative effect on comprehension scores of 18-month-old infants.

Overall, these correlation analyses suggest that properties of the bilingual exposure may affect vocabulary development in 11-month-olds, in particular for the AL. As these analyses are exploratory, the effects remain to be confirmed with new data. A larger sample size (especially regarding the AL comprehension scores, for which only half of the participants could be included) will allow for further investigation of the relative contribution of each factor to bilinguals' vocabulary in each language.

4 | GENERAL DISCUSSION

We examined properties of the input and the environment that characterize bilingual exposure in 11-month-old infants growing up in Paris, France, and their possible effects on receptive vocabulary size. In order to capture these properties, we used a language diary method: Caregivers kept a record of

TABLE 2 Correlations between measures of French direct input and French vocabulary scores

Language exposure measure	Correlation with French vocabulary (<i>N</i> = 56)	
	<i>r</i>	<i>p</i>
Total percentage of French exposure	-.04	>.10
Number of French-only blocks	.00	>.10
Percentage of French in mother's input	-.18	>.10
Number of French speakers	-.11	>.10
Average within-block language purity (WBLP)	.12	>.10
Average within-speaker language purity (WSLP)	.36	.007

TABLE 3 Correlations between measures of AL direct input and AL vocabulary scores

Language exposure measure	Correlation with AL vocabulary (<i>N</i> = 24)	
	<i>r</i>	<i>p</i>
Total percentage of AL exposure	.21	>.10
Number of AL-only blocks	.40	.05
Percentage of AL in mother's input	.57	.004
Number of AL speakers	.18	>.10
Average within-block language purity (WBLP)	.07	>.10
Average within-speaker language purity (WSLP)	.35	.09

their infant's language input every half-hour throughout 2 days. This information was complemented with a Language Environment Questionnaire.

Our sample of 58 infants differed from those in previous diary studies by Place and Hoff (2011, 2016) in two main aspects. First, the participating families did not belong to a large homogeneous bilingual community. The infants in our study were regularly exposed to their community's majority language (French) and one of a number of additional—minority—languages (AL). Second, in our study infants were much younger than those in previous research. Thus, this study provides information regarding the environment and input of early bilinguals in a population that has not been previously described in depth.

Bilinguals' environments differed in many dimensions. In particular, we observed variability in the proportions of each language in the infants' input (covering a wide range from mostly French to mostly AL), in how often they encountered both languages within the same half-hour, in how frequently speakers in their environment used both languages within the same half-hour, and in the number of caregivers and frequent speakers who provided input in each language. However, we also found some characteristics that were common to most infants. The first one, and perhaps the least surprising, is that the community language had a strong presence in their environment: Infants encountered significantly more speakers of French than of their AL, most of them heard a great majority of French while out of home, and on average, infants heard more French than AL in both their direct and indirect input. In general, exposure to a majority language is likely to be stable and similar across infants who are raised in a bilingual context within a monolingual community. The presence of the majority language may become even more dominant as children grow up, representing a potential risk for the development and maintenance of their AL, as observed for instance in Welsh–English

bilinguals growing up in an English-dominant region (Gathercole & Thomas, 2009). While this was not part of our research question, the observation that a majority of the siblings of our participants spoke only French suggests that this may be the case.

A second, and less expected characteristic of our sample, is that a majority of the participating families adhered to the one-parent–one-language (OPOL) approach, although in practice many parents used both languages to address their infants from time to time. It should also be noted that they generally did not respect OPOL while talking to other people in the presence of the child, sometimes even introducing a third language. Although direct input is usually the focus of studies on language development, it is likely that infants exploit at least part of their indirect input, and therefore, bilinguals will have to deal with an imperfect language separation. Moreover, the co-occurrence of both languages within the same half-hour block was generally very common, even in direct speech.

Are these observations typical of the general bilingual population in Paris, or are they peculiarities of our sample? While there are no similar studies to compare our results to, we can examine whether our pool of participating families is overall representative of the target population. A study based on the census conducted by the French National Institute of Statistics and Economic Studies (Insee) in 1999 showed that 18% of adults in France heard both French and an additional language (foreign or regional) from their parents at the age of 5, and 8% heard only the AL at home and French out of home (Clanché, 2002). Out of those who heard both languages at home, 55% heard mostly French and occasionally AL; 33% heard mostly AL and occasionally French; and the rest heard both frequently. This indicates that, as in our sample, a balanced bilingual exposure is not very common, and that French has a more dominant presence than AL. The dominance of the community language is also observed in the rate of transmission of the AL to the next generation: 70% of adults who heard AL during their childhood spoke only or mostly French to their own children.

Regarding the separation of languages, the same study revealed that approximately 28% of bilinguals (5% of the total population) heard a different language from each parent. Thus, the percentage of families included in the census taking an OPOL approach is significantly lower than that in our sample (84%). This difference could be due to the fact that the number reported in the census takes into account not only foreign languages but also regional ones, whose speakers are much more likely to be bilingual. Thus, it is possible that a separate count only for foreign languages reveals a higher proportion of OPOL. However, it is also possible that the particular families who volunteered to participate in our studies were somehow different from the general bilingual population. Indeed, these families had previously shown interest in our babylab, by registering in our database of prospective participants in reply to a letter sent to all parents living in Paris at the time their infant is 5 months old. Furthermore, registered families typically have a mid-to-high socioeconomic background. The particular families who volunteered to participate in our study might have shared a more positive attitude toward bilingualism and a particularly active role in their children's dual language development than those found in the general population. Their strong interest in transmitting both languages may have led them to adhere to the OPOL approach, which is perhaps the most well-known pedagogical method for early bilingualism. While our sample might not be representative of all heterogeneous bilingual populations, our results contribute new observations to a growing pool of studies aimed toward the discovery of similarities and differences across the diversity of bilingual contexts in early childhood.

A second goal of our study was to examine potential effects of the dual input (using the measures derived from the diaries) on the size of infants' French and AL receptive vocabularies. The correlation results revealed some positive associations of language exposure with AL vocabulary, namely the proportion of the maternal input in AL, and—as non-significant trends—the number of AL-only half-hour blocks and the within-speaker language purity (WSLP). In the case of French vocabulary, a (positive) correlation was only found for WSLP. Evidence for an impact of language separation by

speakers, as reflected by the latter correlation, is in line with previous results using a parental language mixing questionnaire in 1.5-year-olds (Byers-Heinlein, 2013) but not in older toddlers (Place & Hoff, 2016).

A question that emerges from these results is whether an association between WSLP and vocabulary scores could be indirectly reflecting the influence of other factors. For instance, the input of infants attending daycare may be expected to contain less language mixing and more French, which could positively impact their French vocabulary. However, a post hoc analysis reveals no difference in WSLP (Wilcoxon ranked sum test; $W = 508.5$, $p > .1$) or % of French exposure ($W = 328$, $p > .1$) between the infants who attended daycare ($N = 28$) and those who did not ($N = 28$). Furthermore, as discussed earlier, language purity in direct input does not seem to be significantly influenced by parents' balance in their language skills. Nevertheless, many other properties of mixed language input, such as the frequency of intra-sentential code switching and the amount of accented speech, could correlate with WSLP and may thus be underlying the association of this measure with vocabulary scores. A more detailed description of the input would be necessary to answer this question; follow-up studies should therefore include more fine-grained measures of the properties of language mixing, which can be obtained either through naturalistic recordings or through additional questions to be included in the diaries and the LEQ.

The absence of other effects often found in the literature (such as the overall proportion of the input in each language, or the number of speakers), especially on French vocabulary, may have several explanations. Firstly, different properties of dual language input may have different effects at different ages. Indeed, the infants we examined were much younger than those in previous studies. The importance of language separation in their input may be especially important during the first year of life, as infants need to discover the presence of two languages in their environment, but decrease over time. Meanwhile, other input factors found to be relevant for older infants could play a smaller role early on and become more relevant at a later stage. Future research would thus benefit from a longitudinal study evaluating the effects of a range of factors on bilinguals' lexical acquisition at different developmental stages.

Secondly, some input measures defined in this study may not be the most relevant factors in language development for this specific population of bilingual French–AL infants living in a French-dominant region. For instance, the previous finding that the number of speakers of each language and the number of one-language-only half-hour blocks affect vocabulary size in Spanish–English 2-year-olds (Place & Hoff, 2011, 2016) might be due to the fact that those studies were conducted in a region with a large number of monolingual Spanish, monolingual English, and bilingual Spanish–English speakers (i.e., South Florida, USA). That is, contrary to infants in our sample, infants in those studies were more likely to be exposed to a large range of speakers and of one-language-only situations for both languages. Furthermore, for our population, French vocabulary may be relatively easier to acquire than AL vocabulary—regardless of the amount of input in French from regular speakers—given that it is the majority language. Indeed, previous studies with pre-school and school-aged children have found that the community's dominant language has an advantage over the minority language, with higher success rates in acquisition (e.g., De Houwer, 2007; Gathercole & Thomas, 2009; Yamamoto, 2001). In contrast with French, our bilinguals' AL development depends solely on the input coming from the few AL speakers in their environment, which are coincidentally among the main caregivers, possibly making the measures derived from the language diaries more relevant for AL than for French vocabulary acquisition.

Thirdly, the language diaries may have been too noisy to observe certain effects. While we found our diary estimates of language exposure to be in good agreement with the LEQ, it is possible that more fine-grained differences in infants' backgrounds are not observable with only two diaries. Moreover,

the real contribution of each speaker to the total input may have been misrepresented, as diary reports were separated by speaker and then weighted equally across speakers in a given time block. Note also that even if speakers' relative contributions were not misrepresented, the diaries do not capture infants' absolute input, which might be a better predictor of vocabulary outcomes (Marchman et al., 2017). As suggested by De Houwer (2018), complementing diary information with audio recordings may be necessary to get better estimates of the infant's input.

Finally, the outcome measure, which consisted of a short version of the CDI (containing less than 100 words), may be inappropriate to measure individual differences in bilingual infants' vocabulary. Mean vocabulary size was far from the upper limit (and no infant was reported to know all the words in the questionnaire), but whether an infant knows a word or not may be up to subjective impressions. In previous studies with 2-year-olds (Place & Hoff, 2011, 2016), only productive vocabulary was used as outcome measure, which is arguably easier to estimate than receptive vocabulary. This is not an option for 11-month-olds, though, since they hardly produce any words.

Beyond these caveats, it is worth recalling that infants in our study were exposed to a wide range of languages with varying cross-linguistic similarities with French. Language distance has been suggested to modulate language acquisition, affecting, for instance, phonological development (Bosch & Sebastián-Gallés, 2003; Sundara & Scutellaro, 2011), acquisition of translation equivalents (Bosch & Ramon-Casas, 2014), and grammatical structures (Döpke, 2000; Hulk & Müller, 2000; Müller & Hulk, 2001). However, it has only recently been taken into account in studies of lexical acquisition with heterogeneous populations (Floccia et al., 2018; O'Toole et al., 2017). For instance, Floccia et al. (2018) investigated the effect of language distance on vocabulary development in 400 toddlers learning English and an additional language in the UK. Based on measures of phonological overlap, morphological complexity, and word order typology, they showed a cross-linguistic effect of language distance on vocabulary size: Close language pairs resulted in higher vocabulary sizes in the additional language.

These recent results raise the question as to whether there are differences in vocabulary size across the language pairs included in our sample. While for most language pairs we only had one or two participants, we did have fair amounts of infants exposed to French plus either another Romance language (Catalan, Italian, Portuguese, Romanian, and Spanish; $N = 26$) or a Germanic language (English, German, and Swedish; $N = 18$). Interestingly, those in the former group seemed to have overall higher vocabulary scores in French (mean: 32, $SD = 15$) than those in the latter (mean: 19, $SD = 10$). Out of all language pairs, French–Spanish had the highest French vocabulary scores (mean: 38). Although we cannot draw conclusions from these observations, they are in line with previous studies suggesting that the specific language pair may play a role in bilingual language development. Particularly, language pairs from the same family (here, the Romance languages), which are likely to share a large number of cognates and structural similarities, resulted in higher French vocabulary scores. It should be kept in mind, however, that these results may be due not only to cross-linguistic differences (or similarities), but also to cultural differences. In conclusion, in studies comparing bilingual infants with a variety of language pairs, it might be necessary to include a larger sample that allows one to adequately measure language pair effects.

Given these large differences in vocabulary between language families, one might wonder if the strong effect that we observed of within-speaker language purity would still hold within each language group. We thus recalculated separately for the French + Romance and the French + Germanic groups the correlation of WSLP with French vocabulary scores. While both groups have a positive correlation coefficient, only in the French + Romance group is the correlation significant (Romance: $r(24) = .50$, $p = .01$; Germanic: $r(16) = .33$, $p = .19$). Again, as these are post hoc analyses with small sample sizes, this finding calls for a replication study. However, if confirmed, this might mean that by-speaker

language separation is only relevant to infants learning close language pairs. Further research should thus address the interaction of language distance and environmental language separation.

A final remark should be made regarding methodology. The language diary technique has several advantages and disadvantages over other methods of examining language input. On the one hand, in comparison with parental questionnaires that estimate the global percentage of exposure to each language, the language diaries offer more local estimates, which may not reflect the child's whole experience since birth. On the other hand, as parents report language use as it takes place, the diaries are less affected by their memory and biases, therefore giving a more accurate and detailed picture of their child's language exposure. Furthermore, as discussed earlier, bilinguals' language skills may be sensitive to changes in the amount of each language's input (David & Wei, 2008; De Houwer, 2009). The diaries thus offer the possibility to analyze the role of current exposure on language outcomes. While in this study we have only asked parents to fill in the diary during 2 days, this method can be used over longer periods of time, as has been done before by Place & Hoff, 2011, 2016 (one diary per week for 7 weeks), and De Houwer, 2011 (one diary per week over 15 months). An alternative technique to examine language exposure is the use of naturalistic audio or video recordings (e.g., De Houwer, 2014; De Houwer & Bornstein, 2016; Marchman et al., 2017; Ramírez-Esparza, García-Sierra, & Kuhl, 2017). While recordings provide an even more accurate measure of language input (it allows one, for instance, to compute the number of words per hour in each language, and to evaluate the amount of code switching), their processing is highly resource- and time-consuming. Existing semi-automatic systems, such as LENATM, have been mainly designed for monolingual speech and may not cope well with multilingual speakers. Current advances in speech technologies may provide a better automatized solution to this problem in the near future. Until then, the language diaries offer a good cost-benefit ratio and could easily be adapted to request additional information, such as how much each person talked or code-switched, to get a better picture of the child's language environment.

ACKNOWLEDGMENTS

This research was funded by a doctoral grant from *Ecole des Neurosciences de Paris* to M.J.C, and by grants from the *Agence Nationale de la Recherche* (ANR-17-CE28-0007-01 and ANR-17-EURE-0017). We would like to thank Erika Hoff for sharing the Language Diary used in Place and Hoff (2011) with us, Anne-Caroline Fiévet for assistance with recruitment, and all participating families.

CONFLICT OF INTEREST

The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Maria Julia Carbajal  <https://orcid.org/0000-0001-8275-1147>

Sharon Peperkamp  <https://orcid.org/0000-0001-5985-8878>

REFERENCES

- Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: The one-parent-one-language approach*. Clevedon, UK: Multilingual Matters.
- Barton, K. (2019). *MuMin: Multi-Model Inference*. R package (Version 1.43.6). Retrieved from <https://CRAN.R-project.org/package=MuMin>

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- 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. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33–69. [https://doi.org/10.1016/S0010-0277\(97\)00040-1](https://doi.org/10.1016/S0010-0277(97)00040-1)
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language specific vowel contrast in the first year of life. *Language and Speech*, 46(2–3), 217–243. <https://doi.org/10.1177/00238309030460020801>
- Bridges, K., & Hoff, E. (2014). Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied Psycholinguistics*, 35(2), 225–241. <https://doi.org/10.1017/S0142716412000379>
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16(1), 32–48. <https://doi.org/10.1017/S1366728912000120>
- Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreccx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language and Communication Disorders*, 49(6), 649–671. <https://doi.org/10.1111/1460-6984.12082>
- Clanché, F. (2002). *Langues régionales, langues étrangères: de l'héritage à la pratique*. Insée Première, no 830, February.
- David, A., & Wei, L. (2008). Individual differences in the lexical development of French-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 11(5), 598–618. <https://doi.org/10.1080/13670050802149200>
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411–424. <https://doi.org/10.1017/S0142716407070221>
- De Houwer, A. (2009). *Bilingual first language acquisition*. Clevedon, UK: Multilingual Matters.
- De Houwer, A. (2011). Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, 2, 221–240.
- De Houwer, A. (2014). The absolute frequency of maternal input to bilingual and monolingual children. *Input and Experience in Bilingual Development*, 13, 37–58.
- De Houwer, A. (2018). The role of language input environments for language outcomes and language acquisition in young bilingual children. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields*, (pp. 127–153). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- De Houwer, A., & Bornstein, M. (2003). Balancing on the tightrope: Language use patterns in bilingual families with young children. In *4th International Symposium on Bilingualism*, Tempe, AZ.
- De Houwer, A., & Bornstein, M. H. (2016). Bilingual mothers' language choice in child-directed speech: Continuity and change. *Journal of Multilingual and Multicultural Development*, 37(7), 680–693. <https://doi.org/10.1080/01434632.2015.1127929>
- DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, 59(6), 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234
- Döpke, S. (2000). Generation of and retraction from cross-linguistically motivated structures in bilingual first language acquisition. *Bilingualism: Language and Cognition*, 3(3), 209–226. <https://doi.org/10.1017/S1366728900000341>
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, 21(1), 95–116. <https://doi.org/10.1017/S0142716400001053>
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248.
- Floccia, C., Sambrook, T., Delle Luche, C., Kwok, R., Goslin, J., White, L., ... Mills, D. (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).

- Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., & Vigário, M. (2015). *Questionários MacArthur-Bates (CDI) para o português europeu: Formas reduzidas (8 aos 30 meses). II Jornadas-Comunicação e Desenvolvimento da Linguagem.*
- García-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., ... Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, 39(4), 546–557. <https://doi.org/10.1016/j.wocn.2011.07.002>
- Gathercole, V. C. M. (2002a). Command of the mass/count distinction in bilingual and monolingual children: An English morphosyntactic distinction. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children*, (pp. 175–206). Bristol, UK: Multilingual Matters.
- Gathercole, V. C. M. (2002b). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children*, (pp. 207–219). Bristol, UK: Multilingual Matters.
- Gathercole, V. C. M. (2002c). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children*, (pp. 220–254). Bristol, UK: Multilingual Matters.
- Gathercole, V. C. M. (2014). Bilingualism matters: One size does not fit all. *International Journal of Behavioral Development*, 38(4), 359–366. <https://doi.org/10.1177/0165025414531676>
- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language take-over, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12(2), 213–237. <https://doi.org/10.1017/S1366728909004015>
- Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin and Review*, 22(1), 147–155. <https://doi.org/10.3758/s13423-014-0649-7>
- Grosjean, F. (2010). *Bilingual*. Cambridge, MA: Harvard University Press.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(4), 215.
- 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. <https://doi.org/10.1017/S0305000910000759>
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- Hoff, E., Welsh, S., Place, S., & Ribot, K. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. *Input and Experience in Bilingual Development*, 13, 119–140.
- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), 227–244. <https://doi.org/10.1017/S1366728900000353>
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236. <https://doi.org/10.1037/0012-1649.27.2.236>
- Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. (2013). Short-form versions of the Spanish MacArthur-Bates communicative development inventories. *Applied Psycholinguistics*, 34(4), 837–868. <https://doi.org/10.1017/S0142716412000045>
- Juan-Garau, M., & Perez-Vidal, C. (2001). Mixing and pragmatic parental strategies in early bilingual acquisition. *Journal of Child Language*, 28(1), 59–86. <https://doi.org/10.1017/S03050009000004591>
- Kern, S., Langue, J., Zesiger, P., & Bovet, F. (2010). Adaptations françaises des versions courtes des inventaires du développement communicatif de MacArthur-Bates. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, 107(108), 217–228.
- Lyon, J. (1996). *Becoming bilingual: Language acquisition in a bilingual community*. Clevedon, UK: Multilingual Matters.
- Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, 20(1). <https://doi.org/10.1111/desc.12425>
- Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4(1), 1–21. <https://doi.org/10.1017/S1366728901000116>

- Nakamura, J. (2016). Hidden bilingualism: Ideological influences on the language practices of multilingual migrant mothers in Japan. *International Multilingual Research Journal*, 10(4), 308–323. <https://doi.org/10.1080/19313152.2016.1206800>
- O'Toole, C., Gatt, D., Hickey, T. M., Miękisz, A., Haman, E., Armon-Lotem, S., ... Kern, S. (2017). Parent report of early lexical production in bilingual children: A cross-linguistic CDI comparison. *International Journal of Bilingual Education and Bilingualism*, 20(2), 124–145. <https://doi.org/10.1080/13670050.2016.1179258>
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1), 41–58. <https://doi.org/10.1017/S0142716400009863>
- 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>
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 1/2-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041.
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, 17(1), 57–70. <https://doi.org/10.1177/1367006911431198>
- R Core Team (2017). *R: A language and environment for statistical computing (Version 3.3.3) [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish-English bilingual infants. *Child Development*, 88(4), 1216–1234. <https://doi.org/10.1111/cdev.12648>
- Ronjat, J. (1913). *Le Développement du Langage Observé chez un Enfant Bilingue*. Paris, France: H. Champion.
- Silven, M., Voeten, M., Kouvo, A., & Lunden, M. (2014). Speech perception and vocabulary growth: A longitudinal study of Finnish-Russian bilinguals and Finnish monolinguals from infancy to three years. *International Journal of Behavioral Development*, 38(4), 323–332. <https://doi.org/10.1177/0165025414533748>
- Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, 39(4), 505–513. <https://doi.org/10.1016/j.wocn.2010.08.006>
- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72(3), 748–767. <https://doi.org/10.1111/1467-8624.00313>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426–445. <https://doi.org/10.1177/1367006911403202>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- Wong Fillmore, L. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly*, 6(3), 323–346. [https://doi.org/10.1016/S0885-2006\(05\)80059-6](https://doi.org/10.1016/S0885-2006(05)80059-6)
- Yamamoto, M. (2001). *Language use in interlingual families: A Japanese-english sociolinguistic study*. Clevedon, UK: Multilingual Matters.

How to cite this article: Carbajal MJ, Peperkamp S. Dual language input and the impact of language separation on early lexical development. *Infancy*. 2020;25:22–45. <https://doi.org/10.1111/inf.12315>

APPENDIX
EXAMPLE OF A PAGE OF THE LANGUAGE DIARY

Time range	Person or people who were with the infant	Languages spoken to the infant	Languages spoken to other people in the room
7:00 to 7:30	Person #1: Who?	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____
	Person #2: Who?	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____
	Person #3: Who?	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____
	Person #4: Who?	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____
	Other people: How many? Who?	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____	<input type="checkbox"/> Only French <input type="checkbox"/> Mostly French <input type="checkbox"/> Both equally <input type="checkbox"/> Mostly English <input type="checkbox"/> Only English <input type="checkbox"/> None <input type="checkbox"/> Other: _____
Place and activity:			
Comments:			