Accessing word meaning: Semantic word knowledge and reading comprehension in Dutch monolingual and bilingual fifth-graders
Cremer, M.

Citation for published version (APA):
Chapter 3

Do word associations assess word knowledge? A comparison of monolingual and bilingual, child and adult word associations

Differences in word associations between Dutch monolingual and bilingual minority children can reflect differences in how well seemingly familiar words are known. In this chapter, we will compare monolingual and bilingual, child and adult free word associations. As briefly explained in section 2.5, comparing language and age groups in this task allows for an indication of cognitive and lexical differences in word association behaviour (research questions 1 and 2). The chapter starts with a brief introduction to the word association paradigm and relevant earlier studies (section 3.1). It introduces the idea of semantic networks (section 3.1.2) and the difference between contextual and semantic knowledge (section 3.1.2). Section 3.1.3 on categorization and elicitation procedures is followed by a discussion of the research methodology used (section 3.2). The results are presented in section 3.3 and discussed in section 3.4.

1 This chapter is an adapted version of 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. The study reported on was set up as a preparatory study to generate association norms for the studies in Chapters 4 and 5. Part of the research was carried out as course work by the second and third authors under the supervision of Prof. dr. Jan Hulstijn.
3.1 Background

Word knowledge develops from experiences with the world around us. Apart from learning labels for items or objects (a complicated process in itself), a language learner has to learn the exact extension of a word’s meaning (i.e. which other objects can be called the same name) and has to see in what way words fit together; in other words, he or she has to discover relations between words and thus build up a semantic network (Aitchison, 2003; Henriksen, 1999). Understanding relations between words means having a rich and a densely interconnected mental lexicon, which is considered an important feature of developing language proficiency. Word knowledge is often described in terms of a lexical and a semantic level (Kroll, Michael, Tokowicz, & Dufour, 2002; Potter, 1979; Snodgrass, 1984). The lexical level consists of the words in all of a speaker’s languages; the underlying concepts are stored at the abstract, semantic level (Kroll and Stewart, 1994). Knowledge of relations between words can be studied through a word association task, which asks people to respond with the first word that comes to mind for a number of stimulus words. Such associative behavior is presumed to reflect how words and concepts are structured and interrelated in the mind (Deese, 1965; Szalay & Deese, 1978). The word association paradigm has long been used in psychology to screen patients for possible idiosyncratic associations (Lukavsky, 2004; Mohr, Graves, Gianotti, Pizzagalli, & Brugger, 2001; Slechta, 2001). In L1 studies, it has been used to investigate children’s lexical and semantic development (Cronin, 2002). Schmitt (1998) claims that word associations reflect the strongest mental connections between words in the mind. They result from experience with the associated words. Somewhat more recently, word association tasks have been used to compare the organizational principles of the L1 and L2 lexicon (Fitzpatrick, 2006; Sheng, McGregor, & Marian, 2006; Zareva, 2007), and they can also be seen as reflecting lexical and conceptual processing (Van Hell & De Groot, 1998). Differences in types of association responses are sometimes taken as an indicator of language proficiency (Zareva, 2007), but proficiency level has not consistently been found to predict response patterns (Kruse, Pankhurst, & Sharwood Smith, 1987). So far, studies have failed to find consistent behavior patterns. Some have found more or
less stable association patterns for native speakers showing that responses become more meaning-based after late childhood. More varied responses have been found for nonnative speakers (Meara, 1983). It remains uncertain to what extent L1 or L2 speakers behave as homogeneous groups, or, how differently dispersed their responses are across different response categories.

In (young) children lexical and conceptual development are intertwined. As the first language is acquired, young children learn words (lexical labels for concepts in the world) and they learn concepts (what concepts in the world mean, how they relate to other concepts, where concept boundaries lie, and so on) (Verhallen, 1994). When we compare monolingual L1 and bilingual L2 development in a given language, it is not always clear whether differences found have a lexical and/or conceptual origin (Verhallen & Schoonen, 1993). Some of the L2 ‘problems’ may not occur in bilingual children’s L1, suggesting L2 language difficulties; others may show in their L1 too, suggesting broader, conceptual problems (Verhallen, Özdemir, Yuksel, & Schoonen, 1999; Verhallen & Schoonen, 1998).

In the present study, we investigate differences in word knowledge of child and adult first and second language speakers of Dutch; we compare their Dutch word associations. A combined analysis of child and adult association behavior allows us – to some extent – to tease apart language background effects (monolingual versus bilingual) and cognitive development effects (child versus adult). To prevent a possible influence from respondents with different L1 backgrounds, we focus on a ‘homogeneous’ bilingual group, that is, Turkish L1 speakers with Dutch as an L2.

3.1.1 Word knowledge: lexical and semantic network
Researchers have distinguished different dimensions of lexical knowledge, ranging from knowledge of a word’s form, position, semantic network, collocations and associated words, to polysemy, formal definitions and its receptive and productive use, to name but a few (Cronbach, 1942; Henriksen, 1999; Nation, 2001; Richards, 1976; Wescie & Paribakht, 1996). Depth of word knowledge can comprise all of the
aforementioned aspects of word knowledge. Read (2004) focuses on word meaning and suggests that we can understand depth as learners’ ability to distinguish semantically related words and, more generally, their knowledge of how individual words are linked to each other.

Most models of multilingual word processing distinguish between a lexical and a semantic level. In the Bilingual Interactive Activation Plus (BIA+) model (Dijkstra and Van Heuven 1998, 2002), lexical (and sublexical) levels involve orthographic and phonological information processing and a semantic level contains semantic/conceptual representations. Connections at and between the levels lead to dynamic activation spread between levels during language processing. The Revised Hierarchical Model (RHM), developed by Kroll and Stewart (1994), also contains a lexical and a conceptual store. Both lexical and semantic representations are shared across languages (Kroll et al. 2002); that is to say, a speaker’s languages ‘tap into’ one common lexical store and one common conceptual store. This means that words from a speaker’s languages are linked in terms of, for example, orthography and phonology but also in terms of meaning and conceptual features. These connections underlie word processing. Even though both conceptual and lexical links are active in semantic memory, the strength of such links differs as a function of (L2) fluency (Dijkstra, 2007). Whereas in the RHM L1 word use is conceptually mediated from the start, L2 learners acquire lexical links between L2 and L1 before they are able to conceptually mediate L2 word use. Thus, for a beginning learner an L2 word will most likely be translated into its L1 word and with that, activate its semantic/conceptual underpinnings.

Quality of word knowledge is often seen as a feature of the semantic network (Aitchison, 2003; Henriksen, 1999). Assuming that words get their meaning from their relations to other words, acquiring more meanings and related words leads to a deepening of word knowledge: the network becomes denser and highly interconnected. Henriksen states that understanding the relations among the items is a prerequisite for a more precise understanding of each individual item (Henriksen, 1999). Similarly, Meara and Wolter emphasize the interdependence of vocabulary size and depth and organization (2004). The idea of networks also grounds Read’s
Word Associates Format (1993) in which test takers have to find relations between a stimulus word and a given set of other words. This approach is also the focus of the multiple-choice, word association test format for children (WAT) developed by Schoonen and Verhallen (2008).

3.1.2 From context dependent to meaning based

Children build their lexicon from scratch by adding and reorganizing their lexical semantic network, a never-ending process. Links between words and concepts, from one or several languages, change as a function of salience, exposure to various uses, fluency and such. Words that co-occur in the language encountered will be stored together. Thus, associative relations between words are based on the likelihood of two words co-occurring in the language (Wettler, Rapp, & Sedlmeier, 2005). As language is learnt, word knowledge is gradually abstracted and refined from children’s functional, idiosyncratic knowledge (Nelson, 1977, 1982) into more meaning-based, semantic connections. As mentioned in section 2.1.2, Petrey (1977) describes this knowledge development as a dramatic shift from episodic to semantic. Using responses from Entwisle’s (1966) free word association data, she shows that “[e]arly associations ... depend not on semantic content but on episodic experience” (1966: 65). Very young children respond with contextually related words (example from a medical context: examine – needle) rather than with inherent semantic associations (examine – look (at), check), which become common after third grade (Petrey, 1977). Episodic kindergarten responses to dark, such as bed and sleep (both indirectly related to dark), disappear entirely by third grade, when light has become a common response. This illustrates the abstraction to more context-independent associations. This semantic development was also shown by Verhallen and Schoonen (1993), who set up an interview task and found that nine-year-old children mentioned significantly more context-dependent meaning aspects than eleven-year-olds, who mentioned relatively more abstract, context-independent meaning aspects.

Similar differences in quality of knowledge have been found for monolingual and bilingual children for, apparently, known words. This is remarkable, as all young children seem to acquire word knowledge rapidly, even
children who have been raised in another language than the language used at school. Verhallen (1994, Verhallen & Schoonen, 1993) has shown that children with Dutch as a second language not only have less semantic knowledge of familiar words compared to their monolingual peers, their semantic knowledge is also more context-dependent and more subjective. Most of these children who speak Dutch as an L2 have an immigration background and do not speak Dutch at home. This causes a mismatch between home environment and school setting in terms of language use. The lacking school achievements of these bilingual minority children may lie at the lexical (word) and/or at the semantic (conceptual) level. Because of this mismatch between home and school, children with Dutch as an L2 may have a delay in semantic development and abstraction in their L1 as well: their mental lexicon may not (yet) contain (strong) connections between semantically related words but rather between subjectively, context-dependently related words. This is a feature of the semantic level and, consequently, will also show in their L1. Verhallen and colleagues (Verhallen et al., 1999; Verhallen & Schoonen, 1998) indeed found that children’s semantic knowledge about (Turkish) L1 words was even more limited than their relatively limited semantic knowledge in L2.

Alternatively, if word knowledge displayed in association tasks is only explained by linguistic contiguities and associative learning (Wettler et al., 2005), Dutch bilingual minority learners may not have had enough opportunities for associative learning, because of less exposure.

Insights into the organization and development of language learners’ mental lexicons serve both assessment and research purposes. If we consider word association tasks as a kind of language test, we should bear in mind that association responses reveal subconscious connections in a constantly changing mental word base; connections between words and concepts are constantly being added, strengthened and weakened. Free word associations may therefore be an indicator of the language learners’ current states of mind, rather than a deliberate and reliable expression of their word knowledge.
3.1.3 Assessing word associations: categorization and elicitation

When using word associations to assess word knowledge, the elicitation procedure used, the selection of stimulus words, and the classification of responses (see Fitzpatrick, 2006) have to be approached carefully.

The different kinds of association given by participants can be categorized in different ways. Using a meaning-based approach, De Groot (1980) distinguishes objective versus subjective semantic relations. Objective semantic relations hold for all users of the language; examples include superordinates, subordinates and synonyms. Subjective semantic relations hold for individuals; an example could be vase-mother. De Groot further distinguishes phonologically related responses, word extensions and idiomatic responses. Another, commonly used, way of categorizing association responses is in terms of the syntagmatic-paradigmatic distinction. This distinction is mainly based on word class. Paradigmatic associations are those in which stimulus word and response word belong to the same word class (e.g., noun–noun); syntagmatic associations have a word class other than the stimulus word (e.g., noun–verb). Often the substitutability of paradigmatic words is contrasted with syntagmatic words occurring together with stimulus words in discourse (Cronin, 2002). The paradigmatic-syntagmatic approach to categorizing associations has received criticism since the 1970s. Petrey (1977) labels the categories paradigmatic and syntagmatic inadequate for describing changes with age in word associations, “because [they] can designate nothing but syntactic properties” (1977: 69). She prefers the terms episodic and semantic (see section 2.1.2). When given the word add, children gave the episodic responses flour (syntagmatic) and cook (paradigmatic); both are related in the same way to add (they come from the same situation) but would be misclassified as opposites by S-P terminology. She adds that, “syntagmatic responses properly so-called are highly problematic because either semantic or episodic storage could produce them” (1977: 66). Nissen and Henriksen (2006) criticize the syntagmatic-paradigmatic shift as they find a surprisingly strong native speaker preference for syntagmatic responses. According to Fitzpatrick the syntagmatic-paradigmatic distinction imposes artificial constraints on the exploration of response types, as their contents are difficult to define. Many studies
do indeed add an ‘other’ category (sometimes included in the phonological) for
difficult to classify responses, but this might be inherent to any classification
scheme. In her categorization scheme Fitzpatrick uses four main categories:
meaning-based, position-based, form-based and erratic associations. Each main
category is divided into subcategories, detailing the type of relation between
stimulus and response word. Fitzpatrick advises the use of subcategories, as only
general main categories do not allow a very precise method of categorization (2006).
In addition to categorizing the associations, it is insightful to know how dispersed
the responses are across the association categories, in other words, do (groups of)
speakers give many different associations or do they have preferences? Since L2
associations are often claimed to be less predictable than L1 associations, L1 and L2
speakers may show differently dispersed responses. L1 and L2 associations may for
example cluster in different categories.

Different studies use different procedures of eliciting word associations and
findings are diverse. Schmitt (1998) advocates eliciting multiple responses. He uses
word association responses to determine whether a participant’s associative network
is native-like. Although the assumption is that the participant responds with the
strongest associate, Schmitt questions whether this will always be the case (1998:
391). He pleads for multiple responses to one stimulus word as this might better
capture the richness of a subject’s associative network which can then be compared
to a native speaker’s network. Others prefer the single-response procedure as it
provides a reliable index of a word’s strongest associates (Nelson, McEvoy, &
Dennis, 2000). Nelson and colleagues found that second responses added new but
weak items to the set, and, when the primary associate was not produced as the first
response, it tended not to be produced on the second. Furthermore, the first response
may act as a stimulus for the second response thus eliciting a chain of responses and
“such responses are not independent of the first” (2000: 891). Therefore, the present
study adopts the single-response paradigm.

Finally, the kind of stimulus words used, influence the associations elicited.
Many researchers found an influence of word class on response types (Cronin, 2002;
Nissen & Henriksen, 2006). Nouns, for example, elicit paradigmatic responses more
often than verbs or adjectives do. There are inherent differences between word classes that need to be taken into account when comparing studies. Also word frequency affects the probability of response types (Cronin, 2002). Highly frequent words tend to elicit predictable words, especially among native speakers. This homogeneity of native speaker responses, compared to those of L2 speakers, tends to disappear when words are less frequent (Fitzpatrick, 2007). De Groot (1989) looked at imageability of words and found that concrete words elicit more homogeneous responses than abstract words. In our study, the list of stimulus words consisted mostly of nouns; words had varying frequencies but all occurred in a Dutch corpus of words used in primary school materials (Schrooten & Vermeer, 1994).

Differences in word knowledge between speakers might show in free word association tasks. However, the question is to what extent word association tasks can be considered assessment tools for quality of word knowledge. Can we assume a continuous development towards more and more (abstract) meaning-related word associations, preferably in one category, for example superordinates? We should bear in mind that word associations are not necessarily the result of abstract, semantic knowledge. Language use, such as amount or recentness of exposure to co-occurring words, contributes to the likelihood that one word triggers the other in a spontaneous word association task. In this chapter, we address structural differences in word associations at two levels. Firstly, we investigate possible differences in degree of dispersion of word associations across different types of associations between L1 and L2 children and L1 and L2 adult speakers of Dutch (research question 1). Secondly, we investigate differences for these speakers in distributional patterns of word associations across four categories, and within those main categories across subcategories, thereby assessing whether possible differences between L1 and L2 speakers hold for both children and adults (research question 2).

Differences in word associations between L1 children and L2 children might be caused by differences both at the lexical and at the semantic developmental level. These two levels can be investigated in the comparison of child L1 and L2 and adult L1 and L2 speakers. If differences between L1 and L2 groups hold for children
and for adults, they are most likely due to lexical differences (too little exposure to 
the words to come up with native like associations); if differences ‘disappear’ with 
age, they most likely reflect semantic, conceptual differences. Thus, comparison of 
L1 and L2 speakers from two distinct age groups could further reveal the role of 
word associations in the assessment of word knowledge.

3.2 Method
To answer the foregoing questions, the free word associations of over 400 children 
and a smaller number of adults were analyzed and compared. This study was 
originally set up to generate child word association norms. The adult associations 
were collected later by the second author as part of course work. For this reason, the 
number of adults is smaller than the number of children.

3.2.1 Participants
There were 422 children and 54 adults participating in the study. Children were 
tested in February 2007 as part of a larger study; adults were tested in an exploratory 
follow-up in March and April 2008. The original number of children tested was 534. 
Of the 534 original respondents, 112 were excluded from the current analyses as 
they spoke first languages other than Turkish (or Dutch). The 422 children were 
either monolingual speakers of Dutch (n = 389), or bilingual speakers with Dutch as 
a L2 and Turkish as the L1 (n = 33). Ages ranged from eight to thirteen. Children 
were pupils from seven primary schools in the Randstad, the western part of the 
Netherlands. Socio-economic background of schools varied. A third of the children 
were in grade six, the final year of primary school (eleven – twelve years old), a 
third in grade five (ten – eleven years old), and a third in grade four (nine – ten years 
old).

The adult participants spoke Dutch either as their L1 (n = 41) or L2 (n = 
13). Again, second language speakers of Dutch spoke Turkish as their first language. 
The age of the adults ranged from 17 to 59. L2 adults were recruited by the second 
author through a Dutch learning institute located at the university, through Turkish 
student societies, and through personal connections. The adult sample included 32
women and 22 men. Level and amount of education among adults varied from lower level tertiary education to technical school to university education. Most L2 adults were born in the Netherlands; one man had lived there since a year.

3.2.2 Materials

A set of 118 stimulus words was used. Word selection was inspired by the Word Association Task (Schoonen & Verhallen, 2008) and determined by the frequency of occurrence in a word corpus based on reading materials for children in primary school (Schrooten & Vermeer, 1994). Words were 108 nouns, two verbs and eight adjectives of varying length. The words varied in frequency but all words occurred in the Schrooten and Vermeer corpus; high-frequent words from the corpus were selected as much as possible to ensure participants’ familiarity with the stimuli. The complete word list with the 118 randomized words was completed by all adults. Because of their shorter attention span, the children did not respond to as many as 118 words. Each child completed a 59-word list. The original list was split into two lists (I and II, see Appendix A) each of 59 words; both lists were randomized into five different orders, resulting in ten versions. This was done to prevent order effects and to prevent cheating during classroom administration. Both lists had been piloted with two children in the target age to assess the suitability of words and of the list length.

The list was preceded by a written instruction. At the end of the word list, a number of questions were included regarding participants’ background. These were somewhat different for children and adults. The children answered questions about the country in which they had been born, which language they had learnt first and which language(s) their mother and father, respectively, spoke with them at home. The adults were asked additional questions about their schooling and profession.

3.2.3 Procedure

All children were tested at school during a morning class with the paper-and-pencil version of the task. Adults were tested individually. Only Dutch was spoken. Participants were instructed to give one response only. To ensure that all child
participants understood the instruction a number of examples were discussed with
them in class. After completing the association task, the responses of a few children
from each class were investigated by means of a retrospective interview. For
practical reasons, this was not done for adults. All responses were entered into a
database where they were categorized manually.

Categorisation scheme
The coding scheme used to categorize the responses is adapted from the
classification systems used by Petrey (1977), De Groot (1980), and Fitzpatrick
(2006) as discussed in Section 2.1.2. In order to trace the prominence in the lexicon
of meaning aspects of word knowledge, our coding scheme has as its main
distinction responses that are directly (inherently) related to the meaning of the
stimulus word and responses that are only indirectly or subjectively related to the
meaning of the stimulus word. This contrasts with Fitzpatrick’s system, but is
similar to that of De Groot. Direct meaning-related words are related irrespective
of the context in which they appear. Thus, in our study, Fitzpatrick’s examples lecture–
university, stability–baby, and cultural–cathedral (2006: 133) would not be direct
meaning-related. In addition to the two meaning-based categories, we identify a
form-based and an ‘other’ category. The last two are not based on a meaning
relation. Form-based responses contain both responses based on orthography and on
phonology; it thus includes the traditional ‘clang’ (sound-based) category. Clang
associations are often characteristic of the associations of beginning learners (Meara,
1983). The ‘other’ category contains non-classifiable responses, repetitions and null
responses. The four main categories are subdivided into more specific subcategories
allowing more detailed analyses, as is shown in Table 3.1. Table 3.1 shows
examples of each category.
### Table 3.1 Categorization scheme

| 1. Direct meaning-related | 1 coordinate (*dog-cat*)  
|                          | 2 subordinate (*car-porsch*)  
|                          | 3 superordinate (*deer-animal*)  
|                          | 4 antonym (*fake-real*)  
|                          | 5 partonym 1 (part-whole) (*teeth-mouth*)  
|                          | 6 partonym 2 (whole-part) (*elephant-trunk*)  
|                          | 7 context-independent characteristic (*sledge-snow*)  
|                          | 8 goal/target (*knife-cut*)  
|                          | 9 synonym (*simple-easy*)  
| 2. Indirect meaning-related | 1 subjective association (*motor bike-cool*)  
|                          | 2 composite word (*apple-tree*)  
|                          | 3 context-dependent characteristic (*strong-muscles*)  
| 3. Form-based association | 1 change of affix (*dog-dogs*)  
|                          | 2 similar form, other meaning (*hash-harsh*)  
| 4. Other | 1 non-classifiable (*volcano-fish*)  
|          | 2 repetition (*ocean-ocean*)  
|          | 3 no response

**Categorizing responses**

In our scheme, collocations can belong to various categories depending on the collocation and the exact relationship between the two parts. A collocation such as *bloem – plukken* (cf. English *pick – [a] flower*) belongs to category 2.3. The stimulus-response pair *neus – snuiten* is a collocation (cf. English *blow [your] nose*) belonging to category 1.7. Functional associations such as goal-target pairs (category 1.8) are taken as aspects of abstract, semantic knowledge (a *knife* cuts, irrespective of the context). Not all associations were clear-cut for classification. In Dutch *kaars-vet* (candle-wax/greasy), it was unclear whether the response *vet* is
meant to be the second part of the composite word *kaarsvet* (candle wax, category 2.2) or whether it is a separate adjective referring to a candle as being greasy (category 1.7). Yet, the unlikely co-occurring of *kaars* and *vet* separately compared to the occurrence of the word *kaarsvet* decided for the composite word option.

Difficult cases show that relatedness in meaning has to be weighed for each individual word pair. Two words can be related on a continuum from clearly not related through indirectly related to clearly related (i.e., prototypical category members). One pair of context-independently related words can be more context-independently related than another (compare *sledge-snow* and *ocean-fish*).

To ensure reliable categorization, part of the collection of association responses was categorized twice by independent raters and interrater reliability was determined. The responses to 28 words (out of 118, i.e., 23.7%) were categorized twice. This was only done for the child data. The raters assigned 98.5% of responses to the same subcategory.

### 3.2.4 Analyses

All word associations were classified into one of the 17 categories (Table 3.1). As a first step in assessing respondents’ word associations, we addressed variation in kinds of associations given. Do L1 and L2 speakers prefer one or two types of associations or do they respond with all types of associations? For this, we computed Gini’s concentration index (Wickens, 1989: 130ff.) for each participant. The lower the index, the less the associations are spread across the 17 categories. A participant who uniformly responds with the same type of association scores 0. Maximum spread across the categories is achieved when 3–4 associations occur in each category for children (59/17) or 6–7 for adults (118/17). In these cases the concentration index is maximum, that is .94. To check whether the two sets of stimulus words in the two lists for the children were similar enough to treat them as interchangeable, we assessed Gini’s concentration index in the adult data for the words of list I and list II separately. Remember that adults responded to all words,

---

2 Gini’s concentration index is defined as: \( C_r = 1 - \sum \pi_j^2 \), with \( \pi_j \) referring to the proportion of associations in category \( J \).
whereas the children responded to the words of one of the lists. It turned out that the correlation between the two lists was .81, which is a reassuring split half reliability.

Next, analyses involving log linear model fitting were used to analyze the distributional patterns of word associations across the categories for both L1 and L2 speakers, children and adults at once (see later). First, response patterns are studied at the level of the four main categories; then, the subcategories in each main category are analyzed.

In a log linear analysis, observed frequencies are described in terms of main effects and interaction effects (see Everitt, 1977; Wickens, 1989) in the most parsimonious way possible. Our unit of analysis is the word association. Word associations are aggregated across words and participants. Within a group (adults vs. children) all participants responded to the same number of words (118 and 59 respectively) and thus contributed equally to the group’s distributional pattern. The three-dimensional table of Category by Age by Language Background is analyzed in terms of effects of each of the dimensions and their interactions, as is done in an ANOVA. Main effects indicate that the levels of a dimension are not equally frequent. In our analyses this is a trivial finding: it indicates that there are more L1 associations than L2 associations, and more child associations than adult associations, and more associations in one category than in another. The first two are artifacts of the design, because we have more L1 than L2 participants, and more children than adults. The latter effect is not surprising either, because there is no theoretical reason to expect an equal distribution across the categories involved. The interesting effects are those in which categories interact with person variables (age and language background), which means that certain groups of participants responded with word associations more frequently in a certain category than other groups. A three-way interaction of Age, Language Background and Category would imply that different associative behavior between L1 and L2 speakers is mediated by age, meaning that differences change over time, or that age differences differ for L1

---

3 An alternative approach is to compute the percentages of responses or mean score for a certain category per person, and then to compare these ‘scores’ across groups (see Fitzpatrick, 2006). However, the comparisons to be made are interdependent: a high percentage of associations for one category implies low percentages for the other categories.
The outcomes of the log linear model analyses for main categories and subcategories will be interpreted in two stages: first the search for the best data description will be described, which is a trade-off between fit and parsimony, and second the interpretation of the parameters for age, language background, and category in the selected model will be discussed. Because of the large number of word associations and $\chi^2$'s sensitivity to sample size, models will easily be rejected. To allow for a more comprehensive evaluation of models, descriptive measures will be reported, beside the $\chi^2$ test. A normed fit index (NFI) $\Delta$ indicates the relative increase in fit compared to a base model and the perfect fit of a saturated model (Bonett & Bentler, 1983). As a starting point, we took a base model including the main effects and the interaction between Age and Language Background, as these effects are inherent to the design of the study.

### 3.3 Results

A total of 31,270 association responses was categorised, of which the larger part comes from children (24,898), especially L1 children (22,951) (see Table 3.2).

<table>
<thead>
<tr>
<th>Age</th>
<th>Language Background</th>
<th>Main Categories</th>
<th>Total N (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Meaning-related</td>
<td>Indirect Meaning-related</td>
<td>Form-based</td>
</tr>
<tr>
<td>Child</td>
<td>L1</td>
<td>47.9</td>
<td>45.3</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>43.1</td>
<td>40.2</td>
</tr>
<tr>
<td>Adult</td>
<td>L1</td>
<td>46.1</td>
<td>53.0</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>39.9</td>
<td>57.2</td>
</tr>
</tbody>
</table>

The general pattern is that most of the reported associations are, directly or indirectly, meaning-related (93.6%). Few associations are form-based or unclear (‘Other’, 6.3%); these are almost exclusively given by the L2 children. This indicates that respondents were familiar with most stimulus words, as we had
intended. The adults have a small preference for Indirect rather than Direct Meaning-related associations (mostly expressed by the L2 adults). L2 children give less meaning-related responses than L1 children, in both the direct and indirect category. The relations between types of associations, language background, and age will be analysed in more detail when we fit log linear models.

3.3.1 Degree of dispersion

Analyses of Gini’s concentration index show that participants’ word associations spread across the 17 response categories. All groups score high, that is, all groups give associations of all types (see Table 3.3).

<table>
<thead>
<tr>
<th>Age</th>
<th>Language Background</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L1</td>
<td>L2</td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>M</td>
<td>.81</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>(.056)</td>
<td>(.091)</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>389</td>
<td>33</td>
</tr>
<tr>
<td>Adult</td>
<td>M</td>
<td>.78</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>(.064)</td>
<td>(.115)</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>41</td>
<td>13</td>
</tr>
</tbody>
</table>

A two-way ANOVA with Age and Language Background as fixed factors reveals a significant effect of Age (F = (1,472) = 21.2, p < .001, ηp = .043) and of Language Background (F = (1,472) = 6.7, p = .010, ηp = .014): children’s responses are more dispersed than those of adults; L1 responses are more dispersed than L2 responses. The age effect is larger than the language background effect. The standard deviations show that the groups of L2 speakers differ slightly more among themselves in their dispersion than the L1 speakers.
3.3.2 Response patterns for main categories

Log linear analyses show in which respect response patterns are different for the groups. The fit of different models is summarized for main categories and subcategories (Tables 3.4 and 3.5 respectively).

<table>
<thead>
<tr>
<th>Models</th>
<th>χ²/df</th>
<th>p</th>
<th>Δ NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. [A<em>L][C</em>L]</td>
<td>507.5/6</td>
<td>&lt;.001</td>
<td>.31</td>
</tr>
<tr>
<td>3. [A<em>L][C</em>A]</td>
<td>315.0/6</td>
<td>&lt;.001</td>
<td>.57</td>
</tr>
<tr>
<td>5. [C<em>A</em>L]</td>
<td>0/0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

C=Category, A=Age, L=Language Background; [ ]=interactions including lower order interactions and main effects.

The model fit indexes in Table 3.4 show that statistically speaking a model without the three-way interaction (i.e., model 4) should be rejected (χ²(3) = 18.8, p < .001). However, when considering the NFI, we see that the model with just the two-way interactions with Category describes the data nearly as well as the saturated model (model 5). Dropping any of the two two-way interactions with Category causes a noticeable loss in model fit (i.e., model 2 and 3). A closer look at the relevant partial associations, between Category and Age, and Category and Language Background, shows that the Age differences are more prominent than the Language Background differences (χ²(3) = 631.3, p < .001 and χ²(3) = 233.4, p < .001 respectively. Not shown in the table). This is also indicated by the NFI: the Age by Category interaction (model 3) improves the base model fit with 57 per cent; the Language Background by Category interaction (model 2) brings about an improvement of 31 per cent.

When we zoom in on the two two-way interactions with Category, it shows that the main differences appear in the two meaning-related categories: adults’ word
associations are more meaning-related than children’s associations (whose responses also include some form-based or unclear (‘Other’) associations), and L1 speakers’ word associations are meaning-related more so than those of L2 speakers which is mostly caused by the L2 children.

3.3.3 Response patterns for subcategories

Similar analyses were performed for the distribution within the four main categories. However, we should bear in mind that these comparisons are relative to the number of associations within the main category (and not to a group’s grand total).

The nine direct meaning-related subcategories contain a total of 14,679 associations. Again, the three-way interaction is statistically significant ($\chi^2(8) = 21.0, p = .007$; see Table 3.5), but otherwise negligible. Two-way interactions also provide a very adequate description of the data. According to the partial associations, the major of the two two-way interactions with Category is Age by Category ($\chi^2(8) = 369.8, p < .001$), rather than Language by Category. The effects related to language background are minor. This is corroborated by the NFI indicating that the base model together with only the Category by Age interaction covers 83% of the possible improvement in fit. If we consider the parameter estimates, we see that the differences contributing most to the age effect are the relatively frequent mentioning by the children of superordinates, partonyms (part-whole related words) and functionally related words (goal-target), and their infrequent mentioning of subordinates and antonyms compared to the adults.

The analysis of the 14,610 Indirect Meaning-related associations shows a similar pattern (see Table 3.5). The three-way interaction is statistically significant; this effect is mainly caused by a relatively higher number of composite words of the L2 children. Of the two two-way interactions with Category, the Age by Category interaction is by far the largest, covering 86 % of the possible improvement from base model to saturated model. The children respond with context-dependent characteristics of stimulus words more frequently than the adults.

Relatively few word associations are Form-based, namely 234. Their distribution across the eight cells of Age by Language Background by two Form-
Based subcategories shows a significant, and according to the NFI substantial, three-way interaction effect ($\chi^2(1) = 5.1, p = .024$). However, we should be cautious interpreting this interaction because the (L2) adults had hardly any word associations in this category. For the L2 children 33 word associations were classified, all of which were Changes of affix. For the L1 children on average four-fifths of their form-based associations were Changes of affix. One fifth were clang associations. Going by the two-way interactions, Language Background seems to be more related to Category than Age is. However, because of the few adult responses, it remains uncertain whether this possible Language Background effect is mediated by Age, as the three-way interaction suggests.
Table 3.5 Model fit for subcategories

<table>
<thead>
<tr>
<th>Model (Direct Meaning-related)</th>
<th>$\chi^2$/df</th>
<th>p</th>
<th>Δ NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. [A<em>L][C</em>L]</td>
<td>410.8/16</td>
<td>&lt;.001</td>
<td>.07</td>
</tr>
<tr>
<td>3. [A<em>L][C</em>A]</td>
<td>77.1/16</td>
<td>&lt;.001</td>
<td>.83</td>
</tr>
<tr>
<td>4. [A<em>L][C</em>A][C*L]</td>
<td>21.0/8</td>
<td>.007</td>
<td>.95</td>
</tr>
<tr>
<td>5. [C<em>A</em>L]</td>
<td>0/0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model (Indirect Meaning-related)</th>
<th>$\chi^2$/df</th>
<th>p</th>
<th>Δ NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. [A<em>L][C</em>L]</td>
<td>178.2/4</td>
<td>&lt;.001</td>
<td>.05</td>
</tr>
<tr>
<td>4. [A<em>L][C</em>A][C*L]</td>
<td>20.3/2</td>
<td>&lt;.001</td>
<td>.89</td>
</tr>
<tr>
<td>5. [C<em>A</em>L]</td>
<td>0/0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model (Form-based)</th>
<th>$\chi^2$/df</th>
<th>p</th>
<th>Δ NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. [A<em>L][C</em>A][C*L]</td>
<td>5.1/1</td>
<td>.024</td>
<td>.67</td>
</tr>
<tr>
<td>5. [C<em>A</em>L]</td>
<td>0/0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model (Other)</th>
<th>$\chi^2$/df</th>
<th>p</th>
<th>Δ NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. [A<em>L][C</em>A]</td>
<td>56.0/4</td>
<td>&lt;.001</td>
<td>.06</td>
</tr>
<tr>
<td>4. [A<em>L][C</em>A][C*L]</td>
<td>0.9/2</td>
<td>.623</td>
<td>.98</td>
</tr>
<tr>
<td>5. [C<em>A</em>L]</td>
<td>0/0</td>
<td>1.00</td>
<td>1</td>
</tr>
</tbody>
</table>

---

4 Due to the relatively small number of observations in the Form-based category some cell frequencies were very low (0 or 1), causing some inconsistency in the Pearson $\chi^2$ for nested models. Therefore, the log-likelihood ratio $\chi^2$ was computed for this subcategory.
The three ‘Other’ subcategories included a total of 1747 word associations (only 5.6%). It is remarkable that the three-way interaction is clearly non-significant. The partial associations support what is indicated by the NFI: the Language Background by Category interaction has more effect than the Age by Category interaction; both are significant ($\chi^2(2) = 59.6, p < .001$ and $\chi^2(2) = 9.5, p = .008$). The relatively small number of associations of the adults should make us cautious again. The adults did not repeat any of the stimulus words whereas the children did. The non-responses were most frequent among the L2 participants, which might indicate that those stimulus words were not familiar enough to them.

In sum, the analyses show that all respondents give mostly meaning-related associations: L2 children less than L1 children and L2 adults clearly more Indirect Meaning-related associations (Table 3.2). Responses of the children and of the L1 speakers are more dispersed than of adults and of L2 speakers. The in-depth log linear analyses show that the distributional patterns of association responses differ among the groups. The age groups differ most. The adults give more meaning-related responses than the children: children give some Form-based (e.g., changes of affix) and Other responses while adults hardly do. The L1 speakers give more meaning-related responses than the L2 speakers: this is mostly caused by the L2 children’s relatively high number of Form-based and Other word associations. Within the Direct Meaning-related category, the adults tend to respond more with subordinates and antonyms, while the children prefer partonyms, functionally related words and superordinates. Within the Indirect Meaning-related category, children mention context-dependent characteristics more often than adults. The few Form-based associations were mostly given by the children; the few Other associations were mostly given by the L2 children.

### 3.4 Discussion and conclusion

In this study, we compared word associations of Dutch child and adult monolinguals and bilinguals with Turkish as their L1. We looked at differences between language and age groups with respect to degree of dispersion of word associations (research question 1) and distributional patterns of responses across response categories.
The general patterns of responses are the same for L1 and L2 children and adults: most associations are meaning-related. Differences between the groups of speakers are generally small. Adults’ responses are more meaning-related than children’s responses: this is mainly due to the relatively many ‘Other’ associations of the L2 children. This may be because of their possible unfamiliarity with some of the stimulus words. However, in the retrospective interviews children only incidentally reported not knowing a word. That may also contribute to the fact that the L2 children give less meaning-related responses than their L1 peers, both in the Direct and Indirect category. Interestingly, the adult respondents have a preference for Indirect Meaning-related responses. Because the adults do not mainly give associations of the most abstract, conceptually meaningful sort, we cannot consider Form-based, Indirect Meaning-related and Direct Meaning-related as a scale of word knowledge development (see Namei, 2004). However, this preference for less abstract associations need not surprise us if we consider that factors other than semantic, conceptual development (for example, language use) may cause certain associations, not necessarily Direct Meaning-related, to become more prominent in the mental lexicon and to be triggered in a spontaneous association task (see Nelson et al., 2000; Wettler et al., 2005, and also Nissen & Henriksen, 2006).

The effect of language background is relatively small when we look at response patterns within the two Meaning-related main categories. Age is the dominant factor, which may imply that conceptual development is more important, both in L1 and L2: adult L2 learners are most likely to resort to their semantic or conceptual knowledge developed in the L1. In other words, this semantic knowledge may not be very language specific or, at least, it is available for both L1 and L2 associations, provided that the L2 learners are familiar with the corresponding L1 word forms. Age as well as language background are relevant factors in comparing responses within the Form-based and Other categories. Most of these types of associations are given by children; in the Other category mostly by the bilingual children. The differences between monolingual and bilingual speakers may be related to familiarity with the stimulus words. If the stimulus word is unfamiliar to a
participant, he or she has no or not many semantic or conceptual links between that
word and other words and may resort to morphologically related words (Form-
based), repetitions or non-responses (Other). In those cases, the word association
task behaves as a common vocabulary test, which runs counter to the usual purpose
of word association tasks, namely to investigate what kind of knowledge respondents have.

The variation across categories differs significantly between L1 and L2
speakers, especially in the adult group, but the effect is rather small. Also, children’s
responses are more spread out across categories than adults’ responses: adults hardly
give form-based or unclear (Other) associations. As was recently pointed out by
Fitzpatrick (2006), monolingual responses are not as homogeneous as is often
assumed. From an assessment perspective, the idea of homogeneous responses is
problematic. Using word associations in testing presumes some underlying scale
ranging from ‘poor’ to ‘good’. However, the operationalization of this presumed
scale is difficult, if not impossible. Clang and other form-based associations are
often considered indicators of poor development of the mental lexicon. (Direct)
meaning-related associations are considered indicative of well-developed word
knowledge. The discussion about homogeneity of responses usually refers to the
homogeneity across (L1) speakers. In this way, homogeneity is a feature of a group
rather than of an individual participant. In contrast, if we want to apply the concept
of homogeneity to participants, a participant should stick to a few categories of
abstract direct-meaning relations (as would be shown by a low Gini concentration
index). But a less proficient language learner only responding with one type of form-
based association will also score low on the concentration index. The question is
whether this is what we mean by homogeneous responses: probably not. Another
simple solution would be just to count up the number of ‘good’ associations, for
example the number of superordinates. In this case, presented with the stimulus
word black, the common association white (a coordinate) would not be scored as
‘good’, whereas color (a superordinate) would. The common association made by
adult native speakers would thus remain unrewarded. In our view this illustrates that
responses to word association tasks cannot be measured along a certain ruler of
increasing quality or meaning-relatedness, and as such, free word association tasks do not fit well in the paradigm of (language) proficiency testing (see Kruse et al., 1987). Responses are indicative of (recent) exposure rather than ‘good’ or ‘not so good’ word knowledge, which in itself can be interesting for research. To assess participants’ word knowledge, free word association tasks should be less ‘free’. If the instruction is: “what is the first word that comes to your mind when you see/hear black?” the participant cannot do anything but respond with the first word, be it white, purple or hair, although she may well know that color is the superordinate. If researchers want to tap into the latter kind of knowledge they had better use a more controlled procedure, such as asking for superordinates (e.g., in a definition task) or testing whether participants recognize related words from a limited set of alternatives (see formats developed by Meara and Wolter, 2004; Read, 1993; and Schoonen and Verhallen, 2008).

Limitations
We should bear in mind that this study has some limitations. Respondents should ideally be tested in both their L1 and their L2 to be able to make more conclusive claims about conceptual development and more profoundly explore parallel development in speakers’ L1 and L2. Furthermore, to really know whether differences disappear with age, a longitudinal design would be valuable.

Instead of having respondents write down their responses in a written association task, researchers in subsequent association studies would do well to record spoken responses (Van Hell & De Groot, 1998). A spoken association task, using a voice key for example, is likely to tap into relations between words in the mind in a more direct way, once the intervening writing process is taken out. Moreover, writing down responses could be an obstacle for L2 learners. Additionally, response latencies in an online design could give valuable information about the automaticity of (L2) word processing. Here, the Revised Hierarchical Model would predict longer response latencies for (beginning) L2 speakers, because for them, conceptual knowledge is accessed through the L1 lexicon rather than directly through L2.
The selection of stimulus words for the present study was determined by the goals of a larger research project. All words occurred in a word corpus of primary school language and did so with various frequencies. A precondition for valid word associations is that the words are, at least superficially, known to respondents. The somewhat larger percentages of Other responses for the L2 groups suggest that this was not always the case. An investigation of the Other responses shows that they were distributed rather randomly and were not concentrated around any words in particular. To further explore the relationship between L1 and L2 knowledge in the bilingual group, stricter word selection could be helpful (see Fitzpatrick, 2007). A post-hoc screening of the stimulus words displaying the largest differences in associative behavior between Dutch and Turkish participants did not point to specific word features that might be linked to semantic or translation differences. In future work, the inclusion of stimulus words with different meaning extensions in Dutch and Turkish could provide interesting insights into cross-linguistic association behavior.

Finally, the focus of this study is on word associations of monolingual Dutch and bilingual Turkish–Dutch speakers. We restricted our bilingual sample to speakers with a homogeneous L1 language background (Turkish). This restriction makes the bilingual group limited in size and in L1. Furthermore, the adult bilingual group is rather small, as it proved relatively difficult to find enough speakers in that group. This limits the generalizability of the results, especially those concerning the adult bilinguals. Replication with a larger group of bilingual adults could corroborate our findings.

Our results show that there are small differences between Dutch child and adult monolingual speakers and bilingual speakers with Turkish as their L1 with respect to the amount of dispersion of their word associations across categories: responses of children and of L1 speakers are more dispersed than of adults and of L2 speakers (research question 1). The issue of homogeneity of responses between and within subjects underscores the ‘problematic’ status of the free word association task as an assessment tool (for word knowledge). In the same vein, the results show that Direct Meaning-related associations are not necessarily ‘optimal’ or ‘proficient’
word associations, assuming that L1 adults are the most proficient group (see Nissen & Henriksen, 2006). The lack of such an ‘optimum’ further problematizes the role of word associations in assessment. In spite of these ‘problems’, free word association tasks can give insight into a language user’s current state of mind, that is, the strength of certain word relations, irrespective of the origin of the word associations. Our different findings for monolingual and bilingual children, and for monolingual and bilingual adults in distributional patterns across categories show the influence of bilingualism on the development of word associations (research question 1 and 2). At the same time, the prominent effect of age on meaning-based associations emphasizes the role of conceptual development in word association behaviour: the children did not show the adults’ preference for Indirect Meaning-related responses. Also, the adult data did not show as many Form-based responses as given by the children. With maturity, these shallow associations seem to ‘disappear’. Whether children’s word knowledge develops from episodically structured into conceptually structured or not, is an interesting question, but deserves an investigation using more pre-structured, receptive paradigms.