

BRIEF RESEARCH REPORT

# Vocabulary production in toddlers from low-income immigrant families: evidence from children exposed to Romanian-Italian and Nigerian English-Italian

Chiara BARACHETTI<sup>1\*</sup>, Marinella MAJORANO<sup>1</sup>, Germano ROSSI<sup>2</sup>, Elena ANTOLINI<sup>3</sup>, Rosanna ZERBATO<sup>3</sup>, and Manuela LAVELLI<sup>1</sup>

<sup>1</sup>University of Verona, Italy, <sup>2</sup>University of Milano-Bicocca, Italy, and <sup>3</sup>Pedagogical Coordination of the Nursery Schools and Preschools, Municipality of Verona, Italy

\*Address for correspondence: Chiara Barachetti, Department of Human Sciences, University of Verona, via S. Francesco 22, 37129 Verona, Italy; E-mail: [chiara.barachetti@univr.it](mailto:chiara.barachetti@univr.it)

(Received 11 September 2019; revised 18 June 2020; accepted 26 February 2021; first published online 22 April 2021)

## Abstract

The relationship between first and second language in early vocabulary acquisition in bilingual children is still debated in the literature. This study compared the expressive vocabulary of 39 equivalently low-SES two-year-old bilingual children from immigrant families with different heritage languages (Romanian vs. Nigerian English) and the same majority language (Italian). Vocabulary size, vocabulary composition and translation equivalents (TEs) were assessed using the Italian/L1 versions of the CDI. Higher vocabulary in Italian than in the heritage language emerged in both groups. Moreover, Romanian-Italian-speaking children produced higher proportions of TEs than Nigerian English-Italian-speaking children, suggesting that L1-L2 phonological similarity facilitates the acquisition of cross-linguistic synonyms.

**Keywords:** bilingual vocabulary; toddlers; low-income families; immigrant families

## Introduction

The relationship between first language (L1) and second language (L2) early vocabulary acquisition in bilingual children is still debated in the literature. The earliest theoretical position – that young bilingual children construct their first vocabulary as a single undifferentiated language (Volterra & Taeschner, 1978) – has not been confirmed. More recent research has shown that bilingual first language learners construct two separate lexical systems from the earliest stages of word learning (Bosch & Sebastian-Galles, 2001). However, we need to explore further to what extent L1 and L2 early vocabularies are acquired independently or interdependently.

Results from studies analysing early vocabulary sizes in L1 and L2 diverged when different language measures (comprehension - production), different children's age

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

groups, and different language combinations are considered. For example, significant positive correlations were found between L1 and L2 vocabulary comprehension scores for French–English bilinguals at age 1;4 to 1;6 (Legacy, Zesiger, Friend, & Poulin-Dubois, 2016), whereas the number of words produced in Spanish by Spanish–English bilinguals at age 2;6 was unrelated to their vocabulary size in English (Marchman, Fernald, & Hurtado, 2010). Moreover, the amount of exposure to each language and the condition of ‘dominant language’ may account for L1 and L2 vocabulary sizes more than cross-language abilities. There is growing evidence that in bilingual communities in which one language is dominant, such as immigrant groups, the acquisition of the dominant language may be relatively unproblematic, while the acquisition of the heritage language can be hindered when input is reduced (Oller, Jarmulowicz, Gibson, & Hoff, 2007).

A second contribution comes from a small number of studies on the early vocabulary composition of bilingual children, analysing the distribution of bilingual toddlers’ vocabulary knowledge across different lexical categories in each of the two languages. The results have shown that the proportions and trajectories of different word types in each language were linked more to differences in same-language vocabulary sizes than to general lexical-conceptual knowledge across languages. For example, the dominance of nouns in the early vocabularies of French–English (David & Wei, 2008) and Spanish–English (Conboy & Thal, 2006) bilingual toddlers is greater than it is in French-, Spanish-, and English-speaking monolingual children. The proportion of different word classes in each language in Italian–Spanish speaking toddlers (age range 24–37 months) varied with the size of their vocabulary in the same language, except for closed-class words (Pettenati, Vacchini, Stefanini, & Caselli, 2011). The proportion of verbs in the total vocabulary was higher in Turkish than in German in Turkish–German bilingual toddlers (Rinker, Budde-Spengler, & Sachse, 2017), and in Mandarin Chinese more than in English in Chinese–English bilinguals aged between 1;0 and 4;0 (Levey & Cruz, 2003) (in Turkish and Mandarin, verbs occur more frequently alone or in a salient phrase-final position). Overall, these results suggest that, in bilinguals, early vocabulary composition in each language resembles monolingual children’s patterns of early lexicon organisation and that it is mostly influenced by the specific properties of each language.

A third contribution comes from studies on translation equivalents (TEs). TEs, or cross-language synonyms, are defined as lexical representations that a speaker has in each language for the same concept (e.g., *water* in English and *acqua* in Italian). When assessing the early lexicon of bilinguals, counting TEs is a useful approach to assessing the level of semantic and phonological overlap between the two languages (De Houwer, 2009). Several studies have shown that TEs occur by the end of the second year (Pearson, Fernandez, & Oller, 1995; Byers-Heinlein & Werker, 2013). However, different proportions of TEs in the early total vocabulary of bilinguals have been documented, from 10% found in Turkish–German vocabulary (Rinker et al., 2017) up to 44% in German–English vocabulary (Byers-Heinlein & Werker, 2013; David & Wei, 2008; Junker & Stockman, 2002; Legacy et al., 2017; Poulin-Dubois, Bialystok, Blaye, Polonia, & Yott, 2013). A suggested plausible explanation for differences in proportions of TEs is the phonological proximity between the two languages learned by bilingual children (Schepens, Van Der Slik, & Van Hout, 2013). Indeed, three-year-old bilingual children from immigrant families speaking Swiss German as L2 reported higher scores in TEs when L1–L2 combinations shared more cognates (Gampe, Kurthen, & Daum, 2018), in line with other findings (Bosch &

Ramon-Casas, 2014). Recently, Floccia and colleagues (Floccia *et al.*, 2018) found that a higher L1-L2 phonological overlap (such as in English-Dutch, English-German language combinations) led to higher levels of L1 vocabulary production by 2-year-old bilinguals learning British English as L2. Importantly, linguistic distance contributed unique variance even when other factors, such as exposure to L2 child-directed input, were entered in the same model. The facilitation effect of L1-L2 phonological similarity suggests a relationship between first language and second language vocabulary acquisition.

On the whole, the relationship between L1 and L2 vocabulary acquisition processes and the role of the phonological distance between the two languages need to be investigated further in different language combinations (Hammer, Jia, & Uchikoshi, 2011). Studies on bilingual children acquiring languages other than L1-English (L2) may help us understand better children's early bilingual vocabulary in general and, in particular, to identify those aspects of bilingual development that are shared between languages and cultures and those that are not.

### The present study

The present study examined the expressive vocabulary of two-year-old bilingual toddlers from two linguistic communities of immigrant families, all attending nursery schools, in both the family's heritage language (L1, Romanian or Nigerian English) and the majority language (L2, Italian).

The study addresses the following research questions (RQs):

1. Are there differences in the early expressive vocabulary size between L1 and L2 within each bilingual group and between the two bilingual groups – Romanian-Italian-speaking children and Nigerian English-Italian-speaking children – all from similarly low-income families and with similar amounts of exposure to Italian?
2. Are there differences in the composition of early expressive L2 vocabulary between the two bilingual groups?
3. Are there differences in the proportions of TEs in total vocabulary between the two bilingual groups?

Because Romanian and Italian are both Romance languages, they share more phonological word-forms than do Italian and Nigerian English, a nativised form of English, a Germanic language. On this basis, we expected that Romanian-speaking bilingual children would have greater vocabulary sizes and higher proportions of TEs than Nigerian English-speaking bilingual children. We also expected, based on the hypothesis of a smaller vocabulary size in Nigerian English-speaking bilinguals, that these children will show a higher proportion of nouns and a lower proportion of verbs and closed-class words than Romanian-speaking bilingual children.

### Method

#### Participants

Thirty-nine two-year-old children regularly attending five nursery schools in north-East Italy participated in the study, divided into two groups: 22 children from

Romanian immigrant families (14 females) exposed to Italian and Romanian (ROM bilingual group), and 17 children from Nigerian immigrant families (8 females) exposed to Italian and Nigerian English (NEN bilingual group). The participants were selected from those recruited for a larger longitudinal research project on the lexical trajectories of children from low-income families. Only low-income families – those paying the lowest rate for their children’s nursery schools ( $\leq$ €130) – were eligible for participation in the project. All the families were recruited at the nursery schools. The bilingual group in the larger research project included 74 immigrant families from 9 different countries. Thirteen immigrant families withdrew from the study. The immigrant families included in the current study belong to the two major immigrant groups in the province. The inclusion criteria for children were: having both parents speaking the heritage language at home, being born in Italy, having been exposed to Italian before age 1;6 and producing words in at least one of the two languages at 24 months. All the children were born within two weeks of their due date and were healthy at birth, with normal hearing and had no certified disabilities or developmental disorders.

All parents were native Romanian- or Nigerian English-speakers. Nine parents of the NEN group stated that they also spoke the language of their ethnic group (Edo or Igbo), but never directed to their children. Five immigrant families stated that they did not speak Italian at home (1 in the NEN group, 4 in the ROM group).

The demographic characteristics of the children and their mothers are reported in **Table 1**. No significant differences were found between the two groups in the children’s gender, age, birth order, singleton condition, age of entry at nursery school, age and amount of exposure to Italian. Children in the ROM group spent a significantly higher number of daily hours at nursery school than children in the NEN group. The mothers in the NEN bilingual group had significantly fewer years of formal education, were significantly older, and had been in Italy significantly longer than those in the ROM bilingual group.

All participating nursery schools were state-regulated and funded. All children were cared for in a group with up to 8 children. A total of 24 nursery teachers primarily responsible for the participating children participated in the study and evaluated the children’s language abilities through the Italian version of the CDI. All the teachers were female and native Italian speakers. Five teachers held a degree in Education; 19 of them held a high school diploma in Infant Education or Teaching.

### **Procedure**

Parents were asked to complete a consent form and a brief demographic questionnaire. Then, the researcher conducted a semi-structured interview on language exposure with the parents (usually mothers) at the nursery school. Parent interviews were in Italian and, when necessary, were supported by a cultural mediator in the language of parents’ choice. Parents gave information about their child’s typical weekdays and weekend and the hours that they (or others in regular contact with the child) use L1 vs L2 with the child at home. Based on this information, an estimate of the weekly hours of exposure to Italian (L2) at home was calculated and added to the hours of weekly attendance at nursery school in order to calculate the total amount of weekly exposure to Italian. Then, this amount was expressed as a percentage of a typical number of weekly waking hours (84), following Onofrio and colleagues (Onofrio, Rinaldi, & Pettenati, 2012).

**Table 1.** Children's and Mothers' Characteristics

	ROM bilingual group (n = 22)		NEN bilingual group (n = 17)		Test between groups	<i>p</i>	<i>ES</i>
	<i>f</i> (%)	<i>M</i> ( <i>SD</i> ) range	<i>f</i> (%)	<i>M</i> ( <i>SD</i> ) range			
Child gender (female vs male)	14 (63.6)		8 (47.1)		$\chi^2(1) = 1.07$	.301	$\phi = 0.166$
Child Age (months)		24.3 (0.17) 23.9-24.5		24.2 (0.68) 23.0-25.4	$t(17.57) = 0.345^a$	.734	$d = 0.118$
Birth order (first-born vs not first)	5 (22.7)		3 (17.6)		$\chi^2(1) = 0.15$	.697	$\phi = 0.062$
Singleton (yes vs no)	4 (18.1)		2 (11.8)		$\chi^2(1) = 0.30$	.582	$\phi = 0.088$
Age of entry at nursery school (months)		14.5 (4.50) 6-18		13.5 (3.52) 8-19 <sup>b</sup>	$t(36.98) = 0.801$	.428	$d = 0.255$
Daily attendance at nursery school (hours)		7.6 (0.99) 4.0-8.5		6.9 (1.08) 4.0-8.0	$t(32.99) = 1.952$	.003	$d = 0.634$
Age of exposure to Italian (infant's age in months)		17.8 (7.59) 5-24		18.8 (6.90) 7-24	$t(35.97) = 0.407$	.745	$d = 0.130$
Weekly exposure to Italian (waking time %)		0.49 (0.12) 0.33-0.80		0.57 (0.17) 0.30-0.89	$t(26.83) = -1.700$	.101	$d = -0.562$
Maternal age (years)		32.6 (5.28) 22-46		37.9 (5.91) 24-48	$t(32.42) = -2.911$	.003	$d = -0.947$
Maternal Education (years)		13.3 (3.99) 8-20		9.4 (2.76) 5-14	$t(36.633) = 3.606$	.011	$d = 1.138$
Length of mother's stay in Italy (years)		10.7 (4.20) 3-24		16.3 (5.29) 5-28	$t(29.98) = -3.587$	<.001	$d = -1.175$

Note. <sup>a</sup> Welch's *t*; <sup>b</sup>Only one child entered the nursery school when he was nineteen-month-old.

The children's Italian vocabulary skills were assessed at 24 months using the Italian version of the MacArthur-Bates Communicative Development Inventories (CDI): Words and Sentences-short form (Primo Vocabolario del Bambino-PVB, Caselli, Bello, Rinaldi, Stefanini, & Pasqualetti, 2015; Rinaldi, Pasqualetti, Stefanini, Bello, & Caselli, 2019). The short form of the CDI includes a 100-word expressive vocabulary checklist with lexical categories such as nouns (55), verbs (18), adjectives-adverbs (16) and closed-class words (11). The Italian version of the CDI was completed by the nursery teachers and – after the interview at nursery school – the parents who were familiar with Italian and the child's use of Italian. Since Romanian and Nigerian English adaptations of the CDI were not available, the L1 vocabulary skills of the bilingual children were assessed using the CDI-short form translated into the two languages by cultural mediators in collaboration with a language development researcher. Because the translation of each instrument is problematic (Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005), the cultural mediators were L1 native speakers who could adapt the words of the checklists to the Romanian and Nigerian cultures. One concept in the translated Romanian version and one concept in the translated Nigerian English version were changed because of their low input frequency in the respective languages (for Romanian 'orange soda' was changed to 'milk'; for Nigerian English 'stroller' was changed to 'baby pram').

## Measures

### Vocabulary size

The total number of words checked by parents in L1 served as outcome measures of vocabulary size in the heritage language. For productive vocabulary in Italian, a composite score was computed by providing a single credit for words reported by one or both reporters, because of evidence indicating the utility of combining parents and teachers reports (Pua, Lee, & Rickard Liow, 2017; Vagh, Pan, & Mancilla-Martinez, 2009). Composite inventory scores in Italian served as outcome measures of vocabulary size in the Italian-single vocabulary size. Then, the total vocabulary score was calculated by summing the score in L1 and the composite score in Italian; the total conceptual vocabulary score was calculated with credit being given for knowing a concept, regardless of whether the child knew the word for that meaning in one or both languages.

### Vocabulary composition

We calculated the numbers of nouns, verbs, adjectives/adverbs, and closed-class words as proportions of all words produced in Italian.

### Translation equivalents

First, all the lexical items that each child had in each language for the same concept were identified as TE pairs, based on the words checked in their inventories in L1 and in Italian. We included translation equivalents for content words only (nouns, verbs, adjectives and adverbs) because of the difficulty in assigning translation equivalents for closed-class categories. Second, cognates (TE pairs similar in sound and spelling [e.g. *festa-festa* 'party']) and semi-cognates (TE pairs similar in sound but with slightly different spelling [e.g. *penna-pen*]), were calculated for each child. Then, following Legacy et al. (2017) two calculations of TEs were performed: one to obtain an estimate of the proportion of TEs that included cognates and semi-cognates (total

TEs), and one to obtain the proportion of TEs excluding cognates and semi-cognates. The proportion of TEs including cognates and semi-cognates was calculated by summing the number of identified TE pairs in the inventories and multiplying this score by two. This number was then divided by the child's total vocabulary MINUS non-equivalents (i.e., words that have no translation on the other inventory form) and onomatopoeias. The proportion of TEs excluding cognates and semi-cognates was then calculated by summing the identified TE pairs in the inventories, subtracting all cognate and semi-cognate pairs, and multiplying by two. This number was then divided by the child's total vocabulary minus cognates, semi-cognates, and non-equivalents. Two independent raters identified the TE pairs in the inventories, and they came to an agreement on which words would be selected as pairs and on which TEs would be classified as cognate and semi-cognate pairs.

## Results

First, the normal distribution of the data was assessed using the Kolmogorov-Smirnov test. Most of the measures were found to violate the assumption of normality within groups. We therefore ran all analyses using robust statistical techniques that take into account the violations of the main assumptions required by parametric statistical analyses. Using robust statistics prevents type-I error ('false-positives') inflation (Erceg-Hurn, Wilcox, & Keselman, 2013) and increases the likelihood of discovering genuine differences between groups and associations among variables without increasing the risk of type-II errors ('false-negatives'). The types of corrections usually included in robust techniques are trimmed means and M-estimators. We used recently devised types of M-estimators called MM-estimators (Maechler *et al.*, 2018) for regressions, trimmed means and M-estimators for a robust mixed ANOVA (Mair & Wilcox, 2019), the Welch's correction for t-tests, and Winsorized correlation analyses.

### *Preliminary analyses related to maternal education and age/amount of exposure to Italian*

Since maternal education is known to impact on children's lexical abilities, and the NEN bilingual group differed significantly from the ROM bilingual group in the number of years of maternal education, we controlled for a possible effect of maternal education. Specifically, we used an MM-estimators robust analysis of covariance (with maternal education as covariate) based on regression (Maechler *et al.*, 2018). Furthermore, because it is well documented that language exposure impacts on the vocabulary skills of bilingual children, Winsorized correlation analyses were run for each bilingual group between vocabulary size in Italian (L2) and age (number of months) and amount of exposure to Italian and – as measures of exposure to Italian spoken by native-speakers – age of entry to nursery school and daily attendance there. In the ROM group none of the different measures of exposure to Italian was significantly related to Italian vocabulary size [ExpAge  $r_w(20) = -.322$ ,  $p = .154$ ; ExpAmount  $r_w(20) = .028$ ,  $p = .904$ ; EntrNursery  $r_w(20) = .113$ ,  $p = .619$ ; AttNursery  $r_w(20) = .137$ ,  $p = .548$ ]. Similar results emerged for the NEN group, although in this group the age of exposure to Italian was related to Italian vocabulary size at 6% [ExpAge  $r_w(15) = -.484$ ,  $p = .061$ ; ExpAmount  $r_w(15) = .162$ ,  $p = .540$ ; EntrNursery

$r_w(15) = .105, p = .692$ ; AttNursery  $r_w(15) = .190, p = .473$ ]. Therefore, in the following analyses, we did not control for a possible effect of this factor.

### *Vocabulary size (RQ1)*

First, a robust mixed ANOVA was computed on the single-language vocabularies of bilingual children with language (L1 vs L2) as within-subject factor and group (ROM vs NEN) as between-subject factor, using WRS2:bwtrim in R (Mair & Wilcox, 2019). Then, robust Welch's T-tests were computed to compare the ROM and NEN groups on total vocabulary and total conceptual vocabulary. Descriptive statistics are reported in Table 2. The robust ANOVA showed that L2 vocabulary was significantly larger than L1 vocabulary in both ROM and NEN groups [ $F(1, 17.45) = 8.38, p = .001, d = .96$ ]. The two groups did not significantly differ in their single-language vocabulary sizes [ $F(1, 20.67) = 2.04, p = .168, d = .474$ ], and no effect was found for the interaction between group and language [ $F(1, 20.67) = 0.06, p = .814, d = .079$ ]. T-tests also showed that in their total vocabulary and total conceptual vocabulary sizes the two groups of bilingual children did not significantly differ [respectively,  $t(36.78) = 1.17, p = .248, d = .370$ , and  $t(36.84) = 1.44, p = .159, d = .380$ ].

### *Vocabulary composition (RQ2)*

Four Welch's T-tests with group (ROM vs NEN) as an independent factor were performed on the proportions of nouns, adjectives/adverbs, verbs and closed-class words in Italian. Descriptive statistics are reported in Table 2. No significant differences were found in the proportions of any lexical category in Italian vocabulary between the two groups of bilingual children.

### *Translation equivalents (RQ3)*

Robust Welch's T-tests were conducted to compare the proportions of total TEs and TEs minus cognates and semi-cognates between the two bilingual groups. Descriptive statistics are reported in Table 3. The bilingual children in the ROM group produced a significantly higher proportion of total TEs than those in the NEN group [ $t(32.17) = 2.91, p = .006, d = .964$ ]. No significant differences emerged between the two groups when cognates and semi-cognates were excluded from the count of TEs [ $t(34.31) = 1.55, p = .131, d = .504$ ].

## **Discussion**

The present study examined the expressive vocabulary of two groups of bilingual toddlers from low-income immigrant families in both the majority language of the country (Italian) and their heritage languages (Romanian and Nigerian English), in order to contribute to the debate on the relationship between vocabularies in L1 and L2.

Significantly higher vocabulary in Italian than in the heritage language emerged in both groups. These results suggest that children orientate towards the majority language, as observed in other immigration contexts (Oller, Jarmulowicz, et al., 2007). Indeed, heritage languages of immigrant families are minority languages which are less likely to be supported by the larger community. Therefore, these children might, regardless of their degree of exposure, have experienced that Italian is the language that can be used successfully to communicate in different contexts. At the nursery school Italian was the only language used to communicate, while at



**Table 2.** Expressive Vocabulary Scores (Number of Words), and Proportion of Lexical Categories on Italian Vocabulary by Groups

	ROM bilingual group (n = 22)			NEN bilingual group (n = 17)		
	<i>M</i>	( <i>SD</i> )	range (raw scores)	<i>M</i>	( <i>SD</i> )	range (raw scores)
Italian Vocabulary	30.55	(20.05)	8–74	23.65	(16.61)	1–60
L1 Vocabulary	19.55	(19.65)	0 <sup>a</sup> –67	11.24	(15.09)	0 <sup>b</sup> –52
Total Vocabulary	50.09	(37.41)	14–124	34.82	(28.92)	9–112
Total Conceptual Vocabulary	35.82	(22.44)	14–76	28.24	(16.09)	9–66
Nouns/Italian vocabulary	0.72	(0.14)	0.40–0.93 (6–48)	0.76	(0.14)	0.61–1.00 (1–45)
Adjectives-adverbs/Italian vocabulary	0.10	(0.08)	0–0.27 (0–14)	0.10	(0.08)	0–0.25 (0–10)
Verbs/Italian vocabulary	0.08	(0.06)	0–0.18 (0–10)	0.07	(0.05)	0–0.16 (0–8)
Close-class words/Italian vocabulary	0.09	(0.08)	0–0.33 (0–5)	0.07	(0.05)	0–0.18 (0–4)

Note. <sup>a</sup>Only one child in the ROM group had L1Voc. = 0; <sup>b</sup>Only one child in the NEN group had L1Voc. = 0.

**Table 3.** Proportions of Translation Equivalents (TEs) on Total Vocabulary by Groups

	ROM bilingual group (n = 22)				NEN bilingual group (n = 17)			
	<i>M</i>	( <i>SD</i> )	range	(raw scores)	<i>M</i>	( <i>SD</i> )	range	(raw scores)
Total TEs/total vocabulary	0.43	(0.24)	0–0.80	(0–46)	0.19	(0.27)	0–0.76	(0–38)
TEs/total vocabulary (minus cognates and semi-cognates)	0.29	(0.28)	0–0.84	(0–27)	0.15	(0.28)	0–0.81	(0–31)

home at least one of the two immigrant parents could understand Italian well, and most of the immigrant parents (34 out of 39) said they also spoke Italian at home. In other words, communication could be conducted in Italian in both contexts (home and nursery school), while the heritage language was useful only at home. In this regard, studies comparing L1 and L2 vocabulary size in preschool bilingual children have suggested that early educational settings in which only L2 is spoken may influence the L2 “dominance” in children’s expressive vocabulary (Kan & Kohnert, 2005; Sheng, Lu, & Kan, 2011).

Although the Romanian-speaking children showed larger vocabulary sizes than the Nigerian English-speaking children, especially for the total vocabulary, our first hypothesis was not supported by the findings to a statistically significant extent. Coherently with the lack of differences in their vocabulary sizes, the children’s vocabulary composition in Italian did not differ between the two bilingual groups.

Finally, the differences found in the proportion of total translation equivalents between the two bilingual groups were in line with our third hypothesis. Romanian-Italian-speaking children produced a significantly higher proportion of total translation equivalents (more than double) than Nigerian English-Italian-speaking children. However, when cognates and semi-cognates were excluded from the calculation of the proportions of translation equivalents, no differences were found between the two bilingual groups. This result indicates that most of the translation equivalents produced by Romanian-speaking children were identical or similar cross-linguistic items. This finding could be interpreted in the light of the L1-L2 phonological similarity between Romanian and Italian. Similar evidence of the role of phonological form proximity on TEs in bilingual toddlers’ expressive vocabulary was reported by Bosch and Ramon-Casas (2014) for Spanish-Catalan bilingual children at 18 months of age. They showed that TEs represented 30% of the bilinguals’ total vocabulary size and, interestingly, form-identical TEs (cognates) accounted for 28% of the total vocabulary. Dissimilar words, meanwhile, accounted for 2% and were only present in the largest vocabularies (Bosch & Ramon-Casas, 2014). What is the basis for the facilitation effect of form-similarity on the acquisition of cross-linguistic synonyms? As suggested (Bosch & Ramon-Casas, 2014), children experience a double exposure to words with very similar phonological forms linked to the same concept – therefore, the form-concept pairing may be facilitated since it is heard in both languages. In addition, the ability to produce a word-form may drive the acquisition of other similar word-forms, as documented for monolingual children (Vihman, 1993). A recent study on phonological priming in bilingual preschoolers found a facilitating effect in target recognition when L2 primes were phonologically overlapped with L1 target words (Von Holzen & Mani, 2012). This result provided evidence of interconnectivity between the lexicons in first language and in second language: specifically a cross-language activation based on the phonological form. Learning a word in one language may therefore facilitate the acquisition of a form-similar word in a second language and, in turn, foster vocabulary acquisition in both languages, as suggested by Floccia and colleagues (Floccia *et al.*, 2018).

## Conclusions

The present study makes an original contribution to the literature on the early expressive vocabulary of bilingual children. It extends previous research in two ways. First, few studies examining early bilingual vocabulary development have focused on children speaking Italian as the majority language. Second, we compared bilingual

groups with two heritage languages that differ in their phonological distance from Italian.

The small sample size is a major limitation of this study when interpreting the results. In the light of the small sample size, we used robust statistical analyses that avoid the production of false positives and control the probability of producing false negatives. However, the sample size in future studies needs to be enlarged in order to deepen the findings of the current study. Another limitation is that short versions of a Romanian and a Nigerian English CDI were developed for the present study by translating the short version of the Italian CDI without validating them with a norm group. Therefore, although particular care was taken to develop parallel word-lists and assessment procedures, these measures of Romanian and Nigerian English vocabulary size might not have captured the bilingual children's total L1 vocabulary size.

The study produced evidence of some interconnectivity between vocabulary acquisition in the first language and in the second language, because the phonological form similarity between L1 and L2 emerged as a factor that facilitates the acquisition of cross-linguistic synonyms and supports lexical learning. Moreover, the dominance of the majority language in the expressive vocabulary of young bilingual children from immigrant families regularly speaking their heritage languages suggests that we need to deepen our understanding of the role of educational and cultural contexts in multiple language acquisition.

These results have practical implications. The phonological distance between the two languages should be taken into account in educational contexts, in order to differentiate vocabulary learning opportunities, going beyond the idea that all bilingual children from immigrant families have the same language learning needs.

**Acknowledgements.** Chiara Barachetti, Marinella Majorano, Manuela Lavelli, Department of Human Sciences, University of Verona, Italy; Elena Antolini, Rosanna Zerbato, Pedagogical Coordination of the Nursery Schools and Preschools, Municipality of Verona; Germano Rossi, Department of Psychology, University of Milano-Bicocca, Italy. The authors are grateful to Matilde Mognetti, all the nursery teachers, the children, and the parents who participated in the study.

## References

- Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2, 29–49.
- Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, 38, 317–322.
- Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, 128, 407–416.
- Caselli, M. C., Bello, A., Rinaldi, P., Stefanini, S., & Pasqualetti, P. (2015). *Il Primo Vocabolario del Bambino: Gestì, Parole e Frasi* [The First Vocabulary of the Child: Gestures, words and phrases]. Milano: Franco Angeli.
- Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, 77, 712–735.
- David, A., & Wei, L. (2008). Individual differences in the lexical development of French–English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 11, 598–618.
- De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.
- Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). Robust statistical estimation. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods. Volume 1: Foundations* (pp. 388–06). Oxford, UK: Oxford University Press.

- Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., ... Plunkett, K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance: III: Analyses and results for Study 1: Estimating the effect of linguistic distance on vocabulary development. *Monographs of the Society for Research in Child Development*, 83, 43–60.
- Gampe, A., Kurthen, I., & Daum, M. M. (2018). BILEX: A new tool measuring bilingual children's lexicons and translational equivalents. *First Language*, 38(3), 263–283.
- Hammer, C. S., Jia, G., & Uchikoshi, Y. (2011). Language and literacy development of dual language learners growing up in the United States: A call for research. *Child Development Perspectives*, 5, 4–9.
- Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology*, 11, 381–394.
- Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, 48, 372–378.
- Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or chien? Translation equivalents in the receptive and expressive vocabularies of young French-English bilinguals. *Journal of Child Language*, 44, 881–904.
- Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2016). Vocabulary size, translation equivalents, and efficiency in word recognition in very young bilinguals. *Journal of Child Language*, 43, 760–783.
- Levey, S., & Cruz, D. (2003). The first words produced by children in bilingual English/Mandarin Chinese environments. *Communication Disorders Quarterly*, 24, 129–136.
- Machler, M., Rouseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., ... di Palma, M. A. (2018). Robustbase: Basic Robust Statistics R package version 0.93-3. Downloaded from <http://CRAN.R-project.org/package=robustbase>, July 10, 2019.
- Mair, P., & Wilcox, R. (2019). Robust statistical methods in R using the WRS2 Package. *Behavior Research Methods*, 52, 464–488. <https://doi.org/10.3758/s13428-019-01246-w>
- Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish-English bilinguals. *Journal of Child Language*, 37, 817–840.
- Oller, D. K., Jarmulowicz, L., Gibson, T., & Hoff, E. (2007). First language vocabulary loss in early bilinguals during language immersion: A possible role for suppression. In H. Caunt-Milton, S. Kulatilake, & I. Woo, (Eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development* (pp. 474–484). Somerville, MA: Cascadilla Press.
- Onofrio, D., Rinaldi, P., & Pettenati, P. (2012). Il primo sviluppo del linguaggio in bambini che imparano più lingue: una proposta per la valutazione e l'interpretazione del profilo linguistico. *Psicologia Clinica dello Sviluppo*, 16, 661–669.
- Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language*, 22, 345–368.
- Pettenati, P., Vacchini, D., Stefanini, S., & Caselli, M. C. (2011). Parole e frasi nel primo vocabolario di bambini bilingui italiano-spagnolo. *Rivista di Psicolinguistica Applicata*, 11(1–2), 49–67.
- 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, 57–70.
- Pua, E. P. K., Lee, M. L. C., & Rickard Liow, S. J. (2017). Screening bilingual preschoolers for language difficulties: Utility of teacher and parent reports. *Journal of Speech, Language, and Hearing Research*, 60, 950–968.
- Rinaldi, P., Pasqualetti, P., Stefanini, S., Bello, A., & Caselli, M. C. (2019). Long and short form of Gesture and Word - MB-CDI Italian version McArthur Bates: Norming data and validity study. *Journal of Child Language*, 46, 546–566.
- Rinker, T., Budde-Spengler, N., & Sachse, S. (2017). The relationship between first language (L1) and second language (L2) lexical development in young Turkish-German children. *International Journal of Bilingual Education and Bilingualism*, 20, 218–233.
- Schepens, J., Van Der Slik, F., & Van Hout, R. (2013). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin, & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 199–230). Boston, MA: De Gruyter.
- Sheng, L., Lu, Y., & Kan, P. F. (2011). Lexical development in Mandarin-English bilingual children. *Bilingualism: Language and Cognition*, 14, 579–587.

- Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J.** (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development, 80*, 1545–1563.
- Vihman, M. M.** (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61–82.
- Volterra, V., & Taeschner, T.** (1978). The acquisition and development of language by bilingual children. *Journal of Child Language, 5*, 311–326.
- Von Holzen, K., & Mani, N.** (2012). Language non-selective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology, 113*, 569–586.
- Widenfelt, B., Treffers, P., de Beurs, E., Siebelink, B., & Koudijs, E.** (2005). Translation and Cross-Cultural Adaptation of Assessment Instruments Used in Psychological Research with Children and Families. *Clinical Child and Family Psychology Review, 8*, 135–147.

---

**Cite this article:** Barachetti C, Majorano M, Rossi G, Antolini E, Zerbato R, Lavelli M (2022). Vocabulary production in toddlers from low-income immigrant families: evidence from children exposed to Romanian-Italian and Nigerian English-Italian. *Journal of Child Language* **49**, 408–421. <https://doi.org/10.1017/S0305000921000222>