

ORIGINAL ARTICLE

The structure of developing semantic networks: Evidence from single and multiple nominal word associations in young monolingual and bilingual readers

Tessa Spätgens¹  and Rob Schoonen² 

¹University of Amsterdam and ²Radboud University, Nijmegen

*Corresponding author: Email r.schoonen@let.ru.nl

(Received 10 August 2017; revised 29 June 2020; accepted 13 July 2020; first published online 14 October 2020)

ABSTRACT

The present study focuses on the effect of an important methodological choice in word association studies in children: the elicitation of single versus multiple responses. This choice has been shown to affect the numbers and types of associations adults produce, however, little is known about how it affects children's word associations. A total of 11,725 associations to 80 nouns from 207 monolingual and bilingual minority children were classified according to a detailed coding system, and differences between the semantic characteristics of first, second, and third responses were examined. We show that in children as well, the multiple association task elicits more and qualitatively different responses, resulting in more diversified semantic networks surrounding the stimulus nouns. On the speaker level, reading comprehension scores were related differently to initial and later responses, suggesting a more complex measure of semantic knowledge emerges from the multiple word association task. No differences were found between monolingual and bilingual children's associative preferences. We argue that the multiple association task produces more detailed data on language users' semantic networks than the single association task, and suggest a number of ways in which this task could be employed in future research.

Keywords: word association; reading comprehension; vocabulary knowledge; monolingual and bilingual children

Introduction

The network metaphor for the storage of semantic knowledge in the mental lexicon is evoked in many studies on lexical knowledge, and the studies using word associations to probe the exact nature of this network are numerous (Fitzpatrick, 2012). Assumed to reflect the strongest connections in the mental lexicon's network structure (e.g., Schmitt, 1998), word associations provide us with valuable information on both the networks in which specific words reside and the organization of individual

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

language users' semantic knowledge. In other words, word associations give us insight in semantic networks on two levels: on the word level, that is they show us what a semantic network surrounding a specific word may look like based on associations from a group of people; and on the level of the mental lexicon of the individual language user, that is combining word associations to several words from one person can shed light on the types of semantic relations that are more prominent in the individual's mental lexicon (cf. Zareva, 2011). Word association data on the word level are widely used in psycholinguistic studies on lexical knowledge and retrieval (most notably semantic priming studies, cf. McNamara, 2005), while word association data on the level of the language user have been employed with varying success as a tool to measure vocabulary knowledge (e.g., Cremer et al., 2011; Fitzpatrick et al., 2013) and to distinguish different types of language users such as L1 and L2 speakers or bilinguals (e.g., Cremer et al., 2011; Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Izura, 2011; Söderman, 1993; Wolter, 2001; Zareva, 2007; Zareva & Wolter, 2012) and different age groups (Borghgi & Caramelli, 2003; Fitzpatrick et al., 2013).

Most word association studies require their participants to produce a single association to each stimulus word that is presented. It has been argued that only these initial associations are unbiased, while later associations may be primed by the initial association (chaining) or, in other cases, the initial response may block retrieval of any further associations (McEvoy & Nelson, 1982). In a recent, large-scale study, De Deyne and colleagues found evidence of chaining, but the occurrence did not appear to be very frequent, that is in 1% of the cases there was strong evidence for chaining and in 19% only moderate and to some extent doubtful evidence (De Deyne et al., 2019).

Eliciting single responses creates associative networks that only represent very strong associates, while evidence for weaker links is unreliable or even missing altogether (D. L. Nelson et al., 2000). This means that although single associations seem to be the default option in many studies and are used in the larger association databases such as the Edinburgh Associative Thesaurus (Kiss et al., 1973) and the University of South Florida association norms (D. L. Nelson et al., 2004), they may not provide sufficiently extensive information about the many associative connections a word may have.

The qualitative and quantitative consequences of the use of single versus multiple responses has been investigated in great detail in a series of recent studies by De Deyne and colleagues (De Deyne & Storms, 2008a, 2008b; De Deyne et al., 2013; De Deyne & Verheyen, 2015; De Deyne et al., 2019, among others). Using their word association database, which is the largest collected to date (De Deyne et al., 2013), they analyzed associative networks on the word level. Comparing these networks including only first responses versus first, second, and third responses, they discovered that the average set size of nonunique responses, that is responses that were provided by more than one person and thus were not idiosyncratic, increased significantly when multiple associations were included (De Deyne & Storms, 2008b). As a result, more semantic links were uncovered, building more extensive semantic networks around stimulus words and thereby obtaining more accurate association norms. Furthermore, by categorizing their data using an extensive semantic coding scheme, the researchers showed that these additional associations are also qualitatively different from first associations (De Deyne & Storms, 2008a).

Most notably, taxonomic associations such as category subordinates, superordinates, and coordinates showed a clear decline across successive responses, while words associated through context became more prominent across responses. The authors argue that this again shows that the multiple association task provides more accurate association norms because more semantic links are uncovered (De Deyne & Storms, 2008b) and that the resulting associative networks indicate that semantic organization is much more contextually¹ oriented than previously thought (rather than categorically oriented) (De Deyne & Verheyen, 2015; see Santos et al., 2011 for a psycholinguistic explanation for this observation).

The Single Versus Multiple Association Task in Children

The present study aims to replicate and expand upon these findings in a number of ways. Firstly, knowing that age plays an important role in associative behavior (Borghi & Caramelli, 2003; Cremer et al., 2011; Fitzpatrick et al., 2013), we will investigate the effect of eliciting single versus multiple word associations from children. It has been suggested that as children's semantic knowledge develops, they increasingly abstract from concrete experience to more context-independent semantic knowledge (Elbers et al., 1993; K. Nelson, 1977, 1982, 1985, 1991), even though both types of semantic knowledge are present from an early age (Blewitt & Toppino, 1991) and remain important in the adult mental lexicon (De Deyne & Verheyen, 2015). Typically, taxonomic semantic links are considered to be the most abstract, context-independent knowledge (Cremer, 2013; Verhallen & Schoonen, 1993) and because De Deyne and Storms (2008a) demonstrated that especially these semantic links showed different behavior when comparing initial and later associative responses, it will be interesting to see which patterns across responses turn up in children's word associations. It is conceivable that there is a similar but overall reduced pattern, with taxonomic responses being most prominent as first responses, but less prominent overall than in adults. Alternatively, taxonomic links may be less prominent in children than in adults but already developing, and showing up more as second and third associations instead of as first associations. Additionally, the behavior of other types of semantic links such as associations pertaining to features of the stimulus and associations that are related through context may also show different patterns in children than in adults.

As in adults, it is likely that the multiple response procedure allows for more semantic links beyond initial strong associates such as *salt – pepper* to emerge than the single response procedure. Both these qualitative and quantitative differences are consequential for the use of association data on the word level to control for lexical features in experimental research. We expect that as in adults, more and different types of semantic links will be found in children in a multiple word association task compared to a single association task.

Differences between Monolingual and Bilingual Children

Secondly, we want to look more closely at the effects of eliciting multiple responses on the picture that emerges of a speaker's mental lexicon, both regarding the

possibility of distinguishing groups of language users using this type of data and regarding its use as a tool to measure differences in the connectivity in the mental lexicon that may be predictive of other language skills. To start with groups of language users, a large number of studies has attempted to compare L1 and L2 speakers' and monolingual and bilingual speakers' word associations to see whether their associative behaviors are different (e.g., Cremer et al., 2011; Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Izura, 2011; Söderman, 1993; Wolter, 2001; Zareva, 2007, 2010; Zareva & Wolter, 2012). The findings in these studies have been mixed, which is likely due to the variation in types of L2 or bilingual speakers involved but also to the classification methodologies used.

For example, initial L1/L2 association studies such as Söderman (1993), examined differences between L1 and L2 speakers using a traditional classification scheme, focusing on the syntagmatic-paradigmatic contrast, in addition to phonologically related and "other" associations. In this scheme, associations are mainly classified according to word class. Syntagmatic associations belong to a different word class than their stimulus word, while paradigmatic associations belong to the same word class. Söderman (1993) found that both syntagmatic and paradigmatic responses were frequent in L1 and advanced L2 speakers and that as L2 proficiency increased, learners produced more paradigmatic responses compared to syntagmatic responses, and fewer phonologically related and other associations.

However, other researchers have argued that this distinction is not sufficiently detailed because using word class exclusively ignores the range of semantic relations that may exist between concepts. More fine-grained categorization systems have been proposed and resulted in different findings. For example, Fitzpatrick (2006, 2007, 2009) devised a classification system with a meaning-based category, a position-based category, and a form-based category to compare L1 and L2 speakers. She concluded that there may be no such thing as an L2 or even an L1 word association profile. Instead, she found that speakers had personal association profiles, which reflected similar behavior in both the L1 and the L2. Based on Fitzpatrick's work, Cremer et al. (2011) also developed their own classification system, which divides a meaning-based category in a direct meaning-related and indirect meaning-related category, based on the distinction between context-dependent and context-independent semantic knowledge (Schoonen & Verhallen, 2008; Verhallen & Schoonen, 1993). They found that both bilingual children and adult L2 speakers produced more "other" and form-based associations than their monolingual peers, but provided roughly equal proportions of answers in the meaning-based categories in comparison with the respective monolingual groups.

In short, while some researchers find differences between monolingual and bilingual or L2 speakers, which may or may not disappear as L2 proficiency improves, others do not. The studies that employed more fine-grained semantic categories did not find clear differences between monolingual and bilingual or L2 speakers in those categories. This is contrary to what we might expect based on other studies that have found differences in knowledge of various types of semantic relations, such as for example Verhallen and Schoonen (1993), who found that monolingual children produced more context-independent relations than sequential bilingual children in an extended word definition task. Differences in research outcomes may have

various causes, such as the frequency, part of speech or difficulty of the stimulus words, or features of the coding scheme. However, if some studies find differences between monolingual and bilingual associative behavior and others do not, it may also be the case that these differences are simply very subtle. Potentially, the multiple association approach could be a useful tool to detect such subtle differences, which could for example turn up only in later associations rather than initial ones, tapping into different kinds of associative sources (Santos et al., 2011). Furthermore, a comprehensive classification system with categories that are motivated from a semantic and acquisitional perspective may allow for a more detailed comparison between the two groups.

Using the multiple association approach and an extensive coding scheme based on De Deyne and Storms (2008a) and Cremer et al. (2011), we will try to shed more light on these issues. We have developed a classification system largely based on De Deyne and Storm's (2008a) coding scheme with slight adjustments that are related to the distinction between context-dependent and context-independent semantic knowledge (cf. Cremer et al., 2011; Cremer, 2013; Schoonen & Verhallen, 2008; Verhallen & Schoonen, 1993). The classification system and the rationale behind it will be discussed in more detail in the "Method" section.

Word Associations and Other Linguistic Skills: Reading Comprehension

Finally, word association tasks may be useful as a measure of qualitative differences in vocabulary knowledge that may be predictive of other language skills, and initial and later responses may fare differently in this regard as well. Cremer et al. (2011) argue that word associations are not restricted enough to be used as an assessment tool for vocabulary knowledge. They maintain that if some type of association is considered to be the most important or most advanced, for example superordinates, then participants should be asked to produce superordinates to test whether they have this semantic knowledge. Nevertheless, word associations do provide insight into the types of semantic links that are primary in a language user's mental lexicon. Indeed, Fitzpatrick et al. (2013) demonstrated that when tested on separate occasions, individuals produce similar types of associations over time, which means that the prominence of those semantic links is stable. As such, word associations can be a useful tool to investigate the relation between semantic knowledge or preferences and other language skills.

In the present study, we will investigate how individual children's word associations relate to their reading proficiency, and whether there is a difference in the predictive potential of initial and later responses. Because it has been suggested that knowledge of semantic relations and perhaps especially knowledge of context-independent relations contributes to reading comprehension (Bonnotte & Casalis, 2010; Cremer & Schoonen, 2013; Nation & Snowling, 1999; Ouellette, 2006), we expect that children who produce more context-independent associations such as taxonomic associations, are also better comprehenders than those who do not produce these associations as frequently. Depending on the associative patterns that children will show, for example if they, like adults, produce more taxonomic

associations in the first instance than in later responses, the response position of these context-independent associations may be of particular importance.

In summary, the research questions we aim to address in the present study, all focusing on nouns as stimulus words, are:

1. In what ways are the semantic networks arising from word association tasks affected by the elicitation of single versus multiple associations?
2. Do monolingual and bilingual children show different association patterns across responses?
3. How do individual children's word associations relate to their reading proficiency, and is there a difference in the predictive potential of initial and later responses?

Regarding information about individual words, we expect to find more and more semantically diverse links in a multiple association task compared to a single association task, which is of relevance for the measurement of association norms. On the level of the language user, we anticipate that multiple associations may be more sensitive to differences between monolingual and bilingual children. Although both groups can be considered language learners, taking their young age into consideration, the latter group acquired another language as their home language (see following text). We will refer to them as bilingual, but their language experience is different from that of balanced bilinguals.

Method

Participants

Participants were recruited through schools in five different cities in the Netherlands. Only schools in neighborhoods with a mixed population of monolingual and bilingual speakers of Dutch were approached to optimize comparability between language groups. Eight schools gave their permission to administer the test in one or two grade five groups. A passive informed consent procedure was applied and two children did not participate in the tasks because their parents objected. In total, 232 children participated in the study. Seven children with disorders such as dyslexia or autism were identified by the teachers and excluded from the analyses. The data from 16 children were removed because their data were incomplete as they could not be present for all tasks and one child who misunderstood the instructions was also excluded. Finally, one child was excluded because he had not gone to school in the Netherlands from grade 1 onward, thus both language groups had similar exposure to Dutch primary education.

The final data set included the results from 207 children. The descriptives of the participants are provided in Table 1. The mean age of the children was 11;2 (sd 0;6) and ages ranged from 9;10 to 12;10. Based on a language questionnaire, the participants' language status was determined. The majority of the children in the bilingual group spoke both Dutch and another language at home (133 children), and a few children used another language exclusively at home (10 children). A wide range of

Table 1. Participants' descriptives: Language background, Sample size (N), Gender, and mean Age (and standard deviation)

Language group	N	Gender F/M	Age (sd)
Monolingual	64	33/31	10;11 (0;6)
Bilingual	143	69/74	11;3 (0;6)
Total	207	102/105	11;2 (0;6)

languages was reported, with a majority of the children speaking Moroccan Arabic, and/or Berber (57), Turkish (25), or Surinamese (14).

Materials

Word association task

In this task, each child was required to produce three associations to 20 words. A total of 80 nouns were investigated, spread over four lists of 20 words each. Each list was randomized into three orders, to avoid order effects and cheating. All words were selected from the 5,000 most frequent words in Schrooten and Vermeer's corpus of school language, which includes words used in Dutch primary education, based on sources such as schoolbooks and classroom interaction recordings (Schrooten & Vermeer, 1994). Frequent words were used to ensure all participants were familiar with the stimuli. We refrained from testing participants' knowledge of these words, as doing so may affect their spontaneous word associations. Testing the words with another sample of children would not guarantee knowledge of the words in our current sample and would not provide us with any better information than the (large-scale) corpus data of Schrooten and Vermeer (1994). See Zareva (2010) for another approach.

The four lists are provided in Appendix 1. To avoid chaining effects as much as possible, a format based on that used by De Deyne and Storms (2008a, 2008b) was designed. Following De Deyne and Storms, the three associations were filled out from top to bottom, rather than from left to right. In addition, all words were printed three times in a small table, with an open cell under each word for the children's answers. This was done to remind the children of the stimulus word before they made each association.

Reading test

The standardized reading test *Begrijpend Lezen 678* by Aarnoutse and Kapinga (2006) was used to measure reading comprehension skills. It includes a total of 44 multiple choice questions and has been normed in a sample of 42 schools across the Netherlands. It is designed to measure reading comprehension in grades 4, 5, and 6. None of the schools had used this reading test before with the children who participated. The scores on the reading test were somewhat skewed but otherwise normally distributed with skewness $-.720$ ($se = .169$) and kurtosis $-.511$ ($se = .337$).

Language questionnaire

In this short questionnaire, children were asked which language they had acquired first and which languages they spoke at home, and whether they had gone to school in the Netherlands from grade 1 onward.

Procedure

The word association task was administered first. In total, 12 versions of the task were devised (4 lists \times 3 versions), and these were handed out randomly. In each group, each version was given to at least one child. A short instruction and an example were printed on the booklets and read out by the experimenter. It was stressed that there were no right or wrong answers, and that the test would not be checked on spelling mistakes. The participants were urged to write down the first words that came to mind in response to the stimulus words, and to use single words as much as possible. They were also asked to put down a cross for missing associations if only one or two associations came to mind easily, rather than thinking it through too much. Any unfamiliar words could also be crossed out. In the total data set, there were 62 cases in which a stimulus word was unfamiliar to the participant, that is elicited zero associations (1.5%) and in 223 cases a partial response was provided to a stimulus word, that is the second and/or third response was crossed out (5.4%). A maximum of 25 minutes was necessary for all children in a group to complete the task.

Subsequently, the children received the reading test. The use of the answer sheets was explained and an example text with questions was read out and discussed by the experimenter. All children finished the test within the set time limit of 45 minutes. Finally, the children were given the language questionnaire, which took about 5 minutes to complete.

Analysis

Classification of word associations

Based on findings on lexical organization in acquisition and on the relation between different types of semantic knowledge and reading comprehension, it was decided to use a classification system that incorporates the distinction between context-independent and context-dependent semantic relations (Cremer, 2013; Schoonen & Verhallen, 2008; Verhallen & Schoonen, 1993), but as a continuum. See the full classification system with examples in Table 2. On this scale, the most context-independent semantic relations exist between concepts that are related because they share many intrinsic features, and are thus inherently related independent of context. These relations are included in the taxonomic category. A step further down are feature relations, where one concept expresses a necessary or prototypical feature of the other. This is again independent of context, but contrary to taxonomically related items, means only one characteristic is shared. Situational relations exist between concepts that co-occur in given contexts, but are not necessarily or prototypically related. Note that these include features that are nonprototypical. Finally, the most context-dependent category subsumes subjective relations, which are thus tied to an individual's personal context, including, for example, subjective

Table 2. Classification system of word associations

<p>Semantic Context-independent</p>	<p>Taxonomic <u>Coordinate</u>: <i>crocodile – alligator, king – queen</i> (cf. Aitchison, 2012: opposites are members of a small category) <u>Superordinate</u>: <i>dog – animal, trumpet – instrument</i> <u>Subordinate</u>: <i>dog – chihuahua, king – Willem-Alexander</i> <u>Synonym</u>: <i>castle – fort, cook – chef</i></p>
	<p>Feature <u>Context-independent characteristics</u> (necessary or prototypical: normally, X is Y): a. Physical (color, shape, taste, texture): <i>banana – yellow, ocean – deep</i> b. Internal (ability, traits, etc.): <i>snail – slow, cucumber – edible</i> <u>Behavior</u> (typical behavior or action of a stimulus): <i>dog – bark, thief – steal</i> <u>Function</u> (typical function of the stimulus): <i>cot – sleep, trumpet – music</i> <u>Partonym</u> (whole-part and part-whole): <i>coat – sleeve, apple – peel</i></p>
	<p>Situational <u>Co-occurrence</u> (the stimulus and the response occur together in given possible contexts): <i>princess – hat, butterfly – sun</i> <u>Context</u> (the response is an event, location, time that forms a context for the stimulus or vice versa): <i>ocean – rocks, cook – restaurant</i> <u>Action</u> (action performed by or with the stimulus in a situation or script, not typical behavior or main function of a stimulus): <i>bakery – choose, sea – swim</i> <u>Context-dependent attributes</u> (potential, X can be Y, but need not be): a. Physical: <i>car – blue, cook – fat</i> b. Internal: <i>king – calm, island – deserted</i></p>
<p>Context-dependent</p>	<p>Subjective <u>Emotion/evaluation</u>: <i>cheese – ewe, bride – beautiful</i> <u>Personal</u>: <i>hairdresser – my mother, dog – Layla</i></p>
<p>Nonsemantic</p>	<p>Lexical <u>Compound</u> (forward and backward): <i>crocodile – leather</i> <u>Orthographic/phonological similarity</u> (similar form, other meaning): <i>house – hose</i> <u>Morphemic change</u>: <i>cat – catty, dog – dogs</i></p>
	<p>Other <u>Mediation</u> (linked via intermediate concept named earlier): <i>teacher – easy</i> (via <i>lesson</i>), <i>cucumber – fruit</i> (via <i>vegetable</i>) <u>Nonclassifiable</u> (including responses to stimuli that have been misinterpreted as different words): <i>skirt – loud</i> (interpretation of <i>rok</i> “skirt” as <i>rock</i>) <u>Repetition</u> <u>No response</u></p>

evaluations of concepts. Many studies on first language acquisition have shown that children start out learning mainly context-dependent semantic relations between concepts that are bound through context, and abstracting from this situation-specific knowledge to more decontextualized knowledge later on (e.g., K. Nelson, 1977, 1982, 1985, 1991, 2007; Petrey, 1977; Elbers et al., 1993; Lin & Murphy, 2001), which is why we believe this continuum is a useful way of studying knowledge of semantic relations.

Indeed, although the ordering discussed in the preceding text is not used, to a large extent these categories are present in the extensive coding scheme from

De Deyne and Storms (2008a),² which in turn is an extension of classifications of semantic categories by McRae and Cree (2002) and (a prepublication of) Wu and Barsalou (2009). The classification scheme used by De Deyne and Storms was somewhat simplified for our purposes by combining small subcategories into larger ones, and we thus incorporated the explicit ordering from context-independent to context-dependent relations. The only change we made relating to the content, is that we switched the functions subcategory from the situational category to the feature category, with an emphasis on those functions being necessary or prototypical. This was done because the function of an object is inherent to its meaning in the same way as an animate beings' behavior. As is customary in word association studies, a lexical category was included to cover purely form-related associations, which do not fall into the range of semantic associations. Finally, the "other" category includes indirect and unclear links.

In some cases, multiple classifications could apply, for example when stimuli and responses formed a compound. In such cases, a code from the semantic categories was applied if it was most probable, similar to the methodology applied in Cremer et al. (2011). In cases in which multiple semantic categories could be applicable, such as *cat – mouse*, it was chosen to apply the category that was most context-independent, in this case *taxonomic coordinate* rather than *situational co-occurrence*.

The classification was done by the first author and another coder. To make sure there was sufficient agreement, the associations to two sets of words were classified by both coders. A first set of 16 words coded in two rounds of eight words each was used as an initial training set and to identify potential problematic categories. In this set, 82.6% of associations were classified into the same main categories by both coders. Disagreement was resolved through discussion and the category definitions were adjusted where necessary. Another set of 16 words, again in two rounds of eight words each, was used to check whether there was sufficient agreement between both coders to code the rest of the data independently. In this second set, 94.4% were classified in the same main categories by both coders. Disagreement was resolved through discussion. Each coder then classified the associations to half of the remaining 48 words, which were randomly assigned.

Data cleaning and statistical analysis

Before the statistical analyses, all lexical, form-related associations (68 associations, 0.5% of the data) and all "other" associations (5.0% of the data) were removed. In this way, only semantic associations remain in the final data set, consisting of 11,725 associations. For reading comprehension, the number of items correct in the reading test was used.

To answer our first research question about the quantitative and qualitative effects of single and multiple word associations, we will analyze set sizes, that is, the number of nonunique associations per response position, as well as the distribution of semantic association types across response positions. To determine the effect of response position on each of the semantic association types in a regression model, each association was coded 1 or 0 for each of the four main association categories, that is, a taxonomic association would be coded 1 in the variable

“taxonomic,” and 0 in the variables “feature,” “situational,” and “subjective.” Furthermore, each association was coded 1, 2, or 3 for the response position. Separate binomial mixed-effects analyses were performed for each of the four semantic association types using the lme4 package for multilevel and mixed-effects analyses (Bates et al., 2015) in R (R Core Team, 2016). It was chosen to do separate binomial analyses, so that each association type could be inspected in a mixed-effects model that takes into account the multilevel and crossed nature of the data, that is associations being nested within words and children, and children within schools. In the mixed-effects models, the binary coded semantic association type was the dependent variable and the main effect of response position was estimated as the increase or decrease of the probability of an association of the semantic type concerned, in terms of log odds. Response position (1–3) was included as a continuous variable.³ Random intercepts for item, participant, and school were included in the models to take into account random differences between schools, participants, and words. The random effects structure was kept maximal with regard to the grouping variables of interest, items, and participants (c.f. Barr et al., 2013; Linck & Cummings, 2015). We did estimate the more restricted models excluding random slopes for participants and items as well, and verified that for each category, the maximal model indeed was the best fit to the data.

To address the second research question about the effect of language status, we added in the aforementioned models a main effect of language group (monolingual vs. bilingual) and the interaction with response position.

For our third research question regarding the relation between word associations and reading proficiency, we needed scores for children’s preferences for certain types of association. These preferences are represented by the random intercepts for participants in the mixed-effects models. These participant intercepts serve as a score for the tendency of the individual child to produce an association of the association type in question. In short, they are the individual deviations in log odds from the overall mean log odds of producing a certain association type, controlled for item and school variance. For each association category these intercepts were extracted from the respective models. This resulted in four association scores for each child, one for each association type. Linear mixed-effects regression models were estimated to determine the effect of each association type on the reading scores, again controlled for variation between schools. The specifications of each model are discussed in more detail in the “Results” section.

Results

Quantitative and qualitative differences between initial and later responses (RQ1)

Quantitative differences

To establish whether a larger number of different associations is obtained in the multiple association task than there would be if a single association task was used, the average set size per stimulus for each response position and the total set was calculated. Only nonunique responses were counted, that is, responses provided by at least two participants. Note that if a particular response occurs in each response position

Table 3. Average set size and standard deviation of nonunique responses per stimulus word

	Min	Max	Mean (sd)
Response 1	4	19	11.6 (3.2)
Response 2	9	21	14.1 (2.7)
Response 3	8	20	14.2 (2.6)
Total	11	28	18.1 (3.5)

for a particular stimulus, it will be included in each set size. Table 3 shows the set sizes of nonunique associations, indicating that on average, the full response set for a particular stimulus contains about six additional nonunique responses compared to the first response alone. More associative links can thus be identified using the multiple association task. Although it might not seem unexpected that more responses lead to more associative links, we would like to point out that it concerns *new* associative links that had not been mentioned before as an initial association, and it concerns non-unique associations, that is associations mentioned by two or more participants. This is noteworthy as single word associations are widely used, providing only a limited view on participants' mental lexicons.

Qualitative differences

An overview of the qualitative differences between associations across responses is provided in Table 4. Overall, the number of semantically related associations decreases because the number of "other" associations, mainly null responses, increased across response position. Furthermore, the number of taxonomic and feature associations appears to drop, while the number of situational and subjective responses seems to increase, mostly between the first and second response, and to a lesser extent between the second and third response.

To determine the effect of response position statistically, the four separate binomial mixed-effects models were made with each binary coded semantic association type as dependent variable (see "Data Cleaning and Statistical Analyses"). The random effects structure was kept maximal with regard to the grouping variables of interest, items, and participants. Random slopes for response position for schools were not estimated because we did not have enough schools to reliably determine these variances. The estimates for the maximal models for each association type are provided in Table 5.

The main effects of response position confirm the patterns observed in Table 4. The probability of a response being taxonomic or a feature association decreases across response position ($\beta = -0.206$ (0.095), $p = 0.030$, meaning a 19% decrease of taxonomic associations between each response position and $\beta = -0.268$ (0.070) $p = 0.000$, a 24% decrease of feature associations, respectively), while the probability of situational and subjective associations increases ($\beta = 0.433$ (0.060), $p = 0.000$, a 54% increase of situational associations, and $\beta = 0.421$ (0.081) $p = 0.000$, a 52% increase of subjective associations, respectively). Thus, even taking into account variation on the item and participant levels, the effect of response position is significant for all four semantic categories.

Table 4. Counts and percentages of association types by response position

	Counts				
	Taxonomic	Features	Situational	Subjective	Total
Response 1	586	1613	1558	252	4,009
Response 2	361	1224	2045	312	3,942
Response 3	289	1072	2096	317	3,774
Total	1236	3909	5699	881	11,725
	As percentage of response position				
	Taxonomic	Features	Situational	Subjective	Total
Response 1	14.6	40.2	38.9	6.3	100.0
Response 2	9.2	31.1	51.9	7.9	100.0
Response 3	7.7	28.4	55.5	8.4	100.0
Total	10.5	33.3	48.6	7.5	100.0

An interesting exploratory observation is that for each association category, the random intercept and slope variation between items is larger than between participants, with random intercept variance for items ranging from 2.459 to 3.888, compared to the random intercept variance for participants which ranges from 0.201 to 1.629. This suggests that some items are more conducive to associations of a particular type than others, and while participants also vary in how likely they are to provide certain types of associations, this variation is comparatively smaller. This pattern holds for each association type.

In the same vein, comparing the intercept variances between models, that is the variances of the different association types, we see that the by-participant random intercept variance estimates for the taxonomic (1.629) and subjective items (1.400) are noticeably larger than for the feature (0.201) and situational associations (0.396). Children thus differ more in how likely they are to provide taxonomic and subjective associations, than with respect to feature and situational associations. To a certain extent, this may well be linked to the fact that these categories are simply smaller, meaning that variation in log odds is likely to be larger. For example, producing three more associations in a category in which most children produce only a small number, for example, taxonomic associations, makes a larger difference than producing three more associations in a category in which most children produce many associations, for example, feature associations. However, given that this principle should also apply to the variation between items, which are quite similar in terms of intercept variance, this cannot be the only explanation for this pattern.

Differences between monolingual and bilingual children (RQ2)

To determine the effect of language status, we added a main effect of language group and the interaction with response position to the models in Table 5, but neither was significant for any of the four association types. In our sample, the monolingual and

Table 5. Mixed-effects model estimates for the maximal models of each associative category (four analyses), testing the fixed effect of response position on the probability of the occurrence of a word association of the semantic type concerned

		Taxonomic			Features			Situational			Subjective		
Random effects													
Items	Intercept variance	3.888			3.372			3.479			2.459		
	Slope variance for response position	0.382			0.297			0.206			0.126		
Participants	Intercept variance	1.629			0.201			0.396			1.400		
	Slope variance for response position	0.186			0.052			0.053			0.039		
School	Intercept variance	0.004			0.017			0.000			0.050		
Fixed effects		Estimate (se)	<i>z</i>	<i>p</i>									
Intercept		-2.341 (0.267)	-8.76	0.000	-0.362 (0.223)	-1.63	0.010	-0.954 (0.222)	-4.30	0.000	-4.099 (0.269)	-15.26	0.000
Response position		-0.206 (0.095)	-2.18	0.030	-0.268 (0.070)	-3.82	0.000	0.433 (0.060)	7.24	0.000	0.421 (0.081)	5.17	0.000

Table 6. Mean (M) and standard deviation (SD) for reading comprehension scores and raw number of “other” answers by language groups, tested for group differences

	Monolingual (N = 64)		Bilingual (N = 143)		t(df)	p	d
	M	SD	M	SD			
Reading comprehension	32.06	6.43	29.06	5.69	3.210 (109)	0.002	0.49
Number of “other” responses	1.98	2.99	3.50	4.64	-2.808 (179)	0.006	0.39

bilingual children thus do not differ in terms of preferences for any of the four association types, independent of response position, nor in interaction with response position, which means that both groups show similar association preferences and express those at the same response positions in the multiple word association task.

We checked whether the monolingual and bilingual children in our sample did differ in terms of language skill as far as possible by comparing their scores on reading comprehension and the number of “other” responses, which can be considered a proxy of knowledge of the stimulus words. Table 6 shows that there were small but significant differences for both measures: bilingual children produced more “other” responses (Cohen’s $d = 0.39$, a small effect size), suggesting that their vocabularies may have been smaller than those of the monolingual children, and bilingual children performed worse on the reading comprehension scores than their monolingual peers (Cohen’s $d = 0.49$, close to a medium effect size). In other words, although the bilingual children scored lower on reading comprehension and a very rough measure of vocabulary size, they did not show differences in their associative preferences.

Children’s association preferences and reading proficiency (RQ3)

To study whether individual differences in preference for certain associative types are related to reading comprehension, we extracted the random intercepts for each participant from a model including only the fixed general intercept, and random intercepts for item, participant, and school. These individual intercepts can be interpreted as a score for the child’s tendency to produce an association of the semantic type in question.

To investigate whether there are any differences between initial and later associations in their relationship to reading comprehension, we extracted the association preference scores from the full data set, that is with responses 1, 2, and 3, and from two reduced data sets including only first responses, on one hand, and responses 2 and 3, on the other hand. These individual association scores were entered into a series of mixed-effects regression models as continuous predictors for the reading comprehension scores, with school as a random intercept. To compare the predictive value of the later responses relative to the initial responses, we estimated models including the joint score for response 1, 2, and 3, two models including the separate scores for response 1 and responses 2 and 3, and a model including both separate scores.

The preferences for situational and subjective associations did not have any effect on the reading comprehension scores, regardless of whether the full set of associations or either of the subsets of initial or later associations were considered, and therefore these will not be discussed any further here. However, the taxonomic and feature associations did show an effect in each or some of the analyses, which we will discuss in the following text. The models for the taxonomic association scores are presented in Table 7.

Comparing the models for the full set and subsets of associations, an interesting effect of the different response positions becomes apparent. The taxonomic score including all response positions (model 1) has a significant positive effect on the reading comprehension scores: $\beta = 2.854$ (0.892), $t = 3.20$. When children's log odds of associating taxonomically compared to the whole group increased by 1 (i.e., their odds of producing taxonomic associations increased by a factor 2.718), their reading score increased by 2.854. In other words, having a log odds score of 0.35, that is, producing about 42% more taxonomic associations than the average child, results in a 1 point increase on the reading score, which is about 1/6 of the standard deviation for the reading scores.

When considered separately in models 2 and 3, the scores for the first response and those for the later ones show a significant positive effect: $\beta = 1.460$ (0.627), $t = 2.33$ and $\beta = 3.929$ (1.239), $t = 3.17$, respectively. However, when we enter both separate scores in the same analysis, model 4, the effect of the first response is no longer significant: $\beta = 0.800$ (0.675), $t = 1.19$. In this case, only the second and third responses are predictive of the reading scores, apparently including variance explained by the first response in model 1: $\beta = 3.290$ (1.343), $t = 2.45$. This corresponds to a log odds score of 0.30, or producing about 36% more second and third taxonomic responses compared to the group as a whole for a 1 point increase in reading score. We checked whether there was a significant interaction between the scores from response 1 and responses 2 and 3, which was not the case. This suggests that in principle, the children who provide more taxonomic links overall are better comprehenders than the ones who provide few of these relations. However, especially the children who manage to produce taxonomic associations as second and third responses, and thus presumably know more taxonomic links for individual words, are better comprehenders. This illustrates an interesting difference between initial and later associations.

For the feature associations, the same models were estimated, which are presented in Table 8. Here, we see that the combined score in model 1 does not have a significant effect ($\beta = -2.800$ (2.356), $t = -1.19$), nor does the score for the second and third responses in model 3 ($\beta = -0.305$ (1.824), $t = -0.17$). Only the separate score for the first response contributes significantly to the reading score, whether entered separately in model 2 ($\beta = -3.852$ (1.854), $t = -2.08$), or together with the second and third responses in model 4 ($\beta = -3.871$ (1.872), $t = -2.07$). The significant effect is negative, indicating that children who produced more feature associations as first responses, perform worse on the reading comprehension test. In model 2 for example, having a log odds score of 0.26, that is, producing about 30% more feature associations than the average child, was associated with a one point decrease in the reading score. Again, including an interaction between the scores for the initial and later responses did not improve the model.

Table 7. Mixed-effects model estimates for the effects of taxonomic association preference scores on reading comprehension, four separate analyses: all response positions together (1), first responses only (2), later responses only (3), and both first and later responses (4)

		1			2			3			4		
Random effects													
School	Intercept variance	2.10			2.21			1.46			1.72		
Residual		33.72			34.47			33.98			33.81		
Fixed effects		Estimate (se)	<i>t</i>	<i>p</i>									
Intercept		29.869 (0.661)	45.19	0.000	29.860 (0.674)	44.29	0.000	29.871 (0.597)	50.03	0.000	29.853 (0.624)	47.85	0.000
Taxonomic score responses 1, 2, 3		2.854 (0.892)	3.20	0.001									
Taxonomic score response 1					1.460 (0.627)	2.33	0.020				0.800 (0.675)	1.19	0.236
Taxonomic score responses 2, 3								3.929 (1.239)	3.17	0.002	3.290 (1.343)	2.45	0.014

Table 8. Mixed-effects model estimates for the effects of feature association preference scores on reading comprehension, four separate analyses: all response positions together (1), first responses only (2), later responses only (3), and both first and later responses (4)

		1			2			3			4		
Random effects													
School	Intercept variance	1.73			1.71			1.80			1.71		
Residual		35.32			34.83			35.53			34.99		
Fixed effects		Estimate (se)	<i>t</i>	<i>p</i>									
Intercept		29.911 (0.630)	47.45	0.000	29.914 (0.626)	47.76	0.000	29.905 (0.639)	46.81	0.000	29.914 (0.627)	47.69	0.000
Feature score responses 1, 2, 3		-2.800 (2.356)	-1.19	0.235									
Feature score response 1					-3.852 (1.854)	-2.08	0.038				-3.871 (1.872)	-2.07	0.039
Feature score responses 2, 3								-0.305 (1.824)	-0.17	0.867	0.148 (1.824)	0.08	0.936

A possible explanation is that the feature associations are in a complementary relation with the taxonomic associations. In this case, the children who do not produce many taxonomic associations may produce feature associations instead, the next category on the context-independent/context-dependent continuum. It seems that this points to a less well-developed semantic network, which negatively affects reading comprehension. This explanation is supported by the fact that taxonomic associations are especially prominent in the first response position, and it is the feature associations in this position that are negatively correlated with reading comprehension.⁴

Discussion

First, we will discuss our findings with respect to our research questions, focusing on the possible mediating effects of initial and later word associations, in other words comparing the effect of using single versus multiple word associations on the answers to our research questions. The results will also be related to other research findings, often based on adult participants, contrary to our younger participants. Before we draw any conclusions, some limitations of the study will be discussed.

Quantitative and qualitative word level differences between initial and later associations

In this study, we set out to examine how associative networks on the word level and the individual mental lexicon level are affected by the use of a single versus multiple word association task, in monolingual and bilingual minority children. On the word level, we expected to find more and more semantically diverse associations, based on the findings by De Deyne and Storms (2008a, 2008b), which was indeed borne out by the data. Regarding the number of associations, when including second and third associations, more than 1.5 times more nonunique associations were found per stimulus word on average. This means that second and third associations do not represent exclusively idiosyncratic semantic links, which are specific to the individuals producing them (see De Deyne et al., 2019 for similar conclusions). On the contrary, the later associations add useful information about the stimulus word's semantic network. Such information is especially relevant when word association norms are used as a control measure for word selection in experimental tasks, such as priming experiments. When association strength between word pairs needs to be zero or as close to zero as possible, using single word association norms means that word pairs may be included which are associated, even if it is in the second or third instance. Such additional links are especially relevant when a stimulus word has a very strong first associate, which may drown out other associations, for example in the case of *blood* – *red* (De Deyne & Storms, 2008b). In and of itself, this finding is a strong argument for using multiple instead of single word association data, even if that is currently not common practice.

As for the types of semantic links that are found in initial versus later associations, interesting differences were also observed. Recall that for adults, De Deyne and Storms (2008a) observed a clear decline in taxonomic responses in the second and third position, while feature, situational, and subjective associations became

more prominent. In our data, these patterns were mimicked by all categories except the feature category, which decreased across responses, similar to the taxonomic category. On the word level, this suggests that not only do second and third associations add more semantic links, these links are also qualitatively different. Again, this is highly relevant when word association norms are used as control measures. For example, when situational semantic links are studied and need to be controlled for association strength, using norms derived from single associations may miss a substantial amount of associative links that would be found in a multiple association task (e.g., Spätgens & Schoonen, 2018).

Single versus multiple associations in individual language users' mental lexicons

Differences between monolinguals and bilinguals

We also set out to test whether the multiple association task would be more suited than the single task to study differences in monolingual and bilingual associative behavior. Because previous studies have reported mixed results (e.g., Cremer et al., 2011; Fitzpatrick, 2006, 2007, 2009; Söderman, 1993), we hypothesized that potentially, there are very subtle differences between monolingual and bilingual speakers' semantic networks, which may turn up in secondary and tertiary associations if not in initial associations. However, language status was not predictive of any of the four types of semantic associations and did not interact with response position, suggesting that language status did not play a role in our sample's associative behavior.

Of course, not all bilinguals or L2 learners are alike; factors such as age of acquisition and amount of exposure will play an important role in any measure of bilingual language skills. It is therefore not surprising that studies with different populations find different results. Indeed, the bilingual children in this study were almost all born in the Netherlands, all started Dutch elementary school at age four, and the majority spoke Dutch at home in addition to their L1. As such, these bilingual learners have had extensive Dutch language exposure and may therefore have similar Dutch language knowledge compared to their monolingual counterparts. However, looking at the two other measures of language skill that we had at our disposal, the reading comprehension scores and the number of "other," that is null or nonclassifiable responses, we did find small but significant differences between the two groups, with monolingual children outperforming bilingual children. This suggests that there are at least subtle differences in language skill. However, these differences did not extend to the association task, meaning that for those words that the bilingual children seem to know, their associative behavior is similar to that of monolingual children.

Our findings mirror those obtained in the single word association task reported by Cremer et al. (2011), who did not find differences between their monolingual and bilingual learners in terms of production of associations in the various semantic categories. The subtle differences in preferences for different semantic categories we expected to turn up in our multiple task were not borne out by the data. A potential explanation may be that the L2 semantic network structure is more a property of the L2 learner's existing knowledge, than of the individual words being learned. As was proposed by Wolter (2006; see also Zareva, 2010), L2 learners may use the semantic

network from their L1 to structure their L2 knowledge. Perhaps especially in the case of bilinguals and advanced L2 learners, new words may be integrated into this structure fairly quickly, resulting in the apparent incongruity of the number of words known being lower in bilinguals while the association behavior is similar. Furthermore, the stimulus words were all concrete concepts and high-frequency words, selected to be known to all children as much as possible. It may be the case that with more abstract and less frequent words, the monolingual and bilingual children would have shown different associative behavior, an effect that has also been observed in L1 adults who responded to frequent and infrequent words with different types of associations (Stolz & Tiffany, 1972).

Single versus multiple associations and reading comprehension

Finally, to assess how the dense semantic networks arising from multiple association tasks may relate to other language skills, we looked into their effect on reading comprehension scores. We found that situational and subjective associations were not related to reading comprehension, irrespective of response position. This was expected based on previous studies indicating that it is context-independent knowledge and especially taxonomic knowledge that may be predictive of reading comprehension skill (Bonnotte & Casalis, 2010; Cremer & Schoonen, 2013; Nation & Snowling, 1999; Ouellette, 2006). Taxonomic and feature associations were related to the reading comprehension scores, and each showed interesting differences between initial and later responses.

The taxonomic associations' relation to reading comprehension scores showed an interesting difference between initial and later responses. Examining the full set, the first response set and the second and third response set all resulted in positive effects on the reading comprehension scores, suggesting that overall, knowledge of taxonomic relations is associated with better reading comprehension. However, when the two position variables are entered together (model 4 in Table 7), the effect of the first response set scores disappeared, and only the taxonomic scores of the second and third responses had a significant positive effect on the reading scores. Regarding the relation between reading comprehension and knowledge of semantic relations, this suggests that especially children who have a stronger preference for taxonomic links and can produce not only one, but also more taxonomic relations, are better at reading comprehension. It is interesting to see that even this spontaneous and unconstrained vocabulary task can be related to reading comprehension skill. However, the effects appear to be fairly modest, with children having to produce about 42% more primary, secondary, and tertiary taxonomic responses compared to the group as a whole to achieve a 1 point or about 1/6 standard deviation increase in their reading score.

The initial but not the secondary and tertiary feature associations were also related to the reading scores, but negatively so. This seems to go against findings from other studies in which defining characteristics were included among other semantic aspects in measures of vocabulary depth, which were positively related to reading comprehension (Cremer & Schoonen, 2013; Tannenbaum et al., 2006). As was already mentioned in the "Results" section, the presence of this negative relation despite the fact that feature associations are fairly context-independent

in nature may be the result of a complementary relation between taxonomic and feature associations. This hypothesis is strengthened by the observation that the feature associations showed the same decline across responses as the taxonomic associations, a characteristic of context-independent knowledge, which in adults is reserved for the taxonomic responses only. In the development toward a more context-independent network structure, feature associations may form a stepping-stone. Our data suggest that children who are still prone to produce these feature associations early, that is whose semantic network is not as far developed yet, are also those who perform worse in terms of reading comprehension. This reinforces the idea that especially the most context-independent knowledge is associated with good performance in reading comprehension. The difference between our findings and those obtained by Cremer and Schoonen and Tannenbaum et al. is probably also related to the fact that, in our study, the participants were not explicitly asked to produce or select defining characteristics. The preferred routes in the children's semantic networks that are spontaneously accessed in the association task appear to bear a different relation to reading comprehension.

In these analyses of reading comprehension, using only single responses would have provided us with seemingly similar outcomes, namely a positive effect of taxonomic associations and a negative effect of feature associations. However, the fact that we gathered multiple associations allowed us to discover a different and more complex relation between these two types of semantic knowledge and reading comprehension. The shape of the effects provide interesting insights into the relation between the development of the semantic network structure and reading comprehension.

Differences between children and adults

Our findings also have implications for our understanding of language users' semantic networks in general, and more specifically children's mental lexicons. Firstly, as was already found for adults by De Deyne and Verheyen (2015), semantic relations in the mental lexicon are much more context-dependent than previously thought, and this is also the case for the children in our study. Only a minority of associations is taxonomic (10.5% across responses in our study), while most associations are situational in nature (48.9%). As Borghi and Caramelli (2003) have pointed out, it makes sense that the number of context-dependent associations would exceed taxonomic associations, given that they include a far larger set of possible ties, including locations, actions, and a variety of types of entities that may co-occur together in a host of different contexts.

Secondly, we see some interesting similarities and differences between the behavior of our 10- to 11-year-old participants and the adult data presented by De Deyne and Storms (2008a). Starting with the taxonomic associations, it is interesting to see that children exhibit the same decreasing pattern across responses, showing that as in adults, if a taxonomic association is produced, it is most likely produced as a first response. In terms of absolute numbers however, we do see that overall, the children produce fewer taxonomic associations: 10.5% in our study, compared to 19.6%⁵ for the adults in De Deyne and Storms's work. We can conclude that although the children in this age group possess fewer taxonomic links, the available taxonomic

knowledge is already organized similarly to that of adults because it is more prone to being produced early as well.

Another interesting difference shows up in the feature category, the only category in which the association pattern across responses is different from that observed in adults. Mirroring the taxonomic associations, children produce more feature associations in the first instance, while across the second and third responses, this number drops. We would like to tentatively suggest that this also relates to the development from context-dependent to context-independent knowledge: Features may be the “next best thing” on the way to a more context-independently organized semantic network. When context-independent knowledge is not yet fully fledged, feature relations may be more available to adhere to the apparent preference for producing context-independent knowledge earlier. In terms of situational and subjective associations, the children in our study show similar patterns to the adults in De Deyne and Storms (2008a).

On the individual mental lexicon level, the multiple word association data thus provide us with additional information to compare associative behavior in different age groups and examine the process of children’s mental lexicons developing toward the adult standard. In terms of the development from a main focus on context-dependent knowledge toward the extension to context-independent knowledge (c.f. Elbers et al., 1993; K. Nelson, 1977, 1982, 1985, 1991), the children in our age group seem to be fairly close to the adult behavior, but are still expanding their context-independent knowledge and preferences in various ways. An interesting path for future research would be to compare more age groups in the same way and see how not only the proportions of answers in different categories develop during childhood, but also across response positions. A first step could be to sample from the De Deyne and Storms (2008a) word association corpus the responses to the stimulus words we used and to classify these responses according to the coding scheme used in this study (Table 2). This could provide an interesting comparison between children and adults with regard to associations with concrete, high-frequency nouns.

Limitations

There are a few limitations to the present study, some of which are due to its exploratory nature, that should be taken into account in the interpretation of the results and for future studies. Firstly, the number of words and participants included was relatively small—that is each child responded to 20 words, and each word was responded to by (at least) 50 children—compared to the enormous association databases that exist. This means that, for example, the number of additional associations that we found in the second and third answer sets should be interpreted with some caution. However, it is not the case that, because of this, some words only appeared in the second and third instance that would have turned up as first associations in a larger sample: De Deyne and Storms (2008b) found similar set sizes for the first, second, and third responses, also ranging from 11 to 14 on average, but a much larger set size for the collapsed set, namely 31 nonunique responses compared to 18 in our sample. If anything, our small sample might therefore lead us to

underestimate the additional variability that can be detected using a multiple task. The difference between the collapsed sets of adults and children suggests that there was more overlap between the three sets in children than in adults.

The number of items per child may also affect the validity of the association scores for the different semantic types because more accurate associative profiles could be gained from larger samples of items. We controlled for this as much as possible by including random intercepts for items in our analyses, meaning that if certain items were more conducive to certain types of associations, these differences were taken into account in the calculation of the personal association scores.

As was mentioned earlier, the stimuli were confined to nouns only, which limits the generalizability of our findings (see for word class effects, Nissen & Henriksen, 2006), and they were selected to be highly frequent, concrete concepts, which may be a reason why we did not find any differences between the monolingual and bilingual children. It is also likely that with more difficult stimuli, a different effect of associative behavior on the reading task might have emerged. This would be an interesting question for future research. In this study we did not pretest the children's knowledge of these stimulus words not to influence the word associations. However, in future studies a posttest of the children's knowledge of the stimulus words could be considered. Possible differences between the stimulus words in this study, for example due to word-frequency differences, were statistically accounted for in the analyses by including random intercepts for stimulus words. Furthermore, all word associations were collected with a multiple word association task and the results for the single word association task were "simulated" by analyzing the first associations only, instead of collecting word associations with a separate single word association task. However, there is no reason to assume that word associations in a single word association task will be different from the first associations in a multiple word association task. In this respect, we followed the same procedures as De Deyne and Storms (2008a, 2008b).

The analysis of word associations is guided, but also limited by the classification system. The subdivision and interpretations of main categories worked well in this study given the coder agreement, but some of the boundaries between associations could give rise to discussion. For example, subjective appreciations that are shared by many people could be interpreted as a feature (*Mona Lisa – fascinating*), and thus should possibly be valued differently from very personal appreciations, such as *cheese – ew*. This also relates to the questions whether our interpretation of the semantic relations holds for all participants. A superordinate as a taxonomic relation might represent highly valued decontextualized word knowledge, whereas in other cases it might be the outcome of a participants best guess (*chihuahua – animal*). Replication of this study with possibly more qualitative analyses of subcategories of word associations and retrospective interviews could shed more light on these kinds of issues. Finally, the inclusion of a standardized vocabulary size measure to control for vocabulary size effects on the reading comprehension scores and to be able to explicitly distinguish vocabulary size from structure of the semantic network effects, would have strengthened our conclusions regarding the relation between associative behavior and reading comprehension. We investigated highly frequent nouns to counter possible vocabulary size effects in the association behavior of the children, and we considered the "other" associations as a rough measure of

vocabulary size, but it was unsuitable for inclusion in the analyses as such because more than a third of the children did not produce any “other” associations.

Conclusions

Despite the limitations discussed in the preceding text, our study has shown that the use of the multiple association task leads to interesting insights on the structure of the mental lexicon. Similarly to the larger adult studies by De Deyne and Storms (2008a, 2008b), the task resulted in larger and qualitatively different semantic networks, which is highly relevant for our interpretation and use of association norms. The differences and similarities between adults’ and children’s association patterns across responses provide valuable information on the development of the semantic network structure. An interesting topic for future research would be to examine these patterns in more different age groups and with a larger variety of stimulus words such as less frequent or more abstract words, to map the development of semantic links alongside vocabulary growth.

Expanding to other learner groups and linguistic skills, we found that young monolinguals and bilinguals show similar associative behavior, while differences in reading comprehension skill do appear to be affected by associative preferences in various ways (see Tables 7 and 8). Again, the initial and later associations appear to foreground different types of knowledge, relating to reading comprehension skill in different ways. These results open up a fresh perspective for research on the use of associations as measures of vocabulary knowledge, or at least the preeminence of certain relations in the semantic network. As Fitzpatrick (2012) has noted, the large number of studies devoted to this issue remain inconclusive, and our results suggest the multiple association task could be a useful approach. The fact that the results from this unconstrained task are related to such a complex skill as reading comprehension, are also encouraging for the more detailed study of the role of knowledge of semantic relations in reading comprehension. Examining this relation further using more focused tasks such as priming experiments (e.g., Cremer, 2013; Nation & Snowling, 1999; Spätgens & Schoonen, 2018) or self-paced reading tasks involving different types of semantic relations, will be an interesting avenue for future research.

Acknowledgement. We would like to thank Jan Hulstijn for his thoughtful comments and suggestions regarding this research, Paul Boersma for statistical advice, and the anonymous reviewers of our paper for their valuable suggestions. Furthermore we would like to thank the students and teachers for their cooperation.

Notes

1. The authors use the term *thematic*, which relates to the *thematic-taxonomic* distinction (e.g., Jones & Golonka, 2012; Markman & Hutchinson, 1984; Mirman & Graziano, 2011). In this article, we will refer to these categories as occupying spaces on the context-dependent/context-independent continuum, which will be discussed in more detail below and in the Method section.
2. There are a few terminological differences: we prefer the term *features* for their *entity* category, and *subjective* for their *introspective* category.
3. We also considered inclusion of response position as a categorical variable, but this led to convergence issues in the models which include random slopes for response position. We compared the models without

random slopes with the categorical and continuous response position variables, and these showed the same patterns across response position for taxonomic, feature and situational associations. Only for subjective associations, there was no increase between response 2 and response 3 when response position was included as a categorical variable.

4. Note that it is not possible to confirm or reject this hypothesis by calculating the correlation between the taxonomic and feature scores. Since we created the associative variables by recoding each association into four binary types, they are by definition negatively correlated.

5. We recalculated the percentages provided by De Deyne and Storms (2008a) to reflect the percentage of taxonomic associations among the semantic categories, i.e. by leaving the lexical associations out in the same way we did in our study. Including lexical associations, they arrived at 18.9% taxonomic associations.

REFERENCES

- Aarnoutse, C., & Kapinga, T. (2006). *Begrijpend lezen 678 [Reading comprehension grades 456]*. 678 Onderwijsadvisering.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed.). Malden, MA: Wiley-Blackwell.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Blewitt, P., & Toppino, T. C. (1991). The development of taxonomic structure in lexical memory. *Journal of Experimental Child Psychology*, *51*, 296–319.
- Bonnotte, I., & Casalis, S. (2010). Semantic priming in French children with varying comprehension skills. *European Journal of Developmental Psychology*, *7*, 309–328.
- Borghi, A. M., & Caramelli, N. (2003). Situation bounded conceptual organization in children: From action to spatial relations. *Cognitive Development*, *18*, 49–60.
- Cremer, M. (2013). *Accessing word meaning: Semantic word knowledge and reading comprehension in Dutch monolingual and bilingual fifth-graders* (Doctoral dissertation University of Amsterdam). LOT.
- 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*, 187–204.
- Cremer, M., & Schoonen, R. (2013). The role of accessibility of semantic word knowledge in monolingual and bilingual fifth-grade reading. *Applied Psycholinguistics*, *34*, 1195–1217.
- De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods*, *45*, 480–498.
- 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*, 987–1006. <https://doi.org/10.3758/s13428-018-1115-7>
- De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*, 213–231.
- De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*, 198–205.
- De Deyne, S., & Verheyen, S. (2015). Using network clustering to uncover the taxonomic and thematic structure of the mental lexicon. Paper presented at the Word Structure and Word Usage. Proceedings of the NetWords Final Conference, Pisa (pp. 172–176). <http://ceur-ws.org>
- Elbers, L., van Loon-Vervoorn, A., & van Helden-Lankhaar, M. (1993). “Contrastive usage” and the development of lexical organization and innovative labelling. In M. Verrips & F. Wijnen (Eds.), *Amsterdam series in child language development 1: The acquisition of Dutch* (pp. 5–19). University of Amsterdam.
- Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EuroSLA Yearbook*, *6*, 121–145.
- Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*, 319–331.

- Fitzpatrick, T.** (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners* (pp. 38–52). Multilingual Matters.
- Fitzpatrick, T.** (2012). Word associations. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 6193–6199). Wiley-Blackwell.
- Fitzpatrick, T., & Izura, C.** (2011). Word association in L1 and L2. *Studies in Second Language Acquisition*, *33*, 373–398.
- Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. J.** (2013). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, *36*, 23–50.
- Jones, L. L., & Golonka, S.** (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience*, *6*, 1–17.
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J.** (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literacy studies* (pp. 153–165). University Press.
- Lin, E. L., & Murphy, G. L.** (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*, 3–28.
- Linck, J. A., & Cunnings, I.** (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207.
- Markman, E. M., & Hutchinson, J. E.** (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*, 1–27.
- McEvoy, C. L., & Nelson, D. L.** (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, *95*, 581–634.
- McNamara, T.** (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.
- McRae, K., & Cree, G. S.** (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. Humphreys (Eds.), *Category-specificity in mind and brain* (pp. 211–249). Psychology Press.
- Mirman, D., & Graziano, K. M.** (2011). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General*, *141*, 601–609.
- Nation, K., & Snowling, M. J.** (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, *70*, B1–B13.
- Nelson, D. L., McEvoy, C. L., & Dennis, S.** (2000). What is free association and what does it measure? *Memory & Cognition*, *28*, 887–899.
- 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*, 402–407.
- Nelson, K.** (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, *84*, 93–116.
- Nelson, K.** (1982). The syntagmatics and paradigmatics of conceptual development. In S. Kuczaj (Ed.), *Language development volume 2: Language, thought and culture* (pp. 335–364). Erlbaum.
- Nelson, K.** (1985). *Making sense: The acquisition of shared meaning*. Academic Press.
- Nelson, K.** (1991). Concepts of meaning in language development. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 89–115). Erlbaum.
- Nelson, K.** (2007). *Young minds in social worlds: Experience, meaning, and memory*. Harvard University Press.
- Nissen, H. B., & Henriksen, B.** (2006). Word class Influence on word association test results. *International Journal of Applied Linguistics*, *16*, 389–408.
- Ouellette, G. P.** (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*, 554–566.
- Petrey, S.** (1977). Word associations and the development of lexical memory. *Cognition*, *5*, 57–71.
- R Core Team.** (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W.** (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*, 83–119.
- Schmitt, N.** (1998). Quantifying word association responses: What is native-like? *System*, *26*, 389–401.

- Schoonen, R., & Verhallen, M.** (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, *25*, 211–236.
- Schrooten, W., & Vermeer, A.** (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen* [Words in primary education. 15,000 words presented to pupils]. Tilburg University Press.
- Söderman, T.** (1993). Word associations of foreign language learners and native speakers—different response types and their relevance to lexical development. In B. Hammarberg (Ed.), *Problem, process, product in language learning: Papers from the Stockholm-Abo conference* (pp. 157–169). Stockholm University.
- Spätgens, T., & Schoonen, R.** (2018). The semantic network, lexical access and reading comprehension in monolingual and bilingual children: An individual differences study. *Applied Psycholinguistics*, *39*, 225–256.
- Stolz, W. S., & Tiffany, J.** (1972). The production of “child-like” word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behavior*, *11*, 38–46.
- Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K.** (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, *10*, 381–398.
- Verhallen, M., & Schoonen, R.** (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, *14*, 344–363.
- Wolter, B.** (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition*, *23*, 41–69.
- Wolter, B.** (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*, 741–747.
- Wu, L., & Barsalou, L. W.** (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.
- Zareva, A.** (2007). Structure of the second language mental lexicon: How does it compare to native speakers’ lexical organization? *Second Language Research*, *23*, 123–153.
- Zareva, A.** (2010). Multicompetence and L2 users’ associative links: Being unlike nativelike. *International Journal of Applied Linguistics*, *20*, 2–22.
- Zareva, A.** (2011). Effects of lexical class and word frequency on the L1 and L2 English-based mental lexicon. *Journal of Language Teaching and Learning*, *1*, 1–17.
- Zareva, A., & Wolter, B.** (2012). The “promise” of three methods of word association analysis to L2 lexical research. *Second Language Research*, *28*, 41–67.

Appendix 1. Stimulus Words and Translation Equivalents Association Task

List 1		List 2		List 3		List 4	
bruid	<i>bride</i>	uil	<i>owl</i>	karper	<i>carp</i>	hamer	<i>hammer</i>
hert	<i>deer</i>	dief	<i>thief</i>	kaas	<i>cheese</i>	markt	<i>market</i>
duin	<i>dune</i>	woestijn	<i>desert</i>	pop	<i>doll</i>	olifant	<i>elephant</i>
rok	<i>skirt</i>	sloot	<i>ditch</i>	vlinder	<i>butterfly</i>	camping	<i>camp site</i>
krokodil	<i>crocodile</i>	eiland	<i>island</i>	wieg	<i>cot</i>	trui	<i>sweater</i>
kapper	<i>hairdresser</i>	potlood	<i>pencil</i>	visser	<i>fisherman</i>	schilder	<i>painter</i>
kasteel	<i>castle</i>	appel	<i>apple</i>	jas	<i>coat</i>	taxi	<i>taxi</i>
eend	<i>duck</i>	jager	<i>hunter</i>	keuken	<i>kitchen</i>	kat	<i>cat</i>
winkel	<i>shop</i>	ridder	<i>knight</i>	baby	<i>baby</i>	kok	<i>cook</i>
stoel	<i>chair</i>	bakkerij	<i>bakery</i>	regen	<i>rain</i>	zee	<i>sea</i>

(Continued)

(Continued)

List 1		List 2		List 3		List 4	
gitaar	<i>guitar</i>	trompet	<i>trumpet</i>	hond	<i>dog</i>	gang	<i>hallway</i>
hamster	<i>hamster</i>	konijn	<i>rabbit</i>	piloot	<i>pilot</i>	zebra	<i>zebra</i>
juf	<i>teacher</i>	chocola	<i>chocolate</i>	toerist	<i>tourist</i>	jam	<i>jam</i>
komkommer	<i>cucumber</i>	koe	<i>cow</i>	tuin	<i>garden</i>	station	<i>station</i>
oceaan	<i>ocean</i>	bos	<i>forest</i>	banaan	<i>banana</i>	slak	<i>snail</i>
kast	<i>cupboard</i>	kapitein	<i>captain</i>	restaurant	<i>restaurant</i>	boer	<i>farmer</i>
cola	<i>coke</i>	heks	<i>witch</i>	stal	<i>stable</i>	prinses	<i>princess</i>
dokter	<i>doctor</i>	fiets	<i>bike</i>	kanarie	<i>canary</i>	auto	<i>car</i>
chauffeur	<i>driver</i>	krekel	<i>cricket</i>	koning	<i>king</i>	egel	<i>hedgehog</i>
leeuw	<i>lion</i>	zwaan	<i>swan</i>	mus	<i>sparrow</i>	indiaan	<i>indian</i>

Cite this article: Spätgens, T. and 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, 1141–1169. <https://doi.org/10.1017/S0142716420000430>