Research Article

Running head: Semantic convergence in monolingual and bilingual children

Title: Children's Likelihood to Perform Adult-Like in Word Association Test: Effects of Bilingualism and Distributional Properties of Word Relationships

Authors: Boji P. W. Lam[1], Li Sheng[2], and Xian Zhang[3]

Affiliations:
Department of Audiology and Speech-Language Pathology, The University of North Texas[1]

Research Centre for Language, Cognition, and Neuroscience & Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University[2]

Department of Linguistics, The University of North Texas[3]

Acknowledgement:

This project was supported by a fellowship from the National Academy of Education and the Spencer Foundation to the second author. The authors thank the participants for volunteering their time and Diana Cruz, MaryAnn Flores, Aislynn Fulton, Andrea Hinnen, Fang Li, Ashley Pina, Brittany Rasansky, Ruei-An Shen, and other members of the Language Learning and Bilingualism Laboratory for assistance with stimulus design, data collection, and data analyses.

Address for correspondence

Boji P. W. Lam
Department of Audiology and Speech-Language Pathology,
The University of North Texas
907 W Sycamore St, Denton, TX 76201
Phone: 940-565-2653.
Email: pakwingjacky.lam@unt.edu

Keywords: Semantic convergence, word association, Spanish-English bilingual children, Chinese-English bilingual children, the "Small World of Words"

Abstract

Little is known about the effects of bilingualism and distributional properties of word relationships on children's development of semantic convergence, operationalized as children's ability to produce word associates that mirror adults' responses in a word association task. Forty-five Mandarin-English bilingual, 32 Spanish-English bilingual, and 28 English-speaking monolingual children, aged 4 to 7, produced three associates to each of 15 single-word cues in English. Children's productions were compared against adult responses to the same cues in the "Small World of Words" Norm. Three scoring methods comparing similarities of children's responses to adults' showed consistent bilingual disadvantages in producing adult-like responses. Follow-up analyses targeted the three most predominant responses adults produced for each cue and addressed factors predicting children's likelihood to produce these responses. Results showed additional effects of cue-associate relationships measured by co-occurrence and semantic relatedness. The findings highlight the multi-faceted nature of knowledge development of word relationship and semantic convergence.

Introduction

Free word association task (WAT) is commonly used to assess word knowledge and lexical-semantic organization among children (e.g., Entwisle, 1966; Nelson, 1977; for a review, see Wojcik, 2018) and adults (de Deyne et al., 2019; Nelson et al., 2004). In a typical WAT, participants hear or read a lexical cue (e.g., CAT) and report the first word that comes to mind (e.g., DOG). Words are theorized to be stored as nodes in the mental lexicon that are interconnected as a network, within which activation of a node spreads to connected nodes automatically during lexical tasks or daily encounters (Anderson, 1983; Collins & Loftus, 1975; Steyvers & Tenenbaum 2005). During a WAT, the exposure to a cue activates connected neighbors that compete for production, and neighbors that exhibit stronger connectivity and faster activation would be more likely to be reported as the first word that comes to mind (Playfoot et al., 2018). Participants take longer to produce associates that do not exhibit the strongest link to the cue in a WAT (Playfoot et al., 2018), which suggests elevated challenges to suppress and *not* report the first word that comes to mind upon hearing the cue.

WAT performance thus indicates abilities in learning the lexical form of a word, its conceptual referent, and, importantly, the relationships among words. Significant discrepancies between children's and adults' responses to WAT are expected (e.g., Entwisle, 1966; Nelson, 1977) due to differences in conceptual organization and accumulated language experience. An examination of WAT performance would shed light on semantic convergence, a socially mediated process in which children learn to use words and their connotations in a way that mimics the use of adults from the same language community (Adams & Bullock, 1986). Previous work has demonstrated the significant influence of age on WAT performance (e.g., Cremer et al., 2011; Spätgens & Schoonen, 2020; Nelson, 1977). Entwisle (1966) showed that

older monolingual children (e.g., fifth grade) produced more responses commonly shared among group members than younger monolingual children (e.g., preschoolers), suggesting greater within-group homogeneity that characterizes the community. The effect of age on semantic convergence may reflect developmental changes in organizational principles of the mental lexicon (McNeill, 1963; Nelson, 1977; Nelson & Nelson, 1990; Palermo, 1971; Smiley & Brown, 1979; Waxman & Namy, 1997). It has been shown that children between 6 and 9 produce more paradigmatic associations that belong to the same grammatical or taxonomic category (e.g., DOG—CAT) than younger children, whose responses are predominated by syntagmatic associations that are thematically related to or syntactically adjacent to the cues (e.g., DOG—BARK). Such developmental changes in WAT performance, also known as the syntagmatic-paradigmatic shift (Nelson, 1977; Perraudin & Mounoud, 2009), indicate general changes in the guiding principles of conceptual organization (Keith & Nicoladis, 2012).

The age effect on semantic convergence may also indicate knowledge accumulation of word usage patterns based on exposure to language input. Compared to younger children, older children have lengthened exposure to statistical regularities embedded in the ambient language. Increased experience with words may lead to discoveries that word pairs differ in the rates of how often they appear *in proximity* (i.e., co-occurrence) in the same utterance. For example, “sing songs” has a much higher co-occurrence than “sing poems", which may prompt children to intentionally explore the relationship between “SING” and “SONG”. Greater rates of co-occurrence also lead to increased opportunities for processing and producing selected words together, leading to further consolidation of the strength between words. Lengthened language exposures also may lead to discoveries that some words tend to occur in *similar contexts* despite being in different utterances (e.g., Broccoli is GOOD; chips are BAD; My dad is a DOCTOR

and my mom is a LAWYER). The statistical regularities with which words occur in similar contexts and related themes may facilitate the development of taxonomic awareness (e.g., word class, semantic categories, etc.) and semantic relatedness between words.

In a similar vein, bilingualism should exert a significant impact on semantic convergence—the need to learn two languages would effectively limit the amount of language-specific input and thus slow down the rates of knowledge accumulation of word usage patterns. Bilingualism impacts lexical-semantic development (for a review, see Bialystok, 2009) and language-specific performance on vocabulary assessment (for a review, see Bedore & Peña, 2008). Bilingual populations generally control a smaller language-specific vocabulary than monolinguals (e.g., Portocarrero, et al., 2007), though the two language groups may perform comparably on dual-language measures that consider words known in either language (e.g., Anaya et al., 2017; Gross et al., 2014). Divided language exposure reduces the amount of language-specific usage, which leads to weakened linkages between lexical forms and semantic representations that negatively impact word production in the testing language (Gollan et al., 2008; Sandoval et al., 2010).

The impact of bilingualism on semantic convergence in children's WAT performance has received much less attention when compared to other semantic domains, such as picture identification and naming abilities that indicate vocabulary breadth. Notably, existing work on WAT in bilingual children has focused on analyzing qualitative aspects of responses, such as meaning-related versus form-based (Cremer et al., 2011), paradigmatic versus syntagmatic (Holmström et al., 2016; Namei, 2004; Sheng et al., 2006), or context-independent versus context-dependent responses (Spätgens & Schoonen, 2020). Results from these studies are far from conclusive: previous work has found subtle advantages in producing the more advanced

paradigmatic associates in bilingual children (Holmström et al., 2016; Sheng et al., 2006), subtle disadvantages in bilingual children (Namei, 2004) though the effect of language backgrounds was more limited compared to the effect of age (e.g., Cremer et al., 2011), or comparable associative preferences between monolingual and bilingual children (Spätgens & Schoonen, 2020). Highly diverse coding and classification systems (for a review, see Fitzpatrick & Thwaites, 2020) and differences in participant profiles across studies may have led to inconsistent results.

Descriptions of semantic convergence in WAT would be incomplete without considering the effects of word usage patterns in the ambient language on children's performance. Word usage patterns contribute to distributional properties of language input that may be quantified by co-occurrence and degrees of semantic relatedness of words. These parameters may influence the strength of connectivity between words in the mental lexicon—words that are co-activated frequently or share significant semantic features tend to exhibit strengthened connectivity and speeded activation (for a review, see Hutchison, 2003). Co-occurrence, as introduced earlier, refers to the temporal contiguity between words in spoken language or texts. This distributional property of language use can be quantified by mutual information (M. I.) that compares the probability of co-occurrence of two words to the probability of their independent existence (Church & Hanks, 1990). Thus, if two words have a substantially larger probability to co-occur than by chance (suggesting a large M.I.), the two words are likely to have some sorts of syntactic association (e.g., SING—SONG; DOG—BARK). Wettler et al. (2005) derived co-occurrence information from the British National Corpus and demonstrated its correspondence to actual associates produced to selected cues by adult participants. Results have been argued to support paired associative learning of word pairs experienced in proximity (Wettler et al., 2005).

However, human associative behaviors cannot be fully explained by co-occurrence of words (Gruenenfelder et al., 2016). Semantic relatedness, which measures the similarity of contexts surrounding two words, may also explain the associative behaviors. Semantically related words (e.g., CAT and DOG) are connected to numerous shared features instantiated as nodes in the mental lexicon, such as “animals”, “pets”, “fur”, and “paws”. Activation of the node CAT during a WAT would spread out and activate connected nodes, which converge and speed up the activation of the node of DOG. As a result, both co-occurrence and semantic relatedness should predict children’s likelihood to produce associates that mirror adults’ responses.

This study examined the effects of bilingualism on semantic convergence in children, specifically the extent to which monolingual and bilingual children produced the *same* responses reported in the “Small World of Words” association norm (De Deyne et al., 2019) collected from adult speakers. The study has two aims. The first aim focused on children’s responses and addressed the extent to which word associates produced by monolingual and bilingual children are similar to adult responses using measures of stereotypy and association strength within cue-associate pairs. The study used three scoring methods that range from the crudest to the most fine-grained to assess the similarity between child and adult responses. Stereotypy analyses are common in WAT studies (for a review, see Fitzpatrick & Thwaites, 2020) and are typically classified as unweighted and weighted. Unweighted scoring evaluates whether participants’ responses appear in the norm list or not, providing a crude similarity measure. By contrast, weighted scoring additionally attends to the ranking of the selected response in the norm list, with more points allocated to responses that rank higher in the list (i.e., produced by more respondents in the norm). Finally, association strength within a cue-associate pair provides more detailed information about the number of adult participants giving a specific response when

prompted with a cue. The information reflects the conditional probability with which a response is triggered by a cue (De Deyne et al., 2019). As a result, the scoring method based on association strength provides a much more fine-grained measure in assessing children's semantic convergence.

A smaller language-specific vocabulary (Bialystok, 2009; Gross et al., 2014; Sandoval et al., 2010) and divided language input associated with bilingualism should predict a monolingual-bilingual gap in converging to adult-like associative behavior in the target language, manifested as lowered unweighted and weighted stereotypy scores, and weaker associative strength for responses produced by bilingual children. The analyses would shed light on the extent to which the monolingual-bilingual gap in children, if found, applies across scoring systems that focus on matching at both a crude level (i.e., unweighted) and a fine-grained level by considering also prototypicality of adults' responses (i.e., weighted scores and association strength). An interaction between age and bilingualism would indicate a widening or narrowing monolingual-bilingual gap in children's likelihood to produce adult-like associative behavior. Alternately, a stable monolingual-bilingual gap would suggest consistent effects of divided language input on the acquisition of word relationships in young children. This study focused on children of 4- to 7-year-olds because lexical-semantic organization develops rapidly at these ages (e.g., Nelson & Nelson, 1990; Sheng et al., 2013; Sheng & Lam, 2015), which would allow us to examine the stability as opposed to changes in the monolingual-bilingual gap, if it exists.

The second aim followed up on the first question and addressed an under-examined issue in WAT studies—why are certain adult-like associates more likely to be produced by children? Compared to the first question and previous studies that focused on the responses produced by children, the second aim took a closer look at the associates *commonly* produced by adults.

Based on the same association norm (De Deyne et al., 2019) used to address the previous aim, the analyses located the top three most predominant responses adults produced for selected cues and defined these responses as the targets. The second aim focused on the top three predominant responses adults produced because they have been used to define the primary responses when assessing stereotypy of WAT performance (see Fitzpatrick et al., 2015 and Fitzpatrick & Thwaites, 2020 for review). Little is known about factors that predict children's likelihood to produce these responses that define stereotypy and suggest semantic convergence to the adult norm. An examination of word association norms suggests that non-predominant responses usually show declines in the number of respondents producing the word, suggesting increasing idiosyncrasy that may not represent general patterns of lexical experience shared by a language community. For the second aim, the correspondence of co-occurrence in the ambient language to actual associates produced in WAT (Wettler et al., 2005) and the significant role of semantic relatedness in the architecture of the mental lexicon (Collins & Loftus, 1975) predict that adult-like responses that exhibit a stronger relationship to the cue will be more likely elicited from children. Regarding the effects of bilingualism, divided language input may lead children to pay additional attention to word pairs that have a high co-occurrence rate or exhibit heightened semantic relatedness. Word pairs that co-occur infrequently may have a much reduced occurring rate in divided language input, leading to increased saliency for word pairs that co-occur more frequently. Furthermore, the challenges associated with learning two languages may drive children to focus on word pairs that occur in similar contexts and share a great number of semantic features than word pairs that relate remotely in order to catch up or assimilate to the norms of the dominant societal language. If so, distributional properties of word pairs may exhibit a stronger relationship with the likelihood to produce adult-like associates in bilingual

children than in monolingual children. Alternately, both monolingual and bilingual children may show comparable performance in producing cue-associates pairs that exhibit a stronger relationship in the natural language environment because these word pairs are easier to learn for all learners. Addressing the second aim should also consider variables known to influence lexical access (Goh et al., 2016) besides distributional properties of word relationships. We obtained measures of age of acquisition and word frequency for each adult target associate because these variables may influence semantic convergence in monolingual populations. Specifically, members from the same language community tend to produce associates that are acquired earlier in life (e.g., Brysbaert et al., 2000) or occur frequently in daily usage (e.g., Fitzpatrick, 2007; Meara, 1983). These findings highlight the needs to consider age of acquisition and word frequency when examining children's likelihood to produce adult-like associates in WAT.

To examine the generalizability of the bilingual effect on WAT performance, we compared bilingual children who spoke language pairs that differ in language distance to the testing language (English). Mandarin is phonologically and morphologically distinct from Spanish and English (Sheng, et al., 2016). Differences in cross-linguistic features influence language-specific vocabulary development. Young Spanish-English bilingual children exhibited sensitivities to cross-language cognates in vocabulary assessment (Sheng et al., 2016). Compared to Mandarin-English bilingual children, Spanish-English bilingual children exhibited improved awareness of certain English derivational word formation rules that took extended time to develop as well (Lam & Sheng, 2016). However, studies also showed that Mandarin-English and Spanish-English bilingual children performed comparably on measures of vocabulary breadth (Sheng et al., 2016) and taxonomic awareness at the conceptual level (Lam & Sheng, 2020). This

study predicted that the monolingual-bilingual gap in WAT performance, if found, would generalize across bilingual groups because the task does not target specific linguistic features.

This study adopted the adult association norm compiled by De Deyne and colleagues (2019) as the reference model in comparing children's WAT responses. The norm is the latest and largest English association norm developed to-date, and expands on established association norms, such as the University of South Florida association norm (Nelson et al., 2004). The De Deyne et al. (2019) norm included a relatively homogeneous adult sample as most participants had at least a college or master's degree. Furthermore, the norming procedure elicited three consecutive responses from each participant for each cue, with limited chaining effects from initial responses on subsequent associates produced (De Deyne et al., 2019). The current study adopted a similar procedure in eliciting children's responses because eliciting multiple responses provides more details on subjects' semantic networks (Schmitt, 1998; Spätgens & Schoonen, 2020).

Methods

Participants

Forty-nine Mandarin-English bilingual children (23 females), 33 Spanish-English bilingual children (24 females), and 28 English-speaking monolingual children (10 females) between the ages of 4 and 7 years were recruited in the Austin area for the study. The procedures were approved by the Institutional Review Board at the University of Texas. No participants were reported to have any visual, audiologic, or neurologic deficits. Caregivers were interviewed about the children's language profiles that included parent-rated language proficiency and daily language use. The interview was conducted using a questionnaire based on Gutiérrez-Clellen & Kreiter (2003) and Restrepo (1998). The questionnaire inquires about language usages on an

hour-by-hour basis during the waking hours of a typical weekday and a typical weekend day. To ensure enough bilingual exposure, participants in the bilingual groups must have more than 20% daily exposure to each of their two languages at the time of testing. The criteria resulted in the exclusion of four Mandarin–English bilingual children and one Spanish–English bilingual child. To obtain an objective measure of English proficiency, participants completed a 50-item English picture naming task designed by the second author.

Analysis of Variance (ANOVA) confirmed that the three language groups did not differ in age (p=.37) (see Table 1). ANOVA indicated a language-group effect for maternal education: the Spanish-English bilingual children had fewer years than Mandarin-English bilingual children ($F(2, 102) = 35.9$, $p<0.001$, $\eta^2=0.40$; M.D. = -5.9, SE=0.7, $p<0.001$) and English-speaking monolingual children (M.D. = -4.1, SE=0.8, $p<0.001$). No difference in years of maternal education was found between Mandarin-English bilingual children and English-speaking monolingual children ($p=0.053$).

[Insert Table 1 about here]

Materials and procedures

Picture Naming

A 50-item English picture naming task was used to measure vocabulary breadth of child participants in English. The task, which was described in detail in previous studies (e.g., Lam & Sheng, 2020a; Sheng et al., 2016), consisted of 50 concrete familiar nominal concepts derived from a number of semantic categories, such as people, animals, food, tools, and instruments. The 50 target items were represented as colored photographs selected from Google images. For each item, we used established corpus to obtain the frequency-of-occurrence values in oral contexts in English (Brysbaert & New, 2009). The average occurrences per million was 29.3 (SD = 38.3;

range = 1.72 to 103.4 occurrences per million). We also obtained perceived age of acquisition of the item based on adults' rating (c.f. Kuperman et al., 2012). The average age of acquisition was 4.6 years old (SD = 1.3; range = 2.6 to 9.7). Internal consistency of the picture naming task was examined using Cronbach's alpha, which showed acceptable reliability for the task, $\alpha = 0.86$.

Repeated Word Association Test

The stimuli in WAT consisted of 15 early-acquired English cues with 5 adjectives, 5 nouns, and 5 verbs (see Table 4 for full list of cues). The average age of acquisition of the cues was 4.0 (range = age 2.8 to 5.8; c.f., Kuperman, et al., 2012). We sampled early acquired cues from multiple word classes to avoid restricting results to specific word classes. Participants were tested by a native English speaker. The administration of WAT followed the procedure described in Sheng et al. (2006, 2013) with the following instructions given by a trained research assistant.

"When you hear a word, sometimes it makes you think of another word that goes with it. If you hear "MOM," you might think of "DAD." If you hear "BIRTHDAY," you might think of "CAKE." I am going to read a word out loud. Then I will ask you to tell me another word that goes with the word I said out loud. Are you ready?"

After the instructions, the trained research assistant prompted children participants to produce associates for "CAR" with the instruction "If I say CAR, you say____" in three elicitation trials, followed by another practice item "RUN." Feedback was provided during practice. For responses semantically related to the cues (e.g., TRUCK), the research assistant said, "Yes, CAR goes with TRUCK". For responses unrelated to the cues semantically (e.g., producing phonologically related words, such as CAR-BAR, or saying "I don't know"), the research assistant said, "A word that goes with CAR is TRUCK"). Children were reminded to say only one word at a time. Throughout practice and testing trials, the research assistant used

fingers as visual cues to help the participants to keep track of three elicitation trials for each cue. No more feedback was provided during the testing trials.

In summary, the instructions and practice *subtly* guided children toward generating semantically related responses. Such decisions are based on the observation that adults predominantly produce semantically related associates in WAT, even for idiosyncratic responses that are rarely produced by adult participants. Furthermore, the proportion of non-semantically related responses (e.g., producing phonologically related associates) decreases with age (e.g., Namei, 2004). These observations suggest inherent biases towards producing semantically related responses in WAT among adults. When provided with instructions and practice that subtly guide toward generating semantically related responses, failures to do so would shed light on the sophistication levels of the lexical-semantic system of children and their abilities to produce adult-like responses, that is, semantic convergence.

Scoring

For the picture naming task, items were scored as either 1 (correct) or 0 (incorrect). To address the first aim of the present study, children's WAT responses were compared to *all* responses produced by adults for the same cues using the "Small World of Words" Association Norm (De Deyne et al., 2019). Only data from adult speakers who were reportedly based in the United States and spoke English as the native language were utilized because these participants would be geographically similar to the child participants examined in the present study. Children's responses were scored (1) non-weighted, (2) weighted, and (3) using association strength within cue-associate pair. The scoring methods differ in the amount of information provided. Non-weighted scoring focuses simply on whether children produced adult-matching

responses; weighted scoring and association strength provide added information about the prototypicality of a selected response according to the association norm.

In non-weighted scoring, children's responses were scored as 1 (match) or 0 (no match). For example, a child produced "GOOD, CRAZY, HAPPY" in the first, second, and third trial respectively for the cue "BAD". According to the association norm, "GOOD" is among adults' responses in the first trial, while "CRAZY" and "HAPPY" were not produced by adults in the respective trials. The child would, therefore, receive a score of 1, 0, and 0 for the first, second, and third trial respectively, and an average score of 0.33 for the cue "BAD" when aggregating over three trials. Higher non-weighted scores indicate that a child produces an increased number of adult-matching responses and, therefore, performs more similarly as adults.

In weighted scoring, responses were scored as 2 (match and among the top three most predominant responses among adults), 1 (match but not among the top three), and 0 (no match). For the same example above, "GOOD" was, in fact, the most common adults' responses in the first trial. The child thus receives a score of 2, 0, and 0 for the first, second, and third trial respectively, with an average of 0.67 for the cue. In contrast, a child who produced "DOG, CRAZY, HAPPY" for "BAD" would receive a lowered score for the first trial (i.e., 1) and an average score of 0.33 for the cue because "DOG" is produced infrequently in adults.

Compared to weighted scoring that ranks adults' responses in a relatively crude manner, association strength within cue-associate pair provides an even finer evaluation of the similarity between children and adult responses. The association strength informs on the number of adult participants producing a specific response when given a cue in three elicitation trials. For example, "GOOD" is produced by more adult participants than "DOG" as an associate in three elicitation trials combined when they are given the cue "BAD" (c.f. De Deyne et al., 2019). As a

result, “GOOD” exhibits a much elevated conditional probability to be trigged by “BAD” than “DOG” does (0.155 vs 0.017), which indicates a stronger association for the “BAD-GOOD” cue-associate pair. Based on adults’ associative behavior, the average association strength for all responses produced in the three trials for one child could be compared to another child. Summarizing the three scoring methods, non-weighted and weighted stereotypy scoring is trial-specific; association strength speaks to children’s performance according to the general conditional probability with which a cue triggers a selected response in adult participants when they are given three opportunities. Compared to non-weighted scoring, the weighted and associative strength scoring reward children for producing common associates that likely represent the shared behaviors of a language community.

Distributional properties of word relationships, data coding and statistical analyses

To address the second aim, the top three most predominant associates adults produced for each of the 15 cues for the “Small World of Words” Association Norm (De Deyne et al., 2019) were compiled as a list of target responses (15 stimulus cues X 3 target associates = 45 target associates in total). Focusing on the top three most predominant responses allows us to effectively achieve our goal of examining factors that predict convergence to associative behaviors that characterizes a language community. To illustrate, for the current cue words, the top three most predominant responses accounted for almost half of the responses produced by adult respondents in the first trial (46.5%), and 30% of the total responses aggregating over the three trials. The top three predominant associates have also been used to define stereotypy in WAT studies (for a review, see Fitzpatrick & Thwaites, 2020). In contrast, associates produced by few adult participants (e.g., saying “SANTA” for the cue “GOOD”) not only indicate idiosyncrasy but could be erroneous responses less relevant to the goal of the study. Furthermore,

focusing on the top three most predominant responses limit the amount of missing data. Specifically, including responses with increasing idiosyncrasy would lead to elevated methodological challenges in locating variables relevant to the analyses, increasing the likelihood of missing data. For example, should we include "SANTA" in the analysis, which was produced by only one adult respondent for the cue "GOOD" according to De Deyne et al. (2019), the target would be missing age of acquisition rating because the word was not sampled in the age of acquisition norm (c.f., Kuperman,et al., 2012). Finally, focusing on the top three responses is more advantageous than restricting the analysis to the single, most predominant associates by adults—the latter approach may be too narrow in scope and fail to adequately reflect the breadth of semantic networks. Appendix S1 reported, for each target associate, the age of acquisition obtained from a corpus of 30, 000 English words (c.f., Kuperman,et al., 2012) and word frequency based on 51 million English words (c.f., Brysbaert & New, 2009).

Appendix S1 also reported distributional properties of word relationships for each cue-associate pair, which were quantified as mutual information (M. I.) and semantic relatedness. M. I. measures the probability of a word pair appearing in proximity (i.e., co-occurrence) whereas semantic relatedness measures the probability of a word pair appearing in similar contexts. M. I. was calculated by log2(P(AB)/P(A)*P(B)), whereas P(AB) is the probability that word A and word B co-occur together in a corpus and P(A) and P(B) are the probabilities of observing A and B independently. The current study used Corpus of Contemporary American English (COCA, a large corpus with 1 billion word tokens; Davies, 2008) to calculate M. I. indices of the target word pairs. Semantic relatedness was computed using the GloVe model (Pennington et al., 2014), a widely used semantic vector space computational model in natural language processing that measures semantic relatedness between words. GloVe generates word embeddings (i.e.,

vectors that store contextual information of words that allow computational models to evaluate whether two words appear in similar contexts) and uses a weighted log bilinear regression model to compute the probability that a word appears in the context of the target word (e.g., tea and coffee tend to have a higher semantic relatedness index because they share contexts such as "drink", "have", "like", "hate", "enjoy", "a cup of", "hot", "ice", etc., for additional information about the GloVe model, please refer to Pennington et al., 2014). Thus, for the two measures of distributional properties of word relationships, co-occurrence took into account the syntactic association of the word pairs and semantic relatedness considered the contextual association of word pairs.

The 45 target cue-associate pairs were compared against the responses elicited from children. For each child, a target associate was coded as 1 if it was produced in any one of the elicitation trials for the selected cue. Targets not elicited in any of the trials were coded as 0. The coding system was dichotomous (match = 1; mismatch = 0). Repeated responses were counted once only. The analyses applied minimal lemmatization rule that concerned primarily inflectional morphemes that mark the plural forms of nouns (i.e., -s; e.g., gifts → gift).

Analyses

A generalized linear mixed model (GLMER) was conducted in R (R Core Team, 2012) using the package lme4 (Bates et al., 2012) for picture naming performance. The critical dependent variable was children's likelihood to name a target item. The analysis started with language group (English-speaking monolingual children vs. Mandarin-English bilingual children vs. Spanish-English bilingual children), age, and their interaction as fixed effects. English-speaking monolingual children were set as the reference group. Maternal education was included as a covariate.

For WAT, the critical dependent variables for the first aim were (1) non-weighted stereotypy scores, (2) weighted stereotypy scores, and (3) association strength within cue-associate pairs. All analyses used the average scores aggregating over the three elicitation trials because such scores spoke to general levels of WAT performance when multiple opportunities were provided for a cue. The analyses, using linear mixed model (LMER), started with language group, age, and their interaction as fixed effects, similar to the analysis of picture naming performance. Maternal education was included as a covariate. To address the second aim, the critical dependent variable was children's likelihood to produce the 45 target adults' associates. GLMER analyses started with language group, age, and their interaction as fixed effects, similar to the analyses described above. Co-occurrence (i.e., M. I.) and semantic relatedness of target cue-associate pairs and their interactions with language group were included as fixed effects to examine the influence of distributional properties of word relationships on children's likelihood to produce adult-like responses. Word frequency and age of acquisition of target associates were included in the model to consider their influence on lexical access and convergence in WAT performance (Brysbaert et al., 2000; Fitzpatrick, 2007; Goh et al., 2016; Meara, 1983). Finally, maternal education was included as a covariate. An ad-hoc analysis was performed to examine whether English naming performance predicted children's likelihood to produce adult-like associates.

All continuous variables were centered for analyses. For all analyses, by-participant intercepts, by-item intercepts, and a by-item random slope for group were included in the model as random effects. Models were refined for parsimony by removing, one at a time, factors and interactions that exhibited the highest *p*-value, while retaining the hierarchical rule of interactions. Likelihood ratio comparisons were performed to confirm that including a given

factor did not improve the amount of variance explained (Baayen et al., 2008). Post-hoc analyses with Tukey's procedure for multiple comparison correction were performed with version 2.3-00 of the lsmeans package (Lenth, 2016).

Results

Picture naming performance

GLMER indicated a language group effect [χ^2 (2) = 14.37, p<.001], an age effect [χ^2 (1) = 52.75, p<.001], and an effect of maternal education [χ^2 (1) = 19.17, p<.001]. The language group X age interaction was not significant (p=0.68) and was removed from the model. The language group effect showed that English-speaking monolingual children were more likely to name pictures accurately in English than Mandarin-English bilingual children (β= 0.61, SE=0.18, p<.001) and Spanish-English bilingual children (β= 0.75, SE=0.23, p<.001). No differences were found between Mandarin-English bilingual children and Spanish-English bilingual children (p=0.53). The likelihood to name pictures accurately also increased with age (β= 0.05, SE=0.01, p<.001) and more years of maternal education (β= 0.11, SE=0.02, p<.001).

Monolingual-bilingual gap in producing adult-like associative behaviors based on non-weighted and weighted stereotypy, and association strength within cue-associate pairs

Non-weighted scores LMER indicated a language group effect [χ^2 (2) = 22.85, p<0.001], an age effect [χ^2 (1) = 49.69, p<0.001], and an effect of maternal education [χ^2 (1) = 5.98, p<0.001]. The language group X age interaction was non-significant (p=0.25) and was removed from the model. Table 2 showed that English-speaking monolingual children produced more responses that matched with the adults' association norm than Mandarin-English bilingual children and Spanish-English bilingual children, which was confirmed by the language group effect (English monolingual vs Mandarin-English: β= 0.10, SE=0.02, p<0.001; English

monolingual vs Spanish-English: β= 0.12, *SE*=0.03, $p<0.001$). No differences were found between Mandarin-English bilingual children and Spanish-English bilingual children ($p=0.51$). One unit increase in age predicts a 0.006 increase in scores ($p<0.001$). The non-weighted scores were predicted by maternal education as well—one unit increase in maternal education predicts a 0.007 increase in scores ($p<0.001$).

Weighted scores LMER indicated a language group effect [χ^2 (2) = 13.55, $p=0.001$], an age effect [χ^2 (1) = 34.56, $p<0.001$], and an effect of maternal education [χ^2 (1) = 4.51, $p=0.034$]. The language group X age interaction was non-significant ($p=0.36$) and was removed from the model. The language group effect showed that English-speaking monolingual children produced more responses that matched with the adults' association norm than Mandarin-English bilingual children (β= 0.15, *SE*=0.05, $p=0.005$) and Spanish-English bilingual children (β= 0.12, *SE*=0.04, $p=.009$) when prototypicality was considered as well. No differences were found between Mandarin-English bilingual children and Spanish-English bilingual children ($p=0.56$). One unit increase in age predicts a 0.011 increase in scores ($p<0.001$). The weighted scores were predicted by maternal education as well—one unit increase in maternal education predicts a 0.012 increase in scores ($p=0.038$).

Association strength within cue-associate pairs LMER indicated a language group effect [χ^2 (2) = 10.67, $p<0.001$], an age effect [χ^2 (1) = 46.19, $p<0.001$], and a marginal effect of maternal education [χ^2 (1) = 3.73, $p=0.054$]. The language group X age interaction was non-significant ($p=0.47$) and was removed from the model. The language group effect showed that English-speaking monolingual children produced stronger cue-associate pairs than Mandarin-English bilingual children (β= 0.006, *SE*=0.002, $p=0.001$) and Spanish-English bilingual children (β= 0.008, *SE*=0.003, $p=.001$). No differences were found between Mandarin-English bilingual

children and Spanish-English bilingual children (p=0.30). One unit increase in age predicts an 0.0004 increase in association strength within a cue-associate pair (p<0.001). The marginal effect of maternal education suggests more years of maternal education would predict the production of associate with stronger association strength as well (β= 0.0006, SE=0.0002, p=0.059).

[Insert Table 2 about here]

Monolingual-bilingual gap in producing the top three more predominant associates of the "Small World of Words" Association Norm

Table 3 showed that 15.1% of monolingual children's responses matched with the adult-like target associates, which were numerically greater than both Mandarin-English bilingual children (12.4%) and Spanish-English bilingual children (9.4%). Table 3 additionally reported percentage match of children and adult responses separated by word classes, which showed that monolingual children produced more target associates than bilingual children for all word classes.

[Insert Table 3 about here]

Model comparison confirmed a language group effect [χ^2 (2) = 19.05, p<0.001] and an age effect [χ^2 (1) = 58.67, p<0.001]. The language group X age interaction was not significant (p >0.90) and was removed from the model. The monolingual children were more likely to produce adult-like target associates than Mandarin-English bilingual children (β= 0.48, SE=0.14, p<0.001) and Spanish-English bilingual children (β= 0.71, SE=0.17, p<0.001) (see Table 4). The two bilingual groups did not differ from each other (p=.898). An ad-hoc analysis showed that English naming performance predicted children's likelihood to produce adult-like associates [χ^2 (1) = 5.36, p=0.02]. One unit increase in English vocabulary breadth indexed by picture naming performance was associated with a 0.02 higher likelihood to produce adult-like associates

(p=0.02), with the language group effect continued to be significant when English naming was in the model (p=0.002). Finally, one unit increase in age was associated with a 0.04 higher likelihood to produce adult-like associates (p<0.001). Maternal education improved the model as well [χ^2 (1) = 5.48 p=0.019]—one unit increase in maternal education was associated with a 0.05 higher likelihood to produce adult-like associates (p<0.001).

[Insert Table 4 about here]

The effects of co-occurrence and semantic relatedness on the likelihood to produce the top three more predominant associates of the "Small World of Words" Word Association Norm

Analyses further indicated the effects of distributional properties of word relationships [M. I.: χ^2 (1) = 32.23, p<.001; semantic relatedness: χ^2 (1) = 34.23, p<.001]. One unit increase in M. I. was associated with a 0.25 higher likelihood to produce adult-like associates (p<0.001). Similarly, one unit increase in semantic relatedness was associated with a 3.67 higher likelihood to produce adult-like associates (p<0.001). Neither factor interacted with language groups (range of p values = 0.25 to 0.39), and the interactions were removed from the model. The likelihood was also predicted by age of acquisition [χ^2 (1) = 148.91, p<.001] and word frequency of the target associates [χ^2 (1) = 17.22, p<.001]. Adult-like associates were more likely to be produced by children if the target words were acquired earlier (β= -0.75, SE=0.07, p<.001) and occur more frequently (β= 0.64, SE=0.15, p<.001).

Item analysis

An item analysis was performed to examine the extent to which the three groups differed in the most predominant associates they produced for each cue (see Table 5). The analysis showed that nine out of 15 cues elicited the same most predominant associate from the three groups. These cues are evenly distributed across adjectives (bad, big, happy), nouns (cat, doctor,

pants), and verbs (catch, draw, push). The most predominant associates elicited from children also matched the most predominant associates produced by adult respondents for two adjectival cues (bad, happy) and nominal cues (cat, doctor).

[Insert Table 5 about here]

Discussion

Semantic convergence in WAT indicates integration with the target language community. This study examined the monolingual-bilingual gap in children regarding their abilities to produce adult-like associative behaviors with three scoring methods, using an English WAT norm that is the most recent and largest of its kind. Based on the same norms list, the study also located the most predominant associates adults produced for the same cues and investigated the effects of distributional properties of word relationships on children's likelihood to produce these adults' targets in WAT. There were three findings. First, monolingual-bilingual gaps in producing adult-like associative behaviors were found in Mandarin-English and Spanish-English bilingual children, regardless of scoring methods. Second, both co-occurrence and semantic relatedness predicted children's likelihood to produce the adult-like associates. None of the effects interacted with bilingualism. Finally, children's likelihood to produce adult-like associates were influenced by multiple factors besides bilingualism and distributional properties of word relationships, which included maternal education, and age of acquisition and word frequency of the target associates.

The monolingual-bilingual gap in producing adult-like associative behaviors

Previous WAT studies focused on the qualitative aspect of cue-associate relationships (e.g., Cremer et al., 2011; Holmström et al., 2016; Namei, 2004; Sheng et al., 2006; Spätgens & Schoonen, 2020) and generated inconsistent results about bilingual (dis)advantages in WAT

assessment in children. Note that questions have been raised regarding scoring methods devised for qualitative analyses (e.g., paradigmatic vs. syntagmatic responses). For example, though paradigmatic associates are typically deemed as more advanced than syntagmatic associates and indicate maturation of the lexical-semantic system, previous work highlights the existence of late-acquired syntagmatic relationships that are produced by a significant number of adult native speakers in WAT as well (Lam & Sheng, 2020b; Nissen & Henriksen, 2006). This study showed that stereotypy scoring methods with direct response matching common in WAT studies (see Fitzpatrick et al., 2015 for a review) would reveal a bilingual disadvantage, regardless of whether the prototypicality of the associates was additionally considered.

Different scoring methods used in evaluating WAT performance may tap into the distinct dimensions of semantic abilities. Compared to qualitative analyses of cue-associate relationships, the direct matching methods adopted in measures of stereotypy and association strength (i.e., the conditional probability to trigger selected responses given a cue) exhibit a much-elevated specificity for target responses. As a result, the direct matching analyses likely tap into the size of language-specific vocabulary, which has been consistently shown to be smaller in bilingual children as compared to monolingual children in single-language testing (e.g., Bialystok, 2009; Gross et al., 2014). This study shows that the monolingual—bilingual gap in semantic convergence does not decrease with age in children of 4- to 7-year-olds. Reduced language input associated with bilingualism would lead to fewer opportunities (Gollan et al., 2008; Sandoval et al., 2010) to learn language-specific lexical forms and word relationships. The gaps in the language-specific mental lexicon and weakened connectivity between words would likely lead to a lower likelihood to produce adult-like associative behaviors in the testing language. The

relatively stable monolingual-bilingual gap between 4- and 7-years of age highlights the importance of language-specific experience to semantic convergence in young children.

Spanish-English and Mandarin-English bilingual children performed comparably on WAT despite differences in language backgrounds when maternal education was controlled. The results highlight a general lag in semantic convergence in bilingual children in WAT when performance assessment attends to whether children would produce the same responses adults would produce for the same cues. Note that we selected early acquired cues with an aim to capture associative behaviors in young children—six cues were perceived to be acquired before age 4, eight before age 5, and only one before age 6 (c.f. Kuperman et al., 2012). Still, some children might not know certain cue words due to individual differences in vocabulary abilities and, as a result, responded with "I don't know" (DK) or did not respond at all (NR). We performed a descriptive analysis on the amount of DK and NR based on the three age bands of cues (i.e., cues acquired before age 4 vs. age 5 vs. age 6; Appendix S2). For English-speaking monolingual children, the proportion of "DK/NR" remained relatively consistent across the three age bands of cues (7.1 to 9.4 % of total responses). In contrast, both bilingual groups showed comparable and steady increases in the proportion of "DK/NR" (Mandarin-English bilinguals: 5.1%, 9.6%, and 16.3% for cues acquired before age 4, age 5, and age 6 respectively; Spanish-English bilinguals: 6.8%, 11.3%, and 13.5% for cues acquired before age 4, age 5, and age 6 respectively). The differences between monolingual and bilingual children are not likely due to the differences in abilities understanding and completing the task—WAT has been used successfully in monolingual and bilingual children of a comparable range of age (Holmström et al., 2016; Sheng et al., 2006). Furthermore, the three language groups in the present study produced a comparable proportion of "DK/NR" for cues acquired before age 4. The increasing

proportion of "DK/NR" for later-acquired cues in both bilingual groups may reflect decreased familiarity with these cues, affecting children's abilities to produce even associates that are remotely related to the cues, such as responses sharing phonological forms (e.g., FIND-fine; Namei, 2004).

Previous studies showed effects of cross-linguistic differences in tasks that targeted specific language features, such as cognates (Sheng et al., 2016) and derivational word formation rules (Lam & Sheng, 2016). In contrast, tasks measuring the general levels of conceptual development may be more likely to reveal similarities across bilingual groups. Lam and Sheng (2020a) measured taxonomic awareness with category and contrast association tasks and showed comparable coordinate awareness in Mandarin-English and Spanish-English children despite differences in heritage language features. The consistent monolingual-bilingual gap in producing adult-like associates for both bilingual groups in the current study suggests that, like the coordinate awareness task in Lam and Sheng (2020), WAT may be relatively linguistically neutral compared to other vocabulary tasks.

Factors predicting children's likelihood to produce adult-like associates

The development of semantic convergence is multifaceted in nature. Age is a powerful factor because it encapsulates multiple effects that may influence semantic convergence, such as the development of cognitive capacity to categorize world experience, expansion of linguistic abilities, and accumulation of language experience. The study showed that the distributional properties of word relationships additionally predicted semantic convergence in children's WAT performance beyond age. The lack of interaction with bilingualism suggests monolingual and bilingual children may exhibit comparable sensitivities to tightly connected word pairs despite differences in the amount of cumulative language input. An analysis of naturalistic child-direct

speech showed that language input provides rich statistical regularities of word relationships, which support the development of abstract and structured semantic knowledge in computational neural network (Huebner & Willits, 2018). Toddlers (Wojcik & Saffran, 2015) and young children of 4 –5-year-olds (Unger et al., 2020) have been shown to be sensitive to co-occurrence information of words, which in turn influences the development of lexical-semantic organization (Unger et al., 2020) and the formation of word associations.

Both co-occurrence and semantic relatedness predicted semantic convergence. Co-occurrence refers to words bounded by close temporal proximity, which is sometimes addressed as first-order distributional statistics in computational corpus-based studies (Rapp, 2002). The effects of co-occurrence on connectivity strengthening may not be constrained to actual usages of words. Once association between two objects is established upon being experienced together, thinking of an object likely triggers the connected counterpart in covert mental processes (Wettler et al., 2005). Sloutsky and colleagues (2017) further suggested that experiential learning of the statistics of co-occurrences (e.g., LION-ZOO) may not only lead to learning of syntagmatic relationships but also facilitate the development of paradigmatic relationships, which signify the development of a mature lexical-semantic system. Accumulation of co-occurrence information of different word pairs (e.g., LION-ZOO; ZEBRA-ZOO; ANIMAL-ZOO) may allow children to discover and link words that exhibit similar syntagmatic associations. Interestingly, in a study comparing computational model of word associations to actual human performance (Wettler et al., 2005), co-occurrence information derived from corpus predicted not only syntagmatic associates but also paradigmatic associates in WAT. Sloutsky and colleagues' hypothesis suggests a linkage between co-occurrence and semantic relatedness—the accumulation of co-occurrence statistics may facilitate the explorations of semantic relatedness

between word pairs. However, note that understanding of semantic relatedness may tap into more advanced linguistic-cognitive abilities because this distributional property speaks to patterns of word usages across sentences. Indeed, Wojcik and Saffran (2015) found that encoding of word meanings might be facilitated by positional similarities *across* sentences in children as young as 2-year-olds but only in those with advanced grammar abilities. Future studies may examine the extent to which co-occurrence and semantic relatedness differ in their effects on semantic convergence, which is beyond the scope of this study. Nevertheless, findings from this study highlight the important role of experience to the distributional properties of language input in the development of lexical-semantic relationships.

Significant evidence highlights the effects of age of acquisition and word frequency on lexical access (for a review, see Goh et al., 2016)—words acquired earlier or used more frequently are processed faster with improved accuracy. These two lexical factors also influence semantic convergence (Brysbaert et al., 2000; de Groot, 1989; Meara, 1983). Findings from this study extend previous work that focused on the lexical features of the cues to the target associates. The findings that both lexical features predicted children's likelihood to produce the target associates show that word frequency may not be an age-of-acquisition effect in disguise (for a review, see Brysbaert et al., 2000).

Maternal education, which relates to levels of sophisticated word usage (Weizman & Snow, 2001), influences the quantity and quality of children's lexical-semantic experience. Maternal education predicts vocabulary development in children (Pan et al., 2005) and influences children's understanding of word relationships within semantic categories (Sheng & Lam, 2015). The present study shows that maternal education impacts the stereotypy and association strength scores of children's associates in analyses considering all adults' responses. The study also

shows the effects of maternal education on children's likelihood to produce the most predominant adult responses. Mothers with more years of education may be more adept at highlighting similar semantic properties of objects and the relationship between concepts in engaging discourse, which improves the effectiveness in infusing children with a rich lexical-semantic experience that represents the larger language community. Improved effectiveness in concept elaborations may help children grasp how and why two concepts semantically relate, reinforcing the semantic network and the activation of related concepts given a cue. Increased frequency of communicative interactions between parents and children associated with higher social-economic status (Hart & Risley, 1995) may result in more exposures to distributional properties of word relationships, which may highlight the co-occurrence and semantic relatedness of word pairs in sentence contexts. Finally, the relationship between maternal education and semantic convergence may reflect the demographic properties of the adult respondents sampled in De Deyne et al. (2019). Since most of the adult participants in the norm had at least a college degree, the mothers with more years of education in our study may influence their children in a way that they responded more similarly to the norming population.

The significant differences in responses produced by adults and children are expected, given the young age of the child participants of the study. Item analyses highlight significant differences between adults and children even for the responses that are the *most* frequently produced by each group. The most predominant cue-associate pairs produced by the three child groups (see Table 5) were identical for 60% of the cues (adjectives: bad-good, big-small, happy-sad; nouns: cat-dog, doctor-nurse, pants-shirt; verbs: catch-throw, draw-color, push-pull). The two bilingual groups were similar for two additional cue-associate pairs (pretty-ugly; corn-food). These single, most predominant responses accounted for 25 to 30% of responses children

produced in the first trial, and 11 to 13% of the total number of responses when aggregating the three elicitation trials. In contrast, children and adults matched on only four out of these *most* frequently produced associates (bad-good, happy-sad, cat-dog, doctor-nurse). The significant discrepancies between children and adults cannot be explained by the age of acquisition of these associates—some of the frequent adult responses are perceived to be acquired before age 4 (e.g., ball) and yet are still found to be less common in children's WAT responses. Interestingly, the "catch-ball" pair commonly produced by adults would be classified as a syntagmatic relationship, which is less mature than "catch-throw" commonly produced by children that exhibits a paradigmatic relationship. Findings from this preliminary item analysis suggest additional effects on semantic convergence beyond bilingualism and distributional properties between words that are beyond the scope of the study. For example, an elevated likelihood for children to produce the more "mature" associate for the cue "catch" may result from the effects of pedagogy (e.g., verbs are taught systematically in themes of action words, such as catch vs. throw), while adults' performance may be influenced by personal orientation (e.g., baseball is a popular sport in the United States).

Limitations

This study does not aim at teasing apart the effects of semantic relatedness and co-occurrence on WAT performance. As suggested above, co-occurrence and semantic relatedness may exhibit a dynamic relationship (Gruenenfelder et al., 2016; Playfoot et al., 2018)—connectivity is stronger for semantically-related words that are more likely to co-activate in lexical tasks and daily encounters (Collins & Loftus, 1975). Future studies might manipulate co-occurrence and semantic relatedness of target cue-associate pairs orthogonally to better

understand the impacts of distributional properties of word relationships on children's WAT performance.

We derived distributional properties of word relationships and lexical-semantic properties of adults' targets based on established, adult-oriented corpora and rating norms. On the one hand, obtaining these variables based on child-directed language databases (e.g., The Child Language Data Exchange System) may provide additional insights about children's assimilation to the ambient language environment. On the other hand, though child-directed language facilitates language acquisition (e.g., Newport et al., 1977), restricting analyses to child-directed language databases may exclude meaningful forms of language input to which children are exposed, such as adult conversations that children overhear, and TV programs broadcasting in the background at home. One aim of this study was to explore the relationships between children's WAT performance and general patterns of language usages that represent the English communities. The significant effects of the distributional and lexical-semantic variables derived from adult-oriented corpora and norms found in this study suggest children's WAT behaviors may be influenced by factors external to the immediate child-directed environment. The findings may also suggest general social-cognitive principles shared among children and adults that drive the conventionality of language usages (e.g., "happy" and "sad" are two opposite ends of the emotional continuum and, therefore, are strongly associated in children and adults). Future studies are needed to evaluate the extent to which child-directed language input contributes to semantic convergence.

The study is limited by difference in maternal education among children, with lowered maternal education for Spanish-English bilingual children. The importance of maternal education is recognized in our findings and studies of other dimensions of semantic abilities, such as

taxonomic awareness (Lam & Sheng, 2020a; Sheng & Lam, 2015). We performed ad-hoc analyses with maternal education removed from the model. The primary findings of the study hold with such analyses—the monolingual-bilingual gap in producing adult-like associative behaviors is evident in Mandarin-English bilingual children and Spanish-English bilingual children, suggesting a bilingual effect on semantic convergence that generalizes across children speaking different language pairs. The ad-hoc analyses further showed that Spanish-English bilingual children produced fewer adult-matching associates in WAT and named fewer pictures accurately than Mandarin-English bilingual children, when maternal education was not considered. In summary, these findings emphasize the general effects of bilingualism and highlights the additional impacts of maternal education on semantic convergence.

Conclusion

This study demonstrates the multi-faceted nature of semantic convergence development in young monolingual and bilingual children. WAT performance reflects language experience and the acquisition of distributional properties of word relationships based on language environment. The likelihood for children to produce adult-like associates was found to be influenced by the experiential factors (e.g., bilingualism, maternal education), the developmental factor (i.e., age), the distributional properties of word relationships (i.e., co-occurrence, semantic relatedness), and the inherent properties of target associates (age of acquisition, word frequency). A consistent monolingual-bilingual gap in the adult-likeness of children's responses was evident for young children between the ages of 4 years and 7 years. Despite significant differences in language backgrounds (e.g., Spanish vs. Mandarin; monolinguals vs. bilinguals), the distributional properties of word relationships exerted a general effect on children's abilities to mirror adult's associative behaviors.

References

Adams, A. K., & Bullock, D. (1986). Apprenticeship in word use: Social convergence processes in learning categorically related nouns. In *The development of word meaning* (pp. 155-197). Springer, New York, NY.

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of vocabulary testing in bilingual children. *Language, Speech, and Hearing Services in Schools*, *49*(1), 85-97.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261-295.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Scheipl, F. (2012). Package 'lme4'. *CRAN. R Foundation for Statistical Computing, Vienna, Austria*.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism*, *12*(1), 3-11.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215-226.

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22-29.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Cremer, M., Dingshoff, D., de Beer, M., & Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism*, *15*(2), 187-204.

Davies, M. (2008). *The Corpus of Contemporary American English* (COCA). Available online at https://www.english-corpora.org/coca/.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987-1006.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824-845.

Entwisle, D. R. (1966). *Word associations of young children*. Baltimore, MD: Johns Hopkins Press

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*(3), 319-331.

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. J. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, *36*(1), 23-50.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237-274.

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L. C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*, *7*, 976.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574-586.

Gruenenfelder, T. M., Recchia, G., Rubin, T., & Jones, M. N. (2016). Graph-theoretic properties of networks based on word association norms: implications for models of lexical semantic memory. *Cognitive Science*, *40*(6), 1460-1495.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, *24*(2), 267-288.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

Holmström, K., Salameh, E. K., Nettelbladt, U., & Dahlgren-Sandberg, A. (2016). Conceptual scoring of lexical organization in bilingual children with language impairment. *Communication Disorders Quarterly*, *38*(1), 24-34.

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813.

Keith, M., & Nicoladis, E. (2013). The role of within-language vocabulary size in children's semantic development: evidence from bilingual children. *Journal of Child Language*, *40*(4), 873-884.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990.

Lam, B. P. W., & Sheng, L. (2016). The development of morphological awareness in young bilinguals: Effects of age and L1 background. *Journal of Speech, Language, and Hearing Research*, *59*(4), 732-744.

Lam, B. P., & Sheng, L. (2020a). Taxonomic Development in Young Bilingual Children: Task Matters, and So Does Scoring Method. *American Journal of Speech-Language Pathology*, *29*(3), 1162-1177.

Lam, B. P., & Sheng, L. (2020b). The Nativelikeness Problem in L2 Word-Association Tasks: Examining Word Class and Trials. *English Language Teaching*, *13*(5), 125-138.

Lenth, R. V. (2016). Using lsmeans. *Journal Statistical Software*, *69*, 1-33.

McNeill, D. (1963). The origin of associations within the same grammatical class. *Journal of Verbal Learning and Verbal Behavior*, *2*(3), 250-262.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, *11*(2), 29-38.

Namei, S. (2004). Bilingual lexical development: A Persian–Swedish word association study. *International Journal of Applied Linguistics*, *14*(3), 363-388.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin*, *84*(1), 93-116.

Nelson, K., & Nelson, A. P. (1990). Category production in response to script and category cues by kindergarten and second-grade children. *Journal of Applied Developmental Psychology*, *11*(4), 431-446.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, Id rather do it myself: Some effects and non-effects of maternal speech style.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results 1. *International Journal of Applied Linguistics*, *16*(3), 389-408.

Palermo, D. S. (1971). Characteristics of word association responses obtained from children in grades one through four. *Developmental Psychology*, *5*(1), 118-123.

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. Child Development, 76(4), 763-782.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing*, 1532-1543.

Playfoot, D., Balint, T., Pandya, V., Parkes, A., Peters, M., & Richards, S. (2018). Are word association responses really the first words that come to mind? *Applied Linguistics*, *39*(5), 607-624.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, *22*(3), 415-422.

Rapp, R. (2002). The computation of word associations: comparing syntagmatic and paradigmatic approaches. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1398-1411.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism*, *13*(2), 231-252.

Sheng, L., Bedore, L. M., Peña, E. D., & Taliancich-Klinger, C. (2013). Semantic convergence in Spanish–English bilingual children with primary language impairment. *Journal of Speech, Language, and Hearing Research, (56)*2, 766-777.

Sheng, L., Lam, B. P., Cruz, D., & Fulton, A. (2016). A robust demonstration of the cognate facilitation effect in first-language and second-language naming. *Journal of Experimental Child Psychology*, *141*, 229-238.

Sheng, L., McGregor, K. K., & Marian, V. (2006). Lexical–semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research, 49*(3), 572-587.

Sheng, L., & Lam, B. P. (2015). Slot-filler and taxonomic organization: The role of contextual experience and maternal education. *Journal of Educational and Developmental Psychology*, *5*(1), 128-138.

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, *97*, 1-30.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, *28*(2), 249-257.

Unger, L., Savic, O., & Sloutsky, V. M. (2020). Statistical regularities shape semantic organization throughout development. *Cognition*, *198*, 104190.

Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology*, *33*(3), 555-567.

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265-279.

Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, *12*(2-3), 111-122.

Wojcik, E. H. (2018). The development of lexical–semantic networks in infants and toddlers. *Child Development Perspectives*, *12*(1), 34-38.

Wojcik, E. H., & Kandhadai, P. (2020). Paradigmatic associations and individual variability in early lexical–semantic networks: Evidence from a free association task. *Developmental Psychology*, *56*(1), 53-69.

Wojcik, E. H., & Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition*, *138*, 10-20.

Schmitt, N. (1998). Quantifying word association responses: What is native-like?. *System*, *26*(3), 389-401.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, *28*(2), 249-257.

Spätgens, T., & Schoonen, R. (2020). The structure of developing semantic networks: Evidence from single and multiple nominal word associations in young monolingual and bilingual readers. *Applied Psycholinguistics*, *41*(5), 1141-1169.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41-78.

Appendix S1. Summary of the lexical-semantic features of the 45 most predominant adult respondents of produced for the "Small World of Words" Association Norm for 15 cues. The most predominant associates are defined as responses that are the most frequently produced when examining all responses produced in three consecutive elicitations.

Word Class	Cue	Target Associate[1]	Word Frequency of target associate[2]	Age of Acquisition of target associate[3]	Mutual Information[4]	Semantic Relatedness[5]
ADJECTIVE	bad	GOOD	2610.1	3.6	3.36	0.74
		EVIL	73.2	6.7	1.91	0.53
		BOY	529.8	3.7	3.38	0.44
	big	LARGE	41.5	5.7	0.07	0.66
		HUGE	48.4	4.9	1.99	0.87
		SMALL	125.0	3.2	2.41	0.63
	happy	SAD	63.4	3.2	4.19	0.64
		SMILE	58.0	3.5	2.49	0.60
		GLAD	171.4	5.8	2.29	0.77
	pretty	BEAUTIFUL	279.7	5.7	0.82	0.61
		GIRL	557.1	4.0	3.52	0.51
		FLOWER	22.8	3.1	1.96	0.33
	round	CIRCLE	21.5	3.7	3.81	0.46
		BALL	105.0	2.9	2.65	0.44
		CIRCULAR	1.6	7.7	3.4	0.41
NOUN	cat	DOG	192.8	2.8	6.06	0.80
		FELINE	0.9	8.4	7.44	0.70
		MEOW	2.9	4.0	9.76	0.55
	corn	COB	0.7	6.5	12.07	0.54
		FIELD	70.2	6.1	4.31	0.27
		YELLOW	33.8	3.2	3.67	0.35
	doctor	NURSE	45.0	5.8	5.42	0.69
		MEDICINE	34.2	4.9	3.65	0.69
		LAWYER	79.5	7.8	4.1	0.47
	pants	TROUSERS	5.2	7.9	5.85	0.85
		JEANS	6.6	5.3	5.74	0.84
		SHIRTS	46.4	3.5	7.02	0.79
	window	GLASS	60.7	4.5	5.43	0.54
		PANE	0.5	9.8	8.11	0.67
		DOOR	292.1	3.1	3.63	0.64
VERB	catch	BALL	105.0	2.9	4.72	0.45
		FISH	83.5	4.1	5.72	0.54
		BASEBALL	25.3	4.8	1.2	0.29
	draw	PICTURE	138.5	4.1	4.3	0.34
		PENCIL	9.9	4.1	4.56	0.45
		ART	70.8	6.2	1.15	0.39
	find	DISCOVER	12.0	7.6	0.43	0.75

		SEARCH	48.4	8.7	2.17	0.68
		TREASURE	19.1	6.1	3.26	0.41
	push	SHOVE	13.2	6.7	8.86	0.61
		PULL	146.5	4.8	4.88	0.76
		FORCE	70.7	6.0	1.74	0.57
	sing	SONG	93.7	4.3	7.85	0.72
		MUSIC	151.6	3.8	3.41	0.55
		VOICE	86.2	4.8	1.45	0.57

1. The "Small World of Words" Association Norm (De Deyne et al., 1998, 2004).
2. Word frequency was defined as how often a word occurs per million words in a corpus of 51 million words based on television and film subtitles (Brysbaert & New, 2009).
3. Age of acquisition was based on the age-of-acquisition ratings for 30,000 English words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012)
4. Mutual information was based on Church and Hanks (1990) and Davies (2008)
5. Semantic relatedness was based on GloVe (Pennington et al., 2014)

Appendix S2. The percentage of "I don't know (DK)" or "no responses (NR)" in the word association task performed by child participants. The 15 cues in the present study are divided into three age bands (i.e., cues acquired before age 4, age 5, and age 6). Number of DK/NR and total number of responses are presented in parentheses.

Age band (cues)	Mono-English[1] (N=28)	Bi-Man Eng[2] (N=45)	Bi-Spa Eng[2] (N=32)
Acquired before age 4 (bad, big, cat, happy, pants, sing)	7.9 % (40/504)	5.1% (41/810)	6.8% (39/576)
Acquired before age 5 (catch, corn, doctor, draw, pretty, push, round, window)	9.4% (63/672)	9.6% (104/1080)	11.3% (83/768)
Acquired before age 6 (find)	7.1 % (6/84)	16.3% (22/135)	13.5% (17/96)
Total	8.7% (109/1260)	8.2% (167/2025)	9.7% (139/1440)

1. Mono-English: English-speaking monolingual children
2. Bi-Man Eng: Mandarin-English bilingual children; Bi-Spa Eng = Spanish-English bilingual children

Note: The total number of responses is different between groups due to different number of participants in each group.