results\_markdown

ASM

13/01/2023 last updated 01/05/2023

## Methods

### Participants

Families were contacted from a database of interested families in Montréal, Canada, largely via provincial birth lists, social media, and in-person recruitment, for example at libraries and community events. Data were collected through convenience sampling when they visited Concordia Infant Research Laboratory to participate in various studies examining bilingual language acquisition.

Our dataset consists of demographic and language exposure information collected between the years 2013 and 2020 via parental questionnaires and forms (described below). Caregivers filled out the questionnaires during each visit as part of the lab standard practice. At each visit, the infants also participated in one or more experimental tasks, and parents completed additional questionnaires, but these are not the focus of this research. All parents signed a consent form, and they were given a small thank-you gift for their participation. The current study was approved by Concordia University Human Research Ethics Board (Certification Number 10000439).

The original sample consisted of data from 552 participants. This sample included repeated measures from some families who visited the lab twice (142), three times (54), and four or more times (8). However, for our main analyses we decided to only keep one observation per family, thus we kept families who contributed data once as well as the data from the first visit of families who contributed data during multiple visits. We decided to keep data from the first visit only because that made the data from families with multiple visits and families with a single visit easier to compare.

Most infants in our data were exposed to two languages. We defined the first language (L1) as the language to which they were most exposed which ranged from 47 to 100 in our sample, as some infants were exposed to a small amount of a third language. We therefore defined the second language (L2) as the language to which they were second most exposed. We excluded 66 children who heard more than 10% of a third language from our original sample. Because the purpose of the present study is to explore bilingual family language strategies we also excluded families whose babies were getting monolingual input defined as hearing less than 95% of their L1 or less than 5% of their L2. We excluded a total of 185 families based on this criteria.

After exclusions, the final main analyses dataset consisted of 281 participants aged 4.34 to 30.95 months (M=14.41, SD=6.85). Of whom 138 were female. Our sample was composed of infants who heard 36.2989% of French, 54.8043% of English, and 18.5053% of a heritage language as L1; and 52.669% of which heard French, 34.1637% of English and 18.5053% of a Heritage language as L2. A heritage language was defined as any language other than French or English. There were two reasons behind this decision. First, in Montreal, Canada, both French and English hold a majority language status, making other languages heritage or minority languages. Second, infants who heard a language other than English and French heard one of a wide variety of the following languages: Kabyle, Russian, Romanian, German, Japanese, Arabic, Tamil, Mandarin, Spanish, Italian, Cantonese, Tagalog, Portuguese, Persian, Greek, Creole, Polish, Edo, Armenian, Hindi, Khasi, Berber, Korean, Lithuanian, Wolof, or Yoruba. Since it would have been impractical to divide the sample in so many language groups, and since those groups had few participants, we decided to collapse them into a single heritage language group.

While for the main analyses we excluded the second visit data of families’ who contributed data during multiple visits, we constructed a secondary longitudinal data set (n= 165) which included first and second visit data from these families. We used the longitudinal data for some descriptive analyses (described below). We however excluded the third and fourth visit data from both the main and the exploratory analyses because there were very few data points. We also excluded a second visit that occurred in less than 15 days from the first visit. The length between visits one and two varied between 15 and 655 days (m= 176.97days, sd= 138.88 days).

### Instrument

Information concerning the language environment of the infant was gathered through the Language Exposure Questionnaire (LEQ; Bosch & Sebastián-Gallés, 2001; See appendix A) following the Multilingual Approach to Language Estimates (MAPLE; Byers-Heinlein et al, 2019), in which interviewers walk caregivers through a series of questions designed to help them accurately remember and realize the language environment and exposure of their infant.

The first set of questions for the LEQ with MAPLE interview aim to understand how caregivers use their languages in daily life, particularly when talking to their infant. These questions use a likert type of scale to describe the frequency with which parents use each language when talking to their child (regularly, sometimes, or never). We used the responses from the first section of the LEQ questionnaire to characterize the family language strategy used by each family. As a reminder, parents answer the frequency with which they used each language, so the responses “regularly”, “sometimes”, “never” were given for each language. For example, for a family where both parents indicated they used both their first and second language regularly we assigned the both parents bilingual strategy. By contrast, for a family where each parent spoke a different first language regularly we assigned the one-parent-one-language strategy. In this manner we identified four distinct strategies: both parents bilingual, one parent bilingual, one-parent-one-language, and one language at home (see Table 1). In these subcategories parents that spoke their first language regularly and never spoke another language would be categorized as strict, and parents who each spoke their first language regularly but one of them sometimes would speak another language were categorized as flexible. As we discuss further below, these two categories were ultimately collapsed into a single one-parent-one-language category. Finally, we identified a few families with a single caregiver and we characterized them as single parent families, as we didn’t have enough cases to identify distinct strategies which single parents might use.

The second set of questions for the LEQ with MAPLE interview aim to give a set of estimates of the amount of time babies spend hearing each of their languages throughout their lives based on a detailed calculation made by the interviewer from caregiver responses (cumulative exposure estimate), a rough estimate made by the caregivers themselves (global exposure estimate), and an overall estimate that averages the interviewer and the caregiver estimations (overall exposure estimate). We only used data from the overall exposure estimate to determine how much exposure to each language infants in our sample were getting.

The overall exposure estimate made with LEQ with MAPLE have high reliability (r = .77–.97) when compared to transcriptions of daylong home recordings, which are a direct measure of the infants’ linguistic environment (Orena et al., 2018).

We additionally collected demographic information about our participants and their families via a short paper form.

## Results

### Data Pre-Processing

All the relevant LEQ data were manually entered from LEQ forms into excel spreadsheets. The data were entered twice by two different researchers to to catch errors, and discrepancies were resolved by a third individual who checked the original paper form. Demographic data were exported from Filemaker.

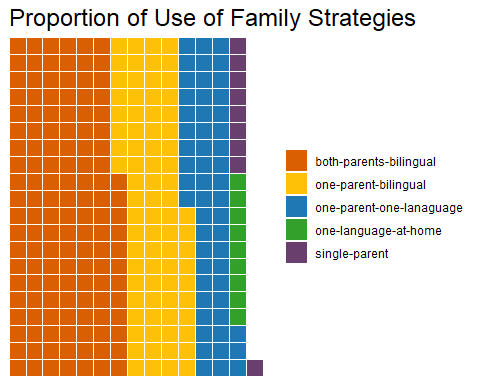
In preliminary analyses, we noticed that the one-parent-one-language strict and flexible categories both categories behaved similarly as predictors when performing statistical tests and other exploratory analyses. Moreover, splitting the categories split the sample size for the one-parent-one-language group, thus decreasing statistical power. Thus we ultimately decided to combine them into a single group.

Table 1

### Analytic approach

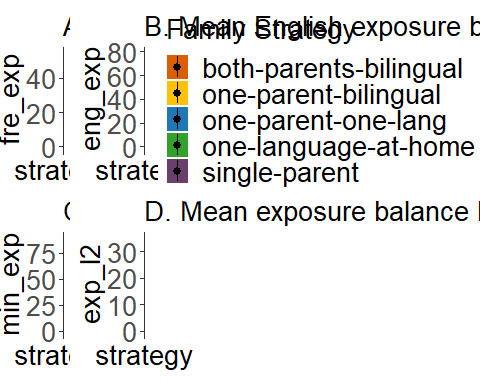
Once characterized, we described the proportions in which family language strategies were used in our sample (See Figure 1). The most common strategy was the both parents bilingual strategy which was used by 47% of the sample, and the least common strategy was the one language at home strategy (27.8%). One-parent-bilingual (27.8%), and one-language-at-home (27.8%) were intermediate. Note that we had 11 single parent families who were not classified in the avobe mentioned strategies.

Figure 1



We performed a chi-square goodness of fit test to determine whether the difference in proportion of usage between strategies was statistically significant, where the null hypothesis was that the strategies were used in the same proportion across the sample. The chi-squared goodness of fit test revealed a statistically significant difference between the distribution of proportions in our data and a distribution in which all the proportions are equal (X= 190.16, p= <.001).

We were also interested to learn how different family language strategies related to the overall amount of language exposure infants were getting. We observed that different family language strategies led to different a different mean exposure to French, English and Heritage languages (See Figure 2a, 2b, and 2c respectively).



To test this statistically, we conducted a one-way analysis of variance (ANOVA test) to explore weather some family language strategies contributed more variance to the amount of exposure babies got in French, English and/or a heritage language, regardless of whether each of those languages was their L1 or L2 language.

In terms of French exposure, we found a statistically significant difference in how much variance the family language strategies contribute (F=3.3972, NA, p= 0.0351, NA). We then conducted a Tukey HSD post-hoc analysis to determine which observed difference between means accounted for more variance(See Table 2). The post-hoc analyses showed that the one parent bilingual strategy was significantly different than the both parents bilingual, one-parent-one-language, and the one language at home strategies. The one parent bilingual strategy was associated with the lowest mean French exposure out all the strategies (see Figure 2a).

In terms of English exposure, we found a statistically significant difference in how much variance the family language strategies contribute (F=7.1862, NA, p= 0.0009, NA). We then conducted a Tukey HSD post-hoc analysis to determine which observed difference between means accounted for more variance(See Table 3). The post-hoc analyses showed again that the one parent bilingual strategy was statistically significantly different than the one-parent-one-language, and the one language at home strategies. However, in the case of English, the one parent bilingual strategy was associated with the highest mean English exposure out of all the strategies (See Figure 2b).This makes sense as most of the families in our sample were French-English bilinguals, and thus English exposure was in most cases a complement to French exposure.

Finally, in terms of exposure to a heritage language, we found a difference that approached statistical difference in how much variance the family language strategies contribute (F=3.5959, NA, p= 0.0347, NA). The fact that the the difference between strategies only approaches significance might be due to a lack of power, since the sample of families who used a heritage language was relatively small. We thus decided to also conduct a Tukey HSD post-hoc analysis. The post-hoc analysis showed that the one language at home strategy was statistically significantly different than the one-parent-one-language strategy. The one language at home strategy was associated with the lowest mean heritage language exposure, while the one-parent-one-language strategy was associated with the highest mean heritage language exposure (see Figure 2c).

Table 2. Tukey HSD Test on the Effect of Strategies on French Exposure

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | estimate | conf.low | conf.high | adj.p.value |
| strategy | OneParentBilingual-BothParentsBilingual | 0 | -6.505 | -15.1343 | 2.123 | 0.1791 |
| strategy | OneParentOneLang-BothParentsBilingual | 0 | 5.014 | -4.4413 | 14.470 | 0.4246 |
| strategy | OneParentOneLang-OneParentBilingual | 0 | 11.520 | 0.9072 | 22.132 | 0.0297 |

Table 3. Tukey HSD Test on the Effect of Strategies on English Exposure

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | estimate | conf.low | conf.high | adj.p.value |
| strategy | OneParentBilingual-BothParentsBilingual | 0 | 10.728 | 2.424 | 19.032 | 0.0072 |
| strategy | OneParentOneLang-BothParentsBilingual | 0 | -4.887 | -14.460 | 4.687 | 0.4520 |
| strategy | OneParentOneLang-OneParentBilingual | 0 | -15.615 | -26.175 | -5.055 | 0.0017 |

Table 4. Tukey HSD Test on the Effect of Strategies on Heritage Langauge Exposure

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | estimate | conf.low | conf.high | adj.p.value |
| strategy | OneParentBilingual-BothParentsBilingual | 0 | -16.211 | -33.90 | 1.476 | 0.0786 |
| strategy | OneParentOneLang-BothParentsBilingual | 0 | 1.985 | -19.05 | 23.022 | 0.9718 |
| strategy | OneParentOneLang-OneParentBilingual | 0 | 18.196 | -1.53 | 37.923 | 0.0762 |

Table 5. Tukey HSD Test on the Effect of Strategies on Balance of Exposure

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | estimate | conf.low | conf.high | adj.p.value |
| strategy | OneParentBilingual-BothParentsBilingual | 0 | -8.335 | -12.2285 | -4.442 | 0.0000 |
| strategy | OneParentOneLang-BothParentsBilingual | 0 | 4.672 | 0.2389 | 9.105 | 0.0362 |
| strategy | OneParentOneLang-OneParentBilingual | 0 | 13.007 | 8.1543 | 17.860 | 0.0000 |

for\_anova\_graphs <- for\_anova\_graphs %>%  
 filter(!strategy == "single-parent") %>%  
 filter(!strategy == "one-language-at-home")  
  
figure\_3\_a <-ggplot(for\_anova\_graphs, aes( x=age , y=fre\_exp, color=strategy, fill = strategy )) +  
 geom\_smooth(alpha=0.2)+  
 scale\_color\_manual (name = "Family Strategy", values =c(my\_orange, my\_yellow, my\_blue)) +  
 theme\_bw() +  
 ylim(0, 100)+  
 theme(axis.text.x = element\_text(size=20), axis.text.y = element\_text(size=20),axis.title = element\_text(size=20), plot.title = element\_text(size=20), legend.position = "none")+  
 labs(title= "A. French exposure changes by age divided by strategy")  
  
figure\_3\_b <-ggplot(for\_anova\_graphs, aes( x=age , y=eng\_exp, color=strategy, fill = strategy )) +  
 geom\_smooth(alpha=0.2)+  
 scale\_color\_manual (name = "Family Strategy", values =c(my\_orange, my\_yellow, my\_blue)) +  
 theme\_bw() +  
 ylim(0, 100)+  
 theme(axis.text.x = element\_text(size=20), axis.text.y = element\_text(size=20),axis.title = element\_text(size=20), plot.title = element\_text(size=20), legend.position = "none")+  
 labs(title= "B. English exposure changes by age divided by strategy")  
  
figure\_3\_c <-ggplot(for\_anova\_graphs, aes( x=age , y=min\_exp, color=strategy, fill = strategy )) +  
 geom\_smooth(alpha=0.2)+  
 scale\_color\_manual (name = "Family Strategy", values =c(my\_orange, my\_yellow, my\_blue)) +  
 theme\_bw() +  
 ylim(0, 100)+  
 theme(axis.text.x = element\_text(size=20), axis.text.y = element\_text(size=20),axis.title = element\_text(size=20), plot.title = element\_text(size=20), legend.position = "none")+  
 labs(title= "c. HL exposure changes by age divided by strategy")  
  
figure\_3\_d <-ggplot(for\_anova\_graphs, aes( x=age , y=exp\_l2, color=strategy, fill = strategy )) +  
 geom\_smooth(alpha=0.2)+  
 scale\_color\_manual (name = "Family Strategy", values =c(my\_orange, my\_yellow, my\_blue)) +  
 theme\_bw() +  
 ylim(0, 50)+  
 theme(axis.text.x = element\_text(size=20), axis.text.y = element\_text(size=20),axis.title = element\_text(size=20), plot.title = element\_text(size=20), legend.position = "none")+  
 labs(title= "D. Exposure balance changes by age divided by strategy")  
  
  
figure\_3<- figure\_3\_a + figure\_3\_b + figure\_3\_c + figure\_3\_d

We also explored the data using beta regression models. While these models are not broadly used in this field, they are especially appropriate for continuous proportions that are bounded between 1 and 0 (REF), which is the case for our outcome variables, as they use a structure that does not assume a normal error structure (REF). Estimation in these kinds of models is performed by maximum likelihood. Beta regressions fit both a conditional model, which models the mean of the response, and a dispersion model, which models the error.

We constructed four beta regression models where the proportion of either English, French, a Heritage language or the exposure to the non-dominant language were predicted both by the strategy used and the age of the baby (e.g.eng\_exp ~ strategy \* age). We decided not to include the one-language-at-home strategy in these analyses because there were too few families who used the strategy which introduced noise to the models. We performed model diagnostics in all of our models and corroborated they met model assumptions (model diagnostics can be found in the Appendices).

We found a significant effect of the both-parents-bilingual strategy on total amount of French and total amount of English exposure in the dispersion models. We also found a significant effect of the both-parents-bilingual strategy on balance of exposure both in the conditional and the dispersion models. We did not find significant effects of any of the strategies on total amount of heritage language exposure.

We found a significant effect of the interaction between age and the one-parent-one-language strategy where for older children the one-parent-one-langauge strategy is more associated with a balanced exposure than for older children. While we did not find statistical effects of age as a separate predictor or in its interaction with family strategies for our other models, the visual inspection of the data (Figure 3) shows a fanning shape where there appears to be no difference between strategies at younger ages and a growing difference with age. With this in mind we do not discard the possibility that age plays a role in all of our models, but the models were under-powered to find it.

Figure 3

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

## Warning: Removed 36 rows containing non-finite values (`stat\_smooth()`).

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

## Warning: Removed 34 rows containing non-finite values (`stat\_smooth()`).

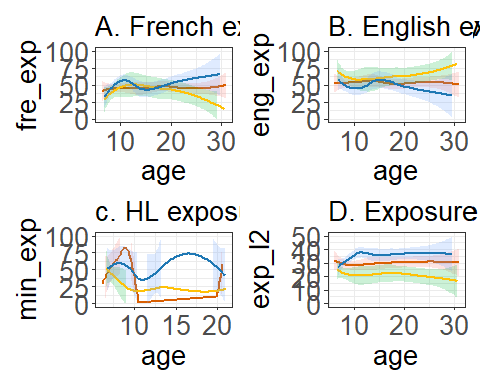
## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

## Warning: Removed 211 rows containing non-finite values (`stat\_smooth()`).

## Warning: Removed 50 rows containing missing values (`geom\_smooth()`).

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

## Warning: Removed 8 rows containing non-finite values (`stat\_smooth()`).



Finally, we wanted to explore whether families switched in their use of family language strategies across time. To explore this, we used a subset of our data that included families that had provided LEQ data two separate visits (i.e. longitudinal data). We plotted the longitudinal data (See Figure 3) to determine the stability of family strategy use across time. As can be observed, most families maintain the same family language strategy throughout both visits. Specifically, 76.3158% kept the same strategy, and 23.6842% changed strategies (See Table 5 for the breakdown of strategy use by visit). The biggest gain between visits was for the both parents bilingual strategy which gained around 6%. The biggest loss between visits was for the one language at home strategy which lost around 5%. The most common change between strategies was from the one parent bilingual strategy, to the both parents bilingual strategy, followed by changing from the one language at home strategy to the one parent bilingual strategy.

Figure 4

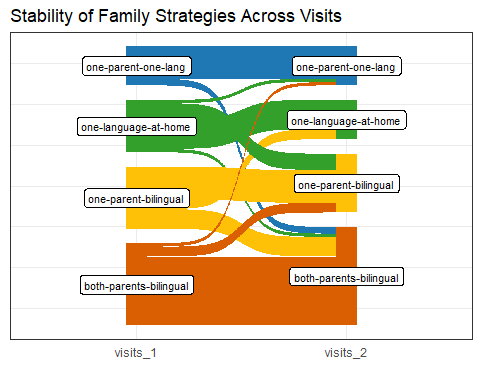


Table 5 Percentage of Family Strategy Use in Each Visit

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| strategy | n\_first\_visit | n\_second\_visit | percentage\_first\_visit | percentage\_second\_visit |
| both-parents-bilingual | 25 | 30 | 34.72 | 41.67 |
| one-language-at-home | 16 | 12 | 22.22 | 16.67 |
| one-parent-bilingual | 19 | 18 | 26.39 | 25.00 |
| one-parent-one-lang | 12 | 12 | 16.67 | 16.67 |