Supplemental Materials

Rachel Ka-Ying Tsui[1], Jessica E. Kosie[2], Laia Fibla[1], Casey Lew-Williams[2], and & Krista Byers-Heinlein[1]

[1] Concordia University

[2] Princeton University

Author Note

Correspondence concerning this article should be addressed to Rachel Ka-Ying Tsui, Department of Psychology, 7141 Sherbrooke St. West, Montreal, QC, Canada, H2T1V2. Rachel Ka-Ying Tsui is now at Laboratory for Language Development, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako-shi, Saitama, Japan, 351-0198. Email: rachelkytsui@gmail.com. Experiment material, analysis scripts, data set, and supplemental materials are available at https://osf.io/8vk3b/. E-mail: rachelkytsui@gmail.com

Supplemental Materials

Detailed descriptive statistics of the children's proficiency scores

As mentioned in the main paper, out of the 49 children with proficiency data: 14 children had equal comprehension proficiency in both languages (8 French–English, 6 Spanish–English); 17 children were more proficient in English comprehension than French/Spanish (12 French–English, 5 Spanish–English); and 18 children were more proficient in French/Spanish than in English (10 French–English, 8 Spanish–English). Proficiency data was missing from 1 French–English and 3 Spanish–English children. Table S1 lists the detailed descriptive statistics of the proficiency scores.

Table S1. Language proficiency scores across the two groups of bilingual children. Proficiency data were missing for 1 French–English and 3 Spanish–English bilingual children.

	French–English			French–English		
	N (%)	Mean (SD)	Range	N (%)	Mean (SD)	Range
Overall	30 (97%)			19 (86%)		
English		8.6 (1.71)	3 – 10		8.32 (2.16)	3 – 10
French/Spanish		8.23 (2.14)	3 – 10		9.16 (0.9)	8 – 10
Equal proficiency	8 (26%)	9.25 (0.89)	8 – 10	6 (27%)	9.67 (0.82)	8 – 10
More proficient in English	12 (39%)			5 (23%)		
English		9.42 (0.9)	8 – 10		9.6 (0.55)	9 – 10
French/Spanish		6.42 (2.27)	3 – 9		8.2 (0.45)	8 – 9
More proficient in French/Spanish	10 (32%)			8 (36%)		
English		7.1 (2.02)	3 – 9		6.5 (2.2)	3 – 9
French/Spanish		9.6 (0.7)	8 – 10		9.38 (0.74)	8 – 10

A linear regression model with proficiency score as the dependent variable was run to compare proficiency scores between the French–English and Spanish–English bilinguals. In the linear regression model, fixed effects included language community (French–English vs. Spanish–English) and language (English vs. French/Spanish), as well as their interaction[1]:

proficiency ~ language * lang_community

We found no significant effect of language, language community, nor their interaction (ps $>$ 0.11. Therefore, there was no significant difference between the French–English and Spanish–English children in terms of their proficiency in English or their proficiency in the other language (i.e., French for the French–English children and Spanish for the Spanish–English children).

Detailed descriptive statistics of the children's exposure to each language

As mentioned in the main paper, in addition to children's language proficiency, we also asked caregivers to report their child's percentage global exposure to the two languages[2]. Overall, French–English children had an average global exposure of 49% to English (SD = 20.32, range = 15 – 80) and 48% to French (SD = 20.25, range = 15 – 85); whereas Spanish–English children had an average global exposure of 48% to English (SD = 23.81, range = 5 – 90) and 48% to Spanish (SD = 23.05, range = 10 – 95). A linear regression model with global exposure as the dependent variable was run, with fixed effects of language community (French–English vs. Spanish–English) and language (English vs. French/Spanish), as well as their interaction:

exposure ~ language * lang_community

[1] We tried running a linear mixed-effects model with a random intercept for participants; however, the model returned a singular fit.

[2] The question asked was "if you think about your child's entire life, what percentage of the words that your child has heard are in English/French/Spanish/another language".

We found no significant effect of language, language community, nor their interaction (ps > 0.97. Therefore, the two groups of bilinguals did not significantly differ in terms of their global exposure to English versus to the other language (i.e., French for the French–English children and Spanish for the Spanish–English children).

Detailed descriptive statistics of the caregivers' code-switching ratings

Table S2. Caregivers' self-reported codeswitching rating across the two groups of bilingual children.

	French–English		Spanish–English	
	Mean (SD)	range	Mean (SD)	range
Within-sentence codeswitching	1.87 (1.59)	0 – 6	1.47 (1.26)	0 – 5
Between-sentence codeswitching	2.7 (1.78)	0 – 6	2.58 (1.57)	0 – 6

Caregivers were also asked to rate themselves on a scale of 0 to 6 how often they switched between the two languages — both between- and within-sentences — when interacting with their children. A rating of 0 represents that they never switched between the two languages, and a rating of 6 represents that they switched between the two languages very frequently. Table S2 lists the detailed descriptive statistics of ratings across the two bilingual groups. A linear regression model with self-reported codeswitching rating as the dependent variable was run. Fixed effects included language community (French–English vs. Spanish–English) and pattern (within-sentence vs. between-sentence), as well as their interaction:

rating ~ pattern * lang_community

There was a significant effect of pattern, suggesting that overall caregivers switched languages between sentences more often than within sentences (Estimate = -0.94, SE =

0.33, t = -2.87, $p < .01$). However, we found no significant effect of language nor language community (ps > 0.61).

Exploratory analysis of the main paper

In the main paper, we reported the preregistered analyses on accuracy — our primary dependent variable in determining bilingual children's word learning in touching the labeled target object on each test trial. In this supplemental materials, we report two preregistered exploratory analyses: (1) looking at the effect of age, (2) using response time as the dependent variable, and an additional analysis on the effect of language proficiency, language exposure, and caregivers' codeswitching rating.

Effect of age

Our accuracy analysis revealed that bilingual children in both communities successfully learned the novel words regardless of the language switching patterns used during the learning blocks. As previous research has shown that children's ability to learn words may improve with age (e.g., Read et al., 2021; Scaff et al., 2022), in our preregistration we also expected older children to show a greater accuracy than younger children. Therefore, we additionally compared models with and without age as a predictor variable. Bilingual children's age in months was scaled and centered for ease of interpretation. The final model specification was:

accuracy ~ condition * lang_community * age_in_months + (1 + condition|participant) + (1|item)

When added to the model, neither the main effect nor interactions with age were significant (all ps > 0.30; see Table S3 for the coefficient estimates from this model). Moreover, a model comparison with and without age as a variable indicated no significant improvement in model fit, $\chi^2(4) = 2.26$, $p = 0.69$. Overall, the pattern in this model was consistent with the main accuracy analysis reported in the paper where there was no significant difference in terms of condition or language community, suggesting that bilingual children successfully learned the novel words regardless of age.

Table S3. Coefficient estimates from the logistic mixed-effects models predicting accuracy in the test blocks with age_in_months as an additional fixed effect.

	Estimate	SE	z	p
Intercept	1.240	0.221	5.610	<.001
condition	0.118	0.307	0.386	0.7
lang_community	0.237	0.384	0.618	0.537
age_in_months	0.109	0.181	0.604	0.546
condition * lang_community	0.020	0.592	0.034	0.973
condition * age_in_months	0.088	0.294	0.298	0.766
lang_community * age_in_months	-0.256	0.361	-0.709	0.479
condition * lang_community * age_in_months	0.609	0.590	1.030	0.302

Response Time

In addition to accuracy, we explored response time on each correctly-responded test trial as a dependent variable; a total of 853 correct trials were included in this analysis. The decision to use response time as an additional measure was driven by a possibility that this measure might be able to better capture more individual variability in terms of the speed of children's lexical comprehension performance, since children were generally very accurate in the task. Moreover, response time may also be more sensitive to age effects, as children answer more quickly as they get older and gain more expertise with language (Scaff et al., 2022). On average, French–English bilingual children had a mean response time of 1997ms in the *immediate-translation* condition (SD = 1064.2; range = 451.86 – 5130.5) and 1788ms in the *one-language-at-a-time* condition (SD = 880.09; range = 339.33 – 4323.33).

On the other hand, Spanish–English bilingual children had a mean response time of 2620ms in the *immediate-translation* condition (SD = 1353.25; range = 764.33 – 6166.5) and 2303ms in the *one-language-at-a-time* condition (SD = 1455.58; range = 457.25 – 5550.6).

A linear mixed-effects model was run on response time. To correct for issues of non-normality, raw response time was log-transformed. Condition, language community, and age, as well as their interactions, were entered as fixed effects in the model; a random slope of condition by participant and a random intercept of stimulus item were also entered:

log_rt ~ condition * lang_community * age_in_month + (1 + condition|participant) + (1|item)

Table S4. Coefficient estimates from the linear mixed-effects model predicting log-transformed response time in the test blocks.

	Estimate	SE	t	p
Intercept	7.220	0.076	94.600	<.001
condition	-0.098	0.079	-1.250	0.22
lang_community	0.164	0.146	1.130	0.265
age_in_months	-0.094	0.071	-1.330	0.189
condition * lang_community	-0.122	0.157	-0.775	0.443
condition * age_in_months	0.010	0.078	0.125	0.901
lang_community * age_in_months	0.072	0.142	0.505	0.616
condition * lang_community * age_in_months	-0.281	0.157	-1.790	0.08

The coefficient estimates from this model are shown in Table S4 and Figure S1 visualizes this model. The only terms that approached significance was the three-way interaction. Following up on this three-way interaction, separate linear mixed-effects analyses were run for the French–English bilinguals and the Spanish–English bilinguals. The models revealed that the effect of age approached significance in the *immediate-translation* condition for the French–English bilinguals (Estimate = -0.19, SE = 0.10, t = -1.76, p = 0.09), suggesting that the reaction time for French–English bilinguals decreased significantly in the *immediate-translation* condition across age. However, the effect of age was not significant for the Spanish–English bilinguals (Estimate = 0.02, SE = 0.11, t = 0.13, p = 0.90). Moreover, we found no significant interaction across age between the two conditions in either group of bilinguals. Overall, similar to the patterns reported in the first set of analysis on accuracy, we did not observe any significant difference in terms of condition or language community, nor in their interaction. Therefore, consistent with the accuracy analysis, bilingual children in both communities performed similarly in word learning across both the *immediate-translation* and *one-language-at-a-time* conditions.

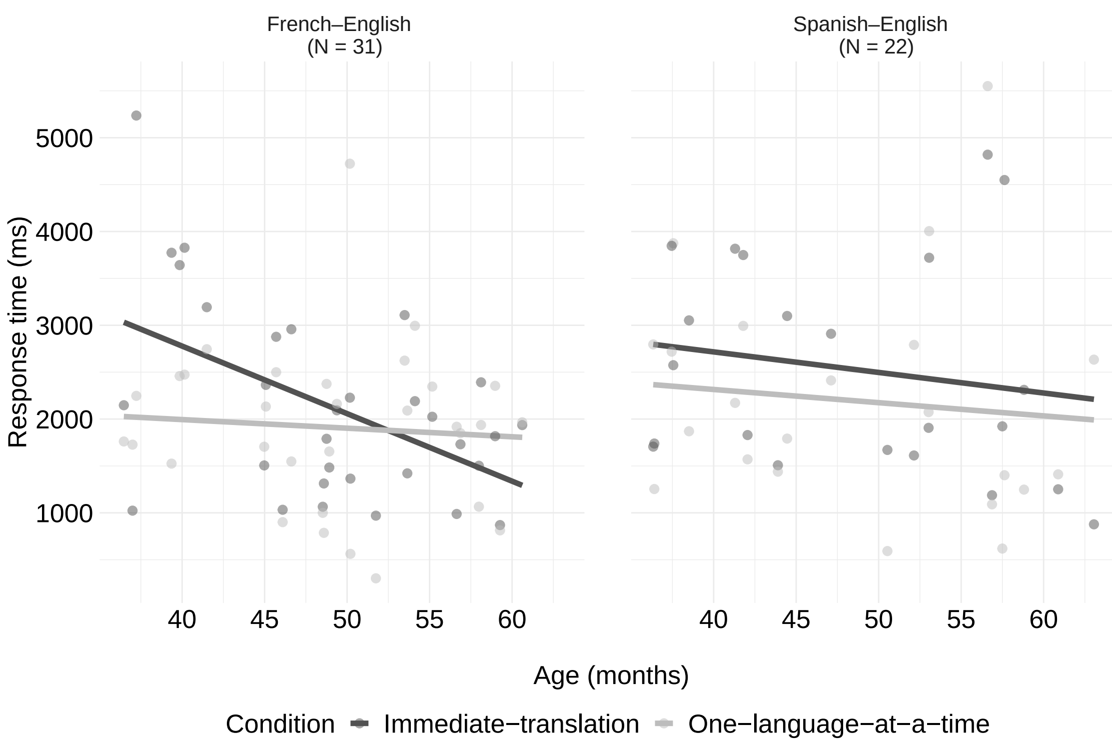

Figure S1

Response time by condition, language community, and age in the test block. Individual dots plot the data from each individual participant.

Language Proficiency

In addition to the preregistered analyses, we explored the effect of language proficiency on bilingual children's learning of the novel cross-language words. Previous research has revealed mixed evidence as to whether language proficiency interacts with bilingual children's word learning ability during different types of bilingual book reading sessions (e.g., Brouillard et al., 2022; Read et al., 2021). Therefore, it is plausible that language proficiency may have an effect on how bilingual children learn from different language switching patterns.

Building upon the logistic mixed-effects model used in the main accuracy analysis, we added a variable of proficiency score to the model. This created a continuous variable with the caregiver-reported proficiency rating in each language, which we used to predict children's performance on trials in that same language. In this analysis, we excluded data from one French–English and three Spanish–English children participants who were missing their proficiency information; data from 30 French–English and 19 Spanish–English children remained in the analysis. The initial specification was:

accuracy ~ condition * lang_community * proficiency + (1 + condition|participant) + (1|item)

However, as the initial model did not converge, we first removed the random slope for condition and the random intercept for item. Moreover, since we did not find any significant difference between the two communities in the main accuracy analysis, we performed a model comparison between the model with language communities and the one without. This comparison also indicated no significant improvement in model fit, $\chi^2(4) = 4.74$, $p = 0.32$, so we further pruned the effect of language community from the model. Therefore, the final model was:

accuracy ~ condition * proficiency + (1|participant)

Table S5. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test blocks with language proficiency scores.

	Estimate	SE	z	p
Intercept	0.533	0.468	1.140	0.255
condition	0.652	0.693	0.940	0.347
proficiency	0.076	0.051	1.480	0.138
condition * proficiency	-0.064	0.079	-0.802	0.423

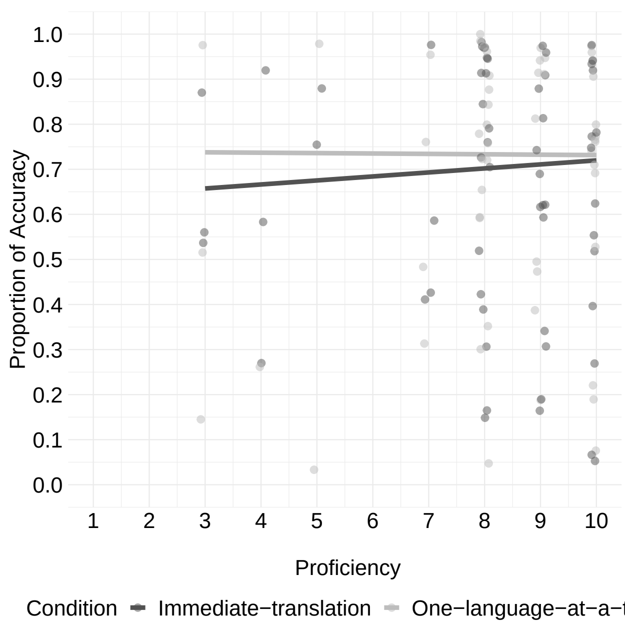

Figure S2

Proportion of accuracy by condition and language proficiency in the test blocks. Individual dots plot the data from each individual participant.

The coefficient estimates from this model are shown in Table S5 and Figure S2 visualizes this model. Similar to the main analysis, no significant difference between conditions was found. Moreover, we did not observe any significant effect of language proficiency. As can be seen in Figure S2, the level of proficiency did not hugely impact children' accuracy in our experiment. Note that we also ran another model including all the children who participated in our experiment (i.e., including those who did not initially meet our language proficiency criteria; please refer to the section below for the detailed statistics).

Global language exposure

In addition to language proficiency, we also explored the effect of global language exposure on bilingual children's learning of the novel cross-language words. Similar to the analysis on language proficiency, we first added a variable of global exposure to the logistic mixed-effects model used in the main accuracy analysis. Global exposure,a continuous variable with caregiver-reported global language exposure in each language, was used to predict children's performance on trials in that same language. Data from 30 French–English and 19 Spanish–English children remained in the analysis (i.e., after excluding one French–English and three Spanish–English children participants who were missing their language exposure information). The final model specification after pruning random intercept of item and the random slope for condition for model convergence was:

accuracy ~ condition * lang_community * exposure + (1|participant)

The coefficient estimates from this model are shown in Table S6 and Figure S3 visualizes this model. Similar to the main analysis, no significant difference between conditions was found. While the effect of exposure was not significant, we observed a significant effect of language community which was moderated by a significant interaction between language community and exposure. As can be seen in Figure S3, the French–English bilingual children who had a higher percentage of global language exposure to the target language were more accurate in learning the novel words. On the other hand, Spanish–English children's performance was not significantly related to their global language exposure to the target language.

Table S6. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test blocks with global language exposure.

	Estimate	SE	z	p
Intercept	0.994	0.251	3.960	<.001
condition	0.082	0.377	0.216	0.829
lang_community	1.500	0.499	3.010	<.01
exposure	0.005	0.004	1.420	0.155
condition * lang_community	-0.096	0.754	-0.128	0.898
condition * exposure	0.001	0.007	0.095	0.925
lang_community * exposure	-0.023	0.007	-3.100	<.01
condition * lang_community * exposure	0.003	0.014	0.199	0.842

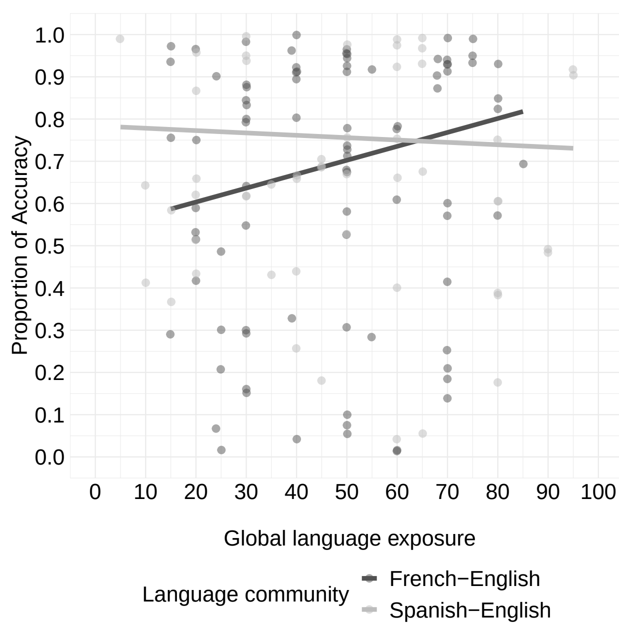

Figure S3

Proportion of accuracy by condition and global language exposure in the test blocks. Individual dots plot the data from each individual participant.

Caregivers' self-reported code-switching rating

Finally, we also explored the effect of caregivers' code-switching practices on bilingual children's learning of the novel cross-language words. Since our experiment only tested between-sentence switches, this analysis also focused on caregivers' self-reported between-sentence code-switching practices. We added a variable of caregivers' between-sentence code-switching rating to the logistic mixed-effects model used in the main accuracy analysis. Data from 30 French–English and 19 Spanish–English children remained in the analysis (i.e., after excluding one French–English and three Spanish–English children participants who were missing their language exposure information). The final model specification after pruning random slope for condition to achieve model convergence was:

accuracy ~ condition * lang_community * codeswitch_rating + (1|participant) + (1|item)

Table S7. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test blocks with caregivers' self-reported between-sentence code-switching ratings.

	Estimate	*SE*	*z*	*p*
Intercept	1.160	0.350	3.310	<.01
condition	0.170	0.291	0.582	0.56
lang_community	0.748	0.684	1.090	0.274
codeswitch_rating	0.011	0.109	0.104	0.918
condition * lang_community	1.120	0.583	1.930	0.054
condition * codeswitch_rating	-0.015	0.093	-0.164	0.87
lang_community * codeswitch_rating	-0.154	0.218	-0.708	0.479
condition * lang_community * codeswitch_rating	-0.407	0.187	-2.180	<.05

While model comparison between the model with code-switching rating and the one without indicated no significant improvement in model fit, $\chi^2(4) = 5.78$, $p = 0.22$, we listed the coefficient estimates from this model in Table S7 and visualized this model in Figure S4. Similar to the main analysis, no significant difference between conditions was found. However, there was a significant 3-way condition * language community * code-switching rating interaction. As can be seen in Figure S4, the two groups of bilinguals showed opposite patterns of performance when caregivers' code-switching practice was taken into account. The French–English children showed greater accuracy under the one-language-at-a-time condition if their caregiver switched languages very frequently; whereas Spanish–English children were less accurate under the same condition. The reverse pattern was observed in the immediate-translation condition. The French–English children were relatively less accurate the more frequently their caregiver reported switched languages, whereas Spanish–English children were more accurate under the same condition.

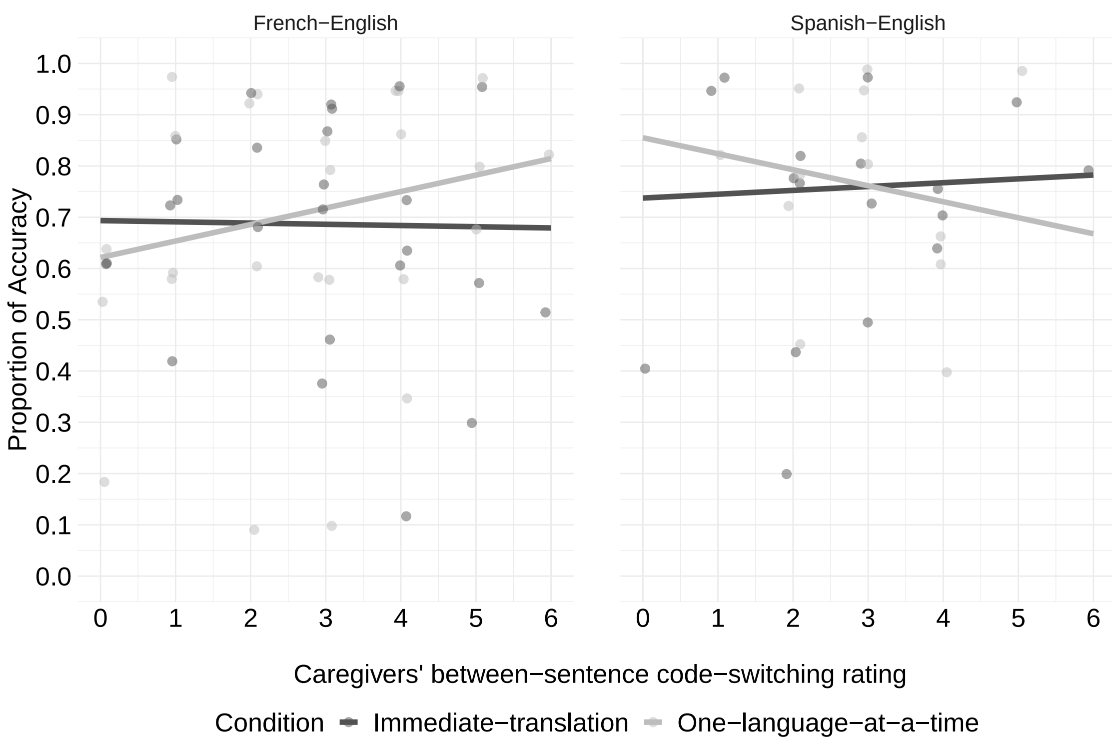

Figure S4

Proportion of accuracy by condition and caregivers' self-reported between-sentence code-switching rating in the test blocks. Individual dots plot the data from each participant.

Supplemental analysis using the preregistered exclusion criteria

Analyses reported in the main paper deviated from the preregistered language exclusion criteria, as it resulted in exclusion of a higher than anticipated number of children and thus led to a smaller sample size and decreased statistical power. For transparency, this supplemental material reports the analyses using the more stringent preregistered exclusion criteria.

Following the preregistered language exclusion criteria, a total of 10 French–English and 12 Spanish–English children were excluded. When additional exclusion criteria were applied (see participants section in the main paper), the remaining sample consisted of 22 French–English children (13 girls; Mean age = 4.04 years, SD = 0.56, range = 3.04 – 4.94) and 14 Spanish–English children (7 girls; Mean age = 4.21 years, SD = 0.75, range = 3.03 – 5.26).

Language proficiency information was missing for 1 French–English and 3 Spanish–English children; 13 children had equal comprehension proficiency in both languages (7 French–English, 6 Spanish–English); 6 French–English children were more proficient in English comprehension than French/Spanish; and 13 children were more proficient in French/Spanish than in English (8 French–English, 5 Spanish–English). Table S8 contains descriptive statistics of the proficiency scores. We also ran a linear regression model with proficiency score as the dependent variable to compare proficiency scores between the French–English and Spanish–English bilinguals[3]. There was no significant effect of language, language community, nor their interaction (ps > 0.30). Similar to the sample reported in the main paper, there was no significant difference between the French–English and Spanish–English children in terms of their proficiency in English as

[3] In the linear regression model with proficiency scores as the dependent variable, fixed effects included language community (French–English vs. Spanish–English) and language (English vs. French/Spanish), as well as their interaction.

well as their proficiency in the other language (i.e., French for the French–English children and Spanish for the Spanish–English children). In this more restricted data set, 68% of the mothers in Montreal and 64% of the mothers in New Jersey had completed a university degree or higher.

Table S8. Language proficiency scores across the two groups of bilingual children whose proficiency scores were available.

	French–English			French–English		
	N (%)	Mean (SD)	Range	N (%)	Mean (SD)	Range
Overall	21 (95%)			11 (79%)		
English		9 (1.1)	7 – 10		8.91 (1.14)	7 – 10
French/Spanish		9.19 (0.93)	8 – 10		9.64 (0.67)	8 – 10
Equal proficiency	7 (32%)	9.29 (0.95)	8 – 10	6 (43%)	9.67 (0.82)	8 – 10
More proficient in English	6 (27%)			0 (0%)		
English		10 (0)	10		—	—
French/Spanish		8.33 (0.52)	8 – 9		—	—
More proficient in French/Spanish	8 (36%)			5 (36%)		
English		8 (0.76)	7 – 9		8 (0.71)	7 – 9
French/Spanish		9.75 (0.71)	8 – 10		9.6 (0.55)	9 – 10

In the following analyses, we followed the same procedure as in the main paper, where we first explored the preregistered analyses with proportion of accuracy as the dependent variable and then performed exploratory analyses: (1) on the effect of age, (2) with response time as the dependent variable, and (3) the effect of proficiency, language exposure, and caregivers' code-switching rating.

Accuracy

Familiar word block. French–English bilingual children showed a mean accuracy of 0.98 in the familiar English-word trials ($SD = 0.05$; range = 0.83 – 1) and 0.98 in the familiar French-word trials ($SD = 0.05$; range = 0.83 – 1). Meanwhile, Spanish–English bilingual children showed a mean accuracy of 1 in the familiar English-word trials ($SD = 0$) and 1 in the familiar Spanish-word trials ($SD = 0$). As children's performance was almost at ceiling with little variance, it was not possible to fit the logistic mixed-effects model to compare performance across the two communities. On the other hand, this near-ceiling accuracy also suggests that our preregistered exclusion criteria could be too stringent such that only children with nearly perfect accuracy were included in the analyses.

Test blocks. On average, French–English bilingual children showed a mean accuracy of 0.72 in the *immediate-translation* condition ($SD = 0.25$; range = 0.17 – 1) and 0.68 in the *one-language-at-a-time* condition ($SD = 0.27$; range = 0 – 1). On the other hand, Spanish–English bilingual children showed a mean accuracy of 0.76 in the *immediate-translation* condition ($SD = 0.23$; range = 0.25 – 1) and 0.7 in the *one-language-at-a-time* condition ($SD = 0.24$; range = 0.33 – 1). Separate one-sample t-tests were run on the proportion of accuracy in each condition per community, and confirmed that children from both communities learned the novel words in each condition significantly above the at-chance level of 0.50 (ps $< .01$)[4].

Following the main paper, we ran a logistic mixed-effects model on the proportion of accuracy, with condition and language community as fixed effects, and a random slope of condition by participants and random intercept of item:

[4] For the French–English bilinguals, they performed significantly above the at-chance level in the *immediate-translation* condition ($t(21) = 4.01$, $p < .001$) and the *one-language-at-a-time* condition ($t(21) = 3.14$, $p = .002$). Likewise, the Spanish–English bilinguals performed significantly above the at-chance level in the *immediate-translation* condition ($t(13) = 4.24$, $p < .001$) and the *one-language-at-a-time* condition ($t(13) = 3.03$, $p = .005$).

$$\text{accuracy} \sim \text{condition} * \text{lang\_community} + (1+ \text{condition}|\text{participant}) + (1|\text{item})$$

Table S9. Coefficient estimates from the logistic mixed-effects models predicting accuracy in the test phase.

	Estimate	*SE*	*z*	*p*
Intercept	1.170	0.262	4.460	<.001
condition	-0.247	0.378	-0.654	0.513
lang_community	0.298	0.467	0.639	0.523
condition * lang_community	0.065	0.727	0.089	0.929

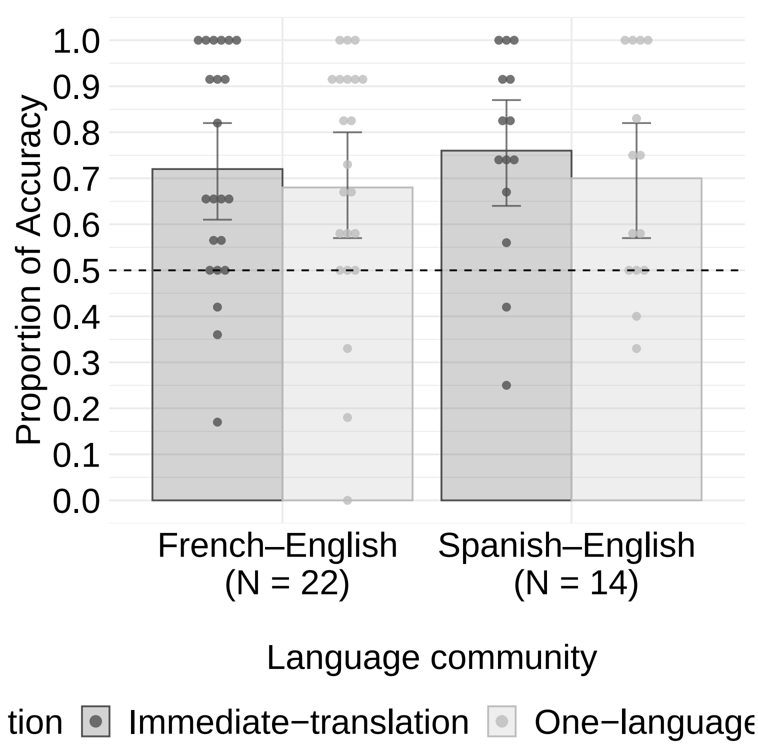

Figure S5

Average proportion of accuracy by condition and language community in the test blocks. Dots plot the data from each individual participant. Error bars indicate 95% confidence intervals, and the black dashed line represents the at-chance accuracy level of 0.50.

The coefficient estimates from this model are shown in Table S9, and Figure S5 visualizes this model. Consistent with the patterns found in the main paper, the model also did not reveal any significant difference in terms of condition or language community, and the interaction between condition and language community was also not significant. Therefore, similar to the main paper, bilingual children in both communities showed strong evidence of word learning in both language switching conditions.

Exploratory Analysis

Effect of age. As an exploratory analysis, we also ran a logistic mixed-effects model with age as a fixed effect. Similar to the analysis on the sample reported in the main paper, model comparison with the model without age showed that adding age did not significantly improve the model ($\chi^2(4) = 5.55$, $p = 0.24$). Moreover, we had to prune the random slope for condition from the model since it would not converge, the final model specification was:

accuracy ~ condition * lang_community * age_in_months + (1|participant) + (1|item)

The coefficient estimates from this model are shown in Table S10 and Figure S6 visualizes this model. The model did not reveal a significant effect of age, but there was a significant three-way condition * language community * age interaction. This was due to the crossover interaction in the French–English bilingual children, where their performance significantly improved on the *immediate-translation* condition with age. In contrast, the Spanish–English children showed similar performance in the two conditions and across age.

Table S10. Coefficient estimates from the logistic mixed-effects models predicting accuracy in the the test phase with age_in_months as an additional fixed effect.

	Estimate	SE	z	p
Intercept	1.060	0.235	4.490	<.001
condition	-0.328	0.177	-1.850	0.065
lang_community	0.236	0.420	0.562	0.574
age_in_months	0.201	0.192	1.050	0.295
condition * lang_community	-0.054	0.354	-0.153	0.878
condition * age_in_months	-0.298	0.179	-1.670	0.095
lang_community * age_in_months	-0.286	0.383	-0.747	0.455
condition * lang_community * age_in_months	0.975	0.357	2.730	<.01

Response time. We also included response time on each correctly-responded test trial as a dependent variable. There were a total of 573 trials in this analysis. On average, French–English bilingual children had a mean response time of 2101ms in the *immediate-translation* condition (SD = 1101.21; range = 451.86 – 5130.5) and 1803ms in the *one-language-at-a-time* condition (SD = 765.06; range = 339.33 – 3101). On the other hand, Spanish–English bilingual children had a mean response time of 2410ms in the *immediate-translation* condition (SD = 1108.38; range = 764.33 – 4549) and 2033ms in the *one-language-at-a-time* condition (SD = 1371.97; range = 457.25 – 5525.8).

We ran a linear mixed-effects model with condition, language community, age, as well as their interactions, entered as fixed effects. A random slope of condition by participants and random intercepts of item were also entered. To correct issues of non-normality in response time, raw response time was log-transformed. The final model specification was:

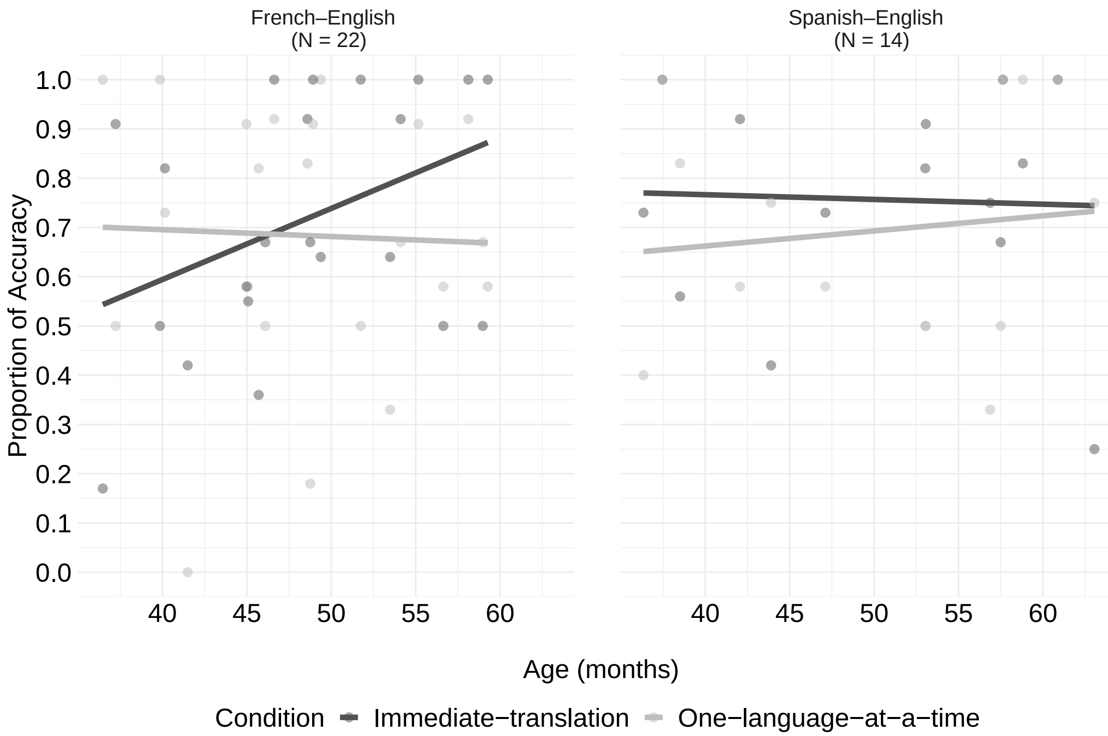

Figure S6

Proportion of accuracy by condition, language community, and age in the test phase. Individual dots plot the data from each individual participant.

$$\text{log\_rt} \sim \text{condition} * \text{lang\_community} * \text{age\_in\_months} + (1 + \text{condition}|\text{participant}) + (1|\text{item})$$

The coefficient estimates from this model are shown in Table S11 and Figure S7 visualizes this model. Overall, visualization of the model suggests that, across both conditions and language communities, older children generally responded faster than younger children. However, the model did not reveal any significant effects or interactions of condition, language community, or age. Therefore, consistent to the patterns reported for the sample included in the main paper, in terms of response time, bilingual children in both communities performed similarly in word learning across both the *immediate-translation* and *one-language-at-a-time* conditions.

Table S11. Coefficient estimates from the linear mixed-effects model predicting log-transformed response time in the test phase.

	Estimate	SE	t	p
Intercept	7.180	0.085	84.800	<.001
condition	-0.138	0.090	-1.530	0.137
lang_community	0.066	0.168	0.394	0.696
age_in_months	-0.129	0.082	-1.570	0.126
condition * lang_community	-0.171	0.180	-0.953	0.349
condition * age_in_months	0.036	0.089	0.403	0.69
lang_community * age_in_months	0.076	0.164	0.466	0.644
condition * lang_community * age_in_months	-0.189	0.178	-1.060	0.298

Language proficiency in all participants. For a more inclusive analysis on the effect of proficiency, this analysis also included data from children who were previously eliminated for not fulfilling the language criteria — either the criteria reported in the main paper or the more stringent criteria reported in this document. In total, there were 30 French–English children and 22 Spanish–English children who were born full term and without any reported language problems or participated without any problem. Among this group of children, the French–English children had a mean proficiency score of 8.42 (SD = 1.93, range = 3 – 10) and the Spanish–English children had a mean proficiency score of 8.34 (SD = 2.07, range = 2 – 10). To explore the effect of proficiency, we added a variable of proficiency score to the logistic mixed-effects model. The initial model specification was:

accuracy ~ condition * lang_community * proficiency + (1+ condition|participant) + (1|item)

However, as the initial model could not converge, we removed the random slope for

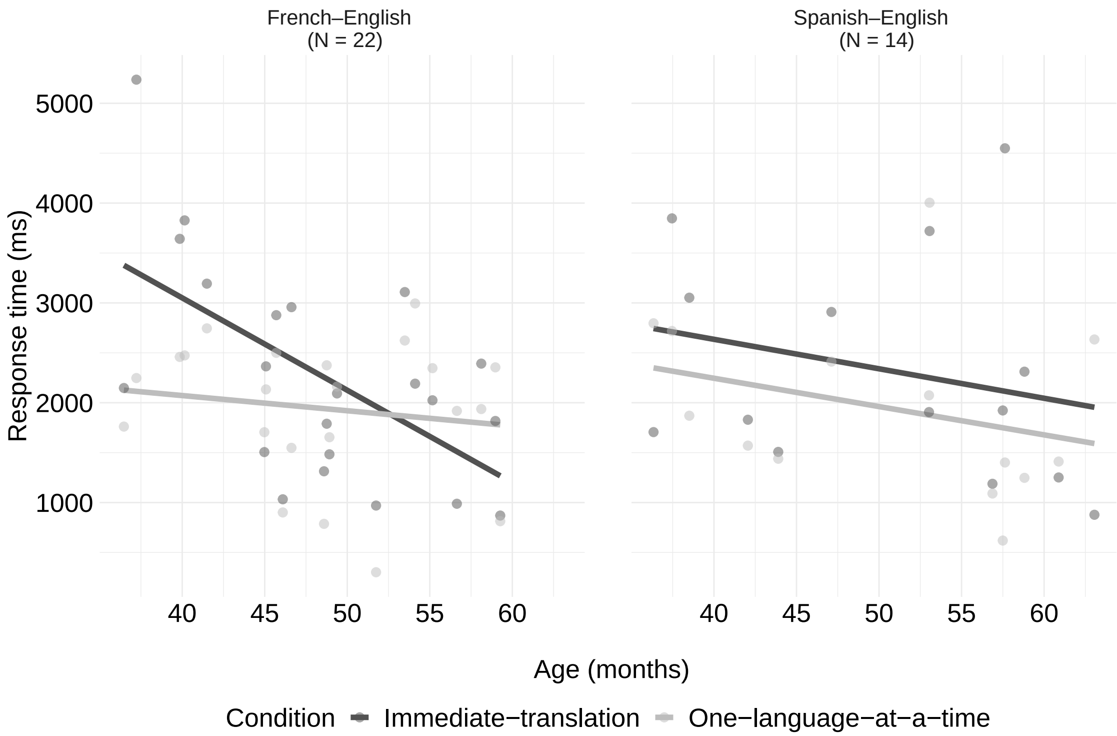

Figure S7

Response time by condition, language community, and age in the test phase. Individual dots plot the data from each individual participant.

condition and the random intercept for stimulus item. Moreover, since we did not find any significant difference between the two communities in the main accuracy analysis, we further pruned the effect of language community from the model. Note that model comparison between the model with language community and the one without also indicated no significant improvement in model fit, $\chi^2(4) = 7.44$, $p = 0.12$. Therefore, the final model was:

$$\text{accuracy} \sim \text{condition} * \text{proficiency} + (1|\text{participant})$$

The coefficient estimates from this model are shown in Table S12 and Figure S8 visualizes this model. The model revealed a significant effect of proficiency, suggesting that bilingual children were overall more accurate for trials in which they had a higher level of language proficiency in the trial language. In contrast to the patterns reported for the sample included in the main paper, this model here revealed that the effect of proficiency may be more evident when a wider range of language proficiency level was included in the analysis. In other words, the sample reported in the main paper could have possibly focused on children who were relatively proficient. Yet, the lack of a significant interaction points to the possibility that proficiency level did not affect children's accuracy in learning words under different language-switching patterns.

Table S12. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test phase with proficiency scores among all the children.

	Estimate	SE	z	p
Intercept	0.109	0.416	0.262	0.794
condition	-0.428	0.596	-0.717	0.473
proficiency	0.121	0.046	2.630	$<.01$
condition * proficiency	0.049	0.070	0.699	0.485

Global language exposure in all participants. Similar to the proficiency analysis, we also included data from all children in this analysis. Among this group of 30 French–English children and 22 Spanish–English children, the French–English children had a mean global exposure of 49% to English (SD = 20.32, range = 15 – 80) and 48% to French (SD = 20.25, range = 15 – 85), and the Spanish–English children had a mean global exposure of 52% to English (SD = 24.72, range = 5 – 90) and 45% to Spanish (SD = 23.60, range = 10 – 95). To explore the effect of global language exposure, a variable of global language exposure was added to the logistic mixed-effects model. The final model specification after removing the random slope for condition to achieve model convergence was:

accuracy ~ condition * lang_community * exposure + (1|participant) + (1|item)

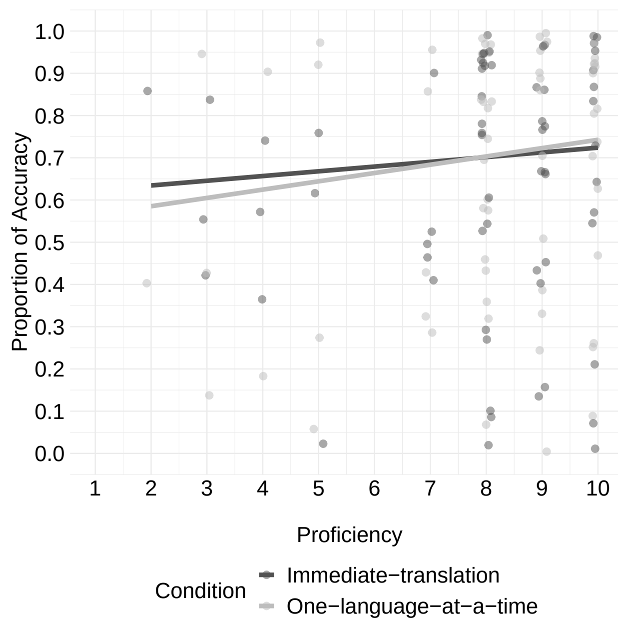

Figure S8

Proportion of accuracy of all children by condition and proficiency in the test blocks. Individual dots plots the data from each individual participant.

The coefficient estimates from this model are shown in Table S13 and Figure S9 visualizes this model. We observed a significant effect of exposure and the effect of language community also approached significance. Moreover, similar to the results reported for the sample included in the main analysis, the model revealed a significant interaction between language community and exposure. As visualized in Figure S9, similar to the results reported in the main analysis, the French–English bilingual children with a higher percentage of global language exposure were more accurate in learning the novel words. On the other hand, Spanish–English children's performance was not significantly related to their global language exposure.

Table S13. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test phase with global language exposure among all the children.

	Estimate	SE	z	p
Intercept	0.702	0.257	2.730	<.01
condition	-0.125	0.353	-0.353	0.724
lang_community	0.905	0.488	1.850	0.064
exposure	0.008	0.003	2.370	<.05
condition * lang_community	-0.485	0.705	-0.688	0.491
condition * exposure	0.002	0.007	0.286	0.775
lang_community * exposure	-0.015	0.007	-2.280	<.05
condition * lang_community * exposure	0.004	0.013	0.320	0.749

Caregivers' self-reported code-switching rating in all participants. Finally, we also explored the effect of caregivers' self-reported code-switching practices on all children's novel word learning. Among this group of 30 French–English children and 22 Spanish–English children, caregivers of the French–English children reported a rating of 1.87 for within-sentences code-switch (SD = 1.59, range = 0 – 6) and 2.70 for between-sentence code-switch (SD = 1.78, range = 0 – 6), and caregivers of the Spanish–English children reported a rating of 1.36 for within-sentences code-switch (SD = 1.22, range = 0 – 5) and 2.27 for between-sentence code-switch (SD = 1.67, range = 0 – 6). To explore the effect of caregivers' code-switching practice, we focused only on their between-sentence code-switching rating since our study only tested between-sentence language switches. Following the model used in the main analysis, we added a variable of caregivers' between-sentence code-switching rating to the logistic mixed-effects model. The final model specification was:

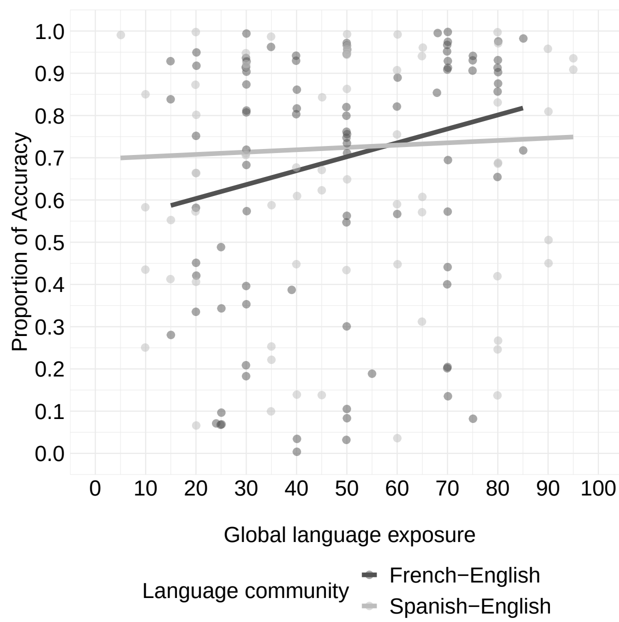

Figure S9

Proportion of accuracy by condition and global language exposure in the test blocks. Individual dots plot the data from each individual participant.

accuracy ~ condition * lang_community * codeswitch_rating + (1|participant) + (1|item)

The coefficient estimates from this model in Table S14. Different from the analysis reported for the sample included in the main analysis, we observed no significant effects or interactions.

Table S14. Coefficient estimates from the logistic mixed-effects model predicting accuracy in the test blocks with caregivers' self-reported between-sentence code-switching ratings.

	Estimate	SE	z	p
Intercept	0.842	0.320	2.630	$<.01$
condition	-0.364	0.250	-1.460	0.146
lang_community	0.159	0.612	0.261	0.794
codeswitch_rating	0.098	0.101	0.965	0.335
condition * lang_community	0.059	0.500	0.119	0.905
condition * codeswitch_rating	0.133	0.084	1.570	0.116
lang_community * codeswitch_rating	0.016	0.203	0.078	0.938
condition * lang_community * codeswitch_rating	-0.115	0.169	-0.678	0.498