

Language-specific noun bias: evidence from bilingual children*

LEI XUAN

University of Texas Southwestern Medical Center at Dallas

AND

CHRISTINE DOLLAGHAN

University of Texas at Dallas

(Received 7 October 2011 – Revised 9 April 2012 – Accepted 7 September 2012 –
First published online 9 November 2012)

ABSTRACT

Most evidence concerning cross-linguistic variation in noun bias, the preponderance of nouns in early expressive lexicons (Gentner, 1982), has come from comparisons of monolingual children acquiring different languages. Such designs are susceptible to a number of potential confounders, including group differences in developmental level and sociodemographic characteristics. The aim of this study was to quantify noun bias in bilingual Mandarin–English toddlers whose expressive lexicons in each language contained 50–300 words. Parents of fifty children (1;10–2;6) reported separately on their English and Mandarin expressive lexicons. The mean percentage of Mandarin nouns (38%) was significantly lower than the percentage of English nouns (54%) and was robust to analyses of twelve potential covariates. Analyses of the most frequently reported words suggested that lexical reduplication could be considered as a potential influence on vocabulary composition in future studies. Results suggest that characteristics of the input significantly shape early lexicons.

[*] This research was based on the first author's dissertation at the University of Texas at Dallas and was supported by a dissertation grant. We thank William Katz, Mandy Maguire, Virginia Marchman, and Robert Stillman for their invaluable comments and support. We also thank Christina Worle for assistance with reliability coding. Finally, we are indebted to the parents who participated in the study. Address for correspondence: Lei Xuan, Department of Clinical Sciences, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9169, USA. e-mail: lei.xuan@UTSouthwestern.edu

INTRODUCTION

The hypothesis that early lexical acquisition is shaped to a significant degree by a universal preference for nouns over other word types was first proposed by Gentner (1982), and considerable cross-linguistic evidence consistent with this ‘noun bias’ hypothesis has accrued since that time (Bates *et al.*, 1994; Bornstein *et al.*, 2004; Gentner & Boroditsky, 2009; Goldfield, 2000). The noun bias effect could imply a universal order of lexical acquisition resulting from fundamental cognitive constraints (Gentner & Boroditsky, 2001). As Gentner (1982: 301) stated: “... the linguistic distinction between nouns and predicate terms, such as verbs and prepositions, is based on a preexisting perceptual-conceptual distinction between concrete concepts such as persons or things and predicative concepts of activity, change-of-state, or causal relations; and ... the category corresponding to nouns is, at its core, conceptually simpler or more basic than those corresponding to verbs and other predicates.”

Some languages, however, including Mandarin Chinese and Korean (e.g. Au, Dapretto & Song, 1994; Choi & Gopnik, 1995; Tardif, Gelman & Xu, 1999) appear to be relatively more ‘verb friendly’ than others (Gentner & Boroditsky, 2009: 8). Such cross-linguistic variation has spurred a number of hypotheses about potential influences on the composition of early lexicons, including cognitive factors (Gentner, 1982; Gentner & Boroditsky, 2001; Maguire, Hirsh-Pasek & Golinkoff, 2006; Nelson, Hampson & Shaw, 1993), linguistic factors (Gentner, 1982; Slobin, 1985), and input factors (Gentner, 1982; Goldfield, 1993; Gopnik, Choi & Baumberger, 1996; Tardif, Shatz & Naigles, 1997).

Interpreting the literature on noun bias is complicated by significant heterogeneity among studies of the phenomenon (Gentner & Boroditsky, 2001). Studies have varied with respect to participant characteristics (e.g. age, level of linguistic development, sociodemographic background), vocabulary sampling methods (e.g. parent report, spontaneous production, elicited production), the criteria for identifying nouns (e.g. common nouns, proper nouns), and the formulas used to calculate noun percentages (e.g. types or tokens, relative to total vocabulary or to subsets of vocabulary). In addition, most data concerning cross-linguistic differences in noun bias have come from between-subjects research designs in which groups of monolingual children acquiring different languages are compared. Because such groups may differ not only in their native languages but also with respect to many other variables that could influence vocabulary development, such as cognitive skills and sociodemographic characteristics (e.g. Caselli, Casadio & Bates, 2001; Tardif *et al.*, 1999), it is difficult to confidently attribute noun bias differences to the language factor alone.

Studies of bilingual children employing within-subject designs in which factors such as cognitive development and sociodemographic background

are consistent within each child have the potential to yield stronger evidence concerning cross-linguistic variations in noun bias (Dale, Dionne, Eley & Plomin, 2000; Marchman, Martínez-Sussmann & Dale, 2004). However, only a few such studies have been reported (Conboy & Thal, 2006; Lucas & Bernardo, 2008; Marchman, Xuan & Yoshida, 2005a), and in none of these does it appear that participants had comparable levels of vocabulary development in each of their two languages. This is an important consideration because in monolingual children the degree of noun bias varies considerably at different vocabulary sizes, as illustrated in Figure 1. This forest plot represents the percentage of nouns reported from studies of typically developing children, aged 1;6 to 2;6, who were acquiring one of five languages: English ($N=1803$; Bates *et al.*, 1994), Spanish ($N=68$; Jackson-Maldonado, Thal, Marchman, Bates & Guitierrez-Clellen, 1993), Italian ($N=581$; Caselli *et al.*, 1995; Caselli *et al.*, 2001), Korean ($N=90$; Pae, 1993), or Mandarin ($N=1056$; Tardif, Fletcher, Zhang & Liang, 2002). Vocabulary counts in all of these studies were based on the MacArthur-Bates Communicative Development Inventories: Words and Gestures for infants and Words and Sentences for toddlers (Fenson *et al.*, 1993) or their culturally equivalent adaptations, and in all studies nouns were defined as common nouns (Bates *et al.*, 1994). Further, in all of these studies the mean percentage of nouns as a function of total vocabulary size either was reported or could be estimated from the lexical curves at different vocabulary sizes. The forest plot shows the mean percentage of nouns at each vocabulary size and its corresponding confidence interval, either as reported by the investigators or, when possible, calculated from the raw data; the size of the datapoint reflects the number of children contributing to the value.

Figure 1 suggests that noun bias varies among these five languages and among the monolingual children acquiring them, as indicated by the width of the confidence intervals. More importantly, however, the percentage of reported nouns also differed at different vocabulary sizes, within and between languages. For example, the largest difference (15%) between the percentage of nouns in English (54%) and the percentage of nouns in Mandarin (39%) occurs within the vocabulary range from 50–300 words. But the size of this difference is tied to this particular vocabulary size window; the difference is smaller at other vocabulary sizes. In addition, although noun percentages within a single language vary considerably over the full range of vocabulary sizes, there is reasonable consistency within each language in the 50–300 word range. For example, although the percentages of nouns in English range from 16% to 55% in vocabularies of 1–5 words and 100–200 words, respectively, within the 50–300 word range the percentages vary by just 4%. Similarly, although the percentage of Mandarin nouns varies from 16% to 39% over the full range of vocabulary sizes, it varies by only 8% within the 50–300 word interval.

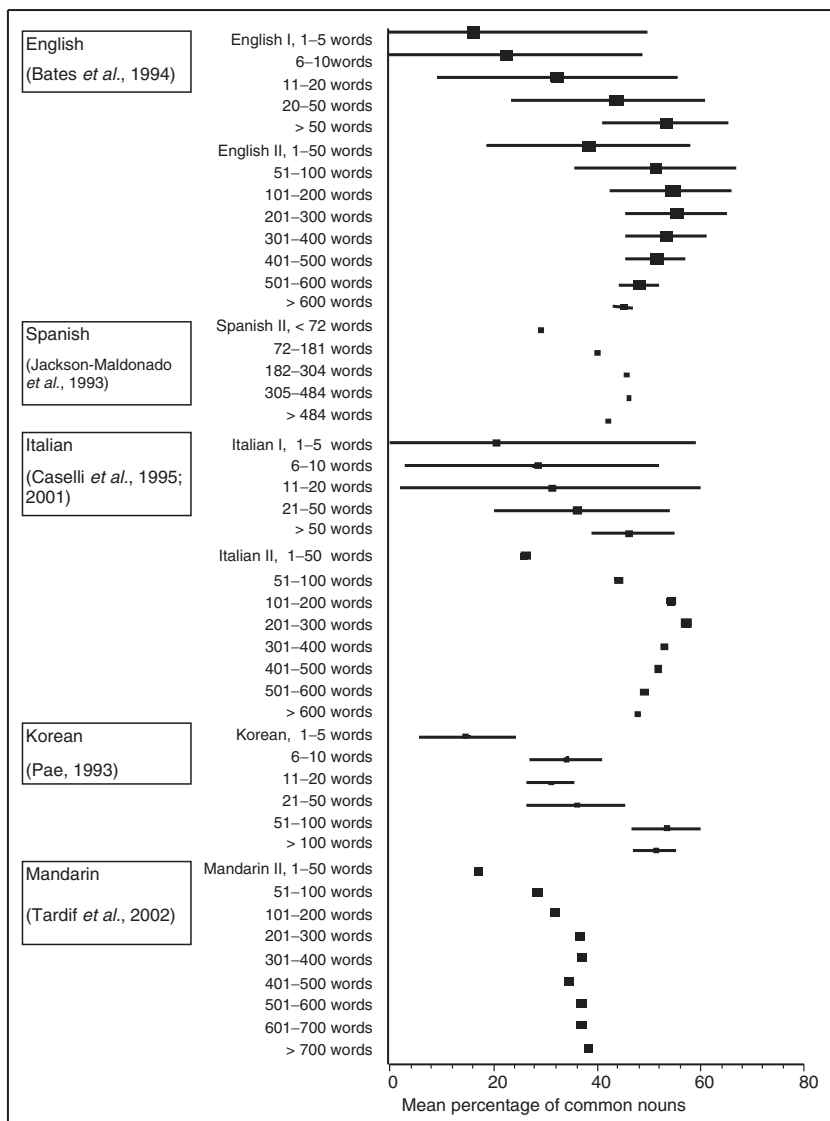


Fig. 1. A forest plot of the mean percentages of nouns in five languages, based on the MacArthur-Bates Communicative Development Inventories I: Words and Gestures for infants and II: Words and Sentences for toddlers (Fenson *et al.*, 1993) and their culturally equivalent adaptations. Sample size at each vocabulary level, reflected in the size of the datapoint, ranged from $n = 12$ (Korean, 51-100 words) to $n = 155$ (English, 401-500 words). Confidence intervals are indicated by the horizontal line through each datapoint. The 95% CI was reported only for the Korean data; for English and Italian data, the 10th-90th percentiles are shown.

For the present study, we examined noun bias in young bilingual children acquiring English and Mandarin, two strikingly different languages that are believed to fall near the extreme ends of the noun bias continuum. By studying bilingual children we hoped to minimize the threat of confounding due to individual differences in cognitive and sociodemographic factors. By focusing on children whose parent-reported vocabularies in both English and Mandarin fell within the 50–300 word interval, we hoped to reduce the impact of variations in noun bias at different vocabulary sizes. Our objective was to provide the clearest test to date of the hypothesis that children acquiring these two languages manifest significantly different degrees of noun bias. We hypothesized not just that the two languages would differ significantly, but that the mean percentage of nouns in English would exceed the mean percentage of nouns in Mandarin by 15%, a value selected based on the forest plot analysis described above.

In addition to enabling a test of the noun bias hypothesis, the parent-reported vocabularies from these bilingual toddlers provided an opportunity to explore the relationship between the words that were most frequently reported in Mandarin and in English. Accordingly, we also examined the characteristics of the words that occurred most often in the Mandarin and English lexicons of these children.

METHOD

Participants

Selection procedures. To identify a sample of typically developing bilingual children with parent-reported vocabulary sizes in the target range in each language (50–300 words) while minimizing the amount of data collected from parents whose children would not meet these criteria, participant selection occurred in two stages. First, parents of 117 bilingual English–Mandarin children completed a brief screening questionnaire concerning the child's age, sex, medical and developmental history, birth order, primary caregiver, ages at which the child's exposure to English and to Mandarin began, percentages of time that the child was exposed to English and to Mandarin (totaling 100%), and the estimated number of words spoken by the child in each language (fewer than 50, 50–300, or more than 300). The last question was included in an effort to minimize the number of parents whose children would be enrolled in the study but would later be found not to meet the vocabulary size criterion using parent reports. Parents also reported their own ethnicity and education level.

Based on the screening questionnaires, no children were excluded due to age, medical or developmental problems, or significant exposure to more than two languages, but four children were excluded because their estimated vocabulary size in at least one language exceeded 300 words. Thirty-one

children were eligible for inclusion immediately by virtue of parent estimates that their vocabulary sizes were between 50 and 300 words. The remaining eighty-two children had estimated vocabulary size(s) of fewer than 50 words but could become eligible as their vocabularies increased over the next few months. Parents of these eighty-two potentially eligible children consented to estimate their child's vocabulary size in each language every month until the child became eligible for the study or reached the age of 2;7 (years; months); of these, seventy-two became eligible within the study's duration.

Accordingly, parents of 103 children (31 who were eligible immediately and 72 who became eligible later) completed the MacArthur-Bates Communicative Development Inventory: Words and Sentences (CDI; Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007) and its Mandarin Chinese adaptation, the Putonghua Communicative Development Inventory: Words and Sentences (PCDI; Tardif, Fletcher, Zhang, Liang & Zuo, 2008) no more than one week after a child was identified as eligible for the study. Parents were instructed to complete the forms in an order that had been determined randomly and to take a break of at least 10 minutes between completing the two forms. They were instructed to complete the forms independently, although they were allowed to ask family members or teachers about any specific words about which they were unsure.

Data from two parents were excluded due to forms being filled out incompletely or incorrectly (i.e. filling out a single form to report on vocabulary in both languages). For fifty of the remaining 101 children, reported expressive vocabulary sizes in both English and Mandarin fell within the target range of 50–300 words. These fifty children constituted the sample for the present study; their characteristics are described below.

Subject characteristics. Table 1 summarizes participant characteristics ($N=50$). Their ages ranged from 1;10 to 2;6 ($M=26.9$ months, $SD=2.2$); there was a higher percentage of girls than boys and a higher percentage of later-born children than of first-borns. Most of the children (66%) either were enrolled in daycare or were receiving childcare from a non-parent caregiver at the time of the study. Although all of the children resided in the US and were reportedly exposed to both English and Mandarin at birth or soon after birth, on average their parents estimated that they were exposed to relatively more Mandarin than English.

All mothers, and 90% of fathers, of children in the sample were first-generation immigrant Chinese in the US who reported Mandarin as their native language; the other 10% of fathers reported their ethnicity as White and their native language as English. All parents had graduated from high school and almost all parents (94% of mothers, 92% of fathers) had graduated from college. The majority of respondents completing the vocabulary measures were mothers.

TABLE I. *Sample characteristics* ($N = 50$)

Sex	
Female	62%
Male	38%
Birth order	
First born	42%
Later born	58%
Primary caregiver	
Mother at home	34%
Daycare or non-parent	66%
Respondent to CDI and PCDI	
Mother	88%
Father	12%
Mean (SD) age in months	26.9 (2.2)
Mean (SD) % daily exposure to Mandarin (of 100)	63% (20%)
Mean (SD) % daily exposure to English (of 100)	37% (20%)

Vocabulary measures

Vocabulary was measured by parent report, using the vocabulary checklists of the CDI (Fenson *et al.*, 2007) and its Mandarin counterpart PCDI (Tardif *et al.*, 2008). Such measures have been used widely in studies of vocabulary composition (Bates *et al.*, 1994; Tardif *et al.*, 1999; Tardif *et al.*, 2002). Parent-report measures have several advantages over direct observational measures such as spontaneous language analysis for sampling young children's expressive vocabularies. Several investigators have noted that naturalistic speech must be sampled extensively, over lengthy intervals and in varying contexts, to estimate expressive vocabulary (e.g. Pine, Lieven & Rowland, 1996). Evidence also suggests that naturalistic data are highly variable across contexts such as book reading versus toy play (Tardif *et al.*, 1999). By contrast, with the CDI recognition (checklist) format parents can validly report expressive vocabulary in as little as 30 minutes (Dale, Bates, Reznick & Morisset, 1989; Dale, 1991; Pine, 1992). Parent-report measures also avoid the problem of young children's reluctance to speak in unfamiliar settings and with unfamiliar people, and may reduce the impact of variations in the behaviors of children's conversational partners during naturalistic language sampling. In addition to having been designed for children in the same age range, the checklist format and administration instructions are the same in the CDI and PCDI, and their checklists are organized into semantic categories that generally map onto syntactic categories, making it relatively easy to identify words in certain categories as nouns. Finally, evidence suggests that the CDI and PCDI are valid measures of lexical development in bilingual as well as monolingual children (Marchman & Martínez-Sussmann, 2002; Marchman, Xuan & Yoshida, 2005b).

The English CDI contains 680 words grouped into twenty-two semantic categories; the Mandarin PCIDI contains 799 words grouped into the same twenty-two categories along with two additional categories that are specific to Mandarin, 'Classifiers' and 'Sentence Final Particles'. On each form the child's total vocabulary size was the total number of words that the child was reported to use. It should be noted that both the CDI and the PCIDI list include some words as members of more than one category; for example, on both forms the word *water* appears in both 'Outside Things' and 'Food & Drink' categories and the word *fish* appears in both the 'Animals' and 'Food & Drink' categories. In addition, there are some minor inconsistencies in category assignments between the two forms; for example, the word *hi* is listed in 'Games & Routines' on the CDI while the word translated as *hello* appears in 'Sound Effects' on the PCIDI. Such variations did not affect the total vocabulary size, however.

Consistent with several large-scale studies (e.g. Bates *et al.*, 1994), in the present study the total number of nouns was the sum of words reported in eight semantic categories: 'Animals', 'Vehicles', 'Toys', 'Food & Drink', 'Clothing', 'Body Parts', 'Small Household Items', and 'Furniture & Rooms'; words in the categories 'People', 'Outside Things', and 'Places to Go' were not included. The percentage of nouns in each language was calculated by dividing the total number of nouns in that language by the total number of words in that language.

To facilitate comparisons with previous studies contrasting the relative frequencies of nouns and verbs (e.g. Lucas & Bernardo, 2008; Tardif *et al.*, 1999), we also calculated for each language the number of verbs, defined as words in the 'Action Words' category.

Reliability. To determine inter-rater reliability for word counts, 20% of the CDI and 20% of the PCIDI report forms were randomly selected for independent coding by a trained research assistant blinded to the research questions. The point-by-point percentage of agreement for the individual words entered by each coder into a spreadsheet was >99%.

Statistical analyses

A within-subject paired sample design was used to test the hypothesis that the mean percentage of nouns in English would exceed the mean percentage of nouns in Mandarin by at least 15%, a value selected based on the forest plot described earlier. With this specified mean percentage difference and a Cohen's *d* (effect size) of 0.6, the sample size needed to detect a difference of this magnitude with *p* (one-tailed) < 0.05, power > 0.80 and a correlation of $r = -0.20$ (Marchman *et al.*, 2005a) was at least forty participants (Cohen, 1988). With the obtained sample size of fifty, statistical power for this analysis was 0.88.

TABLE 2. Mean (*M*), standard deviation (*SD*), and range for total number of words, total number of nouns, and total number of verbs reported on the English CDI and Mandarin PCDI forms

Language	English			Mandarin		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Total words	121	62	51–300	190	71	50–299
Total nouns	65	33	21–157	73	31	14–137
Total verbs	13	10	0–41	45	21	6–91

In addition to testing the primary hypothesis, we conducted a repeated measures ANOVA with 2 levels of language (Mandarin, English) and 2 levels of word type (Nouns, Verbs).

Word frequency analyses

A list of the most frequent words used in each language was compiled across the sample by tallying the percentage of children who were reported to use each of the words on the CDI and on the PCDI. The fifty words with the highest percentages in each language were compared in an effort to identify features that might have contributed to their high frequency ranking in each language.

RESULTS

Statistical analyses

Table 2 shows descriptive statistics concerning the raw frequencies of words, nouns, and verbs reported in each language. Relative to monolingual children aged 2;2 acquiring English (Fenson *et al.*, 1993) or Mandarin (Tardif *et al.*, 2008), vocabulary levels of these bilingual children in each language fell at approximately the 10th percentile. When English and Mandarin words were combined (Pearson & Fernández, 1994), the vocabulary size of children in the present sample fell at approximately the 40th percentile relative to monolingual English-speaking children (Fenson *et al.*, 1993). In short, it appears that the bilingual participants in this study had lexical skills consistent with their ages.

Table 2 also shows that, on average, the children's Mandarin vocabularies were larger than their English vocabularies and that nouns outnumbered verbs in both languages. In addition, the size of the noun–verb gap was larger in English than in Mandarin, consistent with previous studies (Tardif *et al.*, 1999). These observations were supported by the results of the 2 (Mandarin, English) by 2 (Nouns, Verbs) repeated measures ANOVA

of the raw frequencies shown in Table 2, which showed a significant main effect of language ($F(1, 49) = 16.21$, $p < 0.001$, $\eta^2 = 0.25$) and of word type ($F(1, 49) = 277.18$, $p < 0.001$, $\eta^2 = 0.85$), as well as a significant interaction ($F(1, 49) = 21.44$, $p < 0.001$, $\eta^2 = 0.30$).

With respect to the primary hypothesis that the percentage of English nouns would exceed the percentage of Mandarin nouns by at least 15%, the mean percentage of nouns in these bilingual children's English vocabularies was 54% ($SD = 9\%$, 95% CI [52%, 57%]); the mean percentage of nouns in their Mandarin vocabularies was 38% ($SD = 7\%$, 95% CI [36%, 40%]). This mean difference between languages, 16%, was significant (paired $t(49) = 12.29$, $p < 0.001$, $d = 1.73$). Results were similar in the subset of ten children whose vocabulary size in each language fell within the narrower range of 100–200 words; in this group, the percentage of English nouns exceeded that of Mandarin nouns by 20% (paired $t(9) = 7.21$, $p < 0.001$, $d = 2.28$). The difference between English and Mandarin in percentage of nouns was robust when twelve covariates that have been linked to word learning were taken into consideration. Neither children's demographic variables (i.e. age, sex, birth order, number of children at home, English exposure level, and Mandarin exposure level) nor parents' variables (i.e. mother's and father's ethnicity, mother's and father's educational level, primary caregiver, and questionnaire respondent) accounted for significant amounts of variance (all p values > 0.05).

As noted earlier, the procedures for defining nouns and for calculating the percentage of nouns have varied across studies (Caselli *et al.*, 2001; Gentner, 1982; Gentner & Boroditsky, 2009; Tardif *et al.*, 1999). To facilitate comparisons with these alternative approaches, we conducted two additional analyses. First, we calculated the percentage of nouns relative to the sum of nouns and verbs rather than to the total number of words. Doing so increased the relative percentage of nouns in both languages, to 84% for English and to 62% for Mandarin, maintaining the original finding of a substantially greater percentage of nouns in English than in Mandarin.

We also examined the extent to which including words in the 'People' category as nouns would alter the results of the original analysis. Counting 'People' words as nouns increased the mean number of nouns from 65 to 71 in English and from 73 to 86 in Mandarin; the corresponding percentages of nouns relative to total words were 59% and 46%, while the corresponding percentages of nouns relative to the sum of nouns and verbs were 86% and 67%, respectively. Although the size of the gap between English and Mandarin in the percentage of nouns relative to total words (13%) and the size of the gap in the percentage of nouns relative to the sum of nouns and verbs (19%) were slightly smaller when 'People' words were included than in the original analysis (16% vs. 22%), the finding of a substantially higher percentage of nouns in English than in Mandarin was maintained.

TABLE 3. *Most frequent 55 English words and their CDI categories, ranked by percentage of parents (N=50) reporting them; the 33 nouns are **bolded** and the sole verb is marked with an asterisk (*)*

Word	CDI category	%	Word	CDI category	%
apple	Food and Drink	96	eye	Body Parts	74
no	Games/Routines	96	nose	Body Parts	74
hi	Games/Routines	94	moo	Sound Effects	72
dog	Animals	92	bear	Animals	72
bye	Games/Routines	92	duck	Animals	72
fish	Animals	90	cookie	Food and Drink	72
car	Vehicles	90	ice cream	Food and Drink	72
ball	Toys	90	milk	Food and Drink	72
baby	People	88	shoe	Clothing	72
banana	Food and Drink	86	ear	Body Parts	72
yes	Games/Routines	86	pig	Animals	70
car	Animals	84	bubbles	Toys	70
mommy	People	84	pizza	Food and Drink	70
juice	Food and Drink	82	yum yum	Sound Effects	68
thank you	Games/Routines	82	bus	Vehicles	68
bird	Animals	80	balloon	Toys	68
daddy	People	80	cake	Food and Drink	68
child's name	People	80	choo choo	Sound Effects	66
baa baa	Sound Effects	78	ouch	Sound Effects	66
meow	Sound Effects	78	cow	Animals	66
uh oh	Sound Effects	76	tiger	Animals	66
train	Vehicles	76	star	Outside Things	66
book	Toys	76	tree	Outside Things	66
elephant	Animals	74	water	Outside Things	66
monkey	Animals	74	peekaboo	Games/Routines	66
fish	Food and Drink	74	please	Games/Routines	66
water	Food and Drink	74	go*	Action Words	66
diaper	Clothing	74			

Word frequency analyses

Tables 3 and 4 list the fifty words most frequently reported across all participants in English and Mandarin, respectively; due to ties at the 50th rank, fifty-five English and fifty-six Mandarin words are shown. The fifty-five top-ranked English words included thirty-three nouns and one 'Action' word; the fifty-six top-ranked Mandarin words included twenty nouns and twelve 'Action' words. As shown in Tables 3 and 4, only twenty words were reported for both languages; the majority of the most frequent words were reported for one language but not the other.

Perceptual-cognitive characteristics of the top fifty words

To better understand the perceptual-cognitive characteristics of the top fifty reported English and Mandarin words in this study, we attempted to

examine them using the four features of the SICI continuum defined by Maguire *et al.* (2006): distinctive shape (S), easy individuation (I), concreteness (C), and imageability (I). Operational definitions are not available for all of these features (Ma, Golinkoff, Hirsh-Pasek, McDonough & Tardif, 2009), and some of them appear to be more relevant to concrete count nouns than to other kinds of words. However, even a cursory examination of Tables 3 and 4 suggests that although the referents for many of the fifty most frequently reported words in English and Mandarin would fall at the high end of the SICI continuum by virtue of exhibiting all four SICI properties (e.g. 鞋 ‘shoe’), roughly one-third of the frequently reported words in each language exhibit fewer, or none, of the SICI features (e.g. 谢谢 ‘thank you’). For example, the Mandarin and English words for ‘water’ and ‘milk’ refer to substances that are concrete and imageable, but such substances do not have a distinctive shape nor are they easily individuated, at least as these features are typically defined. Similarly, the lists contain a number of words referring to events such as sounds, games, routines, and actions that are not readily characterized by any of the four SICI features, at least as these have generally been conceptualized. In short, although there is some support for the hypothesis of general cognitive-perceptual predispositions (Gentner, 1982) or specific cognitive-perceptual features (Maguire *et al.*, 2006), the most frequent words in the early vocabularies of the bilingual children studied here also included many words that were not salient on these grounds. Thus, it appears that factors other than cognitive and perceptual constraints also must be considered in understanding children’s early word learning.

Verb semantics. Twelve verbs appeared on the most frequent list of Mandarin words, but only one verb was found among the most frequent English words (Tables 3 and 4). It is interesting to note that eleven of the twelve early Mandarin verbs (e.g. 吃 ‘eat’, 坐 ‘sit’, 画画 ‘draw’, and 打 ‘hit’) would be characterized as being relatively semantically heavy, in the sense that they encode fairly specific meanings (Clark, 1993; Tardif, 2006) and/or arguments (Brown, 1998). Only one Mandarin verb, 要 ‘want’, would be characterized as semantically light, as is the only verb on the most frequently reported list of English words, *go*. The role of heavy and light verbs in language development and disorders is controversial (e.g. Gentner & Boroditsky, 2009; Maouene, Laakso & Smith, 2011). Some investigators (e.g. Clark, 1993) have observed that light or multi-purpose verbs dominate the early verb lexicons of monolingual children learning English, but heavy verbs are reported to be common in the early lexicons of other languages, such as Tzeltal (Brown, 1998).

Because there were so few verbs among the most frequent English words, we also examined the English and Mandarin verbs that were reported for at least 50% of the children (Table 5). For Mandarin, this resulted in an

TABLE 4. *Most frequent 56 Mandarin words, with English glosses and their PCDI categories, ranked by percentage of parents (N = 50) reporting them; the 20 nouns are **bolded** and the 12 verbs are marked with an asterisk (*)*

Word [gloss]	PCDI category	%	Word [gloss]	PCDI category	%
妈妈 [Mommy]	People	100	开(门) [open]*	Action Words	84
爸爸 [Daddy]	People	100	眼睛 [eye]	Body Parts	84
抱 [hold]*	Action Words	98	嘴(口) [mouth]	Body Parts	84
水 [water]	Food/Drink	98	一 [one]	Quantifiers	84
再见 [bye bye]	Games/Routines	96	奶奶 [grandma]	People	82
花 [flower]	Outside Things	96	姐姐 [elder sister]	People	82
吃饭 [have a meal]	Games/Routines	94	哥哥 [elder brother]	People	82
狗 [dog]	Animals	94	吃 [eat]*	Action Words	82
车 [vehicle]	Vehicles	94	要 [want]*	Action Words	82
自己名字 [own name]	People	92	耳朵 [ear]	Body Parts	82
猫 [cat]	Animals	92	鱼 [fish]	Animals	82
汪汪 [woof woof]	Sound Effects	90	水 [water]	Outside Things	82
要 [want]	Games/Routines	90	喂 [Hello]	Sound Effects	80
洗 [wash]*	Action Words	90	坐 [sit]*	Action Words	80
睡觉 [sleep]*	Action Words	90	打 [hit]*	Action Words	80
手 [hand]	Body Parts	90	看 [watch, look]*	Action Words	80
喵(猫叫) [meow]	Sound Effects	88	二 [two]	Quantifiers	80
谢谢 [thank you]	Games/Routines	88	阿姨 [aunt]	People	78
牛奶 [milk]	Food and Drink	88	喝 [drink]*	Action Words	78
饭 [meal]	Food and Drink	88	嘎嘎 [quack quack]	Sound Effects	76
蛋 [egg]	Food and Drink	88	妹妹 [younger sister]	People	76
鼻子 [nose]	Body Parts	88	拉屎 [go potty]	Games/Routines	76
球 [ball]	Toys	88	画画 [draw]*	Action Words	76
鞋 [shoe]	Clothing	88	鱼 [fish]	Food and Drink	76
不要 [don't want]	Games/Routines	86	头发 [hair]	Body Parts	76
洗澡 [wash]	Games/Routines	86	鸭子 [duck]	Animals	76
电话 [telephone]	Small Items	86	烫 [hot]	Descriptive Words	76
走 [walk]*	Action Words	84	书 [book]	Toys	76

additional twenty-one verbs, for a total of thirty-three verbs; for English this resulted in four additional verbs (*kiss, jump, hug, sit*), all of which would be characterized as relatively heavy. In short, verbs were generally rare in the English lexicons of these bilingual children, but the majority of those that occurred could be described as semantically heavy, consistent with the kinds of verbs that dominated their Mandarin lexicons. Accordingly, the theoretical distinction between light and heavy verbs does not appear to be related in any obvious way to the differences in the verb lexicons of these bilingual children.

TABLE 5. *English verbs (n=5) and Mandarin verbs (n=33) reported for at least 50% of the bilingual children*

English verbs				
<i>go</i>	<i>kiss</i>	<i>jump</i>	<i>hug</i>	<i>sit</i>
Mandarin verbs [gloss]				
抱 [hold]	洗 [wash]	睡觉 [sleep]	走 [walk]	开(门) [open]
吃 [eat]	要 [want]	坐 [sit]	打 [hit]	看 [watch, look]
喝 [drink]	画画 [draw]	打开 [open]	来 [come]	玩 [play]
拿(过来) [bring]	穿 [wear]	飞 [fly]	亲(一个) [kiss]	起来 [stand up]
哭 [cry]	关 [close]	跑 [run]	开(车) [drive]	尿 [go potty]
背 [carry]	上 [go up]	下 [go down]	摸 [touch]	给 [give]
有 [have]	跳 [jump]	游泳 [swim]		

Lexical reduplication. The lists of the most frequently reported words in English and in Mandarin (Tables 3 and 4) both include forms consisting of two identical syllables, such as *choo choo* in English and 奶奶 ‘grandma’ in Mandarin; in some but not all cases these forms comprise a root word and an exact repetition of it. This phenomenon, known as reduplication (Li & Thompson, 1981: 28–36), has been observed in adult Mandarin Chinese and also in prelinguistic canonical babbling (e.g. /nananana/) but it has rarely been analyzed in lexical acquisition. Tables 3 and 4 suggest that lexical reduplication may be more common in Mandarin than in English, and also that reduplicates in Mandarin are distributed over a broader range of semantic categories, including ‘Action Words’, than English reduplicates. Specifically, of the most frequently reported fifty-five English words, eight words (14%) can occur as reduplicated forms (although not all of these appear as reduplicative forms on the CDI): *no no*, *bye bye*, *baa baa*, *moo moo*, *yum yum*, *choo choo*, *mama* ‘mommy’, and *dada* ‘daddy’. All of these words are in the ‘Games & Routines’, ‘Sound Effects’, or ‘People’ category.

In the fifty-six most frequently reported Mandarin words, by contrast, sixteen of twenty words in the noun categories (80%) and all twelve verbs are acceptable when reduplicated by adults or by children. Only four of the other twenty-four words in Table 4 lack the potential to occur as lexical

reduplicates (一 ‘one’, 二 ‘two’, 再见 ‘bye bye’, 不要 ‘don’t want’); one word (自己的名字 ‘child’s own name’) may or may not be reduplicated depending on the child’s name.

DISCUSSION

This study was designed to provide strong observational evidence on the question of whether the noun bias in early lexical development appears to be language-general or language-specific. We compared the percentages of English and Mandarin nouns in a sample of fifty bilingual toddlers whose parent-reported vocabulary in each language contained 50–300 words; the sample size was chosen to ensure adequate statistical power for the comparison. The mean difference between the percentage of English and Mandarin nouns was 16%, exceeding the minimum effect size that had been specified a priori. Additional analyses showed that nouns predominated the total number of words as well as the fifty most frequently produced words in both languages, but the list of the most frequent fifty words in English included substantially more nouns and substantially fewer verbs than did the most frequent words in Mandarin, consistent with the main finding of a lower level of noun bias in Mandarin.

Confounding is not a likely explanation for the noun bias differences between English and Mandarin that were observed in this study. First, the within-subject bilingual design controlled for child-specific cognitive differences and at least some differences in children’s environments that could have influenced their vocabulary acquisition. Second, because analyses concerned children whose vocabulary size in each language fell within a relatively narrow range, cross-linguistic differences in noun bias are unlikely to reflect developmental variation. Third, the sample size was sufficiently large to ensure that statistical power was adequate to detect the language difference of interest. Fourth, consistent measures, definitions, and formulas were used in calculating the percentage of nouns in each language. Finally, the same informant completed the vocabulary checklist form in each language, presumably applying the same criteria and standards for identifying words in the child’s vocabulary. Thus, the present findings of a significantly lower level of noun bias in Mandarin than in English appear robust.

Analyses of the most frequently reported Mandarin verbs showed that nearly all of them encoded specific meanings, and that all of them could be reduplicated in adult-to-child speech. Li and Thompson (1981) studied adult Mandarin Chinese and reported that reduplication is a common morphological process that can apply in several form classes. In a sense, lexical reduplication can be viewed as effectively doubling a child’s exposure to each of the word’s repeated components. If such reduplication is found to

occur more frequently in Mandarin verbs than in English verbs, it is reasonable to ask whether this phenomenon might contribute to the relatively higher percentage of Mandarin verbs than English verbs in early vocabularies. However, it appears that lexical reduplication may also apply more often to Mandarin nouns than to English nouns. Accordingly, additional research is needed concerning cross-linguistic variations in lexical reduplication experienced by young children and the potential role of lexical reduplication processes in cross-linguistic differences in early vocabulary composition.

The design used in the current study, in which within-subject comparisons were made in bilingual children whose lexicons in each language were of similar size, suggests a fruitful approach to understanding early word learning in bilingual children. For example, words from different languages that have very similar meanings are known as translational equivalents (Pearson, Fernández & Oller, 1995). We are currently exploring noun bias in several vocabulary subsets that are unique to bilingual children, including words having various translational relationships across languages. Comparisons among such word sets in bilingual children with comparable vocabulary sizes may provide additional insights into the factors that influence lexical acquisition within and between languages.

Of course, the observational nature of this study does not permit causal inferences about the reasons for the language-specific differences in noun bias that we observed; experimental studies (e.g. Chan, Tardif, Chen, Pulverman, Zhu & Meng, 2011) are very much needed. In addition, to investigate developmental trends in the noun bias phenomenon across languages, studies of bilingual children with vocabulary sizes outside the 50–300-word window are needed. Finally, more detailed information on the language experiences of bilingual children is essential. Although all children in the present study were reportedly exposed to both English and Mandarin shortly after birth, and their vocabulary sizes in the two languages were similar by the time they participated in the study, the particular bilingual circumstances that bear on lexical acquisition were complicated and dynamic. A more detailed description of the variations in exposure to each language over time would improve efforts to understand the relationship between the input and the early lexicons of bilingual children.

In conclusion, findings from the present study extend understanding of the noun bias phenomenon (Gentner, 1982; Gentner & Boroditsky, 2001), revealing (i) a noun bias for both Mandarin and English, and (ii) a significantly higher percentage of English than Mandarin nouns in individual bilingual children with comparable expressive lexicons in each language. These results converge with previous findings from monolingual children (e.g. Bates *et al.*, 1994; Tardif *et al.*, 1999) and suggest that variables specific to the language input experienced by a child should be considered in understanding the preference for nouns during early word learning.

REFERENCES

- Au, T. K. F., Dapretto, M. & Song, Y. K. (1994). Input vs. constraints: Early word acquisition in Korean and English. *Journal of Memory and Language* **33**, 567–82.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J. & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* **21**, 85–123.
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., Pêcheux, M. G., Ruel, J., Venuti, P. & Vyt, A. (2004). Crosslinguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development* **75**, 1115–39.
- Brown, P. (1998). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics* **36**, 713–53.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L. & Weir, J. (1995). A crosslinguistic study of early lexical development. *Cognitive Development* **10**, 159–99.
- Caselli, M. C., Casadio, P. & Bates, E. (2001). Lexical development in English and Italian. In M. Tomasello & E. Bates (eds), *Language development: The essential readings*, 76–110. Oxford: Blackwell.
- Chan, C. C. Y., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L. & Meng, X. (2011). English- and Chinese-learning infants map novel labels to objects and actions differently. *Developmental Psychology* **47**, 1459–71.
- Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: A crosslinguistic study. *Journal of Child Language* **22**, 497–529.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conboy, B. T. & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development* **77**, 712–35.
- Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research* **34**, 565–71.
- Dale, P. S., Bates, E., Reznick, J. S. & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language* **16**, 239–49.
- Dale, P. S., Dionne, G., Eley, T. C. & Plomin, R. (2000). Lexical and grammatical development: A behavioral genetic perspective. *Journal of Child Language* **27**, 619–42.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S. & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: Words and sentences*. Baltimore, MD: Brookes Publishing Co.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (ed.), *Language development: Vol. 2. Language, thought and culture*, 301–334. Hillsdale, NJ: Erlbaum.
- Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (eds), *Language acquisition and conceptual development*, 215–56. Cambridge: Cambridge University Press.
- Gentner, D. & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. C. M. Gathercole (ed.), *Routes to language*, 5–36. New York, NY: Psychology Press.
- Goldfield, B. A. (1993). Noun bias in maternal speech to one-year-olds. *Journal of Child Language* **20**, 85–99.
- Goldfield, B. A. (2000). Nouns before verbs in comprehension vs. production: The view from pragmatics. *Journal of Child Language* **27**, 501–520.
- Gopnik, A., Choi, S. & Baumberger, T. (1996). Crosslinguistic differences in early semantic and cognitive development. *Cognitive Development* **11**, 197–227.

- Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E. & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language* **20**, 523–49.
- Li, C. N. & Thompson, S. A. (1981). *Mandarin Chinese – A functional reference grammar*. Berkeley, CA: University of California Press.
- Lucas, R. I. G. & Bernardo, A. B. I. (2008). Exploring noun bias in Filipino–English bilingual children. *Journal of Genetic Psychology* **169**, 149–63.
- Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C. & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language* **36**, 405–423.
- Maguire, M. J., Hirsh-Pasek, K. & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (eds), *Action meets word: How children learn verbs*, 364–91. New York, NY: Oxford University Press.
- Maouene, J., Laakso, A. & Smith, L. B. (2011). Object associations of early-learned light and heavy English verbs. *First Language* **31**, 109–132.
- Marchman, V. A. & Martinez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research* **45**, 983–97.
- Marchman, V. A., Martinez-Sussmann, C. & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science* **7**, 212–24.
- Marchman, V. A., Xuan, L. & Yoshida, M. (2005a). Vocabulary composition mirrors monolingual patterns in children learning two languages: A study of English- and Mandarin-learning toddlers in the US. Poster presentation at the Society for Research in Child Development Biennial Meeting, Atlanta, Georgia, April 2005.
- Marchman, V. A., Xuan, L. & Yoshida, M. (2005b). The validity of parent report measures of vocabulary in Mandarin- and English-speaking toddlers living in the US. Poster presentation at the Society for Research in Child Development Biennial Meeting, Atlanta, Georgia, April 2005.
- Nelson, K., Hampson, J. & Shaw, L. K. (1993). Nouns in early lexicons: Evidence, explanations and implications. *Journal of Child Language* **20**, 61–84.
- Pae, S. (1993). Early vocabulary in Korean: Are nouns easier to learn than verbs? Doctoral dissertation, University of Kansas.
- Pearson, B. Z. & Fernández, S. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning* **44**, 617–53.
- Pearson, B. Z., Fernández, S. & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language* **22**, 345–68.
- Pine, J. M. (1992). How referential are ‘referential’ children? Relationships between maternal-report and observational measures of vocabulary composition and usage. *Journal of Child Language* **19**, 75–86.
- Pine, J. M., Lieven, E. V. M. & Rowland, C. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language* **23**, 573–90.
- Slobin, D. I. (ed.) (1985). *The crosslinguistic study of language acquisition*, Vols. 1–2. Hillsdale, NJ: Lawrence Erlbaum.
- Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Hirsh-Pasek & R. M. Golinkoff (eds), *Action meets word: How children learn verbs*, 477–98. New York, NY: Oxford University Press.
- Tardif, T., Fletcher, P., Zhang, Z. & Liang, W. (2002). Nouns and verbs in children’s early vocabularies: A crosslinguistic study of the MacArthur Communicative Development Inventory in English, Mandarin, and Cantonese. Poster presented at the Joint Conference of the IX International Congress for the Study of Child Language and the Symposium on Research in Child Language Disorders, Madison, Wisconsin, July 2002.

- Tardif, T., Fletcher, P., Zhang, Z., Liang, W. L. & Zuo, Q. H. (2008). *The Chinese Communicative Development Inventory (Putonghua and Cantonese versions): Manual, forms, and norms*. Beijing: Peking University Medical Press.
- Tardif, T., Gelman, S. A. & Xu, F. (1999). Putting the 'noun bias' in context: A comparison of English and Mandarin. *Child Development* **70**, 620–35.
- Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language* **24**, 535–65.