A new word association test for Norwegian

Implications for theories on the mental lexicon, and on language and ageing

Malene Bøyum

MA thesis in linguistics

LING4190

Department of Linguistics and Scandinavian Studies

The University of Oslo

June 2016

A New Word Association Test for Norwegian

Implications for theories on the mental lexicon, and on language and ageing

Malene Bøyum

MA thesis in linguistics

Department of Linguistics and Scandinavian Studies
The University of Oslo
June 2016

A New Word Association Test for Norwegian:
Implications for theories on the mental lexicon, and on language and ageing

Malene Bøyum

http://www.duo.uio.no/
Trykk: Reprosentralen, Universitetet i Oslo

Synopsis

This thesis investigates on how data from a new word association (WA) test for Norwegian may shed light on theories on the mental lexicon, and on language and ageing.

A new WA test for Norwegian was developed, based on the methodology of Fitzpatrick et al. (2015). The WA test consists of 100 cue words, and the participants were instructed to write the first word that came to mind for each of them. WA data was collected from 173 participants from two age cohorts: 20-30 year olds (n=122), and over 60 year olds (n=51). Norms lists were compiled for both cohorts.

Comparisons of the word associations from the two cohorts show clear differences between the age groups, both in the norms lists, and in the response category patterns. The results suggest that the observed age-related differences are best explained as a combination of generational effects and age-related changes in WA strategies.

Further, the thesis aimed to explore how the collected WA data might shed light on, firstly, theories of the mental lexicon and, secondly, theories on language and ageing. A cognitive linguistic approach was adopted for the study. Analyses of the collected data found both inflected forms of words, and multi-word units in the responses. Responses were given on all levels of representation (i.e. meaning, position, and form), and there was much individual variation in the responses. These findings support a non-modular view of language, where linguistic knowledge is organised in a network. The findings also support the role of experience in usage-based theory . Further, significant differences were found between noun cues and verb cues, supporting the fundamental difference between nouns and verbs that is postulated in cognitive grammar.

Finally, the data from the WA test found age-related differences for the number of blank responses and the number of multi-word responses. The results support previous findings of age-related difficulties in the retrieval of and production of words.

Acknowledgements

Firstly, a huge thank you to my two fantastic supervisors, Hanne Gram Simonsen and Marianne Lind, for all your time, encouragement and guidance. Thank you for introducing me to word association tests, for all the interesting discussions, and for making me feel like a part of the research community at MultiLing.

Secondly, this study would not have existed without all of my participants - thank you! Thank you also to the staff at the senior centres for your help, and to the lecturers who kindly let me use some of their teaching time for the study.

I wish to thank Professor Tess Fitzpatrick for generously sharing her expertise on word associations, and to Gözde Mercan for the interesting discussions on the methodology. To Elisabeth Holm, who kindly re-classified parts of my data, and Kristin Hagen at Tekstlaboratoriet who digitalised my norms lists.

Thank you to Bård Uri Jensen and Pernille Hansen for all their help with the statistics. Martin Smidt also helped me with R.

Thank you to Marit Johanne Furunes for proof reading the thesis, and to Aïda Leistad Thomassen for all the help with LaTeX.

Also, thank you to my fellow MA students for way too long coffee breaks, fruitful discussions, and equally necessary silly jokes and collective procrastination.

To Niklas, Emil, Hannah, Siv B. and Siv J. - such a happy home to get back to in the evenings.

Friends and family- thank you for taking my mind of this thesis every now and then. I can't wait to see you all more often from now on.

Finally, a special thank you to Thomas for always being there, for making me laugh even on the days I thought I'd never finish this thesis, and for patiently listening to me talk about word associations all year.

Malene Bøyum
Oslo, June 2016

Abbreviations

A Adjective

Adv Adverb

AoA Age of acquisition

BNC British National Corpus

BNT Boston Naming Test

Eng. English (used in translations of Norwegian examples)

DAT Dementia of the Alzheimer's type

N Noun

L1 First language (native language)

L2 Second language

MMSE Mini-Mental State Examination

RNT Relational network theory

TDH Transmission deficit hypothesis

TOT Tip-of-the-tongue phenomenon

V Verb

WA Word association

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and motivation

Word associations are fascinating. On the surface, they are reminiscent of the games one would play as a child, seeing who could make the longest chain of words. Yet, for over a century, the results from vast amounts of research using word association tests have revealed that word associations are much more than games and word play.

Recent research has provided interesting developments to the methodology and scientific use of word association tests. The role of variation in association behaviour, both between individuals and between different cohorts, is becoming more important, in contrast to previous research where the aims have been mainly to determine association norms for entire language communities. Word association tests have become more task specific and detailed in their methodology, which opens up for new possibilities for what we might learn from them. Yet, despite the promise of recent developments in word association methodology, there are still gaps to fill, and improvements to be made.

Most word association research has been done on English speakers. The current study is the first linguistic study of word associations for Norwegian. In this master's project I will have developed a new word association test for Norwegian, and collected word association data for two age cohorts. I hope that the new test and the norms lists will provide a foundation for more research to come on word associations in Norwegian.

Word association data differ from other sets of linguistic data. The association process is free. There are no restrictions imposed on the participant, and no contexts given for the cue words. Many researchers believe that word association tests involve a process which "gives rise to distinct mental properties that go beyond the information captured in written or spoken text" (De Deyne and Storms, 2015, p. 470). The data collected in this project will not only be used to make norms lists for the associations of native speakers of

Norwegian; it will also be explored to see what it might reveal about the structures and processes of the mental lexicon.

Lastly, although the linguistic development of children has been thoroughly studied using word association tests, little such research has been done on the changes that occur in linguistic skills throughout the lifespan. The time frame of a master's thesis does not allow for any kind of longitudinal study. However, age-related changes in linguistic skills should be more of a focus of word association studies. After all,

> [o]f all aspects of language, the lexical component changes the most over the lifespan, with the acquisition of new words extending throughout adulthood. In that sense, our mental lexicons are never fixed and never cease being linguistic capacities (Jarema and Libben, 2007, p. 3)

In this project I will thus use the word association data from the two age cohorts to investigate both what differences may be seen in the responses, and how the data may support or go against existing theories on changes in linguistic skills as we get older.

1.2 Research questions

Based on word associations collected from two age cohorts (20-30 years, and over 60 years), the study will strive to answer the following sets of research questions:

1. Are the word associations different in the two age cohorts? How do they compare to the results found in Fitzpatrick et al. (2015), and in other previous studies using word association tests across age cohorts?

2. How does the data shed light on theories on the mental lexicon? What might the data tell us about the organisation of the mental lexicon, and of storage and access in particular?

3. How does the data shed light on theories on language and ageing? What might they tell us about potential changes in linguistic skills with advancing age?

The predictions for the three sets of research questions are presented in subsequent chapters. This was done in order to have the predictions as close as possible to the theories and previous research they arose from. The predictions for the first set of research questions can be found in Section 2.1.5. For the second set of research questions, the predictions can be found in Section 3.1.3, and for the third set in Section 3.3.2.

1.3 Outline of the thesis

The thesis is structured as follows: Above, the background and motivation for the project have been presented, along with the research questions for the study. Further, in Chapter 2, previous research on word associations is presented. The theoretical framework for the study is introduced in Chapter 3. Theories on the mental lexicon are presented in Section 3.1, and theories on language and ageing in Section 3.3.

Chapter 4 presents the methodology for the study, and Chapter 5 presents the results from the collected data. These results are discussed in Chapter 6. Lastly, conclusions and ideas for further research are given in Chapter 7.

Chapter 2

Previous research

This chapter will give an overview of previous research on word association tests. As the current study is the first linguistic study of word associations for Norwegian[1], a rather broad and comprehensive introduction to previous research will be given. The chapter firstly gives a brief overview of the earliest use of word association data, where it was used as a tool for psychological diagnosis (2.1.1). The rest of the chapter will focus on the linguistic uses of word association tests, in relation to research on acquired language disorders (2.1.2), language acquisition (2.1.3), language and ageing (2.1.4), and research on the mental lexicon (2.1.6). Lastly, previous Norwegian research on word associations is presented (2.2). Predictions for the first set of research questions (see 1.2), on whether the word associations are different in the two age cohorts, are presented in 2.1.5.

2.1 International research

2.1.1 Word associations and mental state

The study of word associations dates back to the turn of the twentieth century. The focus and use of many of these early word association studies were usually not purely linguistic, but rather the tests were used to indicate general behavior, or to diagnose psychological abnormality (Fitzpatrick et al., 2015). An early major study of word associations and psychological state was carried out by Carl Jung and his assistants (Jung, 1910). Jung saw words as "really something like condensed actions, situations, and things"(Jung, 1910, pp. 223-4), and the responses given in a word association test would reveal something about how the participant reacted to these situations. According to Jung, the association

[1]The other previous large-scale Norwegian word association study presented in 2.2 only surveyed word association norms, without any further analysis of the data.

method gave insight into the emotional processes of the participants, and could thus reveal personality traits, emotional deficiencies, psychological problems, and more (Jung, 1910). For instance, Jung claimed that a person who shows "disturbances to the stimulus words, is in a certain sense really but imperfectly adapted to reality"(Jung, 1910, p. 226). This was taken as a sign of "something morbid in the psyche" (Jung, 1910, p. 226).

The use of word association tests to diagnose psychological abnormality was continued by among others Kent and Rosanoff, who, like Jung, studied the associations of so-called *insane* persons, and compared them to the associations made by healthy subjects (Kent and Rosanoff, 1910). Kent and Rosanoff's study provided large amounts of data on the word associations of different cohorts of the normal population. The cue words they used, known as the Kent-Rosanoff list, has since functioned as a reference point, and the cue words themselves have been frequently used in other studies up until today (e.g. Håseth, 1968; Namei, 2004; Woodrow and Lowell, 1916). The Kent-Rosanoff list has however also been progressively criticised for its method of compilation, and for almost only containing nouns and adjectives (see 6.1.2). A main finding of Kent and Rosanoff's study was the common tendency for normal subjects to produce responses taken from a small group of frequent reactions (Kent and Rosanoff, 1910, p. 14). However, the responses given by those subjects described as insane by Kent and Rosanoff, revealed "evidence of a weakening of the normal tendency to respond by common reactions" (Kent and Rosanoff, 1910, p. 16). Both Jung's and Kent and Rosanoff's studies were based on the idea that there exists a set of common, normal responses for a given group or language, and that one can compare the responses of any person to these norms to discover irregularities in the health and language abilities of the participant.

2.1.2 Word associations and acquired language disorders

Word association data has also been used to study different types of decline in linguistic skills due to acquired language disorders such as aphasia or dementia. The focus of this section will be on studies of word association and dementia. Dementia is a broad term which denotes multiple diseases that have in common that patients suffer a loss of intellectual abilities. This loss is caused by degeneration of brain tissue, and different dementias affect different areas of the brain. Symptoms may thus differ between patients. The section will mainly report on studies with patients suffering from dementia of the Alzheimer's type (DAT).

In the case of dementia, correlations between the severity of dementia and the number of common responses given in word association tasks have been found. Eustache et al. tested two groups of patients diagnosed with DAT in both the Mini-Mental State Examination

(MMSE)[2] and in an oral free word association task[3] (Eustache et al., 1990). 20 French and 19 American patients were all tested twice in the word association task, with a fifteen minute interval in between (Eustache et al., 1990, p. 1317). The participants' responses were scored for commonality, with common responses defined as those occurring with a frequency of 7% or more in the norms list and reference sample used (Eustache et al., 1990, p. 1318). The results after the commonality classification indicated "an association between scores on the screening test sensitive to severity of dementia (Mini-Mental State Examination) and the index of commonality of responses (Most Common Responses)" in both groups of DAT patients (Eustache et al., 1990, p. 1320). Patients were also less likely than healthy subjects to give the same responses at both test times (Eustache et al., 1990, p. 1318).

Results from the performance of patients with dementia on word association tests have also been used to discover in greater detail which areas of linguistic abilities that are more affected by the disease. One such study was carried out by Gewirth et al., who tested the word associations of both dementia patients and aphasia patients, in comparison to those made by healthy participants (Gewirth, Shindler, and Hier, 1984). 22 of the 38 patients with dementia had been diagnosed with DAT, the other 16 with other dementias. The syntagmatic-paradigmatic division was used as a basis for response classification. Syntagmatic responses are those that would co-occur with the cue word in text (e.g. *letter-write*), whereas paradigmatic responses are those that could replace the cue in text without changing the grammaticality of the sentence (e.g. *letter-newspaper*) (Fitzpatrick et al., 2015, p. 38). Gewirth et al. found that with progression of the dementia, the patients produced fewer paradigmatic responses, and instead produced more idiosyncratic responses, or they failed to produce a response. However, the amount of syntagmatic responses remained at about the same level (Gewirth, Shindler, and Hier, 1984, p. 309). Response latencies also increased as the patients' dementia got more severe, which Gewirth et al. link to both slower production and slower retrieval caused by ageing more generally, and additional problems with word retrieval caused by dementia (Gewirth, Shindler, and Hier, 1984, p. 309). Gewirth et al. suggest that the decline in paradigmatic responses "could be due to a progressive loss of semantic markers" (Gewirth, Shindler, and Hier, 1984, p. 310). This would mean that the syntactic knowledge of words is better preserved than the lexical knowledge, for DAT patients. Gewirth et al. suggest that syntagmatic responses are less dependent on access to semantic knowledge, but are rather more based on syntactic knowledge.

[2]The MMSE is a brief mental status examination, which focuses on the cognitive aspects of mental functions. The test consists of eleven questions, and is brief enough (5-10 minutes) to be administered frequently and routinely (Folstein, Folstein, and McHugh, 1975).

[3]A free word association test is one in which the participants write or say the first word that comes to mind in response to a given cue. Free word association tests contrasts to bound word association tests, where the participant is given a cue and a set of potential responses, from which he or she chooses one.

A semantic deficit caused by dementia of the Alzheimer's type has also been proposed by Gollan, Salmon, and Paxton (2006). As in Gewirth et al.'s study, patients suffering from DAT were found to produce different types of associations from the controls (Gollan, Salmon, and Paxton, 2006, p. 297). The participants were asked to give responses to both strong (e.g. *bride-groom*) and weak stimuli (e.g. *body-leg*),[4] and the study found that the strong associations were more affected by the changes in cognitive abilites associated with DAT than the weak associations were (Gollan, Salmon, and Paxton, 2006, p. 294). The patients with DAT produced less common responses only for the strong, and not the weak stimuli. Gollan et al. postulate that weak associations are "less semantic", and thus less reliant on meaning. Compared to Gewirth, Shindler, and Hier (1984), Gollan, Salmon, and Paxton (2006) do not hypothesise that there are additional retrieval difficulties linked to DAT, as "retrieval is clearly equally necessary to produce a response to a weak and a strong stimulus" (Gollan, Salmon, and Paxton, 2006, p. 300).

2.1.3 Word associations and language acquisition

The data from word association tests have been used in studies of both language acquisition in children, and of second language acquisition among adults. Many of the early studies of the word associations of children were closely related to the studies carried out by Jung and others, looking at differences between healthy and mentally ill children (e.g. Eastman and Rosanoff, 1912), or they looked for correlations between intelligence and the responses made in word association tests (e.g. Meumann, 1905). Other studies focused on the differences in associations between children and adults, and on changes with age in the associations of children (e.g. Rosanoff and Rosanoff, 1913; Woodrow and Lowell, 1916). See also Woodrow and Lowell (1916) for a thorough review of these earliest studies on children's word associations.

Woodrow and Lowell (1916) used the Kent-Rosanoff cue words[5] to test the associations of a thousand American school children aged nine to eleven (Woodrow and Lowell, 1916, p. 1). The children's associations were compared to Kent and Rosanoff's norms lists for adults. Furthermore, the norms lists for the children were meant to be standards to compare the responses of individual children against, in order to obtain "pictures of [their] mental constitution" (Woodrow and Lowell, 1916, pp. 1-2). Unlike previous studies on mind and word association though, Woodrow and Lowell emphasised that "the main value of the association test is not in telling us how much mind but what kind of a mind the child has" (Woodrow and Lowell, 1916, p. 2). However, they also suggested that the

[4]Strong stimuli are cue words that tend to elicit a particular response for a very large amount of the responses (Gollan, Salmon, and Paxton, 2006, p. 290).

[5]This list was slightly modified as some of the words from the Kent-Rosanoff list had been shown to be too difficult for children (Woodrow and Lowell, 1916, pp. 27-8).

frequency lists can be used "for comparison with similar frequency tables obtained from children of different social status or race or of defective mentality" (Woodrow and Lowell, 1916, p. 3). In terms of differences between the associations of children, and those of adults, Woodrow and Lowell found that the children's and adults' associations differed in a "striking manner" (Woodrow and Lowell, 1916, p. 97). This is based on both the frequencies of the response words themselves, and differences in frequency between the types of associations given for the two age cohorts (Woodrow and Lowell, 1916, p. 109). They suggest that this is due to the fact that the adults have acquired a number of additional words, and with these additional responses the percentage of similar responses between the two age cohorts declines (Woodrow and Lowell, 1916, p. 97).

The 1960s and 1970s saw an increase in studies using word associations test, many of them to study the changes in the associations for children as they get older. The studies were closely linked to the notion of the syntagmatic-paradigmatic shift. The syntagmatic-paradigmatic shift happens as children's word associations change with age, where syntagmatic responses are gradually substituted by more paradigmatic responses (Entwisle, 1966, p. 2). Ervin (1961) did one of the first studies on the changes with age in the type of associations of children. Ervin expected a decrease in syntagmatic responses, and an increase in paradigmatic responses as the participants got older, in addition to a decrease in clang-responses[6] (Ervin, 1961, p. 362). In her study, Ervin had a total of 184 participants, from kindergarten, first grade, third grade, and sixth grade. They were asked to give responses to both a closed-alternative test, with questions such as "[d]oes *brother* go better with *sister* or *father*?", and a free-association task (Ervin, 1961, p. 363). The results showed that the number of paradigmatic responses increased with age. Ervin suggests that the change in response type patterns might "reflect educational experience," although this is not a fully sufficient explanation, as other studies had found a preference for paradigmatic responses among non-literate adults. Ervin then suggests that education might "hasten changes which occur with experience even without schooling" (Ervin, 1961, pp. 370-1). A decrease in clang-responses was also found (Ervin, 1961, p. 371). These findings are in line with the predictions of the syntagmatic-paradigmatic shift.

Another pioneering study of the word associations of children is Entwisle (1966). Entwisle's study was the first one to investigate the word associations of young children below the fourth grade. She found that "paradigmatic responses increase strikingly over the childhood years, but at different rates for different words" (Entwisle, 1966, p. 7). The shift has several stages following increased exposure to the language (see figure 2.1).

Also an earlier study by Entwisle, Forsyth, and Muuss (1964) supported the hypothesised syntagmatic-paradigmatic shift: "The S-P shift appears to be a large effect over the ages

[6]Clang-responses are responses made on the basis of the sound properties of the cue word, e.g. the cue word *cat* prompting the response *hat*.

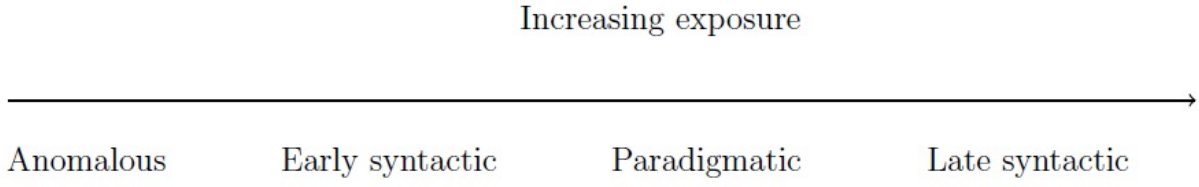


Figure 2.1: Changes in association types due to increasing exposure to the language (Entwisle, 1966). Entwisle uses the terms syntactic *rather than* syntagmatic.

5 to 10,[...] Its existence suggests that clusters of words, presumably related through mediating responses, are formed along different dimentions at different ages" (Entwisle, Forsyth, and Muuss, 1964, p. 27). Entwisle et al. compared their results to previous studies of the words associations of children and the syntagmatic-paradigmatic shift, and noted that the time at which the shift happens seemed to differ between populations, as for instance Woodrow and Lowell (1916) found a later shift, for 9-12 year olds (Entwisle, Forsyth, and Muuss, 1964, p. 27).

The results of Entwisle's study and other studies looking at the changes with age in word associations of children have been transferred to the study of second language acquisition. A persistent hypothesis in multiple studies on second language acquisition has been that the type of responses made (usually classified by the syntagmatic or paradigmatic categories) will become more similar to the response patterns of native speakers as the learner becomes more proficient in the L2. In addition to the response type patterns, the responses themselves will also shift towards those made by native speakers (Wolter, 2002, p. 316).

One such study was carried out by Namei (2004), who investigated whether the L2 mental lexicon in the early stages of learning is primarily phonologically organised, and if the learners at these early stages also have limited semantic knowledge in the L2. With increased proficiency, the learners' lexicons become more semantically organised (Namei, 2004, p. 366). Namei tested Swedish-Persian bilingual children and young teenagers, and monolinguals in both languages in the same age groups. The participants were tested with translated versions of the Kent-Rosanoff list (Namei, 2004, pp. 369-70). Namei found a syntagmatic-paradigmatic shift, but only for individual words, and not for the lexicon as a whole (Namei, 2004, p. 382). Namei's results showed that barely familiar words produce form-based responses, moderately known words produce syntagmatic responses, fairly well-known words give paradigmatic responses, and well-known words produce either paradigmatic or late syntagmatic responses (Namei, 2004, p. 382). Namei suggests that words go through phases of clang responses when they are newly acquired, then syntagmatic responses, before finally "words that are deeply integrated in the mental lexicon are connected paradigmatically to other words" (Namei, 2004, pp. 382-3). Thus, all types of responses may also be found among native speakers; however, second language learners

will have more less familar words.

Politzer tested the word associations of 203 first year French students in American high schools (Politzer, 1978, p. 203). The students were tested in both their L1 (English) and their L2 (French), and the results showed large differences in the types of associations made, mirroring "what one might call heavily unbalanced bilingualism" (Politzer, 1978, pp. 204-5). The number of syntagmatic responses was much higher in the French test, which Politzer suggests is partly due to the teaching method used in the French course, but also the results may be linked to the idea of a syntagmatic-paradigmatic shift in the L2 as one becomes more proficient (Politzer, 1978, p. 205).

However, the validity of the notion of the L2 learners' associations becoming more native-like as they grow more proficient in their L2, has been questioned in several recent studies. Nissen and Henriksen (2006) got results that contradicted previous studies on the syntagmatic-paradigmatic shift in L2 learners. Their study explored the influence of word class on response types, and also tested the hypothesis of the syntagmatic-paradigmatic shift in L2 learners. This was done by testing 25 Danish high school students in a word association task in both their L1 (Danish), and in their L2 (English), which they had been learning at school for 7-8 years (Nissen and Henriksen, 2006, pp. 391-2). After testing the participants in both their L1 and their L2, Nissen and Henriksen found a surprising majority of syntagmatic responses when the participants were tested in their L1 (Nissen and Henriksen, 2006, p. 389). If the participants did not produce a majority of paradigmatic responses in their L1,[7] then the amount of paradigmatic responses when they were tested in their L2 would not reflect their degree of proficiency in the L2. Nissen and Henriksen suggest that the difference in their results compared to previous studies might be due to the distribution of the different word classes in the cue words. Nissen and Henriksen found that nouns elicited more paradigmatic responses than verbs and adjectives. Their study had a more equal distribution of word classes than some previous studies, which had used predominantely nouns as cue words (Nissen and Henriksen, 2006, p. 399).[8] The authors also note that other factors such as word frequency, and "the fact that the number of late syntagmatic responses[9] increase as a function of degree of knowledge of the individual word" might influence lexical and cognitive development and the kind of responses made (Nissen and Henriksen, 2006, p. 404). Nissen and Henriksen conclude that, based on their findings, "[t]he concept of the syntagmatic–paradigmatic shift as an indication of lexical and cognitive development is seriously questioned [...][and] seems to lose its valid-

[7]As the syntagmatic-paradigmatic shift hypothesise that they should, as the participants were high school students, and would have completed the shift in their L1.

[8]A difference in the amount of paradigmatic and syntagmatic responses to different word classes was also reported over fourty years earlier by for instance Deese (1962).

[9]Entwisle's model of the syntagmatic-paradigmatic shift also includes a stage of *late syntagmatic* responses, which comes after the shift to chiefly paradigmatic responses, during increased exposure to the L2 (see figure 2.1). Note that Entwisle used the term *syntactic* instead of *syntagmatic*.

ity as a means of indicating lexical network development" (Nissen and Henriksen, 2006, p. 404).

Like Nissen and Henriksen, Fitzpatrick has also challenged some of the fundamental assumptions that have been prevalent in previous studies on word association and second language acquisition (Fitzpatrick, 2006, 2007, 2009; Fitzpatrick and Izura, 2011). Fitzpatrick (2006) looked at word associations as a tool to access information about the organisation of the L2 lexicon. In this study, Fitzpatrick addressed two problematic factors in previous research: the choice of stimulus words and the categorisation of responses (Fitzpatrick, 2006, p. 121). Fitzpatrick identified several properties of the cue words that might affect the responses given (such as frequency and word class), and for the test methodology (such as the number of cues given, and the number of responses requested per cue word). Fitzpatrick concludes that "it is imperative that lists of stimulus words are compiled in a principled manner" (Fitzpatrick, 2006, p. 124). In terms of the response categories, the study found that a more complex set of response categories (17 in total) revealed differences between participant groups that were concealed when using a traditional set of response categories with fewer, larger categories (Fitzpatrick, 2006, p. 143). Similar conclusions were made based on the study presented in Fitzpatrick (2007), which included using less frequent cue words than had been used in previous studies to avoid predictable responses, and the continued use of a more complex classification system, as in Fitzpatrick (2006). One question that was addressed in Fitzpatrick (2007) was whether native speakers respond to word association tests "in a way that is homogeneous enough for us to refer usefully to 'native-speaker norms' " (Fitzpatrick, 2007, p. 320). With the use of a more complex classification system, the study revealed large differences between participants tested in their L1, and the study concluded that native speakers are not homogeneous, nor predictable in their response behaviour (Fitzpatrick, 2007, p. 326). The results did however show some consistency in individual response types across test times. Fitzpatrick suggests that one should perhaps focus more on individual profiling, and potentially rather examine if L2 speakers move closer to their own L1 profile, rather than comparing their responses to so called native behaviour (Fitzpatrick, 2007, pp. 327-8).

2.1.4 Word association and age differences

Like the changes in word association behaviour throughout childhood (see 2.1.3), changes in association in older age have also been studied using word association tests. In a study by Hirsch and Tree (2001), the word associations of two cohorts of British adults were compared. The two age groups in Hirsch and Tree's study were 21-30 year olds and 66-81 year olds. Hirsch and Tree asked their participants to respond to 90 cue words. These were

chosen either because they were names of concrete, picturable objects, or because it was believed they would frequenctly elicit concrete objects (Hirsch and Tree, 2001, p. 3). The study revealed cohort effects in word associations. Hirsch and Tree found only a moderate overlap of dominant responses between the two cohorts, even when including the top three responses for each of the cue words (Hirsch and Tree, 2001, p. 7). Additionally, the study revealed that the younger participants showed more variability in their responses, but the two cohorts were rather similar in terms of response categories. Hirsch and Tree conclude that these results taken together suggest that there are no differences in the structure of the semantic memory between the two cohorts, but rather differences in the contents of the semantic memory systems of the two cohorts (Hirsch and Tree, 2001, p. 8).

Another study of age-related differences in word association behaviour was carried out by Fitzpatrick et al. (2015). Their study compared the word associations of 48 twin pairs aged 16 years, and 48 twin pairs who were 65 years or older (Fitzpatrick et al., 2015, p. 30). The participants were asked to respond to a list of 100 cue words, by writing down the first word that came to mind for each of them (Fitzpatrick et al., 2015, p. 30).[10] The study found differences in response behaviour between the two age cohorts. The participants performed better when their stereotypy scores[11] were calculated from a norms list created based on response data from their own age cohort, than when scores were calculated from a norms list based on the other age cohort (Fitzpatrick et al., 2015, p. 36). On average the participants gained 6.45 points when compared to a norms list based on data from their own age cohort, compared to when they were scored against the norms list for the other cohort (Fitzpatrick et al., 2015, p. 36). Fitzpatrick et al. suggested three possible reasons for the age-related differences in response behaviour. Firstly, the differences might be due to a change in word associations strategies with advancing age. A second possible reason is that the different generations prefer different sets of vocabulary and/or associations. Lastly, a third possible explanation is that the two cohorts' association behaviours are affected by both ageing and generation (Fitzpatrick et al., 2015, pp. 36-7).

2.1.5 Predictions for age differences in word association data

In line with both Hirsch and Tree (2001) and Fitzpatrick et al. (2015) differences in the dominant responses of the norms lists for the two cohorts would be expected. Since the current study follows a methodology closely related to that of Fitzpatrick et al. (2015), the amount of overlap might be larger with their study than with Hirsch and Tree (2001). However, one should be careful to assume that word associations are necessarily fully

[10] More on the methodology for Fitzpatrick et al. (2015) can be found in Chapter 4.

[11] i.e. the number of responses given by the participant that matched the dominant responses in a given norms list

compatible across languages. In this case, comparing two studies using English, and the current Norwegian study, there is for instance a chance that there might be some foreign language responses, especially English responses (at least for the younger cohort), due to the amount of English used in Norway, both in education and in society at large.

Based on the results of previous studies that found age-related differences in the dominant responses to cue words (e.g. Fitzpatrick et al. (2015) and Hirsch and Tree (2001)), we would expect there to be semantic differences across cohorts in the responses made. In this way, the word association test might contribute to the study of how the meaning and use of given words might change over time.

As some of the cue words in the current study overlap with the only previous large-scale word association study of Norwegian (see 2.2), a comparison with the data from this study (see 4.5.2 for methodology) might also be able to shed light on which of the hypotheses from Fitzpatrick et al. (2015, p. 36) on age-related differences in word association behaviour that is most viable. The participants in Håseth (1968) were around 20 years old when they were tested. Depending on which of these hypotheses that is used as a basis for the predictions, the data from Håseth (1968) should either be more similar to the younger cohort (if changes are a function of changed strategies with ageing), be more similar to the older cohort (if the norms are different due to generational differences), or if these two factors interact, the norms from Håseth's study might be different from both cohorts in the present study.

Further, if one follows the first hypothesis mentioned in 2.1.4, that "certain changes in WA selection strategies occur as a function of ageing" (Fitzpatrick et al., 2015, p. 36), it would also be predicted that one might see cohort differences not just in the norms lists and the actual responses, but also in the response category patterns of the two age groups. However, Hirsch and Tree (2001) found that the two age cohorts in their study produced quite similar response type patterns (p.7). Based on this, their study hypothesised that cohort differences are not due to changes in the structure of semantic memory with age, but rather due to differences in the stored contents in the semantic system (Hirsch and Tree, 2001, p. 8). Hirsch and Tree (2001) used a classification system based on the syntagmatic-paradigmatic division (Hirsch and Tree, 2001, p. 6).[12] The results from the current study might thus allow us to explore further whether the traditional classification system with paradigmatic or syntagmatic responses (or another binary classification system derived from these) might conceal differences in response behaviour that a more detailed system might reveal. Hirsch and Tree (2001) only used noun cues in their association test, which should also be taken into account when comparing the results.

[12]See 4.5.4 for more on the difference between the syntagmatic-paradigmatic division, and a more specific category system.

2.1.6 Word associations and mental lexicon research

The scope of word association data might not always be easy to grasp, as noted by De Deyne and Storms:

> [w]hile the free association task itself is straight forward, its unconstrained nature makes it difficult at first to grasp what association responses actually tell us about the lexicon De Deyne and Storms (2015, p. 468)

Despite this, response data from word association tests are by many researchers assumed to be representative of lexicosemantic connections, and may thus provide insight into the mental lexicon (Fitzpatrick and Izura, 2011, p. 378). For instance, word associations may be used to study factors that may affect storage in, and retrieval from, the mental lexicon, such as imagebility (e.g. de Groot (1989)), word class (e.g. Haastrup and Henriksen (2000)) and frequency (e.g. Higginbotham (2010)). For the role of these factors in theories of the mental lexicon, see 3.1.1.

The associations people give provide insights into the way we acquire words, and the way the lexicon is organised, although Schmitt, like Deyne and Storms, also notes that improvements are still needed in word association methodology before we can establish more precisely just what word association data may tell us (Schmitt, 2010, p. 248). Schmitt suggests that the degree of agreement in responses within a set of native speakers of a language implies a preferred organisation of the lexicon for the given language (Schmitt, 2010, p. 61). This has however been questioned by for instance Fitzpatrick et al. (2015), who found differences both between age cohorts, and between participants within the age cohorts used in the study. Schmitt also presents the recurrent finding that if one assumes that the syntagmatic-paradigmatic shift happens as children acquire their native language, this progression may mirror a gradual change in lexical organisation with age (however, see e.g. Nissen and Henriksen (2006) for a contrasting view) (Schmitt, 2010, pp. 61-2). Participants also tend to give responses that are linked to the cue word based on semantic fields (e.g. *farm-cows*) or common pairs (e.g. *tall-short*), and they often produce a response in the same word class as the cue. This might reflect categorical clusters in the lexicon, which again are linked to other clusters based on shared connections (Schmitt, 2010, p. 62).

The role and origin of word associations have been debated in research, with at least two prevalent points of view (De Deyne and Storms, 2015, p. 470). In terms of the link between memory systems and associations, De Deyne and Storms divide theories of semantic memory and word associations into two different positions. The first point of view sees the origin and role of word association as to mainly reflect word co-occurrence in language. In this view, association data would provide similar statistical patterns to

the patterns found in lexical co-occurrence. Co-occurrence does play a role in association, but studies have found only a moderate correlation between co-occurence in text and in association (De Deyne and Storms, 2015, p. 470). The responses given in word association tests are different from conversation, they are "simply the expression of thought"(De Deyne and Storms, 2015, p. 471), and this is how they might complement data based on spoken and written text (De Deyne and Storms, 2015). In an earlier study, De Deyne and Storms questioned whether word associations can be based purely on the participants' statistical knowledge of language, since common responses such as visual properties of the cue words seem unlikely to co-occur often in conversation whilst they do occur as associations. More research is needed on this topic (De Deyne and Storms, 2008, p. 228). The other way to look at word associations, which is supported by the findings from De Deyne and Storms, also believe that word associations to an extent reflect lexical co-occurence, but the responses also include an additional process which "gives rise to distinct mental properties that go beyond the information captured in written or spoken text" (De Deyne and Storms, 2015, p. 470). De Deyne and Storms (2015) reviewed research on the origin and role of word association and conclude in line with other research suggesting that word association tasks "[do] not reflect authentic language production, but should rather be seen as tapping into the semantic information of the mental lexicon" (De Deyne and Storms, 2015, p. 472).

In the 1960s, there was a change in the way one perceived word association strength. Prior to this, the focus had been on the strength between two words, for instance, there would in these terms be a strong association between *hammer* and *nail*. From the 1960s, research however shifted to studying the strength of associations based on the connections of the word to "the stimulus in the larger network of knowledge" (De Deyne and Storms, 2015, p. 466). Within such a framework, "two words have a similar meaning not because they are associated, but because they have many associates in common" (De Deyne and Storms, 2015, p. 467). In this view, "the meaning of a word is conveyed by the entire set of connections to the stimulus in the larger network of knowledge" (De Deyne and Storms, 2015, p. 466). Using the same framework, one might also quantify how closely linked two words are, based on their shared connections (De Deyne and Storms, 2015, p. 466). In terms of retrieval, the strength of the association (i.e. responses with a high frequency, such as *man-woman*) has been shown to influence the ability to recall words from episodic memory (where stronger associations appear to be memorised better than weak associations), to influence speed of recall, and to be linked to the effects of semantic priming, where strong associations work better as primes (De Deyne and Storms, 2015, pp. 465-6).

Word association data has also been used to model how lexical information is connected in the mental lexicon. One theory that has been used to model this is graph theory (e.g.

Wilks and Meara (2002)). Graph theory is a way to present and explore the connections that link one entity or agent to another. Graph theory has among other things been used to discover differences in density in the networks for L1 and L2 vocabularies (Wilks and Meara, 2002, pp. 306-7). In their study, Wilks and Meara, questioned the previous idea of density as the degree of points that are connected to one another in a network, and concluded that this definition is too simple (Wilks and Meara, 2002, p. 323). Rather they suggest that more factors need to be taken into account when studying network differences between L1 and L2 vocabularies. Wilks and Meara suggested that a measure of density should also take into account "inclusiveness" (i.e. the number of points in the graph that are not isolated), and "sum of degrees" (i.e. the number of links per point in the graph) (Wilks and Meara, 2002). Wilks and Meara suggested that by using a more complex definition of density, difference between two networks, such as an L1 and an L2 network, may be measured more accurately (Wilks and Meara, 2002, p. 323).

In recent research on word association and the mental lexicon, word association data has also been used to model the global structure of the mental lexicon network (De Deyne and Storms, 2015, pp. 465-7). This line of research has resulted in for instance studies that have found the mental lexicon to be structured as a *small world network* (e.g. De Deyne and Storms (2008)). Small world networks are "characterized by sparse connectivity, short average path lengths between words, and strong local clustering" (Steyvers and Tenenbaum, 2005, p. 41). In these kinds of networks, some words are more central and connected to more words than others. It is hypothesised that frequency is one of the factors that determines centrality, and it should thus be viable that there is a link between high frequency and a large number of associates (De Deyne and Storms, 2008, p. 214). De Deyne and Storms (2008) compared Dutch word association data to a set of different research lines on concepts and word associations, to see whether patterns and statistical data from word associations apply cross-linguistically, and, secondly, whether this study could provide insight into the origin and the status of word associations (De Deyne and Storms, 2008, p. 314). De Deyne and Storms (2008) considered three different aspects of word representations, which also characterise association behaviour. These were the influence of cue word class on responses, whether word associations fit into the *small world networks* topology, and lastly what the semantic properties of the association responses were (De Deyne and Storms, 2008, p. 214). De Deyne and Storms (2008) found that, in line with previous research, adjectives and verbs produce a majority of noun responses. Nouns also produce mainly nouns, which means that adjectives and verbs tend to elicit syntagmatic responses, whereas nouns tend to elicit more paradigmatic responses. De Deyne and Storms (2008) used word association data to create association networks for each cue word they used. They found that these networks fit with the small world network structure. In terms of cue centrality, the study found a tendency for central words (or

nodes) to be highly frequent, and to have an early age of acquisition (AoA) (De Deyne and Storms, 2008, p. 222). In terms of the semantic properties of the associations, their study showed that, for instance, taxonomic information becomes available prior to conceptual information (De Deyne and Storms, 2008, p. 228).

2.2 Norwegian research

There has been little published material on word association and association norms for Norwegian. The first and only major published set of norms is Kjell Johan Håseth's *Norske ord-assosiasjonsnormer* (Håseth, 1968). Håseth's study used Norwegian translations of Kent and Rosanoff's 100 cue words as cues. The methodology for the study was set to be closely related to Russell and Jenkins (1954), which is also similar to that of Palermo and Jenkins (1964). The participants were given the list of cue words and were told to write down the first word that came to mind for each of the words as quickly as possible (Håseth, 1968, p. 13). Håseth used 1000 participants in total, 500 women and 500 men, who were all fairly young with a mean age of 19.77 years (Håseth, 1968, p. 16). The participants had different levels of education and/or occupations. Although the study is called Norwegian word association norms,[13] Håseth stresses that the sample of participants cannot be said to be representative of the Norwegian population in general; however, it might be representative for the age group the participants belonged to (Håseth, 1968, p. 17). Håseth's study was purely meant to present word association norms for Norwegian, and no further analysis of the data is given except various ways of presenting the responses and their frequencies. The data from five of the cue words from Håseth (1968) will, as mentioned in 2.1.5), be compared to the data from the same words from the current study. The results from this comparison are presented in 5.1.2, and discussed in 6.1.2.

[13] My translation.

Chapter 3

Theoretical background

This chapter will present a theoretical background for the current study. Word associations have among other topics been used to research and model the mental lexicon (see 2.1.6), and to study age related changes in linguistic abilities (see 2.1.4). The chapter will provide a theoretical framework against which the response data from the word association test will be analysed. This will be done in order to attempt to answer the second and third sets of research questions, on how the data sheds light on theories on the mental lexicon, and on theories on language and ageing, respectively (see 1.2). A theoretical framework for the mental lexicon is presented in 3.1. A model for speech production is presented in 3.2, before theories of language and ageing are presented in 3.3. Predictions related to the second set of research questions are presented in 3.1.3, and for the third set of research questions in 3.3.2. The background for, and predictions for the first set of research questions, on whether the word associations are different in the two age cohorts, are presented in 2.1.4.

3.1 Theories of the mental lexicon

What is the mental lexicon? Jarema and Libben (2007) discuss the dual character of the lexicon, and the difficulties in characterising or defining the object of study:

> The metaphor of a mental lexicon implies a thing - commonly referred to as the dictionary represented in the mind, which allows individual language users to engage in everyday processes of language comprehension and production. Yet, the vast majority of psycholinguistic research on the mental lexicon involves the investigation of lexical processing, from which lexical representation is inferred. Thus, mental lexicon research is in practice the study of lexical activity (Jarema and Libben, 2007, p. 1).

Jarema and Libben (2007), as a potential solution to this issue of definition, propose a broad, unrestrictive definition of the mental lexicon, to focus and frame the core questions for continued research: "The mental lexicon is the cognitive system that constitutes the capacity for conscious and unconscious lexical activity" (Jarema and Libben, 2007, p. 2).

A number of different theories and models of the mental lexicon have been proposed. One line of division is between modular and non-modular theories. Modular theories (e.g. Ullman (2005)) assume two distinct capacities: a mental lexicon and a mental grammar. In modular theories, larger constructions (phrases, sentences and even inflected forms of a word) are not stored in the lexicon, but generated by rules from the mental grammar. The role of the lexicon is thus limited in these theories, as it only stores those forms that cannot be generated by rules. In non-modular theories, there is no division between lexicon and grammar. A usage-based approach (e.g. Bybee (2008, 2010)), is compatible with non-modular theories. In usage-based theory, what would be rules in modular theories are generalisations over a person's experiences with language. The mental lexicon is organised as a network, and in these theories, all word forms are stored, and larger (multi-word) units may also be stored as larger chunks. One of the research questions of the current study is to see how word association data sheds light on theories of the mental lexicon, and what the collected data might tell us about the organisation of the lexicon (see the second set of questions, 1.2). To do this, the study will adopt a cognitive linguistic approach, through usage-based theory and cognitive grammar.

3.1.1 Usage-based theory

Usage-based theory (e.g. Bybee (2008, 2010), Langacker (2000)) sees language as "an embodied and social human behaviour" (Bybee and Beckner, 2010, p. 827). In this theoretical framework, "grammar is viewed as the cognitive organization of one's experience with language" (Bybee, 2008, p. 216). Usage-based theory is maximalist, non-reductive, and bottom-up (Langacker, 2000, p. 1). In usage-based theory, usage has effects on structure, and evidence from language usage and experience are used to understand and explain the cognitive organisation of language (Bybee and Beckner, 2010, p. 827). As mentioned, there is no separation of lexicon and grammar. Rather, all linguistic knowledge is stored in the same associative network (Bybee, 2008, p. 216). Since linguistic abilites are based on experience in usage-based theory, and the theory assumes a minimum of language-specific structures in the brain, "becoming a fluent speaker involves prodigious amounts of actual learning" (Langacker, 2000, p. 2).

Usage-based theory is domain-general, and thus the processes used for language are also involved in other areas of human cognition. Such general processes involved in language

include categorisation, chunking, rich memory storage, analogy, and cross-model association (Bybee, 2010, p. 7). In usage-based theory, learning a language is similar to acquiring for example motor skills, like learning how to ride a bike. Both procedural and declarative memory are used in the language acquisition process. Procedural memory is the memory of how to do things, such as how to tie ones shoes. One is not consciously aware of procedural memories. Declarative memory may be consciously recalled, and are long-term memories of facts and events, such as the fact that Oslo is the capital of Norway, or the memory of one's first day at school. Procedural memory is used for learning through repetition. Implicit knowledge of frequency and other statistical patterns in language are slowly built up through the use of domain-general processes. Declarative memory is also somewhat involved in language, as the meaning of lexical items is linked to declarative knowledge; however, one cannot access or use these items appropriately without procedural knowledge (Bybee, 2008, p. 220).

As mentioned, usage-based theory is nonreductive and non-minimalist, and the theory assumes that detailed experiences of language are stored as exemplars, side by side with generalisations and schemas (Bybee and Beckner, 2010, p. 833). Exemplars are "built up from tokens of language experience that are deemed to be identical" (Bybee, 2010, p. 7). Additionally, non-linguistic memories may also affect cognitive representations (Bybee, 2010, p. 8).

> [I]n the current framework we are interested in how the specific experiences speakers have with language combine to yield more general patterns, and how the specific and general interact in acquisition, processing, and language change (Bybee and Beckner, 2010, p. 833).

A speaker's cognitive representation of language is mapped in a network, based on experience. This process happens through categorisation, where, if a previously experienced exemplar is experienced again, it becomes strengthened. New exemplars are placed in relation to existing exemplars, based on similarities and differences (Bybee and Beckner, 2010, pp. 833-4). Exemplar representations may potentially contain "all the information a language user can perceive of linguistic experience", from phonetic detail, to meaning, context of use, and more (Bybee, 2010, p. 14). The network has relations between words along both phonetic, semantic, and pragmatic dimensions (Bybee, 2008, p. 216). Each word in the lexicon has a "cloud or cluster of phonetic exemplars representing the phonetic variants of the word," in addition to information on linguistic and social context for each variant (Bybee and Beckner, 2010, p. 833). Likewise, there are also meaning clusters, with each word linked to exemplars for the meanings and contexts of each token of use (Bybee and Beckner, 2010, p. 833). The exemplars are connected along various levels and dimensions (Bybee, 2010, p. 22). A word might be linked to other words in different ways, through exemplars in the phonetic exemplar cluster linked to the word,

or in the semantic exemplar cluster linked to the word (Bybee, 2010, p. 22). In this way, for example the word *beds* might be linked phonologically to a word such as *legs*, through their shared phoneme representing the plural morpheme, in [bed**z**] and [leg**z**]. The same two exemplars are also linked through sharing the same vowel, /e/, and through both consisting of only one syllable. The two are also semantically linked through by being plurals, and additionally also likely linked semantically through *bed legs* in a given person's network. However, *beds* would also have an additional set of different connected meaning exemplars or context exemplars than *legs*. In such a network, the production of a word consists of choosing the meaning one wants to convey and a word that expresses this meaning, and then finding the phoentic exemplar that best suits the intended context of use (Bybee and Beckner, 2010, p. 833).

The *construction* is the unit for morphological and syntactic representation in usage-based grammar, and constructions may be stored as whole units in the network. A construction in this theory is defined as a "form-meaning pairing that has sequential structure and may include positions that are fixed as well as positions that are open" (Bybee, 2010, p. 9). Constructions can be any strings of morphemes or words, from simple lexical words (*book, angry, run*), to varying degrees of abstract constructions, such as partially filled idioms, for instance the *drive someone crazy* construction, which can be changed as long as it follows the following structure:

SUBJECT + [DRIVE] + X + [adjective or prepositional phrase meaning *crazy*]

Examples of this might be *They drive me mad* or *The sound of the vacuum cleaner used to drive Louise crazy.* Constructions are produced by *chunking*, which is a property of procedural knowledge. Chunking is the process when

> sequences of actions that occur together repeatedly are chunked into a single action[...] In usage-based theory, constructions are chunks —neuromotor routines—with moveable parts (Bybee, 2008, p. 220).

In the network, all types of constructions may be represented, so that both compositionally non-transparent multi-word units such as idioms, and conventionalised, transparent constructions, so-called "prefabs",[14] may all be represented by exemplars (Bybee and Beckner, 2010, pp. 834-5). The structure of the network and its links is such that one might access these constructions as a whole, or access any of their components. For example, one might access the idiom *a hot potato* as a whole, or just access *hot* or *potato* (Bybee and Beckner, 2010, p. 836). An idiom stored as a whole unit will also be linked to a different set of other exemplars than would *hot* or *potato* on their own.

[14]Examples of prefabs are for instance *look up*, or *applied for*, or *I can remember when.* Prefabs are established by repetition as a conventional way to express a certain notion, however they do not need to be highly frequent (Bybee, 2008, 2010).

As linguistic knowledge is based on linguistic experience in usage-based theory, frequency naturally has an effect on linguistic representation. In terms of network representation, frequency is an important factor influencing both which exemplars are represented in the network, and for how they are connected. *Token frequency* is the number of times a unit appears in running text. A unit may be small, as a single consonant, or larger, as an entire sentence (Bybee, 2008, p. 218). According to usage-based theory, repetition has the same effect on language as it has on other areas of cognition (Bybee and Beckner, 2010, p. 829). People have a cognitive ability to learn which events that tend to co-occur, and whether certain combinations of events are more common than others. This can also be found in our linguistic abilities, where people are able to learn linguistic patterns based on relatively little input (Bybee and Beckner, 2010, p. 830). As mentioned, the network that stores one's cognitive representation of language is built up by categorisation of experienced exemplars. Highly frequent exemplars will have a stronger representation in the network (Bybee and Beckner, 2010, p. 833). Frequent exemplars as also likely yo be linked to more other, similar exemplars in the network than exemplars with a lower frequency.[15] Words that have a high frequency of co-occurence, such as *salt* and *pepper*, could both be stored as a unit, and as two separate exemplars that would be connected by strong links due to the two often appearing together in the same context. As each instance of experience with an exemplar affects the network, it will change throughout the lifespan, and although the effect of a new token of experience is likely to have a greater effect for children, or for someone acquiring a second language, all tokens have some effect on representation throughout adult life (Bybee, 2010, pp. 21-2). If a word continues to be frequently experienced throughout the life of a person, it would potentially result in a very large set of connecting exemplars in the network for the given word.

3.1.2 Cognitive grammar

Cognitive grammar (e.g. Langacker (2008) and Taylor (2002)), is based on the idea of language as symbolic, i.e. that language is "essentially a means for relating semantic and phonological structures" (Taylor, 2002, p. 23). Cognitive grammar is usage based, and compatible with the presented usage-based theory (3.1.1). The usage-based orientation of cognitive grammar means that the theory is sceptical of underlying structures, it is very much surface oriented, and any abstract entities might only exist if they are abstracted from actual language usage (Taylor, 2002, p. 28).

In cognitive grammar, essential grammatical notions, such as nouns and verbs, are characterised semantically, both at the prototype level and at the schema level (Langacker,

[15] See also 2.1.6 for the findings from De Deyne and Storms (2008) that central nodes in a small world network constructed from word association data tend to be highly frequent (and to have an early AoA).

2008, p. 103). In terms of prototypes, nouns and verbs have contrasting basic properties. For one, archetypical nouns are physical objects, residing primarily in space, without any particular location in time, where they might continue to exist indefinitely. Verbs, on the other hand, are not physical, but change and transfer energy. They are temporally bounded, and have a diffuse location in space (Langacker, 2008, p. 104). Due to these differences, nouns are conceptually autonomous, and may be conceptualised on their own. Verbs, however, are conceptually dependent and can only be conceptualised if those participants who interact in the event, are also conceptualised (Langacker, 2008, p. 104).

In cognitive grammar, words are symbolic units where a phonological structure is associated with a semantic structure. Both of these structures may be part of schema-instance relations (Taylor, 2002, p. 165). For example, the word *dog* consists of the concept [DOG] and the phonological structure [dɒg]. [DOG] is schematic for such concepts as [POODLE], [BEAGLE] and [GREYHOUND], or a specific dog. [DOG] is also itself an instance of other schematic concepts, such as [ANIMAL]. In terms of phonology, [dɒg] is an instance of [SYLLABLE], and is also schematic for variations of pronunciations of that syllable (Taylor, 2002, p. 165).

Because the theory assumes that one learns language through experience, one will encounter different usage events that might differ in context, pronunciation and/or concept. The usage-events are generalised upon to produce a schematic concept and a schematic phonological representation for a given word. "The word *dog*" is an example of such a decontextualised schematic unit (Taylor, 2002, p. 165). It is however very rare that a symbolic unit has a single semantic representation, and the symbolic unit should thus perhaps rather be regarded as "an association of a network of semantic representations with a network of phonological representations" (Taylor, 2002, p. 461). Through language experience, extensions are added to the prototype concept for a symbolic unit, and with extensive experience this may result in highly complex networks (Taylor, 2002, p. 465). Given the manner in which the network is constructed, one would anticipate nouns and verbs to occur in different kinds of networks, with different types of associations.

Storage and processing of words are also affected by other factors such as imageability (e.g. Paivio, Yuille, and Madigan (1968) and Prado and Ullman (2009)). Imageability is the ease or difficulty with which a word arouses a mental image, sound, or other sensory experience (Paivio, Yuille, and Madigan, 1968). A person's imageability rating for a given word will also differ based on experience. Imageability ratings have been shown to differ between nouns and verbs, where nouns tend to be more imageable than verbs (Simonsen et al., 2013, p. 436). This fits into the characteristics of the two words classes in cognitive grammar, where nouns are conceptually autonomous and self-contained, in addition to prototypically denoting things (Simonsen et al., 2013, p. 443).

3.1.3 Predictions on word associations and the mental lexicon

On the basis of usage-based theory, several predictions can be made in relation to the second set of research questions: How does the data collected in the current study shed light on theories on the mental lexicon? What might the data tell us about the organisation of the mental lexicon, and of storage and access in particular? (1.2). Firstly, the role of experience in usage-based theory should lead to individual differences in the network, and thus also to individual differences in responses. One would predict to see this both in terms of actual response words, and also in terms of response categories, as the contexts of linguistic experience would affect the structures of the participants' networks. This would also be in line with the findings from Fitzpatrick et al. (2015), who found variation between participants in the types of response categories produced, but consistency over time for individuals. It is also likely that this same mechanism may show generational differences across the cohorts, as certain exemplars will have had different frequencies of use across time.

The importance of frequency in usage-based theory is additionally likely to be mirrored in the dominant responses in the norms lists. Certain exemplars have a high frequency of co-occurrence, and will be strongly represented in the networks of many of the participants. Thus, if presented with for instance the cue word *bread*, it is likely that the participant has experienced the word often in a collocation with the word *butter*. In this way, according to usage-based theory, the exemplar for *bread* in a collocation with *butter* will be strong. It is predicted that the strength of the exemplar might affect access and retrieval, and that *butter* thus will be more likely to be given as a response than other connected exemplars.

As multi-word constructions might be stored as units in the network, this might also lead to the participants' giving multi-word responses if for instance an idiom is a fairly frequent replacement for a single word, such as for instance *kick the bucket* and *die*. As mentioned in 2.1.6, strong associations can also be used as an indicator of the number of connections between words in the network (De Deyne and Storms, 2015, p. 466). This will be further discussed in Chapter 6.

In terms of how the structure of the network might be reflected in the results, the following predictions are made: Firstly, as the nodes in the network may be linked on several different levels (i.e. phonological, semantic, and so on), one would expect to see associations on different levels for the same cue word. Thus, if *couch* was given as a cue, one would expect to see semantic responses such as *sofa* (synonym), or *chair* (lexical set), but perhaps also morphologically based responses, (*couch**es***), or perhaps even in some cases phonologically based responses, such as *pouch*. Associations might also reflect

co-occurence frequency, with responses such as *potato* ('couch potato'.[16]) Secondly, the tendencies presented by Schmitt (2010) (see 2.1.6) that a preference for certain response types might reflect categorical clusters in the network (Schmitt, 2010, p. 62), should lead to meaning-based responses being the most frequent response categories, as these common responses are often based on semantic properties, such as to belong to the same semantic field, or belong to the same word class. Thirdly, the fundamental differences between nouns and verbs in cognitive grammar are assumed to lead to several kinds of differences between the responses given to noun cues compared to those given to verb cues. For one, as nouns are conceptually autonomous, one would expect to see more meaning-based responses for nouns than for verbs, at least lexical sets and collocations. For synonyms, however, one might expect to see a higher number of verbs than nouns, as noun-synonyms are rare, but verbs on the other hand might overlap in meaning, at least in specific contexts. Since verbs need other participants to be conceptualised, one would expect to see more position-based responses to verbs. This would also be likely to mean that nouns would elicit more noun responses, than verbs elicit verb responses, as the other participants needed to conceptualise a verb would necessarily belong to other word classes. Similar hypotheses have also be made by e.g. Nissen and Henriksen (2006), who assumed that noun cues would trigger more paradigmatic responses than verbs or adjectives due to the way nouns are often organised in hierarchies (p. 402).

Additionally, properties of the cue word, such as imageability, are also predicted to affect responses. It may be assumed that a high imageability rating might cause less blank responses than cue with a low imageability rating, as highly imageable words might be easier to access, and associate from due to their richer representation. It may also be predicted that the imageability ratings of the cue words might cause different response category patterns, as there might be storage differences between words with different levels of imageability.

3.2 Modelling language processing and production

Word associations involve both processing and production of language: firstly the processing of the cue, then the activation of the associated response, and lastly the production of the response. Theories of lexical access and/or production have usually taken one of two positions: *discrete* theories or *cascaded/interactive* theories (Schiller and Verdonschot, 2015, p. 487). Discrete theories (e.g. Levelt (1989)) assume that the stages of the process happen in temporal succession. Cascading models (e.g. Dell (1986)) on the other hand postulate that different levels can be activated from early on in the process

[16]I.e. someone who spends leisure time idly sitting around, often watching a lot of television.

(Schiller and Verdonschot, 2015, p. 483). Research has found empirical evidence for both discrete and cascading models, which has led some researchers to proposing *limited cascading* models (e.g. Bonin et al. (2012) and Roelofs (2008)) (Schiller and Verdonschot, 2015, p. 487).

Based on the predictions made above in 3.1.3, that associations will occur on all levels of representation, a cascading model appears to be most fitting. As associations are the first word that comes to mind when presented with a cue, this postulates that all levels should be able to be activated very early in the process. Two cascading models will be presented here, Lamb's model for relational network theory (RNT) (Lamb, 2000), and Dell's spreading activation model (Dell, 1986). Additionally, a logogen model (Patterson and Shewell, 1987) will be presented first, as it provides a schematic overview of all components needed from reading or hearing a cue, to writing or uttering a response.

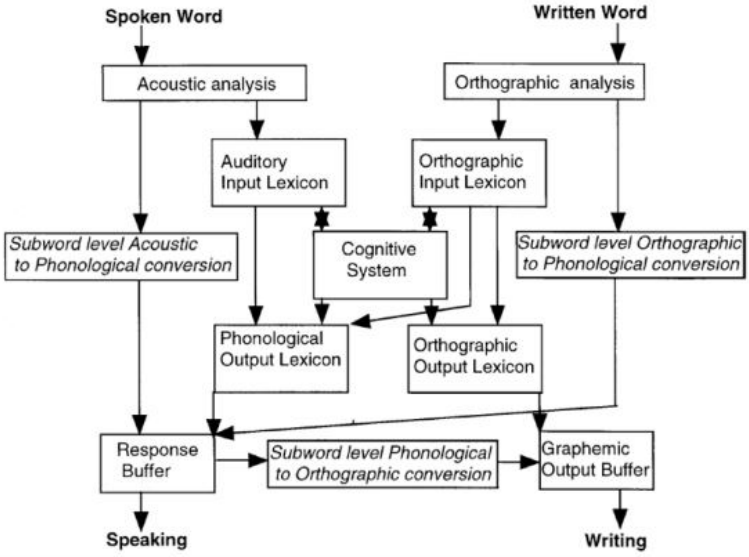


Figure 3.1: Language processing model for single words, from Patterson and Shewell (1987)

The logogen model of language production for single words, proposed in Patterson and Shewell (1987), as mentioned, incorporates the process from seeing or hearing a word, to speaking or writing a word in response. As Whitworth et al. point out, such models are underspecified, and each box in the model might contain a whole set of new processes (Whitworth, Webster, and Howard, 2014, p. 6). However, for the current purposes, the model provides a useful basis for the further discussion of the collected data, as it, for one, focuses on the processing and production of *words*, rather than larger units such as sentences,[17] and secondly, that it models the process all the way from a word is read or heard, to the response is written or spoken. This is valuable as word association studies might be performed both written and orally. There are however many factors needed to

[17]Although it is likely that some responses will be larger, multi-word units, most of the responses will be single words.

explain word associations that are not part of Patterson and Shewell's model. For one, as previously mentioned, the components of the model are not explained to any further extent than that they are included because they are believed to be needed in the process (Whitworth, Webster, and Howard, 2014, p. 4). For the purposes of the current study, the processes within the cognitive system are particularly important. Word associations have been shown to happen on different levels of knowledge of a word (phonological, semantic, syntactic, and so on), and associations have also been shown to differ based on various properties of the cue words given in the word association test (see 2.1.4 and 2.1.6). A more detailed model is thus needed to shed light on the processes within the cognitive system.

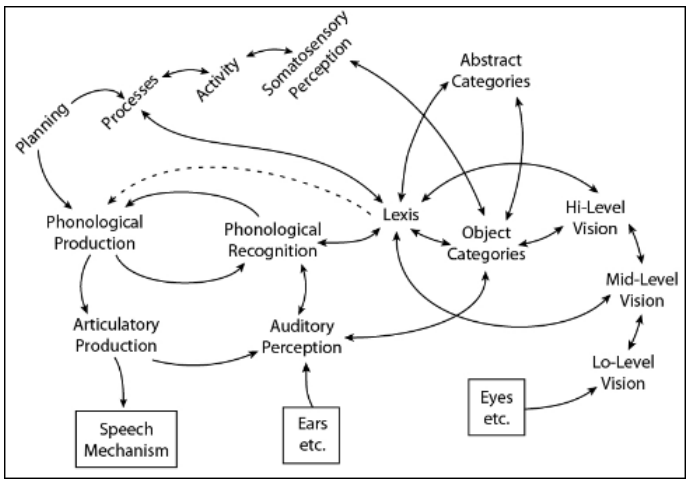


Figure 3.2: The linguistic system and related cognitive subsystems, from Lamb (2000)

Lamb's relational network model (RNT) is a neurocognitive, usage-based model, that accounts for both comprehension and production for speaking as well as writing (Lamb, 2000). The linguistic system works in cooperation with other related cognitive systems in the brain (see figure 3.2) (Lamb, 2000, p. 111)). The model assumes that linguistic information is stored in the connectivity of the network, and learning consists of "building new connections, strengthening existing connections, and adjusting thresholds" (Lamb, 2000, p. 101). A relational network is thus "the end-result of a long learning process" (Lamb, 2000, p. 104). The network is organised with *lines* and different types of *nodes* as the elements on one level. On the next level, the lines and nodes make up *nections* (see figure 3.3). The network is built up through connections between modules, where the modules are nections (Lamb, 2000, pp. 101-2). In the process of understanding and producing speech or writing, both incoming and outgoing activation is based on the strength of the incoming activation, controlled by a threshold function. The threshold is adjusted based on experience of the given item. There is no postulated limit for the amount of nections that may work in parallel (Lamb, 2000, pp. 101-3). As the model is usage-based and cognitive, with no separation of lexicon and grammar, simultaneous processing on different levels is less demanding, as there is no need for serial processing

within a level to compile units (Lamb, 2000, p. 104).

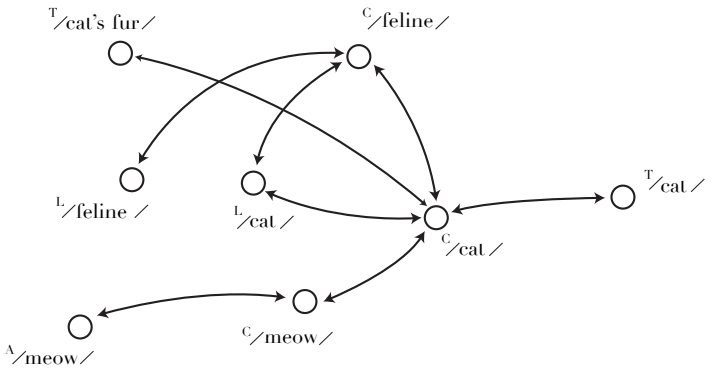


Figure 3.3: A nection for the conceptual category CAT, with some connections to other nections, from Lamb (2000). The subscripts denote the following: A- auditory, C-conceptual, L-lexical, T-tactile, and V-visual.

The relational network model is bidirectional. Understanding speech or reading and producing speech or writing have different subsystems, but they are also interconnected (Lamb, 2000, p. 113). Activation in the network goes back and forth between levels of integration. This means that connections are strengthened not just through experience, but also through downward activation in the network (Lamb, 2000, p. 113). In this bidirectional activation, in particular in the backward activation, a whole set of nections representing the properties of a category might be activated, where only one or some properties are relevant for the current activation. This might explain why a person might choose a "false property" for a given item (Lamb, 2000, p. 115).

A similar relational network model is Dell's spreading activation model (Dell, 1986) (figure 3.4). Whereas the RNT provides a cognitive view of the entire process from understanding to producing language, Dell's model focuses on language production. It is included here because it focuses in particular on slips or errors, which are assumed to occur in the data, for instance in two-step associations, where something might have gone wrong in either the understanding of the cue or the production of the response. Dell's model also has a particular focus on the simultaneous activation of the different levels of linguistic knowledge. This is in line with the prediction made in 3.1.3 that associations would occur on all levels. The model presented in Dell (1986) is a parallel model, which assumes four levels of processing that take place simultaneously during language production: semantic, syntactic, morphological, and phonological (Carroll, 2008, p. 204). The levels have separate representations of what the speaker wants to convey, but they still work in parallel. When a node gets activated it may activate other nodes, both on its own, and on other levels (Carroll, 2008, p. 204). The representations on a given level might be envisaged as "a collection of order tags that are attached to nodes in the lexical network, dictating

the contents of the representation and their order" (Dell, 1986, p. 287). Activation in the model is bidirectional, and there might thus be interaction between the semantic and the phonological properties of the word (Fromkin and Ratner, 1998, p. 337). Errors may occur when two lexical items share semantic and/or phonological representations, and the spreading activation *slips* to the unintended item, such as represented by figure 3.4.

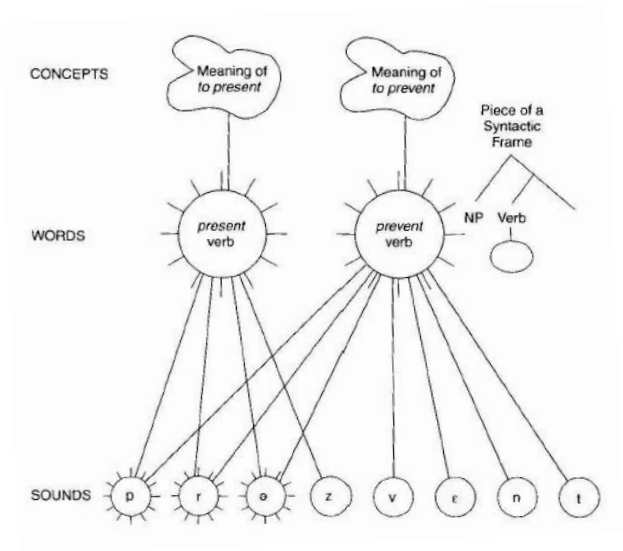


Figure 3.4: A case of spreading activation which shows how present *might slip to* prevent*, based on the spreading activation in Dell (1986). Figure taken from Fromkin and Ratner (1998, p. 337).*

Dell's spreading activation model should also, in light of the aims of this thesis to shed light on theories on language and healthy ageing, be compared to the transmission deficit hypothesis (TDH) presented in 3.3.1. The TDH assumes, in production, a process more similar to that predicted by discrete models of lexical access and language production, as it proposes that firstly nodes in the semantic system are activated, prior to the activation of nodes in the phonological system. Data from the current study will be compared to both the two cascading models presented above, and to the TDH in the discussion in Chapter 6.

The models presented in this section will be used in the discussion of the data from the word association test (chapter 6) to try to shed light on what processes take place in the network when a person is asked to associate to a given cue word.

3.3 Language and ageing

With age, changes in language are seen in lexical retrieval, in sentence processing, and one may see changes in discourse patterns (Obler and Pekkala, 2008, p. 351). Older adults may

also experience problems linked to ageism[18] (Clark-Cotton et al., 2007, p. 1). Although studies have found linguistic decline in normal ageing, Clark-Cotton et al. emphasise that most these changes are often barely, if at all, noticeable outside of an experimental setting, and that they usually do not affect communication (Clark-Cotton et al., 2007, p. 2). Word-retrieval difficulties are though, as presented below, often reported by elderly persons, and might be experienced in everyday life. It should also be pointed out that the changes that happen in speech during adulthood are much smaller than those that happen during childhood language acquisition, or when one acquires a new language (Mortensen, Meyer, and Humphreys, 2006, p. 238).

Word-retrieval difficulties are often reported among and by the elderly, especially problems with retrieving nouns. Research on language and normal ageing has thus often used various object naming tasks, such as the Boston Naming Test (BNT), in order to reveal what changes occur in linguistic abilities as we get older (Nicholas et al., 1997). In general, older adults have been found to both be less accurate, and to be slower in naming test performance (Mortensen, Meyer, and Humphreys, 2006, p. 252). In naming tests there is no discourse to help the participant retrieve the word that is being looked for, and no way to avoid any difficult words, and thus retrieval issues are made highly visible. In the BNT, the participants are asked to name items depicted by line drawings, with different levels of difficulty. If the participant is unable to name an object, standardised phonemic or semantic prompts might be given. Phonemic prompts involve giving the participant the initial sound of the word, and semantic prompts involve giving for example the category membership of the item (Nicholas et al., 1997, p. 167). Studies using the BNT across age cohorts have found that older adults name fewer pictures accurately than younger participants. The older group also has been shown to need more phonemic prompts. However, once they are given the prompt, they are equally helped by it as the younger group. The fact that semantic and/or phonemic prompts tend to aid retrieval shows that knowledge of the word is intact, even if one cannot access its phonological form. This suggests that semantic abilities remain intact throughout healthy ageing (Clark-Cotton et al., 2007, pp. 2-4). Similarly, studies have shown that difficulties with lexical retrieval mainly affects the ability to retrieve specific words, such as in naming tests, whereas if asked to give a definition for a given word, few age-related changes are seen (Obler and Pekkala, 2008, p. 351). This also supports the hypothesis that semantic abilities do not generally decline in older age. Note should be made that not all studies using naming tests have found statistically significant age-related changes in naming; however, the older participants have performed worse than the younger ones in those studies, too (Nicholas

[18] An example of this is the use of *elderspeak*; i.e. when older adults are spoken to in the same manner as one would speak to young children, with slower and simplified speak, and exaggerated intonation. These efforts might of course be helpful, too, but the older person might also find the use of elderspeak patronising (Wingfield and Stine-Morrow, 2000, pp. 403-4).

et al., 1997, p. 168).

In terms of the slope of the decline in normal ageing, studies have found a slight decrease throughout adulthood (Wingfield and Stine-Morrow, 2000, p. 374), perhaps from as early on as the late 30s (Goral et al., 2007). The decline in correct performance in picture-naming tests has been shown to become more rapid with increasing age (Goral et al., 2007, p. 231).

Also linked to lexical retrieval problems is the tendency for older adults to experience more of the tip-of-the-tongue phenomenon (TOT), i.e. failure to retrieve a word that is temporarily inaccessible, while still being aware of certain properties of the word. The TOT phenomenon has frequently been studied in relation to difficulties with word retrieval in ageing. Older adults tend to experience more TOTs, and they tend to have less access to information about the word they are trying to access (Nicholas et al., 1997, p. 169). TOTs are further discussed below in the section on hypotheses on language in ageing (3.3.1).

In terms of discourse production, older persons tend to produce more words than younger people. In particular in narrative production tasks, older participants tend to give longer, less dense responses, with more irrelevant information included (Obler and Pekkala, 2008, p. 352). Older adults have also been shown to produce more disfluencies, which has been linked to retrival problems (Mortensen, Meyer, and Humphreys, 2006, p. 247). The use of more wordy responses among older adults has been suggested to be a way to "mask difficulties in conceptual planning or in selecting or retrieving individual words" (Mortensen, Meyer, and Humphreys, 2006, p. 251). It has also been suggested that the linguistic changes seen in older age are due to a decrease in the ability to inhibit irrelevant information (Obler and Pekkala, 2008, p. 356).

Comprehension problems by age are seen in complex structures, such as at the sentence level and text level. These problems appear to be caused by not only linguistic factors, but to be related to a combination of linguistic changes and other cognitive changes due to ageing (e.g. attention, memory and processing speed) (Obler and Pekkala, 2008, p. 352).

3.3.1 Hypotheses on language in ageing

Burke, MacKay, and James (2000, p. 208) divide theories applicable for language and ageing into two main categories: information-universal and information-specific. Information-universal theories include *general slowing* theories, where "the speed of executing cognitive operations decreases with ageing regardless of the task or the mental operations involved in the task" (Burke, MacKay, and James, 2000, p. 208). In terms of language, Burke,

MacKay, and James see general slowing theories as problematic for explaining age-related language change due to two issues. Firstly, deficits in language performance are not universal, and secondly, statistic analysis has shown that the slowing factor varies, even within a domain (Burke, MacKay, and James, 2000, p. 209). Another information-universal theory is the *inhibition deficit* hypothesis, which assumes that age weakens inhibitory processes, which causes older adults to both activate, and later be unable to suppress, irrelevant information more often than younger adults. Inhibition is weakened in all cognitive areas (Burke, MacKay, and James, 2000, pp. 209-10). However, studies have proved inconsistent in proving the hypothesis (Burke, MacKay, and James, 2000, pp. 210-11).

Burke, MacKay, and James (2000) suggest that unlike in the information-universal theories, asymmetries in linguistic ageing may be accounted for by the use of information-specific theories. One such theory is the *transmission deficit hypothesis* (TDH) (e.g. Burke, MacKay, and James (2000)) (see figure 3.5). The hypothesis attempts to account for the asymmetric effects of ageing on linguistic abilities, with comprehension remaining largely stable throughout the lifespan, whereas production tends to show a decline (Burke, MacKay, and James, 2000, pp. 204-5). Tests involving semantic priming have shown that the semantic processes needed for comprehension remain intact throughout the lifespan. This has been found for both direct priming (*student*—> *teacher*), and through an unpresented word as a link between the prime and the target (*lion*—> (*tiger*) —> *stripes*) (Burke, MacKay, and James, 2000, p. 205). According to the hypothesis, "language perception and production depend on how fast and how much priming can be transmitted across the connections linking representational units, called nodes, in the language-memory system" (Burke, MacKay, and James, 2000, p. 212). These links make a chain from concepts to lemmas to word forms. Concepts and lemmas are represented by nodes in the semantic system, whereas word forms are represented by nodes in the phonological system (for spoken words) and the orthograpic system (for written words) (Burke and Shafto, 2004)

The TDH explains the age-related linguistic problems by a weakening of the strength in the links between the nodes in the various stages of language production. For speakers of all ages, the connections between nodes might become weaker due to infrequent or non-recent use (Nicholas et al., 1997, p. 175). However, for older persons, the TDH states that there is an additional age-related general weakening of these connections. This leads to reduced priming, which affects retrieval (Mortensen, Meyer, and Humphreys, 2006, p. 242). In the semantic system there are multiple connections to the lexical nodes; however, from the lexical nodes (the lemma form) to the syllable nodes (the lexeme/word form), there are only single connections to each syllable needed, and thus this step of the activation is more vulnerable when the connections get weaker. Thus, with the weaker connections due to age, this might prevent word form retrieval (Mortensen, Meyer, and

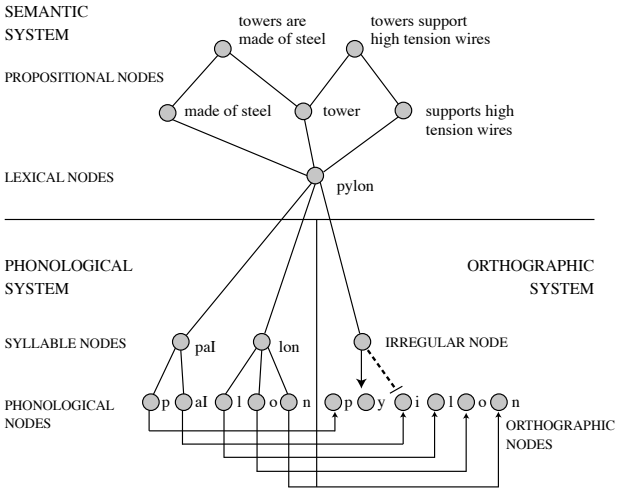


Figure 3.5: A model of the transmission deficit hypothesis *from Burke and Shafto (2004)*

Humphreys, 2006, p. 242). Proper names are even more vulnerable than other words because a proper name lemma does not have direct connections to conceptual features, but rather, the retrieval has to go via a proper name phrase. In this way, proper names are more vulnerable to the weakened connections, because there are more connections that can fail to be activated, and additionally, there is only one connection between the name phrase and the lemma (Mortensen, Meyer, and Humphreys, 2006, p. 242).

The fact that older people experience more tip-of-the-tongue-effects may also be explained by the weaker connections between nodes (Nicholas et al., 1997). One feels that one knows the word because the semantic and lexical representations are activated, but one cannot activate the phonological information about the word (Burke and Shafto, 2004, p. 22). The hypothesis also predicts a parallel effect in spelling to what is seen in phonological production, due to the fact that both phonology and orthography are connected to the lemma in the same manner (Burke and Shafto, 2004, p. 23). Nicholas et al. also suggest that the hypothesis might explain retrieval differences between nouns and verbs. The two word classes are likely to have a different sets of associate connections, with verbs being linked to more words from different word classes, which again should lead to a greater chance of success in retrieving verbs than nouns (Nicholas et al., 1997, pp. 175-6).

Importantly, the TDH may also explain why comprehension does not decline when production does. According to the hypothesis, the transmission deficits that happen with age do not affect comprehension and production in the same way. Comprehension is less vulnerable because there are several connections from phonemes going to the lemma

(Mortensen, Meyer, and Humphreys, 2006, p. 274). Mortensen, Meyer, and Humphreys reviewed research on age-related effects on speech production. They concluded that in naming in older age, object recognition, lemma selection, and phonological representations of word forms stay intact, but there are issues in the transfers from lemma to phonological form (Mortensen, Meyer, and Humphreys, 2006, pp. 271-2). These findings fit with the TDH, as do the results in Goral et al., who link their findings of age-related decline in picture naming to the postulated production problems with age in TDH (Goral et al., 2007, p. 231).

3.3.2 Predictions on word associations and theories on language and ageing

As presented in the section above, a main linguistic issue in older age is lexical retrieval, as seen in for instance object naming. The word association data collected for the present study might be used to research this further (see question 3 in 1.2). Context and prompts have been shown to aid retrieval; however, as the cue words in the association test occur in isolation, one would assume that there would be a higher frequency of blank responses among the older cohort. As previous studies have shown that semantic abilities remain intact in older age, difficulties in the retrieval of a specific word might also lead the older cohort to produce more multi-word responses to explain the words they are unable to retrieve. However, multi-word responses might also be expected to occur more frequently among the older cohort based on the tendency for older people to be more verbose.

Like the findings of the studies presented above, the TDH assumes that semantic skills remain intact in older age, but that production is affected. Thus, the TDH would also be supported by a finding of more blank and/or multi-word responses in the older cohort, caused by the same reasons as stated above.

Nicholas et al. (1997) suggested that the TDH might explain the observed differences in retrieval between nouns and verbs, and it may also explain how nouns are more prone to retrieval difficulties, as they have a smaller number of connections. However, as mentioned, the TDH may also explain how comprehension is not impaired, due to more connections going from phonemes to the lemma, than in the opposite direction (Mortensen, Meyer, and Humphreys, 2006, p. 274). It is thus predicted, based on the TDH, that the word class of the cue in itself should *not* affect the number of blank or multi-word responses produced. However, if the prediction made in 3.1.3 that nouns will elicit mostly nouns holds, then there might be a difference in the number of blank and/or multi-word responses after all.[19]

[19]For this to happen, nouns will also not just need to trigger a high proportion of noun responses, but noun cues must also trigger more noun responses than any of the other cue word classes.

Additionally, already mentioned in 3.1.3, and as will be discussed in 6.2.2, it is predicted that other properties of the cue word, such as imageability, will affect the number of blanks and multi-word responses. This means that there is a chance that it might be difficult to determine the exact cause of any potential differences in the rate of occurence of multi-word or blank responses.

Chapter 4

Methodology and data collection

The following chapter presents the methodology used in the collection and processing of data in the current study. This methodology is largely based on the methodology used in Fitzpatrick et al. (2015). Although it was an aim to follow Fitzpatrick et al. (2015) as closely as possible, certain parts of the methodology was emphasised above others in order to best explore aspects relevant to the research questions for the present study (see 1.2).

The present chapter is structured as follows: The first section (4.1) describes the process of selecting the cue words for the word association test. Section 4.2 presents the questionnaire presented to all the participants. A pilot study (described in 4.3) was carried out prior to the actual data collection. The methodology for the data collection process for the study proper is presented in (4.4). Lastly, the processing of the collected data is presented in a series of subsections in Section 4.5.

4.1 Cue word selection

A number of considerations have to be made when selecting cue words for a word association task. As pointed out by Fitzpatrick et al., "it is imperative that lists of stimulus words are compiled in a principled manner"(Fitzpatrick et al., 2015, p. 124). For one, if selection is done in the same manner across tasks in different languages, it allows for potential cross-linguistic uses of the test.[20] Also, studies have revealed that factors such as word class and cue strength can affect response behaviour. Strong cues tend to elicit clearly dominant responses, i.e. a large part of the participants produce the same response for the given cue word. Examples of cue words likely to induce predictable responses are

[20]However, as mentioned in 2.1.5, one should also be careful to assume compatibility between sets of word association data for two languages.

man, which is likely to influence the participant to produce the response *woman*, or *salt*, which likely gives the response *pepper*. Thus these factors need to be taken into consideration when selecting cue words, for example when deciding on which frequency range to select from, as there is a link between high-frequency words and more homogeneous responses (Fitzpatrick, 2007, p. 323). The selection of cue words is also dependent on a number of other factors, such as who the participants in the study will be, whether one is testing the participants in a first or a second language, or whether one wishes to use the word association data to study a specific linguistic phenomenon. The current study aimed to follow the principles for cue word selection in Fitzpatrick et al. (2015) closely, however with some modifications, as will be presented below in 4.1.2.

4.1.1 Test design in Fitzpatrick et al. (2015)

The word association test in Fitzpatrick et al. (2015) was based on data from the British National Corpus (BNC) (Fitzpatrick et al., 2015, p. 30). The BNC consists of words from samples of written and spoken English from a wide range of sources. 90 % of the words come from written sources, which include for instance newspaper articles, fiction, and specialist journals.[21] Specifically, the cue words for the test were selected from the BNC baseword lists in Paul Nation's Range programme.[22,23] The 100 cue words for the word association test were randomly selected from the 2k and 3k bands of the baseword lists, i.e. from the 1000-1999, and the 2000-2999 most common words, respectively.

The baseword lists in the Range programme are based on *word families* created by Bauer and Nation's levels system (Bauer and Nation, 1993). The cue words for the word association test were selected among the headwords of the word families.[24] In Bauer and Nation's terms, "a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (Bauer and Nation, 1993, p. 253). The word families are linked to a system of levels, where each level includes an increasing amount of inflection and affixation. Thus, with each level, each word family increases in size. For instance, at level 1, all words are considered separate word families (i.e. *remember* and *remembers* would be two separate word families at this level), but at level 2, the base plus all inflections are considered to belong to the same word family. Up until level 6 more and more potential affixes are added to each family (Bauer and Nation, 1993, pp. 253-4). Thus, when all levels are included, this results in potentially very large word families, with multiple word classes represented in each family.

[21] *What is the BNC?*, accessed from http://www.natcorp.ox.ac.uk/corpus/index.xml, 31.01.2016.

[22] The Range programme, with additional information, can be downloaded from http://www.victoria.ac.nz/lals/about/staff/paul-nation, accessed 31.01.2016.

[23] Personal correspondence with Professor Fitzpatrick, via email 8th September 2015.

[24] Personal correspondence with Professor Fitzpatrick, via email 8th September 2015.

For instance, the word family *nation* consists of 27 members. The headwords (roots) of the families were used as cues for Fitzpatrick et al.'s word association test (e.g. *bright* would be the headword for the word family that if all levels are included will consist of the words *bright, brighter, brightest, brightly, brightish, brightness* and *brighten*).[25]

Although not explicitly stated in the article, a comparison between the list of cue words used in Fitzpatrick et al. (2015), and the 2k and 3k bands of Nation's baseword lists reveal that approximately every 20th word from the baseword lists were chosen. Words from the 1k band were not used due to the fact that there is, as mentioned, "some indication that high-frequency words produce more predictable responses" (Fitzpatrick, 2007, p. 324), and the use of the most common words in the list of cues could hide differences between participants (Fitzpatrick et al., 2015, p. 30). Nation's baseword lists are, like the BNC, divided into bands of a thousand basewords (with their belonging 'word-families'). The basewords have been alphabetised within each band. For the cue word list in Fitzpatrick et al. (2015), the cue words are presented in the order they were selected, with words from the 2k band on one page, and words from the 3k band on the other. Due to the baseword lists being alphabetised within each band, the words on each page are thus also in alphabetical order (Fitzpatrick et al., 2015, pp. 32,47).

4.1.2 Modifications to the test design methodology

Although this project aims for the Norwegian word association test to mirror the test developed by Fitzpatrick et al. (2015) as closely as possible, some modifications were made to the selection methodology.

An issue with the cues from the word association test used in Fitzpatrick et al. (2015) is that they are not marked for word class, which leaves many of them ambiguous. Examples of this includes *attack* (noun/verb), *click* (noun/verb), and *liquid* (adjective/noun) (Fitzpatrick et al., 2015, p. 47). This is probably due to the nature of the word families which the baseword frequency lists are based on. In these word families it is not possible to determine the word class of the headwords, as the families include words from several word classes, and in many instances the headwords are left ambiguous in terms of word class. The word class of the cue words has been shown to affect the responses they elicit (see for instance Nissen and Henriksen, 2006). The lack of word class classification of the cues means that one is not able to look further into this factor when analysing the results, and one cannot for instance, for those cue words left ambiguous, check if responses seem to be influenced by the word class of the cues. Although Norwegian has fewer words that are ambiguous in terms of word class than English, there are some, e.g. *ønske* (Eng: *wish*

[25]Personal correspondence with Professor Fitzpatrick, via email 8th September 2015.

(noun or verb)) and *klemme* (Eng: *hug/squeeze* (verb) or *clip/clamp* (noun) or *difficulty* (noun)). Some instances of these ambiguous words occured when selecting the cue words for the word association task, and these were then removed from the task in order to make it possible to analyse the responses for word class influence.

Another modification that was made was the decision to use frequency lists based on lemmas rather than word families. As mentioned, word families produce very large groups of words, with potentially great discrepancies in frequency between the members of each family. For instance, the word family *nation* in Paul Nation's baseword lists includes words as different as *nations*, *internationalism*, and *nationhoods*. It seems unrealistic to assume that a person learning a given language as an L2 would master all members of a word family if he or she masters the headword of the family. Previous studies also support this, for instance, Schmitt and Zimmerman (2002), a study of the acquisition of word families among non-native speakers of English at different educational levels. Although evidence suggests that the members of a word family are grouped together in the mind (see e.g. Bertram, Baayen, and Schreuder, 2000), studies have also indicated that "a strong facilitative effect does not operate in the productive mode, or at least not one that works equally well for all word classes" (Schmitt and Zimmerman, 2002, p. 150). Schmitt & Zimmerman found that learners of English tended to know only some members of the word family. Nouns and verbs were more likely to be known than adjectives and adverbs. This was true also for the word families of words that the participants claimed to know well (Schmitt and Zimmerman, 2002, p. 158). In fact, not even the native speakers of English used as controls in the study showed a "complete productive knowledge of the derivational morphology" (Schmitt and Zimmerman, 2002, p. 160). Lemmas also appear to be more in line with the criteria for later analysis of the responses. Following what was done in Fitzpatrick et al. (2015), the responses were lemmatised before creating norms lists.

4.1.3 Test design for the Norwegian word association task

The words for the Norwegian word association task were chosen from the NoWaC corpus (for more information on the corpus, see Guevara (2010). A main reason for choosing the NoWaC corpus was its size, as it contains 700 million words, and it is by far the largest corpus available for Norwegian. The corpus consists of texts taken from a vast number of web documents from the *.no* domain, which means that a number of genres are represented, including texts with a more colloquial style that will partly be able to replace the spoken part of the BNC. Unlike the baseword lists in Paul Nation's Range Programme, the frequency lists for the NoWaC corpus are tagged for morphological and syntactic properties, which is practical for both response type categorisation, and to check

for correlation between word class and response type. The NoWaC corpus has two different frequency lists, one for word forms and one for lemmas. The cues for the Norwegian test were chosen from the lemma frequency lists, as this is closer to the baseword lists used by Fitzpatrick et al., yet avoiding the problems of word class ambiguity, as discussed above.

The words for the word association task were, as for the English test, randomly chosen from the 2k and 3k bands of the lemma frequency list from the NoWaC corpus.[26] Every 20th word from the 1000th to the 2980th most frequent word in the corpus were selected as cues in the word association test. However, several words chosen this way were not suitable for the purposes, for one of a number of potential reasons, for instance due to being a homonym, or being a function word. A full list of unsuitable words can be found in table 4.1. If a word chosen in this manner was considered not suitable for the task, the closest suitable word was used instead, trying first the word directly below in the frequency list, and then the word directly above. In most cases, either the word below or the word above were suitable, however, in some cases the final word was further away from the initially selected cue word in the frequency list. To mirror the test in Fitzpatrick et al. (2015) as closely as possible, the cue words were alphabetised within each of the 2k and 3k bands. The full list of cue words for the Norwegian test can be found in Appendix D, and their English translations in Appendix E.

For the pilot study, an additional 10 words were added, in order to have pre-tested words available if any of the originally chosen words had to be substituted after analysing the data from the pilot. These ten words were chosen in the same manner as the original words were chosen, only at a different interval. Every 200th word was chosen from the 1050th to the 2850th most frequent lemmas. The same guidelines for excluding words considered unsuitable as for choosing the original 100 cue words applied.

After finalising the selection of the cue words, the list was checked for frequency correspondence with the English cue list developed by Fitzpatrick et al. (2015). As many of the English cue words are ambiguous in terms of word class, and each word might require several possible translations to be checked, the words were checked for frequency correspondence by translating the Norwegian cue words into English, and checking if the English translation equivalent(s) could be found in the 2k and 3k bands of Nation's baseword lists. Translations were made using digital Norwegian-English dictionaries available from Ordnett,[27].

Where several English translations of a cue word were possible, a number of these were checked for frequency. In total, 136 English translated equivalents to the 100 Norwegian

[26]Frequency lists from NoWaC can be found at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/nowac-frequency.html, accessed 31.01.2016.

[27]Accessed from http://www.ordnett.no in September and October 2015.

Words considered unsuitable as cue words:	
1.	Names of people or places
2.	Numbers, measuring units, names of months or weekdays
3.	Function words
4.	Punctuation marks
5.	Adverbs (also excluded from Fitzpatrick et al. 2015)[28]
6.	Words that were a synonym to a word already selected as a cue, or very close in meaning to a word already selected. E.g. *ligne* (Eng: *resemble* (V)) was discarded when *likhet* (Eng: *similarity* (N)) had already been selected
7.	Words that partly overlapped (in form and/or meaning) with words already selected as cues. For instance, *kjær* (Eng: *dear* (A)) was excluded because *kjærlighet* (Eng. *love* (N)) was already part of the test.
8.	Homonyms. Examples include *merke* (Eng: *brand/mark/badge* (N), or *mark/notice* (V)), *bygg* (Eng: *building (N)/ barley(N)/ build* (imperative)), and *feste* (Eng: *party/attach/lease* (V))[29]
9.	Words in other languages. Due to being a web based corpus there are for instance a number of English words present in the frequency lists

Table 4.1: Words considered unsuitable as cue words

cue words were checked for frequency, i.e. whether they could be found in the baseword lists used in Fitzpatrick et al. (2015). The comparison showed that the vast majority of the translated equivalents could be found in the 1k, 2k and 3k bands of the baseword lists. Approximately half of the translated equivalents (63 out of 136 checked words) were in the 1k band of the baseword lists. This is likely due to the fact that the NoWaC frequency lists are based on lemmas, whereas the English baseword lists are based on headwords. Each entry in the English baseword lists may contain several words, making the total mount of words in the English lists much greater than the NoWaC lemma lists. In total, 85% of the translated equivalents were found within the first 3000 basewords of the English baseword lists. 15% were found to have a lower frequency than the English cue words.

The number of ambiguous words in the English test leaves it impossible to count precisely the distribution of words belonging to each words class. However, it was an aim to make sure the Norwegian test had a fair representation of both verbs, nouns and adjectives. Based on findings from for example Nissen and Henriksen (2006) that cue word class

[28]Personal correspondence with Professor Fitzpatrick, via email 8th September 2015.

[29]Where multiple English translations would be possible for each meaning of the words, only one has been included here, as the purpose is only to exemplify the number of different meanings for each homonym.

influences the results of word association tasks, it was considered important to have a list of cue words that were unambiguous in terms of word class. This would allow for investigation into the effects of cue word class. To have unambiguous cue words also matters in the response type classification process, where one might for example only be able to positively identify synonyms when the word class of the cues are clear. The final version of the test, after the pilot study, is made up of 55 nouns, 26 verbs and 19 adjectives.[30]

4.1.4 *Bokmål* or *Nynorsk*?

For the purposes of this study, *Bokmål* was decided to be the most natural choice for which written standard of Norwegian to use in the word association test. Firstly, the data would be collected in Oslo, where *Bokmål* is by far the most commonly used written standard. Secondly, in Norway as a whole, *Bokmål* is by far the most widely used written standard, and a word association test in *Bokmål* would thus be applicable in many more parts of Norway, and to far more speakers than a test in *Nynorsk*.[31] Thirdly, due to being the most common written form, *Bokmål* is also what most people learning Norwegian as a second language would learn. Thus, as the test and norms lists may be used in bilingual contexts in the future, to develop and administer the test in *Bokmål* was considered the most useful.

Ideally, one would have two versions of the test available, one in *Bokmål* and one in *Nynorsk*. However, within the time limits of a master thesis project it was not feasible to design, administer, and analyse data from two different tests. For future studies, comparisons should be made between the results of word association tests in both written standards to look into potential differences in results and association behaviour.

The response data from those participants who in the questionnaire (see 4.2) reported to predominantely use *Nynorsk* when writing were transcribed into a separate spreadsheet. However, for several reasons these responses were not used in further analysis in the present study. Firstly, there was a tendency for these participants to mainly "translate" the cue words in *Bokmål* into their corresponding *Nynorsk* forms, rather than to freely associate. Secondly, the fact that the participants (also those who did associate more freely) were responding to cue words written in another written standard than the one they predominantly use is problematic on its own. One does not know how the association process would proceed between the two forms, and whether it is more similar to same-

[30]The complete word association test can be found in Appendix D, with English translations in Appendix E.

[31]No official data exist for the population as a whole, but Statistics Norway report that approximately 12 % of the pupils in primary and lower secondary school had *Nynorsk* as their chosen language form in 2015 *Elevar i grunnskolen, 1. oktober 2015* accessed from, http://www.ssb.no/utgrs/ 19th January 2016.

language association, or may also share traits with second language association. Thirdly, it is difficult to decide on whether there is a complete semantic overlap between the *Bokmål* word and the *Nynorsk* word, even for seemingly equivalent forms.

4.2 Questionnaire

All participants were asked to fill in a questionnaire (see Appendix C) prior to undertaking the word association test itself. The questionnaire aimed to register aspects of background information on the participants that would be useful in the analysis of the word association data. The questionnaire included questions on age, gender, education level, work history, potential previous or current neurological illness, and linguistic background.

Some of the questions asked, about gender, age, and educational level, were included to be able to group the participants based on these three factors when later analysing the word association data. Norms lists and response category patterns could then be contrasted and compared across these factors, and combinations of them, to be able to discover how they might influence the responses given in the test.

The participants were asked about potential neurological illness because this may affect one's linguistic abilities. Only the word association responses from those who reported that they had not had any neurological damage or illness were included in the results, as the aim of the study was to map out norms and response patterns for healthy, native speakers of Norwegian.

The last section of questions in the questionnaire were linked to the participants' linguistic background. One question related to which of the two official written forms of Norwegian (*Bokmål* or *Nynorsk)* that was the main form used by the participant. As previously mentioned, since the test was only administered in *Bokmål* in this study, the responses for those who reported to mainly use *Nynorsk* were discarded from the study (see Sections 4.1.4 and 4.5).[32] The participants were also asked which Norwegian dialect they speak. This question was not asked in order to discard anyone's responses, but knowing the participants' dialect background might be useful when categorising the responses given. Norwegian dialects vary in terms of grammar, syntax, pronunciation, and most important for this study, vocabulary. If some responses were words unknown to me when categorising the responses, it was useful to know the dialect background of the participant to see if this could be an explanation. The dialect given by the participant in the questionnaire would also in most cases tell something about what part of Norway the participant originated

[32]However, their responses were still kept in a separate excel sheet, so that they might be used in potential follow-up studies.

from, and thus might also explain responses linked to certain areas of Norway (e.g. places, companies, etc).

The second part of the questions on linguistic background were linked to other languages than Norwegian known by the participant. The participants were asked to name any other languages they knew, how old they were when they acquired the language(s), whether they had used the language(s) in connection with work or studies, and to state how well they knew the language(s). This information was partly gathered for the same reason as the participants' dialect background, to be able to map out the reasons for some given responses. In addition, these questions would uncover if any of the participants had been bilingual their entire life, in which case their responses would also have to be discarded.

4.3 Pilot study

Following the selection of the cue words, a pilot study was performed in early November 2015. The pilot study took place at the offices of a special needs education service in Eastern Norway. 14 employees at the office volunteered to participate in the pilot study. The group consisted of two men and twelve women, with an age span from 24 to 67 years. The mean age was 46 years. Some of the participants in the pilot study would have fitted into one of the cohorts to be used in the actual study (20-30 years old, and over 60 years), and some were in between these two age groups.

The aims of the pilot study were firstly to check whether any of the 100 originally selected cue words should be discarded from the final version of the word association test. This could be due to the words commonly being misunderstood,[33], or words that would tend to produce very strong responses, here defined as eliciting a dominant response with a response rate of over 70%. Secondly, the pilot study would give an idea of how much time would be needed to complete the test.

The analysis of the response data collected in the pilot study showed that four words had a dominant response with a rate of occurence of above 70%. As these four words would be likely to continue to produce the same pattern of a dominant response in the actual study, they were excluded from the final test. They were: *hovedstad* (Eng: *capital (city)*), *eiendom* (Eng: *property*), *vesentlig* (Eng: *essential, considerably*), and *gripe* (Eng: *catch, grab* (verb)). On the other end, one word, *tilgang* (Eng: *access, supply* (noun)), produced only idiosyncratic responses, and there was little or no consistency in the responses given. Thus *tilgang* was also excluded, as the erratic responses might be due to the participants

[33]For instance, Fitzpatrick et al. (2015) reports that in their study, *miner* was commonly mistaken for *minor*, and responded to as such.

misunderstanding the cue. These five words were replaced by five of the 10 extra words added for the pilot study. The first five of these words, excluding those of the extra words that produced very strong responses (above 70% for the dominant response), were included. They were: *produsere* (Eng: *produce, manufacture* (verb)), *opprinnelig* (Eng: *initial* (adjective)), *sunn* (Eng: *healthy* (adjective)), *forventning* (Eng:*expectation* (noun)) and *motstander* (Eng: *opponent* (noun)). The final list of 100 cue words for the word association test can be seen in Appendices D and E.

In terms of the time needed to complete the test, the participants in the pilot study used on average 12 minutes and 25 seconds, but the differences in time used between the participants were however quite large, with a span from 6 minutes and 45 seconds to 19 minutes and 57 seconds. As the participants in the study proper would include people over a rather vast age span, even larger than in the pilot study, it was decided not to set an upper time limit, but to emphasise clearly that the test should be performed as quickly as possible, without going back to previous words, or thinking for too long about any of the cue words. Differences in for example writing speed was observed already during the pilot study. This was assumed to be even more of an influence when including the oldest participants in the proper study, some of which were likely to be in their 80s or 90s.

The pilot study also provided some comments on the questionnaire, which were taken into consideration before deciding on its final version. The changes made were the inclusion of an option to say that one has studied *in* a given second language,[34] in addition to the option to say that one has studied the language. Another change made was to expand the number of second languages each participant could include, as the pilot showed that several participants knew more than three languages apart from Norwegian.

4.4 Data collection

Data for the study was collected between December 2015 and February 2016. Participants from two age groups were recruited for the study. The younger group included participants between 20 and 30 years, and the older group included participants over 60 years. The lower age limit for the older cohort was set to follow the United Nations agreed cutoff of 60 years of age,[35]. By including two age cohorts, the study would be able to reveal potential age related differences in responses, in line with the first set of research questions (related to whether the word associations are different in the two age cohorts) and the third set

[34]This means that one has mainly used the language to read (and write) academic texts, and perhaps participated in academic discussions in the language.

[35]World Health Organization: *Definition of an older or elderly person* accessed from the following web address on the 16th of May 2016: http://www.who.int/healthinfo/survey/ageingdefnolder/en/#.VTnvzCwz4oE.email.

of research questions (related to how the data sheds light on theories on language and ageing) (see 1.2).

Testing took place across a number of different locations, sometimes testing one person at a time, sometimes testing smaller or larger groups. Separate summaries for the collection for each of the two age cohorts are presented in 4.4.1 below. In general, the recruitment process was more difficult and more time consuming than expected. A large part of these difficulties was due to the time to took to participants, and in particular to find groups of participants willing to be tested at the same time. For the younger age cohort this was somewhat easier, and some large group test sessions were conducted. For the older group, however, it was a challenge to find settings where multiple participants would be interested in or able to take part in the test simultaneously. Thus, due to only testing one, or a small number of participants, at the same time, test sessions among the older participant group took longer time, which also resulted in the total number of participants in the older group being smaller than for the younger cohort.

In the test sessions, the participant was first given the information letter, and signed the consent form. Some information about the background of the study was also given orally. After this had been done, the participant was given the questionnaire, and could use as much time as needed to answer all the questions. After the questionnaire had been completed, the participant was both given instructions orally, and told to read the instructions on the front page of the word association test (see Appendixes D and E). All participants were told to work through the test as quickly as possible, writing down the first word that came to mind for each of the 100 cue words. The participants were also told that if nothing came to mind, they could leave the response line for the given cue word blank.

As mentioned in 4.3, no upper time limit was set for the completion of the test. However, the intended methodology for the test sessions included recording time use for all participants. For the older cohort, who largely were tested alone, or in small groups, this was carried out. For the younger cohort however, most of who were tested in large groups in classrooms or lecture halls, time usage was unfortunately not recorded. Time data thus only exist for those in the younger cohort who were tested alone or in small groups. For further discussion of these issues, see Chapter 7.

4.4.1 Participants and test sessions

The younger participant group was recruited from a number of different places. Some were recruited from my own personal network. These test sessions were held in the homes of some of the participants, or at the offices of a students' society in Oslo. In these

sessions, some participants were tested on their own, others in small groups of 2-4 people. In January and February 2016, testing was performed on larger test groups at a time, in lectures at two higher education institutions in Oslo. For two of these, the students who were willing to participate took the test in the break between the two sessions of a lecture. For the last two, testing was done as part of the lecture. All participants, including those recruited from my own network, were either currently enrolled in higher education, or had previously been so.

The older participant group were recruited from different places in Oslo. Part of the participants were recruited from senior centres in the city. These centres are social meeting places for people over 60. For most of them, the seniors still live at home, but come to the senior centres to socialise, to participate in various activities, to get affordable haircuts or physiotherapy, and so on. Thus, these seniors were likely to be healthy, and to have a low rate of neurological illness.

At the senior centres, where it was possible, a separate, quiet room was used for the test sessions. In the cases where it was not possible to use a separate room (or in some cases the participant insisted on performing the test in the main room), the test was performed with other people around. On a few occasions I managed to test larger groups at the same time, and this was then also done in the common areas of the senior centres. Most of the participants read the cue words and wrote down their responses themselves; however, for some participants, the task was performed orally with the instructor writing down the responses consecutively. These participants had hand motoric difficulties, and were unable to write down the responses themselves. The total number of participants who performed the test orally were five, all of them in the older age cohort. The test sessions at the senior centres were performed with the permission of the staff, and the staff were also highly helpful in the recruitment process, and facilitated separate test rooms.

Another group of participants in the older age cohorts was recruited from the staff at a higher education institution in Oslo. These test sessions were performed with one participant at a time, in the participants' offices. In addition to these, a third group of participants in this age group were recruited through the personal networks of my two supervisors and myself. These participants were tested in private homes.

4.4.2 Collected data sets

In total, 223 participants took the test. 50 of the data sets had to be discarded, which left 173 participants whose data sets were used in the study; 122 between 20 and 30 years, and 51 who were 60 years or older. Their distribution in terms of age, education, and gender can be seen in table 4.2. The mean age for the younger cohort was 24 years, and

for the older cohort 75.2 years.[36] For education levels, lower education means completed upper secondary school, or lower non-academic education (vocational training).

	20-30 y/o			Over 60 y/o		
	High ed.	Low ed.		High ed.	Low ed.	
Women	85	0	85	16	15	31
Men	37	0	37	14	6	20
	122	0	122	30	21	51

Table 4.2: Overview of the participants. Full age distributions can be found in Appendix F.

Higher education means higher academic education at university or comparable level, even if no degree has been completed (yet). Due to the fact that the participants were recruited in Oslo, with a high frequency of inhabitants with higher education, it was very difficult to recruit groups of participants with lower education only, at least for the younger cohort. All the recruited participants in the younger cohort had some higher academic education, although not all of them had finished their degrees. This means that all the participants in the younger cohort have been classified as belonging in the higher education group.

Although information was collected on the gender and education level of the participants, it was decided only to use age as an factor in further analysis. As mentioned above, no participants in the younger cohort were classified as having a lower level of education. In terms of both education and gender it was also an issue that the number of participants in the older cohort was only 51, and thus any further division would potentially lead to weaker results. Additionally, in Fitzpatrick et al. (2015), which has provided the basis for the methodology in this thesis, age was also the only factor considered when comparing the two cohorts. The decision to focus on age-related differences is also in line with the first set of research questions (related to whether the word associations are different in the two age cohorts) and the third set of research questions (related to how the data sheds light on theories on language and ageing) (1.2).

An overview of the discarded data sets can be seen in table 4.3. Data sets were discarded either due to reasons previously mentioned in Section 4.2, or if the participant responded to less than 70% of the cue words.

[36]Distributions of participants for different ages within the two cohorts can be seen in Appendix F.

Discarded data sets			
Reason	**20-30**	**60+**	**Sum**
Mainly use *Nynorsk*	2	4	6
Previous neurological illness or damage or did not wish to state	10	2	12
Early bilingualism or other mother tongue than Norwegian	17	2	19
Too many blank responses	4	7	11
Other (e.g. incomplete questionnaire)	2	0	2
	35	**15**	**50**

Table 4.3: Discarded data sets

4.5 Data analysis

The following section will present the way the data collected through the word association tests was prepared for analysis and later analysed. The results of the analysis are presented in Chapter 5.

4.5.1 Preparing the data for analysis

Once collected, each participant was given a six-digit code number, so that the responses would only be stored together with this code number, and details that could potentially reveal the identity of the participant were removed. The list of codes and the names of the participants they belong to were stored separately. The link between the responses and the specific participant was needed in those cases where for example foreign language, or dialect responses, might be explained based on the information provided in the pertaining questionnaire. Even more importantly, it might be needed in potential future, longitudinal studies of the same participants.

The responses from the word association tests were consecutively transferred from the hand written test sheets to excel spreadsheets, one for each age cohort. In some cases, a response could not be interpreted, and these were marked as *[unreadable]* in the spreadsheet. An extract from the spreadsheet can be seen below. The spreadsheet consisted of two sections. The first section (figure 4.1) contained the participants' ID number, the date they performed the task, the time they used to complete the association task,[37] and their gender, age, and education level (H(igh) or L(ow)). The second section (figure 4.2) contained the participants' responses, annotated for word class and response type classification. The capital letters in the cells immediately following the responses denote word

[37]Test times were only reported for those participants time data was available for.

class of the cue words and the responses. The column marked 'RT' denotes response type, and the numbers in the spreadsheet correspond to the numbers of the response categories given in table 4.4.

	A	B	C	D	E	F	G
1	ID	Date	Time	Gender	Age	Age group	Education
2	013902	011215	16:13:00	M	76	2	H
3	014001	011215	10:36:00	F	75	2	H

Figure 4.1: Information on participants in response data spreadsheet

CQ	CR	CS	CT	CU	CV	CW	CX	CY
posisjon	N	RT	prege	V	RT	psykisk	A	RT
stilling	N	1	dominere	V	1	sjelen	N	3
ståsted	N	1	synlig	A	3	sjelelig	A	1

Figure 4.2: Extract from transcribed responses, marked for word class and response type

In terms of correcting spelling, the current study followed the policy of Fitzpatrick et al. (2015) to correct misspelled responses, but only in those cases where the misspelled word could not be a proper word. This meant that e.g. *sjult* was corrected to *skjult* (Eng: *hidden*), as a response to *hemmelig* (Eng: *secret* (adjective)), or that *reddsel* was corrected to *redsel* (Eng: *fright/fear* (adjective)), as a response to *frykte* (Eng: *fear* (verb)). If the given response was a number (e.g. *3* or *10* as a response to *telle* (Eng: *count* (v)), it was transcribed using letters, i.e. *tre* (Eng: *three*), or *ti* (Eng: *ten*).

In some cases, although instructed to write only one response word per cue word, the participants still wrote multi-word responses. In order to analyse these multi-word responses along with the rest of the collected data, they needed to be processed where possible. Fitzpatrick et al. (2015) put their multi-word responses into two main categories. The first one consisted of multi-word responses that were "formulaic sequences with a single, coherent meaning" (Fitzpatrick et al., 2015, p. 31). These responses were left as written by the participant. An example of such a formulaic sequence was the sequence *komme til verden* (Eng: *enter this world*, as a response to *fødsel* (Eng: *birth* (N)). Where multi-word responses were not formulaic sequences, they went through one out of two processes to shorten them (Fitzpatrick et al., 2015, p. 31). For responses made up of two separate one-word responses, separated by punctuation (e.g. *minske/liten*) (Eng: *diminish/small*)), the first one-word response was used (i.e. *minske/liten* would be transcribed as *minske*) (Eng: *diminish*)). Other, non-formulaic, multi-word responses were shortened by deleting function words that were part of the response. This included removing infinitive articles (e.g *å elske* (Eng: *to love*) was transcribed as *elske* (Eng: *love*)), articles (e.g. *et kutt* (Eng: *a cut*) was transcribed as *kutt* (Eng: *cut*)). Two-word responses such as *ikke syk* (Eng: *not ill*), as a response to *sunn* (Eng: *healthy*), was transcribed as *syk*, as the

antonymous relationship between the cue and response is still kept without the negation. Non-words (e.g. *ukoselighet* (Eng: approx. *something that is not nice*, not a real word in Norwegian), were transcribed as written, and later classified as erratic.

To identify the multi-word responses that were formulaic,[38] Fitzpatrick et al. (2015) used an approach developed by Wray, who defines a formulaic sequence as:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray, 2002, p. 9)

Wray and Namba present a set of criteria for identifying whether "a particular wordstring is likely to be formulaic" (Wray and Namba, 2003, p. 24). Their criteria are meant to function as a way to guide the researcher in deciding whether a given wordstring is formulaic or not. The set consists of eleven criteria, each to be evaluated along a five-point scale from *strongly agree* to *strongly disagree*, although all criteria are not applicable to all cases (Wray and Namba, 2003, pp. 27-8). For the purposes of the current study, the criteria that are related to typical linguistic behaviour or linguistic experience for a given person, or those related to the (potential) formulaic units role in a larger context than that given by a single cue word only were unsuitable. The criteria related to the the grammar of a wordstring, such as whether there is something grammatically unusual about it, or that the wordstring lacks semantic transparency, were used as criteria when assessing if wordstrings were formulaic. The list full list of criteria can be found in Wray and Namba (2003, pp. 29-33). This same set of criteria were used for classifying the Norwegian multi-word responses.

However, not all of the multi-word responses could be either classified as formulaic language, or be shortened easily. Thus, some non-formulaic multi-word responses were still left as they were written by the participant. These were classified in the same manner as all other responses, for instance *med store bokstaver* (Eng: *in captial letters*), as a response to the cue *understreke* (Eng: *underline, emphasise*), was classified as a (multi-word) synonym. Likewise, *ta vare på* (Eng: *take care of*), was treated as a synonym to its cue word *beholde* (Eng: *keep*).

[38]Note should be made that a premise for this way of treating multi-word responses is the idea that formulaic language is part of linguistic processing, i.e. contrary to the idea of analytic processing where input is broken down into its smallest components (and analytic production where one creates output starting from the smallest components, and assemble them using rules of grammar) (Wray and Namba, 2003, p. 25)

4.5.2 Norms lists

The norms lists were compiled by counting the instances of each response for a cue word, and then sorting the responses by their number of occurences. Norms lists were made separately for the two age cohorts. When counting responses, all inflectional variants of a word were lemmatised, but not derivational variants. This corresponds with the methodology used in Fitzpatrick et al. (2015). This meant that for instance, *begynnelse* and *begynnelsen* (Eng: *beginning* and *the beginning*) would be considered the same word, whereas *mulig* and *mulighet* (Eng: *possible and possibility*) were considered separate words. Multi-word responses were also included in the norms lists, however, most of them only occurred once, or a couple of times, and thus did not influence the norms lists to a large extent. The dominant responses for each cue word can be seen in tables 5.1 in Chapter 5. The full norms lists for each age cohort can be seen online at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/sprakmateriale.html.

In order to be able to compare the data from the current study with the data from the five overlapping cue words from Håseth (1968) (see 2.1.5) three subsets of norms lists were created, one for each of the two cohorts in the current study, and one for the data from Håseth (1968). Each subset contained the top responses for the five overlapping cue words. These can be seen in Appendix G.

4.5.3 Scoring for stereotypy

Stereotypy scores were used as a measure of the effect of using age appropriate norms lists. As in Fitzpatrick et al. (2015), a "stereotypy point" was given for a response that was the dominant response on the norms list. No other responses on the norms list gave any points (Fitzpatrick et al., 2015, p. 33). The stereotypy scores of each participant in the current study were calculated both for the norms list the participant had contributed to (own norms list), and against the norms list based on the responses from the other age cohort (other norms list). The difference in stereotypy points between these two scores was also calculated, i.e. the score for the participant's own norms list, minus the score from the other list. The results for the stereotypy scores can be seen in 5.1.1.

Due to the scope and time frame of this master's thesis it was not feasible to carry out a full validity test like in Fitzpatrick et al. (2015, p. 34). In their study, the two age cohorts were each split into two, and new norms lists were compared for each of these four groups. Since all participants were pairs of twins, the cohorts were split based on twin pair birth order. All participants were scored against each of the four norms lists, and checked for whether they gained more from being compared to either of the own cohort lists, than to either of the two other cohort lists.

The relatively small number of participants in the older cohort in the current study (n=51) means that to split this group in two would have resulted in two fairly small groups. The small size of these new groups could have affected the results of the validity test, as the results based on these two groups would be easily skewed by erratic responses from a few participants.[39]. Stereotypy scores were thus only calculated for the two cohorts as wholes, each of them scored against the norms list for their own cohort, and for the other cohort. The stereotypy scores are presented in 5.1.1.

4.5.4 Response type categorisation

In Fitzpatrick et al. (2015), the participants' responses were categorised using a set of 14 response categories. This set of classification categories moves away from the previously widely used classification system of paradigmatic, syntagmatic, and clang responses. Paradigmatic responses can replace the cue word without altering the grammaticality of the sentence (e.g. *car-train*), and syntagmatic responses co-occur with the cue word in text (e.g. *car-drive*). Clang responses are made on the basis of the sound properties of the cue word, e.g. the cue-response pair *cat-hat*. Fitzpatrick discusses why the paradigmatic-syntagmatic-clang set of categories is unsatisfactory, and imposes "artificial constraints on the exploration of response types" (Fitzpatrick, 2006, p. 126). For one, many cue-response pairs could be both syntagmatic and paradigmatic, such as for example *business-partnership*, that could either be a collocation (syntagmatic) or synonymous (paradigmatic). Secondly, the categories are broad and difficult to define. Different researchers might then define them differently. Thirdly, the categories do not account for all response types, and often an additional *other* category is needed, too. Lastly, the three categories do not provide precise classifications, and one does not really know what is happening inside them (Fitzpatrick, 2006, pp. 126-7). A more fine-grained system might however be able to reveal both more differences between two cohorts, and with more specific categories it is also possible to reveal more on the nature of different types of associations. The fourteen categories used in Fitzpatrick et al. (2015) also allow for responses to belong in several categories at the same time, such as in categories 11-14 (see table 4.4).

For the Norwegian word association test, the same classification categories were used as in Fitzpatrick et al. (2015). The subcategories from Fitzpatrick et al. (2015, pp. 40-41), with descriptions, and both Norwegian and English examples, can be seen in table

[39]However, it would have made for a stronger validity test, and stronger results to be able to do such a split as in Fitzpatrick et al. (2015) If the older cohort had been larger, and project time had allowed for it, such a test could have been carried out. As the participants in the current study are not twins, each of the two cohorts could have been randomised and split in two to create four new norms lists.

4.4.[40,41]

[40]Fitzpatrick et al. (2015, p. 40) uses *hyponym* rather than *hypernym* in their definition of a lexical set, which appears to be a misprint, as if two words shared a relation to another word, that word would be a hypernym. Thus, in this study, to *share a hypernym* is used as part of the definition for a lexical set.

[41]The definition used here for a lexical set is somewhat broader than that in Fitzpatrick et al. (2015), see 4.5.5.

Response categories			
No.	**Category**	**Definition**	**Example(s)**
1	Synonym	Cue and response are synonymous in some situations	*type-slag* (Eng: *type-kind*), *fort-hurtig* (Eng: *quickly-fast*)
2	Lexical set	Cue and response share a hypernym, or one word in the pair is an example of the other, including antonyms, or the two words belong to the same semantic field	*mamma-pappa* (Eng: *mom-dad*), *dyr-ku* (Eng:*animal-cow*), *få- mange* (Eng: *few-many*), *gård-traktor* (Eng: *farm-tractor*)
3	Other conceptual	Cue and response are related in meaning, but are not synonyms or in the same lexical set	*kjærlighet-ekteskap* (Eng: *love-marriage*), *konsentrere-lekser* (Eng: *concentrate-homework*)
4	Cue-response collocation	Cue is followed by the response in common usage; includes compound nouns	*olje-søl* (Eng: *oil-spill*), *vær-gud* (Eng: *weather-god*)
5	Response-cue collocation	Cue is preceded by the response in common usage; includes compound nouns	*kilde-varm* (Eng:*spring-hot*), *annonse-avis* (Eng: *ad-newspaper*)
6	Cue-response and response-cue collocation	Cue could precede or follow the response in a common phrase(s)	*hard-stein* (Eng: *hard-rock*), *hånd-jern* (Eng:*hand-iron*, *håndjern* is Norwegian for *handcuffs*)

Table 4.4: Response categories, from Fitzpatrick et al. 2015, with Norwegian examples

7	Affix manipulation	Cue is the response with the addition, deletion or changing of an affix	*omfatte-omfattende* (Eng: *comprehend-comprehensive*)
8	Similar in form only	Cue and response are similar in orthography and/or phonology, but do not share meaning	*be-se* (Eng: *ask/pray-see*), *penner-penger* (Eng: *pens-money*)
9	Two-step association	Cue and response appear linked only through another word	*flott-bitt* (Eng: *grand-bite*)(via *flått* (Eng: *tick*))
10	Erratic	The link between cue and response seems illogical. Includes repetition of the cue	*forfatter-mellom* (Eng: *author-between*), *klikk-klikk* (Eng: *click-click*)
11	Lexical set *and* cue-response collocation		*olje-gass* (Eng: *oil-gas*)
12	Lexical set *and* response-cue collocation		*helvete-himmel* (Eng: *hell-heaven*)
13	Synonym *and* cue-response collocation		*vel-bra* (Eng: *well-good*)
14	Synonym *and* response-cue collocation		*bit-rubbel* (Eng: *rubbel og bit* equals English *stock and barrel*)

4.5.5 Classifying the responses

The classification of the responses was done per cue word, so that one would not be influenced by any response type pattern from a participant. Responses were classified in the excel spreadsheet, using the numbers from table 4.4. Due to the number of English responses in the data, an extra set of categories was added, corresponding to the Norwegian ones, with an added *E* behind the number. For instance the response *squint* to the cue *konsentrere* (Eng: *concentrate* (V)), was classified as 3E; English other conceptual. A small number of responses in other foreign languages was also made, but not enough in any one language to justify a separate category. These responses were thus classified using the same number system, with an added *F* for foreign, i.e. *concierto* (Spanish, *concert*), as a response to *konsert* (Eng: *concert* (N)), was classified as 1F; foreign synonym.

To classify the responses, a combination of dictionaries, corpora, and web searches were used in addition to my own intuition. Web and corpora searches were particularly used to check whether a given cue-response combination is commonly used as a collocation. The NoWaC corpus and regular Google searches were used. Google searches were also useful to look up seemingly erratic responses, as the searches sometimes provided conceptual links unknown to me.[42] For synonyms, Gundersen (2000) was used as a point of reference, but responses did not have to be listed among Gundersen's synonyms in order to be classified as such if other material, such as corpora searches, showed that the response could replace the cue (in certain contexts).

As for the norms lists, the responses were treated as lemmas. Unlike for English, Norwegian nouns are classified by gender, with adjectives and determiners inflected in agreement with the gender of the nouns. Definite nouns have suffixed articles in Norwegian. Due to these morphological traits of Norwegian it was necessary to treat responses as lemmas particularly in order to be able to classify them as collocations. For instance, the cue-response pair of *kulturell-mangfold* (Eng: *cultural- diversity*), is a grammatical collocation if *kulturell* is inflected to agree with the noun, *kulturelt*; *kulturelt mangfold* (Eng: *cultural diversity*).

A number of the responses were evaluative (e.g. *mangfold-bra* (Eng: *diversity-good*)), or represented more personal links between the cue and response (e.g *savne-kjæreste* (Eng: *miss-boyfriend* or *girlfriend*)). These kinds of responses were classified as "other conceptual", in line with the classification in Fitzpatrick et al. (2015).[43]

A note should be made that the *lexical set* category in this study was somewhat broader

[42]For instance, one participant gave the response *alakazam* to the cue word *psykisk* (Eng: *psychic, mental* (A)). Although at first seemingly erratic, a web search revealed that Alakazam is a species from the media franchise *Pokémon*, that has psychic abilities.

[43]Personal correspondence with Professor Fitzpatrick, via email 25th February 2016.

than the corresponding category in Fitzpatrick et al. (2015). Their criteria for a lexical set was "[c]ue and response share a hyponym,[44] or one word in the pair is an example of the other; including antonyms" (Fitzpatrick et al., 2015, p. 40). However, for the Norwegian test, lexical sets were defined as the words that describe a *semantic field* (see e.g. Barcroft, Sunderman, and Schmitt (2011, p. 573)). The members of a lexical set defined this way may be defined by their relationship to the other words in the set, but they do not have to be. This means that cue-response pairs such as *gård-bonde* (Eng: *farm/estate- farmer*), although not strictly part of a hyponym-hypernym set, would still be classified as a lexical set.

Although it was attempted to place all responses in one of the fourteen categories, a small number of the responses were ambiguous in terms of their word class, and relationship with the cue. An example of this was the cue-response pair *øl-drikke* (Eng: *beer-drink*), where *drikke* could either be a noun and hypernym (lexical set), or a verb and collocation (response-cue collocation). In these cases the response was classified as either noun or verb, and categorised as a lexical set *and* response-cue collocation.

4.5.6 Assessing validity

To assess the validity of of the response classifications, 10% of the responses were re-classified by a fellow linguist, once I had finished the initial classifications of all the cue-response pairs. The re-classified data consisted of all responses, for both cohorts, for every tenth cue word from the first cue word onwards. This way, it was assumed that the ten re-classified cue words would be a representative cross section of the data.

The second rater was given instructions on how I had defined the response categories when making my classifications, but no examples were given for any of the responses to the cue words the second rater would classify. All fourteen response categories were used for the re-classifications. The responses were re-classified for the cue words presented in table 4.5. As can be seen in table 4.5, there was quite a difference in the percentage of overlap between the different cue words. In total, the percentage of overlap were 68,7% for the younger cohort, and 66,7% for the older cohort.

In Fitzpatrick et al. (2015) the initial classification had an overlap of 76,9% between the two raters, somewhat higher than the results for the re-classified responses in this study. The percentages of overlapping responses for the re-classification in the present study are relatively low, and some notes should be made concerning the implications of this. Due to the <70% of overlap between the initial classification and the re-classification, caution should be taken regarding some of the results for some of the statistical tests involving

[44] See footnote 40.

	20-30 y/o	**Over 60 y/o**
Andel (*share/part* (N))	73,8%	68,6%
Gøy (*fun* (A))	92,6%	98,0%
Kose (*cuddle* (V))	82,8%	64,7%
Prege (*characterise/mark*(V))	33,6%	56,9%
Trolig (*likely/probable*(A))	85,2%	86,3%
Analyse (*analysis*(N))	66,4%	72,5%
Gulv (*floor* (N))	64,8%	51,0%
Konsept (*concept*(N))	70,5%	78,4%
Offer (*victim/sacrifice*(N))	73,0%	60,8%
Sur (*grumpy/sour*(A))	44,3%	29,8%

Table 4.5: Percentage of overlap in re-classifications (for tokens)

the response categories (see 5.6). This should especially be kept in mind if there are cases where the p-values from the Wilcoxon tests or t-tests are close to 0.05. In these cases there is a risk that whether the difference was found to be significant or not might have changes if the data had been classified by another rater. In order to improve the reliability of the results related to response categories, a possible solution could have been to have a second rater classify *all* the responses, and then to exclude the cases where the raters did not agree from the statistical tests. This would however not have been feasible within the time limits of a master's thesis. The suggested solution would also have had to be done with caution, as the re-classification of the ten cue words for the validity test in the current study showed that there is a large difference between the amount of overlap for the different cue words. Thus, if the responses to a given cue word was particularly difficult, this method might exclude many of the responses. One would then potentially face another kind of less strong results, with a smaller sample of responses to perform tests on. Another possibility would be to adjust the analysis, and put the same responses in multiple categories at a time, and see how this affected the results.

The percentage of overlap show how difficult classification might be, especially with a larger, more nuanced classification system that opens up for multiple categories, such as to classify responses as a lexical set *and* a cue-response collocation. Due to the lack of context, it is at times hard to evaluate responses that do not fit neatly into one of the categories. Multi-word responses were particularly tricky to categorise. Additionally, even when using fourteen different classification categories, there were cases when neither of the categories covered all aspects of the cue-response relationship. For instance, a cue-response pair such as *gulv-tak* (Eng: *floor(N)-ceiling/roof (N)*) is both a lexical set *and* a cue-response collocation *and* a response-cue collocation for Norwegian. However, there is no category to include all three. One thus has to decide on whether one should focus on the pair's meaning-based *and* position-based relation, or its double position-based relation. If one decides on a meaning and position-based classification, a further decision has to be made as to whether the cue-response pair is most commonly used as a cue-response or a response-cue collocation, either way losing the dual position-based

possibilities in the classification. Cases such as these are likely to have a risk of being evaluated differently by the raters, thus increasing the amount of non-matching response classifications.

4.5.7 Statistics

All calculations were done with R Studio, version 0.99.879 (RStudio Team, 2015).

The statistical analyses performed on the data from the word association test were principally tests of statistical significance between two populations (see Chapter 5 for the tests performed on the different parts of the data, and their results). The samples to be compared were tested for normality using a *Shapiro-Wilk* test. This test measures how likely it is that the given sample is normally distributed (Baayen, 2008, p. 73). Typically, a returned p-value lower than 0.05 is taken to mean that the sample deviates from normality.

Largely the samples that were tested were not normally distributed. To test whether two samples belong to the same population, a *Wilcoxon test* was performed. The Wilcoxon test is non-parametric, in other words is does not make any assumptions about the distribution of the population the sample is drawn from. Due to this, the Wilcoxon test may be used for samples that are not normally distributed (Baayen, 2008, pp. 76-77). The Wilcoxon test compares the mean ranks of the two samples, and returns a p-value as a measure of how likely it is that the two samples belong to the same population. The higher the p-value, the greater the chance that the samples belong to the same population. A p-value lower than 0.05 is considered significant, with the two samples *not* belonging to the same population. When the compared samples are independent, a Wilcoxon Rank Sum Test (also known as *Mann-Whitney U-test*) is used. For paired samples, the Wilcoxon Signed Rank Sum test is used.

In a few cases, the two samples to be tested were both normally distributed, and a *t-test* was performed. The t-test, like the Wilcoxon test, measures whether the sample means differ. Unlike the Wilcoxon test, the t-test is parametric; it assumes that the samples used are drawn from normally distributed populations. This gives the t-test greater power than the Wilcoxon test, because it "has access to more sophisticated mathematics to estimate probabilities", when the samples are normally distributed (Baayen, 2008, p. 77). As for the Wilcoxon test, there is a difference in the t-test used for independent and paired samples. For paired samples, the paired t-test is used, and conversely, for two independent samples, an unpaired t-test is used.

Statistical significance is marked at the following levels in the tables in Chapter 5:
$*p \leq 0.05$
$**p \leq 0.01$
$***p \leq 0.001$

To counteract the problem of multiple comparisons, and to avoid Type I errors (i.e. to incorrectly reject a true null hypothesis), the p-values from the Wilcoxon tests were adjusted when testing all fourteen response categories (see 5.6.1 and 5.6.2). With a chosen significance level (alpha level) of 0.05, there is a 5% risk of getting $p<0.05$ due to chance variation. A Type I errors happens when such a false effect is reported as a real effect. At the same time, there is a 95% chance of not making a Type I error. However, the chances of making a Type I error increases when one performs multiple comparisons, e.g. by testing two different variables in the same sample (the multiple comparisons problem). When performing fourteen tests, such as for the tests on all response subcategories in Sections 5.6.1 and 5.6.2, the chance of making at least one Type I error is $1\text{-}0.95^{14} \approx 0.51$. To correct for the multiple comparisons problem we must adjust the p-values, to make sure the risk of making a Type I error is still on the 0.05 level. For the 14 tests for the response category subcategories, the p-values from the Wilcoxon tests were multiplied with $(1\text{-}0.95^{14})/0.05 \approx 10.25$, i.e. approximately what is needed to multiply the original risk of making a Type I error (0.05) with in order to get the risk of making a Type I error for 14 tests (0.51).[45]

Cohen's d was used to measure the standardised effect size between the means of two samples. The p-value returned from a test such as the Wilcoxon test just tells whether the difference between the two samples is statistically significant, however the difference between the two might still be very small. An effect size measure such as Cohen's d informs how large the observed difference really is (Walker, 2007). The larger the measured effect size, the more substantial the difference between the means (Walker, 2007). If $d=1$, the difference between the two means equals 1 standard deviation, $d=0.5$ means that they differ by half a standard deviation, and so on (Walker, 2007). How to interpret these values will, like other statistical measures vary based on the object of study. However, a value of $d<1$ will in most cases be a large, pronounced effect size.

4.5.8 Imageability

The imageability of the cue words have been shown to affect response behaviour (see e.g. de Groot (1989)), and may also affect storage and retrieval in the mental lexicon

[45]The description of the formula above is based on descriptions provided by assistant professor Bård Uri Jensen (Personal correspondence via e-mail, May 2016).

(see 3.1.1). It was thus decided to assess the influence of the imageability of the cue words on the given responses. The imageability data used was collected as part of the *Ordforrådet/Norwegian words* project (Lind et al., 2013), a searchable lexical database of Norwegian nouns, verbs and adjectives. The database contains information on different properties of the words, including imageability, frequency, and age of acquisition (AoA). For the current study, all the cue words were chosen from the same two bands of the lemma frequency lists (see 4.1.3), and thus frequency will not be further used as a factor for comparison. In terms of AoA, many of the cue words from the current study were added to the list of words in *Ordforrådet* shortly before the finishing of this thesis. The AoA data were thus unfortunately not ready in time to be used at this time, but they are however an interesting factor to include in later studies on Norwegian word associations.

The imageability data for the cue words in the current study was rated in two rounds. In the first round (presented in Simonsen et al. (2013)), 399 raters took part, and in the second round of data collection, done in 2016, 175 participants took part. All participants rated 100 words each, and gave a score from 1 ("no image") to 7 ("clear image") for each of the words (Simonsen et al., 2013, p. 438). Simonsen et al. (2013) used the definition from Paivio, Yuille, and Madigan (1968) for imageability: "the ease or difficulty with which words arouse mental images or sensory experiences, such as a mental picture or sound" (Simonsen et al., 2013). For the use of the current study, the imageability scores were divided by the median score for all words. Those words rated above the median were classified as having a high level of imageability, and those below the median were classified as having a low level of imageability. Following the predictions in 3.1.3, that imageability would influence the number of blank responses, and the response categories produced, tests were performed to check for correlations between these. As Simonsen et al. (2013, p. 442) found age effects for imageability ratings, age was also used as a factor when checking for potential significant frequency effects for the number of blank responses based on the imageability level of the cue words. Results can be seen in 5.4.4 and 5.6.3.

71 of the cue words from the current study were rated: 12 adjectives, 42 nouns, and 17 verbs. A full list can be seen in Appendix I. An overview of the distribution of imageability ratings within the word classes can be seen in table 4.6.

	Low	High	Total
Verb	16	2	18
Noun	31	10	41
Adjectives	12	0	12
Total	59	12	**71**

Table 4.6: Distribution of imageability for cue words (From the 71 cue words with available imageability ratings. The full list, with ratings for each cue word, can be seen in Appendix I.)

As can be seen from table 4.6, there was a very low number of verbs with high imageability ratings in the material. For the verbs with a high imageability rating the sample was too small (n=2) for it to be possible to make any conclusions based on any test results. I thus decided to not compare the highly and lowly rated verbs in the tests of the effect of cue word imageability on the given response categories (see 5.6.3). Statistical tests were only performed on the two noun samples (highly imageable nouns and lowly imageable nouns), and for the nouns and verbs combined.

Chapter 5

Results

This chapter will present the results from the data collected in the word association test. As mentioned in Section 4.4.2, in the cases where the results are linked to differences between cohorts, the only division made is based on age, so that the two groups compared are the older cohort (over 60 years old, n=51) and the younger cohort (20-30 years old, n=122).

The results will be presented in the following manner: Firstly, in order to explore if the word associations are different in the two age cohorts (cf. the first set of research questions, see 1.2), the norms lists for both cohorts will be presented (5.1). A comparison of the responses to the cue words that were identical in the study by Håseth (1968) and the current study is made in 5.1.2. A further exploration of differences in the responses collected in the present study can be seen in 5.2, on meaning differences in responses, and in 5.3, on foreign language responses. Related to the first set of research questions is also the cross-cohort comparison of response categories in Section 5.6.

In terms of the second set of research questions, on how the data sheds light on theories on the mental lexicon, one set of relevant results is presented in Section 5.5, where differences in responses based on the word class of the cues are presented. Results on the influence of the cue word class on response categories can be found in 5.6.

Additionally, the effects of imagebility on the number of blank responses are presented in 5.4.4. In the exploration of the data sheds light on theories on language and ageing (cf. the third set of research questions, see 1.2), the results for the number of blank and multi-word responses in 5.4 are relevant.

Although it has been attempted to structure this chapter in accordance with which results that are most relevant for each of the three sets of research questions, there are also cases where the same set of results is relevant for more than one set of research questions. The results will thus benefit from being read as a whole.

5.1 Norms lists

Based on the importance of experience in usage-based theory (see 3.1.1), it was predicted that variation would be found both between the cohorts, and between individual participants (see 3.1.3). Norms lists were created for both age cohorts (see 4.5.2 for procedures). The dominant responses for each cue word can be seen in table 5.1. The full norms lists (excluding idiosyncratic responses) can be accessed at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/sprakmateriale.html. The words in the norms lists have been annotated for word class. In the cases where a cue word had several responses with the same prevalence, these were listed alphabetically. In terms of the dominant responses, and the norms lists in general, both similarities and differences can be observed between the two age cohorts. The dominant responses for the two age cohorts overlapped for 41 of the cue words (i.e. 41%). When comparing the top three responses, the percentage of overlap for rose to 45%.

To check the effect of age appropriate norms lists on stereotypy scores (see 4.5.3), cross-cohort comparisons and statistical tests were performed. The results of these can be seen in Section 5.1.1.

Dominant primary responses		
Cue	20-30	60<
andel (N)	del (N)	part (N)
anledning (N)	fest (N)	høve (N)
ansette (V)	jobb (N)	jobb (N)
avis (N)	papir (N)	blad (N) — lese (V)
beholde (V)	ha (V)	ha(r) (V)
direktør (N)	sjef (N)	sjef (N)
eksistere (V)	leve (V)	være/er (V)
formidle (V)	snakke (V)	si (V)
forventning (N)	håp (N)	håp (N)
frykte (V)	redd (A)	redd (A)
gøy (A)	morsom(t) (A)	moro (A)
gård (N)	dyr (N)	bondegård (N)
hverdag (N)	mandag (N)	mandag (N)
hyggelig (A)	snill (A)	koselig (A)
høgskole (N)	utdannelse/utdanning (N)	universitet (N)
innføre (V)	regel/regler (N)	begynne (V)— få (V)
intervju (N)	jobb (N)	samtale (N)
kald (A)	varm (A)	fryse(r)(V) — varm(A)
kjærlighet (N)	hjerte(r)(N)	godhet (N)
konsert (N)	musikk (N)	musikk (N)
kose (V)	klem (N)	hygge (N)
litteratur (N)	bok/bøker (N)	bok/bøker (N)
materiale (N)	stoff (N)	tre (N)
middel (N)	vaske (V)	måte (N)
nyte (V)	mat (N)	kose (V) spise (V)
oppfordre (V)	be(V)— overtale (V)	be (V)
opprinnelig (A)	original(t) (A)	ekte (A)
overføre (V)	penger (N)	gi(V)
planlegge (V)	kalender (N)	tenke (V)
posisjon (N)	stilling (N)	stilling (N)
prege (V)	påvirke(r) (V)	merke (V)
psykisk (A)	lidelse(r) (N)	mental(t) (A)
religion (N)	kristendom(men)(N)	tro (N)
savne (V)	trist (A)	lengte(r) (V)
stolt (A)	glad (A)	kry (A)
størrelse (N)	stor(e)(A)	stor (A)
sunn (A)	frisk (A)	frisk (A)
sykehus (N)	syk (A)	hospital (N)
søt (A)	sukker (N)	pen (A)
telle (V)	tall (N)	regne (V)
trolig (A)	sannsynlig (A)	sannsynlig (A)
understreke (V)	poengtere (V)	presisere (V)
unik (A)	spesiell/spesielt (A)	enestående (A)
usikker (A)	redd (A)	redd (A) — vet ikke (V+Adv)
utvalg (N)	gruppe (N)	del (N)— gruppe (N)— seleksjon (N)
vedtak (N)	lov(er) (N)	bestemmelse (N)
vindu (N)	glass (N)	utsikt (N)
virksomhet (N)	bedrift (N)	arbeid (N)
øl (N)	drikke (N/V)	drikke (N/V)
årsak (N)	grunn (N)	grunn(en) (N)

Cue	20-30	60<
analyse (N)	skole (N)	undersøke (V) — undersøkelse (N)
annonse (N)	avis (N)	avis(er) (N) —reklame (N)
avansere (V)	vanskelig(ere) (A)	stige (V)
bevare (V)	beholde (V)	ta vare på
drømme (V)	sove(r) (V)	sove (V)
fellesskap (N)	venn(er) (N)	samhold (N)
forbund (N)	LO(N)	LO (N)
fremtidig (A)	plan(er) (N)	senere (A)
fødsel (N)	barn (N)	barn (N)
gidde (V)	orke (V)	orke (V)
gulv (N)	parkett (N)	dørk(en) (N)
hemmelig (N)	hviske (V)— skjule (V)	skjule/skjult (V)
hull (N)	tann/tenner (N)	åpning (N)
håndtere (V)	klare (V)	behandle (V)
industri (N)	fabrikk (N)	fabrikk(er) (N)
informere (V)	fortelle (V)	fortelle (V)
initiativ (N)	gjøre (V) — tiltak (N)	tiltak (N)
innvandrer (N)	utlending (N)	utlending (N)
komité (N)	gruppe (N)	gruppe (N)
konsentrere (V)	fokusere (V)	tenke (V)
konsept (N)	idé (N)	idé (N)
kulturell (A)	teater (N)	kunstnerisk (A)
kvinnelig (A)	feminin(t) (A)	feminin (A)
landbruk (N)	gård(er) (N)	bonde/bønder (N) — jordbruk (N)
leverandør (N)	mat (N)	produsent (N)
likhet (N)	brorskap (N)	tvilling(er) (N)
mangfold (N)	mange (A)	flere (A)
midlertidig (A)	nå (adv)	temporær(t) (A)
motstander (N)	fiende(r) (N)	fiende (N)
nabo (N)	hus (N)	granne (N)
offer (N)	drap (N) — voldtekt (N)	gave (N) — stakkar (N)
pappa (N)	far (N)	far (N)
potensiell (A)	mulig (A)	mulig (A)
regelverk (N)	lov(en/er) (N)	lov(er) (N)
rektor (N)	sjef (N)	leder (N)
revisor (N)	økonomi (N)	regnskap(er) (N)
rope (V)	skrike (V)	skrike (V)
servere (V)	mat (N)	gi (V)
struktur (N)	orden (N)	orden (N)
støy (N)	bråk (N)	bråk (N)
sur (A)	sint (A)	gretten (A)
svekke (V)	svak(t/ere) (A)	minske (V)
tempo (N)	rask(t) (A)	fart (N)
tolke (V)	forstå (V)	oversette (V)
transport (N)	bil (N)	bil(er) (N)
uheldig (A)	uflaks (N)	uflaks (N)
vert (N)	fest (N)	vertinne (N)
variant (N)	type (N)	forskjell (N)
vind (N)	kald(t) (A)	sno (N) — storm (N)
økning (N)	mer(e) (Adv)	mer(e) (Adv)

Table 5.1: Dominant primary responses for both cohorts

English translations of dominant primary responses		
Cue	20-30	60<
share/part/portion (N)	part/piece (N)	part/section (N)
occasion/opportunity (N)	party (N)	chance/occasion (N)
employ (V)	job (N)	job (N)
newspaper (N)	paper (N)	magazine/newspaper (N) - read(V)
keep (V)	have (V)	have (V)
director/manager (N)	boss (N)	boss (N)
exist (V)	live (V)	be (V)
convey/impart/arrange (V)	speak (V)	say (V)
expectation/anticitpation (N)	hope (N)	hope (N)
fear (V)	scared/ afraid (A)	scared/afraid (A)
fun (A)	funny/amusing (A)	fun/amusement (A)
farm/estate (N)	animals (N)	farm (N)
everyday (N)	Monday (N)	Monday (N)
nice/pleasant (A)	kind (A)	cosy (A)
academy/ university college (N)	education (N)	university (N)
introduce/import (V)	rule(s) (N)	begin (V) - get (V)
interview (N)	job (N)	conversation (N)
cold (A)	warm (A)	freeze(s) (V) - warm (A)
love (N)	heart(s) (N)	kindness (N)
concert (N)	music (N)	music (N)
cuddle/make pleasant for (V)	hug (N)	comfort/coziness (N)
litterature (N)	book(s) (N)	book(s) (N)
material (N)	fabric (N)	wood (N)
means/resource (N)	wash/clean (V)	manner/means (N)
enjoy (V)	food (N)	cuddle (V) - eat (V)
encourage (V)	ask (V) - convince (V)	ask (V)
initial/natural/original (A)	original (A)	genuine (A)
transfer/transmit (V)	money (N)	give (V)
plan (V)	calendar (N)	think (V)
position (N)	position/posture (N)	position/posture (N)
characterize/mark/stamp (V)	affect(s) (V)	makr (V)
mental/psychic (A)	illness/disease (N)	mental (A)
religion (N)	christianity (N)	faith (N)
miss (V)	sad (A)	long(s) (V)
proud (A)	happy (A)	proud (A)
size (N)	large (A)	large (A)
healthy (A)	fresh/well/healthy (A)	fresh/well/healthy(A)
hospital (N)	ill (A)	hospital (N)
sweet/cute (A)	sugar (N)	pretty (A)
count (V)	number(N)	figure (V)
likely/probable (A)	probable (A)	probable (A)
emphasize/underline (V)	emphasize (V)	specify (V)
unique (A)	special (A)	exceptional (A)
unsure/insecure/risky (A)	scared/afraid(A)	scared/afraid (A)- don't know
selection/sample (N)	group (N)	part (N)- group (N)- selection (N)
decision/resolution (N)	law(s) (N)	decision (N)
window (N)	glass (N)	view (N)
firm/operation/activity (N)	company (N)	work (N)
beer (N)	drink (N/V)	drink (N/V)
cause/reason (N)	reason (N)	reason (N)

Cue	20-30	60<
analysis (N)	school (N)	examine (V) - examination (N)
advertisement (N)	newspaper (N)	newpaper(s) (N) - advertising (N)
advance (V)	(more) difficult (A)	ascend/rise (V)
keep/preserve (V)	keep (V)	take care of
dream (V)	sleep(s) (V)	sleep (V)
community/fellowship (N)	friend(s) (N)	concord/union (N)
association/alliance (N)	LO (N)	LO (N)
future/prosepctive (A)	plan(s) (N)	later (A)
birth (N)	child (N)	child (N)
be bothered (V)	have the energy to	have the energy to
floor (N)	parquet (N)	deck (N)
secret (A)	whisper (V) - hide (V)	hide/hidden (V)
hole/gap (N)	teeth/tooth (N)	opening (N)
industry (N)	factory (N)	factory/-ies (N)
inform(V)	tell (V)	tell (V)
initiative (N)	do (V)- undertaking (N)	undertaking (N)
immigrant (N)	foreigner (N)	foreigner (N)
committee (N)	group (N)	group (N)
concentrate (V)	focus (V)	think (V)
concept (N)	idea (N)	idea (N)
cultural (A)	theater (N)	artistic (A)
female/feminin (A)	feminin (A)	feminin (A)
agriculture/ farming (N)	farm(s) (N)	farmer(s) (N) - farming (N)
supplier/ contractor (N)	food (N)	producer (N)
similarity/equality (N)	fraternity (N)	twin(s) (N)
diversity/variety (N)	many (A)	more (A)
temporary (A)	now (Adv)	temporary (A)
opponent (N)	enemy/-ies (N)	enemy (N)
neighbour (N)	house (N)	neighbour (N)
victim/sacrifice (N)	murder(N) - rape (N)	gift (N) - poor thing (N)
dad (N)	father (N)	father (N)
potential (A)	possible (A)	possible (A)
regulations (N)	law(s) (N)	law(s) (N)
headmaster (N)	boss (N)	leader (N)
accountant (N)	economy (N)	account(s) (N)
shout/scream (V)	scream (V)	scream (V)
serve/wait/deliver (V)	food (N)	give (V)
structure (N)	order (N)	order (N)
noise (N)	noise (N)	noise (N)
grumpy/sour (A)	angry (A)	grumpy (A)
weaken (V)	weak(er) (A)	diminsh (V)
speed/pace (N)	quick (A)	speed (N)
interpret (V)	understand (V)	translate (V)
transport (N)	car (N)	car(s) (N)
unfortunate/unlucky (A)	bad luck (N)	bad luck (N)
host/landlord (N)	party (N)	hostess (N)
variant/version (N)	kind (N)	difference (N)
wind (N)	cold (A)	piercing wind/ storm (N)
increase (N)	more (Adv)	more (Adv)

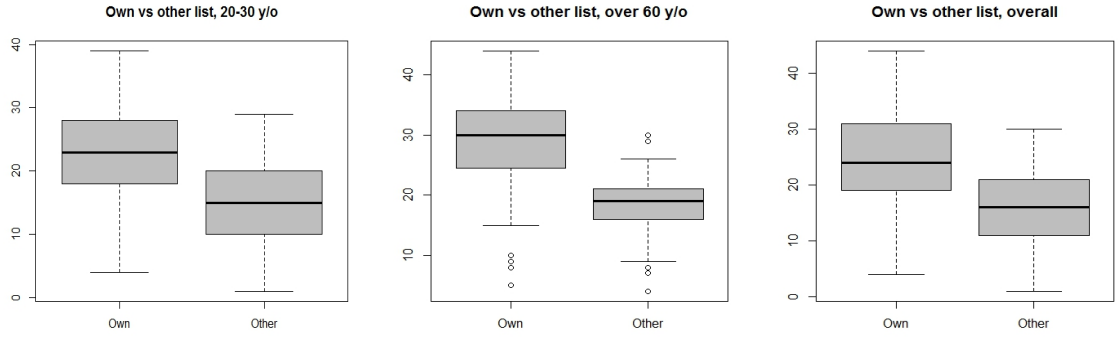
Table 5.2: English translations of dominant primary responses for both cohorts. Note that although some top responses appear to be repetitions of the cue words, these were synonyms in Norwegian, however, the nuance in question does not have a corresponding pair of words in English. Compare to 5.1.

5.1.1 Stereotypy scores and validity assessment

Following the methodology presented in 4.5.2, stereotypy scores were calculated for all participants in both age cohorts. The participants were scored against both the norms list based on their own cohort (*own list*), and against the norms list based on the responses from the other cohort (*other list*). The median values for stereotypy scores for both cohorts, and for all participants taken together can be seen in table 5.3.[46] As can be seen from both table 5.3 and figure 5.1, both age cohorts gained from being compared to an age appropriate norms list, rather than being compared to the norms list for the other cohort.

	Own list		**Other list**		**Difference**	
	median (range)	**mean (SD)**	**median (range)**	**mean (SD)**	**median (range)**	**mean (SD)**
Younger	23 (4-39)	22.80 (7.20)	15 (1-29)	15.04 (6.33)	8 (-5-18)	7.76 (4.51)
Older	30 (5-44)	28.39 (8.76)	19 (4-30)	18.18 (5.45)	10 (0-26)	10.22 (5.94)
Overall	24 (4-44)	24.45 (8.08)	16 (1-30)	15.97 (6.24)	8 (-5-26)	8.49 (5.08)

Table 5.3: Medians, mean scores and standard deviations for stereotypy scores



(a) Younger cohort (n=122), own vs other list *(b) Older cohort (n=51), own vs other list* *(c) Overall (n=173), own vs other list*

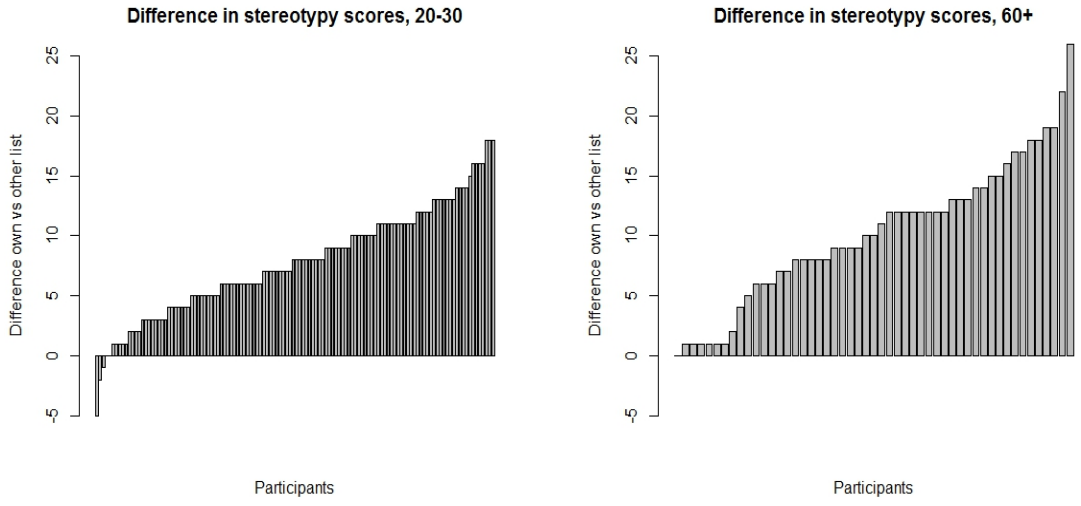
Figure 5.1: Stereotypy scores, own list vs other list

The effect appears to be somewhat larger for the older cohort than for the younger cohort. This might however be due to the fact that the older group had fewer participants, and this resulted in there being more cue words where two (or in one case three) responses shared the spot as the dominant response (n=11); for the younger cohort this was the case for only four cue words. This means that the younger cohort had more responses they could score a point from when being compared to the dominant responses from the

[46]The overall data consisted of all 173 participants scored against the norms lists for their respective age cohort for the *own list* calculations, and against the list for their respective *other cohort*. I.e. the overall *own list* data are the stereotypy scores for all younger participants scored against the norms list for the younger cohort, plus the stereotypy scores for all the older participants scored against the norms list for the older cohort, and vice versa for the overall *other list* data.

norms list for the older cohort. This might have resulted in a smaller average difference in scores for the younger cohort. At the same time, it would also have given the older cohort a greater chance to score stereotypy points for their own cohort, which might have led to the difference between the two cohorts' *difference* scores becoming even larger. See the end of this section for further discussion of potential solutions to this issue.

Figures 5.2a and 5.2b show the difference in stereotypy scores for *own list* compared to *other list* for each participant in the two cohorts.



(a) Younger cohort, n=122 *(b) Older cohort, n=51*

Figure 5.2: Difference in stereotypy scores for own list vs other list, per participant

As can be seen from the figures, all older participants benefited from being scored against an age-appropriate norms list. This is also true for almost all of the younger participants, although a few (n=3) had a higher stereotypy score when scored against norms list for the other cohort, as can be seen from the negative values in figure 5.2a. This might be due to the higher number of shared top spots in the norms list for the older cohort, which are likely to allow participants from the younger cohort to score more stereotypy points when compared to the norms list for the older cohort. However, the results might also be due to the chance variation in the data. As can be seen from figure 5.2b, several of the participants in the older cohort had difference values close to zero. The fact that none of them are below zero might be due to chance, and if the older cohort had had a larger sample, one might have observed negative values for the older cohort too.

After measuring the mean and median differences for the use of age appropriate norms lists compared to lists from the other cohort, statistical tests were used to see whether the differences were significant. A Shapiro-Wilk test (see table 5.4) was performed on the stereotypy scores for *own list* and *other list* for both the younger cohort, the older cohort, and the overall data.

	Own list		Other list	
Younger (n=122)	p=0.5917	W=0.99076	p=0.1047	W=0.98206
Older (n=51)	* p=0.02007	W=0.94526	p=0.1675	W=0.96712
Overall (n=173)	p=0.4732	W=0.99217	p=0.06742	W=0.98537

Table 5.4: P-values for Shapiro-Wilk tests for stereotypy scores

Except for the stereotypy scores for the older cohort checked against their own list (p=0.02007), all Shapiro-Wilk tests gave p>0.05 (table 5.4). A paired t-test showed that there was a significant difference (n=122, p<0.001, t=18.991. Cohen's d≈1.149617) between the own list data and other list data for the younger cohort. For the older cohort, a Wilcoxon Signed Rank test comparing the own list and other list data showed a significant difference as well (n=51, p<0.001, V=1275. Cohen's d≈1.413789). A significant difference was also found for the overall data (n=173, p<0.001, t=21.949. Cohen's d≈1.178878).

As mentioned, a potential weakness in the stereotypy measures is the higher number of shared top spots for the norms list for the older cohort, which may potentially affect the observed differences between the two cohorts. We cannot be sure whether the observed difference between the two cohorts is real, or just an effect of the difference in the sample sizes. After the observations were made of the seemingly larger effect for the older cohort (see table 5.3), a second set of stereotypy measures were calculated where all the cue words that had a difference of two or less occurrences between the two most popular responses were removed. This excluded 30 cue words from the older cohort's norms list, and an additional 15 cue words from the norms list for the younger cohort. The same procedure as for the initial stereotypy score calculations were used for these as well. The results from this second set of calculations can be seen in table 5.5.

	Own list	Other list	Difference
20-30 y/o (n=122)	15.55 (5.34)	11.33 (4.76)	4.22 (2.83)
Over 60 y/o (n=51)	17.90 (5.65)	13.88 (4.34)	4.02 (3.26)
Overall(n=173)	16.24 (5.52)	12.08 (4.77)	4.16 (2.95)

Table 5.5: Stereotypy scores, based on norms lists without cue words with a difference of <2 between the two most common responses. 55 cue words were included in these calculations.

As this second set of calculations was done for only 55 cue words, compared to the full list of 100 cue words in the main set of calculations above, the pure numbers are not compatible. However, as can be seen in table 5.5, the difference scores are very similar between the cohorts, compared to the higher scores for the older cohort in table 5.3. This second set of comparisons points to the potential need for future studies to take

into account factors such as the difference in prevalence between the top responses for a given cue word. However, there are many uncertain factors in the current second set of stereotypy measures (table 5.5), and these results cannot be used to make any firm decisions about what caused the higher difference scores for the older cohort in the initial calculations. For instance, the decision to set the cut off point at <2, was arbitrary.

5.1.2 Comparison with response data from Håseth (1968)

Data for five cue words that were part of both the study presented in Håseth (1968) and the current study were compared (see more on this process in 4.5.2). The norms lists for the five cue words for all three cohorts (Håseth's participants, plus the two cohorts from the current study) can be seen in Appendix G.

In general, a similar amount of overlap between the dominant responses for the five cue words was found for both the younger cohort compared to the data from Håseth (1968) as for the older cohort compared to the data from Håseth's study. The younger cohort shared two top dominant responses with the data from Håseth's study, whereas the older cohort shared one top dominant response.

When the top three responses were compared, there was an overlap of 53% between both the data from Håseth (1968) and the younger cohort, and between the data from Håseth (1968) and the older cohort. For the cue word *søt* (Eng: *sweet/cute* (A)) a small tendency for generational differences can be seen. Both the older cohort from the current study, and the participants from Håseth's study had a higher ratio than the younger cohort of responses linked to *søt* being used to denote something being pretty or beautiful (*pen, nydelig*), or graceful or charming (*yndig*). The younger cohort's most common responses are more linked to *søt* as a kind of flavour, such as *godteri* and *jordbær* (Eng: *candy* (N) and *strawberry* (N)). There are however also cases of *pen* (Eng: *pretty, attractive* (A)) and *fin* (Eng: *nice, good-looking* (A)) among their responses, but less so than for the two other groups. See also 5.2 for more on the differences in the responses to *søt* between the two cohorts in the current study.

5.2 Meaning differences in responses

Response data can tell us something about how the semantics of a word might change over time, or how the pattern of use of a word might change. Word association tests may elicit data on such changes when applied to different age cohorts who respond to the same cues. In the current data, such tendencies of semantic changes were, as mentioned, for instance found for the cue word *søt* (Eng: *sweet, cute, charming* (A)). A change in

meaning also appears to have taken place for *innføre* (Eng: *introduce, import* (V)). For *søt*, there were tendencies for the younger cohort to respond more based on the flavour (the top responses being *sukker, snill*, and *godteri* (Eng: *sugar* (N), *kind*(A) and *candy* (N)). The top response for the older cohort were *pen, nydelig* and *sukker* (Eng: *pretty* (A), beautiful (A) and *sugar* (N)). For the older cohort then, there is more of a link to appearance in the top responses. *Pen* is also the fifth most common response for the younger cohort; however, there still appears to be some difference in the most common usage of this adjective between the cohorts.

For *innføre*, the younger cohort gave very few responses related to import, but rather the responses were linked to policy making or to starting something. The three most dominant responses were *regel/regler* (Eng: *rule(s)* (N)), *ny(tt)* (Eng: *new* (A)), and *starte* (Eng: *start, begin, commence* (V)). The older cohort, in addition to producing the same kind of responses as the younger cohort, also produced a number of responses linked to import and sales, such as *importere* (Eng: *import* (V)), *skip* (Eng: *ship* (N)), and *grossist* (Eng: *wholesaler* (N)). The responses linked to this sense of *innføre* are not present in the data from the younger cohort, and thus that meaning of the word appears to have become less prevalent.

There were also a number of cases where the meaning of the cue word has not shifted per se, but the responses given still mirror generational differences in vocabulary. In the data from the current study, the most striking example of this is the responses given for the cue word *avansere* (Eng: *advance* (V)). For the older cohort, the responses were mainly linked to promotion, or rising, through responses such as *stige* (Eng: *rise, ascend* (V)), *gå fremover* (Eng: *proceed, make progress*), or *forfremme* (Eng: *promote, advance* (V)) (These three were respectively the first, second, and fifth most common response in the older cohort).[47] These responses are not only mainly linked to a sphere of personal achievement, often related to work, but the responses given by the older cohort are also largely positive, in the sense that something is getting better, or increasing. For the younger cohort, however, the responses for *avansere* were to a large extent words with a more negative sense. The two most dominant responses in this cohort were *vanskelig(ere)* (Eng: *(more) difficult/hard/complicated* (A)), and *vanskeliggjøre* (Eng: *complicate* (V)). These two together make up 34% of the responses for the younger cohort. The same kind of responses that were made by the older cohort are also present in the data from the 20-30 year olds, with *bedre* (Eng: *better* (A), or *improve* (V)) being the third most common response, but it only makes up 5% of the responses. It appears that the younger cohort might have to a large extent responded to the participle form of *avansere*, *avansert* (Eng: *advanced*), which may be linked to a sense of something being complicated (however, it

[47]The full norms lists for *avansere* can be found at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/sprakmateriale.html.

might also be used in a more positive sense of something being highly developed). The younger cohort might not be familiar with the base-form of the word, and thus chose to respond to the form they do know, the participle *avansert*. A search in the NoWaC corpus for the two forms show that *avansere* occurs 807 times in the corpus, whereas *avansert* occurs 11 992 times.[48] The large difference in frequency for these two forms supports the hypothesis that *avansere* might be unknown, or of infrequent usage among the younger cohort. The mainly positive responses to *avansere* in the older cohort, and the more negative responses in the younger cohort, suggest that the more negative meaning of the participle is more frequently used than the more positive meaning (i.e. that *avansert* is, for the younger cohort, more often used to denote that something is difficult or challenging, than that something is advanced). Again, in time, this might lead to a shift in meaning for the participle, or even the entire lemma.

The kind of generational differences that were seen for *avansere* were also observed for the cue word *vind* (Eng: *wind* (N)) In the older cohort, *vind* triggered responses such as *sno* (Eng: *piercing wind*), and *blest* (Eng: *wind* (N)), which would likely be seen as old-fashioned among the younger cohort. Similarly, the most dominant response for the cue word *gulv* (Eng: *floor* (N)) was *dørk(en)* (Eng: *(the) deck/floor* (N)), another word that is unlikely to be used by the younger cohort, except maybe as part of the formulaic sequence *gå i dørken* (Eng: *pass out* or more literally to fall to the floor). The reverse can also be seen in the data, with the younger cohort producing responses unlikely to be used by many in the older cohort. The 20-30 year olds had a number of responses related to sexuality, such as the cue-response pair *kose-sex* (Eng: *cuddle*(V)- *sex* (N)), which would probably not be used openly to such an extent by the older cohort. Clear generational differences can also be seen in the responses to *konsept* (Eng: *concept (N)*). Lacking entirely from the older cohort, but represented in both the 3rd and 4th most dominant responses (in addition to two idiosyncratic responses) for the younger cohort were words related to the Norwegian "russ" celebration.[49] Such responses were *russebuss* (Eng: 'russe'*bus*), *russetid* (Eng: 'russe'*period*), or simply *russ*. These celebrations have been happening for over a century in Norway, but the contents of the celebrations have changed with time. In the current case, the word *konsept* itself also appears to have gone through a semantic change. In the case of the responses related to the *russe celebrations*, *konsept* is used to mean something like *theme* for an action, thing, or event. The Norwegian word for theme, *tema*, is also the second most dominant word in the norms list for the younger cohort, but it is not present in the norms list for the older cohort. However, this appears to be an

[48]The selection of cue words for the word association task was based on the lemma frequency list for NoWaC (see 4.1.3). The lemma *avansere* is part of the 3k band in the lemma frequency list.

[49]This celebration takes place for students shortly prior to graduating from high school. It takes place in April and May, where the students dress up in special overalls with different colours. Many of them also buy a bus together, which is used for partying. Each bus will have a *russekonsept*, i.e. a theme/concept. Those celebrating are called 'russ'.

extension to the meaning of the word for the younger cohort, and not a shift in meaning given that the two cohorts shared *idé* (Eng: *idea* (N)) as their dominant response to *konsept*.

Another case of generational differences was seen for the cue word *middel* (Eng: *means/ remedy/ agent/ resources* (N)), where the dominant response for the 20-30 year olds was *vaske* (Eng: wash/clean (V)) (12%), with *vaskemiddel* (Eng: *detergent, washing powder* (N)) as the second most common response with just under 10% of the responses. For the older cohort there are no instances of *vaskemiddel*, and only one instance of *vaske*. Instead, the top responses for the older cohort were *måte* (Eng. *manner/means/process* (N)) and *hjelp* (Eng: *help* (N/V)). These results suggest that perhaps the older cohort have a more abstract understanding of *middel* (and also use mainly other words for detergent than *vaskemiddel*). The younger cohort had only one instance of *måte*, and mostly responses linked to a more concrete understanding of *middel*, such as *såpe* (Eng: *soap* (N)), or *gift* (Eng: *poison* (N)). Similarly, for the cue word *psykisk* (Eng: *psychic, mental* (A)) the younger cohort had *lidelse(r)* (Eng: *illness, disease* (N)) as their dominant response; however, *lidelse(r)* was not given as a response by any participant over 60 years. The top response for the older cohort was *mental(t)* (Eng: *mental* (A)). These differences appears to reflect changes in the way one talks about mental illness in society. The term *psykiske lidelser* has only fairly recently come into use. Earlier, other terms such as *sinnslidelser* (Eng: *mental illness*, more old-fashioned), were used (Malt, 2009).

These generational differences in responses show the need for both age specific, and recently updated norms lists in research using word association data. For one, this will allow researchers to discover which cohort differences that exist in word associations. However, there were also cases where one would have expected differences across cohorts, but no great differences were found. For example, a difference in responses to the cue word *religion* (Eng: *religion* (N)) was expected, due to factors such as increased globalisation and immigration. However, both cohorts have *kristendom* (Eng: *Christianity* (N)) and *tro* (Eng: *faith, belief* (N)) as their two most dominant responses. Likewise, more variation between the cohorts was expected for *innvandrer* (Eng: *immigrant* (N)). For instance, a report by Statistics Norway on attitudes to immigrants and immigration (Blom, 2015), reported that the older participants in the survey were more sceptical towards immigration, and one might assume to see this reflected in the word association responses in the current study, too. Such differences were no found, however, and the two cohorts produced rather similar common responses, with no clear differences in terms of characteristics of the responses given (see table 5.6).

Innvandrer (N)			
20-30 y/o		Over 60 y/o	
utlending (*foreigner*) (N)	10.66%	utlending (*foreigner*) (N)	21.57%
flyktning (*refugee*) (N)	8.20%	fremmed (*foreign(er)* (A/N)	11.76%
mørk (*dark*) (A)	6.56%	immigrant (*immigrant*)(N)	7.84%
ny(tt) (*new*) (A)	6.56%	asylsøker(e) (*asylum seeker(s)*)(N)	3.92%
menneske(r) (*human(s)*) (N)	4.10%	en som kommer (*one who comes*)	3.92%
muslim (*muslim*) (N)	3.28%	innflytter (*immigrant, outsider*) (N)	3.92%
asyl (*asylum*) (N)	2.46%	muslim (*muslim*) (N)	3.92%

Table 5.6: Top responses for innvandrer*, both cohorts*

5.3 Foreign language responses

It was predicted in 2.1.5 that there might be foreign language responses as part of the collected data, and if there were, perhaps most likely English responses.[50] In the collected data, a surprisingly large amount of English responses were found, as although English frequently used in many settings in Norway, the word association test was monolingual, and performed in a monolingual setting. As can be seen in table 5.7, the younger cohort produced 1.27% English responses. The older cohort also produced a number of English responses, but less so than the younger cohort (0.41%). For responses in other foreign languages than English, the opposite pattern was seen, with the older cohort producing more other foreign language responses than the younger cohort (table 5.7).

Foreign language responses)		
	20-30 y/o	Over 60 y/o
English	1.27%	0.41%
Other languages	0.05%	0.55%

Table 5.7: Number of English responses and responses in other foreign languages, both cohorts

A large part of the foreign language responses were simply translations of the cue word into the given other language. This kind of responses included the cue-response pairs *søt-tatli* (Norwegian-Turkish, Eng: *cute/sweet- sweet* (A)), or *fødsel-Geburt* (Norwegian-German, Eng: *birth- birth* (N)). However, for English, a more complex pattern of response types was seen. For instance, English responses were not just given as translations of the Norwegian cue words, but were also given in other types of response categories, such as for the conceptual cue-response pair *konsentrere-squint* (Eng: *concentrate-squint* (V)), or for the lexical set *konsentrere-chill* (Eng: *concentrate-chill* (V)).[51]

[50] At least for the younger cohort, one would expect all of them to have started learning English at an early age, at least by the age of six to eight. One would also assume that many of the older participants would have at least some knowledge of English, and would also most likely have studied (some) German at school, if they have completed (academic) education above the level of elementary school.

[51] There was one instance of a non-translation response in the non-English foreign language responses

Most interesting was however another observed pattern of English influence on the responses. A section of the cue words have the same orthography as a given English word, without sharing their meaning, they are so-called "false friends". It appears that in a number of cases the participants have read the cue words as if they were the English counterpart in the pair of "false friends", and responded to them as such. Such cue-response pairs were *offer-tilbud* (Eng: *victim/sacrifice* (N)- offer/supply (N)), and *rope-tau* (Eng: *shout/call out* (V)- rope (N)). *Offer* thus appears to have been read as it was an English word , (with a meaning linked to sales), and a response linked to this meaning was produced (*tilbud* might either be a translation of the English noun *offer*, or a conceptually related response to the English verb *offer*). This is the same process that appears to have happened for the cue word *rope*, where the participants who produced *tau* as a response appear to have read *rope* as an English word (meaning *a stout cord, and related uses.*[52]) Another interesting example was the Norwegian-English cue-response pair *sur-big* (Eng: *grumpy/sour* (A)- *big* (A)). This pair is particularly interesting due to its complexity. The response makes the pair a response-cue collocation, *Big Sur*, the name of a coastal region and popular tourist destination in California, and it is the name of a novel by Jack Kerouac. Although the word *sur* itself is not an English word (the name of the region is derived from Spanish *el país grande del sur*), the participant appears to still have read the cue word as it was an English word, as the response made was another English word, *big*.

The most interesting aspect of the mentioned responses to *offer* and *rope* is perhaps that they are made in Norwegian, which means that English has either been briefly activated, or that both Norwegian and English have been simultaneously active (see Chapter 6 for more on this). Considering that the test was performed in a monolingual Norwegian setting, the proportion of this kind of responses is surprising. For instance, six participants responded *tau* to *rope*, which was the 86th cue word in the association test, i.e. occuring at a point when the participants had already responded to a rather large number of Norwegian cue words.

5.4 Multi-word responses and blank responses

Two main traits of language and ageing, as presented in 3.3, are object naming difficulties, and a tendency to be more verbose than younger adults. It was predicted in 3.3.2 that the older cohort would produce more multi-word responses than the younger cohort, based on

too; one participant in the older cohort responded with the German word *Gasthaus* (Eng: *pub/inn/tavern* (N)) to the cue word *vert* (Eng: *host/landlord* (N)).

[52] *rope, n.1.* OED Online. Oxford University Press, March 2016. Accessed from http://www.oed.com/view/Entry/167402?rskey=OkVtUR&result=1&isAdvanced=false, 24 April 2016.

either or both of these factors. It was also predicted that naming difficulties would lead to more blank responses in the older cohort. In terms of what word association data might tell us about the mental lexicon, one of the predictions was that the imageability of the cue word would affect access, and thus also affect the proportion of blank responses (see 3.1.3). Clear differences were found between the cohorts both for multi-word responses, and for the proportion of blank responses. The percentages of multi-word responses and blank responses for both cohorts can be seen in table 5.8. The results will be further discussed in 6.2.2, 6.3.1, and 6.3.2.

	20-30 y/o		**Over 60 y/o**	
	median	range	median	range
Multi-word	1	0-22	6	0-58
Blanks	0	0-14	0	0-22

Table 5.8: Multi-word and blank responses for both cohorts

5.4.1 Multi-word responses: cohort differences

For the multi-word responses, a clear difference between the cohorts was observed. As seen in table 5.8, the older cohort produced more multi-word responses than the younger cohort.[53] In order to test whether the difference between the two cohorts was statistically significant, lists of the number of multi-word responses per participant were compiled for both age groups. As can be seen already from the histograms in figure 5.3, the number of multi-word responses were not normally distributed for either of the cohorts.

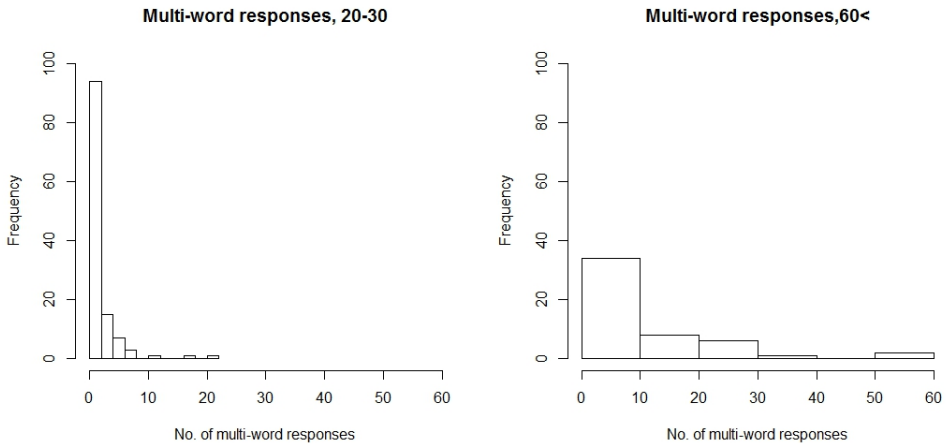


Figure 5.3: Distribution of multi-word responses per cohort. Younger cohort: n=122, older cohort: n=51

[53]Despite the high frequency of multi-word responses for the older cohort, there were also examples of participants carefully observing the instructions to only write one word per response, such as the person who responded *vedsidenav* (Eng. *nextto*) as a response to *nabo* (Eng. *neighbour*)

Shapiro-Wilk tests also clearly supported this for both the older cohort (n=51, p<0.001, W=0.74589), and for the younger cohort (n=122, p<0.001, W=0.56953). A Wilcoxon rank sum test was used to compare the two samples. The test showed that the two populations were nonidentical (p<0.001, W=4978.5, Cohen's $d \approx 1.160132$). A boxplot of the distribution of the multi-word responses can be seen in figure 5.4.

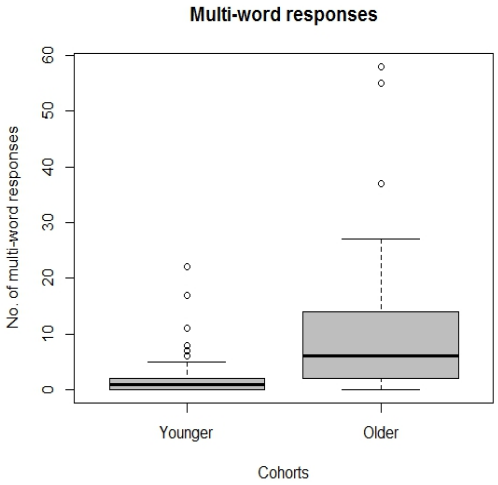


Figure 5.4: Multi-word responses by age

There was, however, also a lot of variation within the cohorts, in particular for the older cohort. The analysis of the data from the older cohort gave a median of 10.24 multi-word responses per participant (see table 5.8). The number of multi-word responses in the older cohort had a span from 0 (n=6) to 58 instances (n=1) per participant. For the younger cohort the median number of multi-word responses was 1. The younger cohort had a span from 0 multi-word responses (n=58) to 22 instances (n=1).

5.4.2 Multi-word responses: cue word differences

Just as there were differences between the participants in the proportion of multi-word responses, there were also differences in the proportion of multi-word responses between the different cue words. A full overview of multi-word responses per cue word for both age cohorts can be seen in Appendix H. Table 5.9 shows the median number of multi-word responses for noun cues and verb cues each of the two age cohorts. Note that the different sizes of the cohorts means that the median values are not directly compatible. The younger cohort had 122 participants, and the older cohort had 51 participants. As can be seen from the medians in table 5.9, the older cohort produced proportionally many more multi-word responses than the younger cohort did.

It was predicted in 3.3.2 that the cue word class itself would not influence the number of

	20-30 y/o			Over 60 y/o		
	median	**range**	**Shapiro-Wilk**	**median**	**range**	**Shapiro-Wilk**
Noun	1	0-5	*** p<0.001, W=0.84978	3	0-10	** p≈0.0081, W=0.93951
Verb	2	0-18	*** p<0.001, W=0.74557	6.5	3-20	*** p<0.001, W=0.81188

Table 5.9: Cue word class and multi-word responses, both cohorts

blank responses or multi-word responses. Shapiro-Wilk tests (see table 5.9) showed that none of the samples followed a normal distribution ($p<0.05$ for both nouns and verbs for both cohorts). Wilcoxon rank sum tests comparing the number of multi-word responses for noun cues and verb cues show that there were statistically significant differences between nouns and verbs for both age cohorts. For the younger cohort (n=122), the Wilcoxon rank sum test returned $p<0.001$ and W=388. Cohen's $d \approx$ -0.9435. For the older cohort (n=51), the Wilcoxon rank sum test gave $p<0.001$ and W=279.5. Cohen's $d\approx$-1.186. See also figure 5.5.

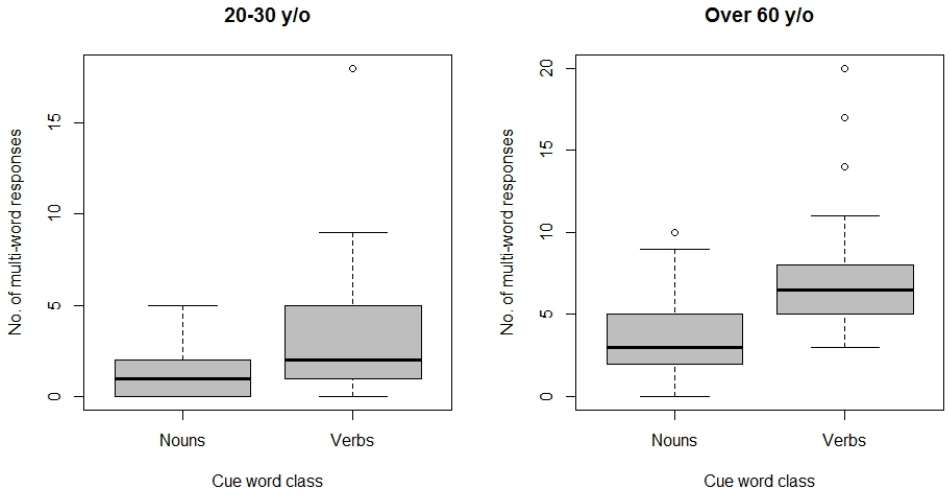


Figure 5.5: Distribution of multi-word responses for noun cues and verb cues

There were significantly more multi-word responses for verb cues than for noun cues for both age cohorts. The results thus go against the predictions made in 3.3.2, as these results suggest that cue word class might influence the number of multi-word responses.

5.4.3 Blank responses: cohort differences

As for the multi-word responses, an age-related difference in blank responses was anticipated (3.3.2). In the same manner as for multi-word responses, lists of the number of blank responses per participant were compiled for each age cohort. Shapiro-Wilk tests showed that the number of blank responses were not normally distributed for either

cohort (Younger cohort: n=122, p<0.001, W=0.29271, older cohort: n=51, p<0.001, W=0.66807). The distributions of blank responses for each of the cohorts can be seen in figure 5.6.

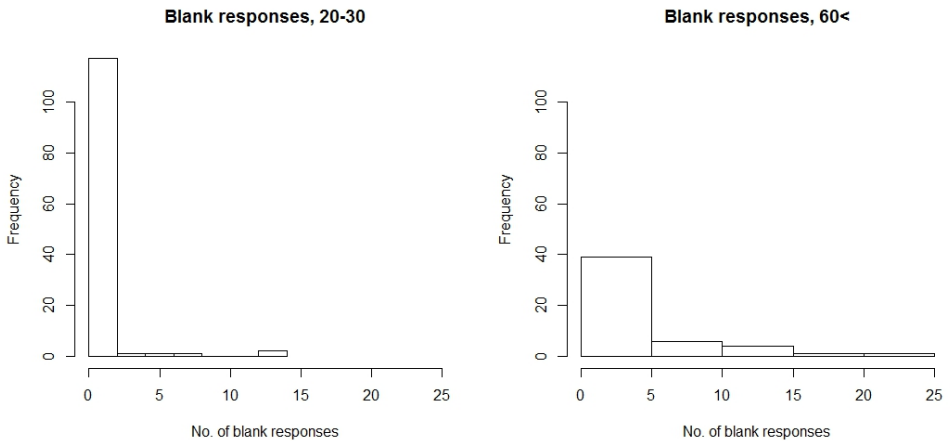


Figure 5.6: Distribution of blank responses per participant, n=122 for the younger cohort, n=51 for the older cohort

A Wilcoxon rank sum test showed that there was a statistically significant difference between the cohorts for the number of blank responses as well (p<0.001, W=4127.5, Cohen's $d \approx 0.7835519$) (see also figure 5.7).

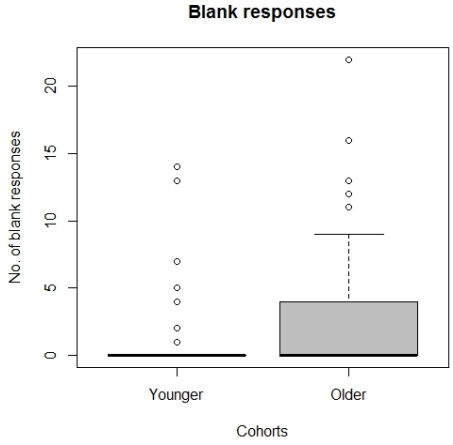


Figure 5.7: Blank responses by age

As for the multi-word responses, there was quite a high degree of spreading within the populations, especially for the older cohort. For the older cohort the observed median was 0, with a range from 0 to 22. The corresponding numbers for the younger cohort were median=0, range=0-14 (table 5.8). These patterns are reminiscent of what was seen for the multi-word response frequencies across the two age cohorts, with a significant difference in the proportion of multi-word responses between the cohorts. There was

a higher proportion of both multi-word and blank responses for the older cohort, and more dispersion in the number of both multi-word and blank responses within the older cohort.

5.4.4 Blank responses: cue word differences

As found for the multi-word responses, there were also differences in the proportion of blank responses for the different cue words. The younger cohort had a span from 0% (n=59) to 3.28% (n=5) blank responses per cue word. For the older cohort there was a span from 0% (n=39), to 15.69% (n=1).[54] A full overview of the number of blank responses per cue word for both cohorts can be seen in appendix H.

A prediction made for the number of blank responses was that the cue word class itself would not influence the number of blank responses (3.3.2). The medians for the number of blank responses for noun cues and verb cues for each cohort can be seen in table 5.10. As for the median values and ranges for multi-word responses for nouns and verbs, these are not directly compatible as the cohorts had different sizes (see 5.4.2). Shapiro-Wilk tests showed that none of the samples were normally distributed (see table 5.11).

	20-30 y/o		**Over 60 y/0**	
	median	**range**	**median**	**range)**
Noun	0	0-4	1	0-8
Verb	0	0-3	1	0-5

Table 5.10: Cue word class and blank responses, both cohorts

	20-30 y/o	**Over 60 y/o**
Noun	*** p<0.001 W=0.65545	*** p<0.001 W=0.77645
Verb	*** p<0.001 W=0.72195	** p≈0.0026 W=0.86325

Table 5.11: Shapiro-Wilk test results for cue word class and blank responses

Wilcoxon rank sum tests were used to compare the number of blank responses for verb cues compared to for noun cues. These tests showed that there were no significant differences between noun cues and verb cues for either of the cohorts (Younger cohort: p=0.8027, W=693, Cohen's d ≈0.0055. Older cohort: p=0.3486, W=626, Cohen's d≈-0.1049).

In 3.1.3 it was predicted that there might be a link between the imagebility rating of the cue word and the number of blank responses triggered is triggers, where a low imageability rating might cause more blank responses than a high rating (see 4.5.8 for more on the imageability data). Table 5.12 shows the median number of blank responses for cues with a high imageability rating and for cues with low imageability ratings. The median values are presented for each of the two age cohorts, and for the total sum of participants.

[54]The proportions of blank responses are given as percentages due to the different in cohort size. The younger cohort had 122 participants, and the older cohort 51 participants.

	20-30 y/o (n=122)		**Over 60 y/o** (n=51)		**Total** (n=173)	
	median	**range**	**median**	**range**	**median**	**range**
Low imageability	0	0 - 4	1	0 - 5	2	0 - 8
High imageability	0	0 - 0	0	0 - 1	0	0 - 1

Table 5.12: Imageability ratings and number of blank responses. Note that comparisons are only made for the 71 cue words that imageability data was available for (see 4.5.8)

The results from the Shapiro-Wilk tests in the samples (see table 5.13) showed that none of them followed a normal distribution.

	20-30 y/o	**Over 60 y/o**	**Total**
Low imageability	*** p<0.001, W=0.72879	*** p<0.001, W=0.85675	*** p<0.001, W=0.9142
High imageability	no blanks	*** p<0.001, W=0.46465	*** p<0.001, W=0.46465

Table 5.13: Shapiro-Wilk test results for the relationship between cue word imageability ratings and the number of blank responses

Wilcoxon rank sum tests were used to assess whether any differences could be found in the proportion of blank responses given to the cue words with low imageability compared to those with high imageability (see table 5.14). The potential influence of imageability on the number of blank responses was tested for both cohorts separately and for all participants together.[55]

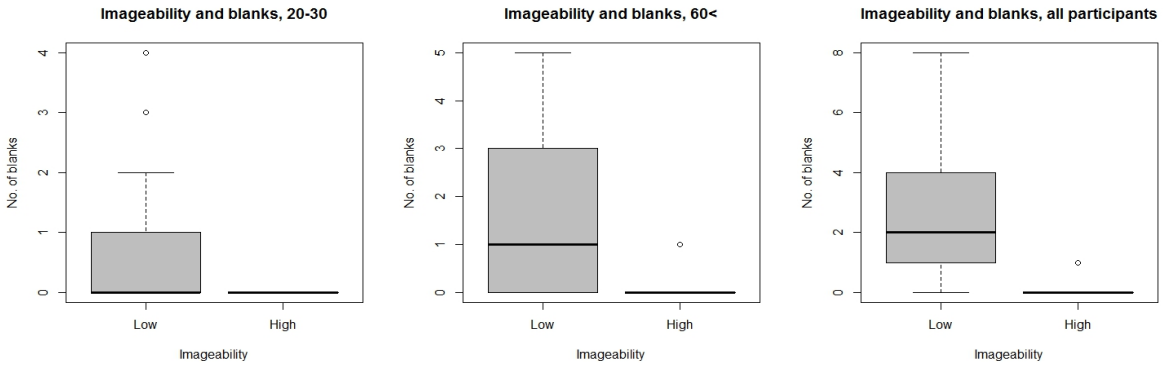
	20-30 y/o	**Over 60 y/o**	**Total**
Wilcoxon (Cohen's *d*)	** p≈0.0055, W=510 (0.794)	** p≈0.0013, W=555 (1.037)	*** p<0.001, W=606 (1.199)

Table 5.14: Wilcoxon rank sum tests and Cohen's d for the relationship between cue word imageability rating and the number of blank responses

As can be seen in table 5.14, there was a significant effect between the number of blanks and the imageability rating of the cue words for each of the cohorts as well as for all participants counted together. Figure 5.8 visualises the very low number of blanks for highly imageable cue words.

However, note should be made that there were only 12 cue words with a high level of imageability, and thus this sample was rather small. Still, there was a clear tendency that low imageability might lead to more blank responses than a high imageability rating. These findings are further discussed in 6.2.2.

[55]Note should also be made that, as presented in 5.4.3, the are significant cohort differences in the number of blank responses given, and imageability effects on blank response were thus checked for both age cohorts individually prior to checking for all participants together.



(a) Relationship between imageability and number of blanks, younger cohort, n=122

(b) Relationship between imageability and number of blanks, older cohort, n=51

(c) Relationship between imageability and number of blanks, all participants, n=173

Figure 5.8: Relationship between imageability and blank responses

5.5 Cue word class influence

In Chapter 3, it was predicted that the fundamental differences between nouns and verbs in cognitive grammar would be reflected in the response data from the current study (see 3.1.1). In the following section, the effects of the cue word class on the word class of responses it triggers are presented. These data will be discussed in Chapter 6 in terms of how they shed light on theories on the mental lexicon (cf. the second set of research questions, Section 1.2). See also 5.6.2 for results concerning the relationship between cue word class and response categories. Based on the assumed fundamental differences between nouns and verbs in usage based theory (see Section 3.1.1), and on findings from previous research on word associations and the mental lexicon (see 2.1.6), it was hypothesised that the tendency to elicit dominant responses in the same word class would be stronger for noun cues than for verb cues.

To evaluate the difference between the word class effects of noun cues and verb cues, data from the younger cohort was used to count the proportion of noun cues the participant responded to with a noun response, and similarily for verb cues, with verb responses. Since the younger cohort has more participants (n=122), it was believed this choice would produce more reliable statistical results than for the older cohort (n=51). The prevalence of nouns elicited from noun cues, and verbs elicited from verb cues (i.e. the proportion of nouns given as responses to noun cues by the participant divided by the total amount of noun cues, and similarly for verbs) were used as the two populations for the testing. Shapiro-Wilk tests to assess whether the samples followed normal distributions returned values of p=0.007036, W=0.96937 (noun-noun), and p=0.1434, W=0.98352 (verb-verb). As the noun-noun data was not normally distributed, a Wilcoxon rank sum test was used to check whether there was a significant difference in the proportion of responses given in the same word class as the cue for nouns compared to for verbs. The Wilcoxon rank

sum test showed that the two populations were nonidentical ($p<0.001$, $W=14207$, Cohen's $d\approx2.465384$), and that nouns elicited significantly more noun responses than verbs elicited verb responses (see also figure 5.9).

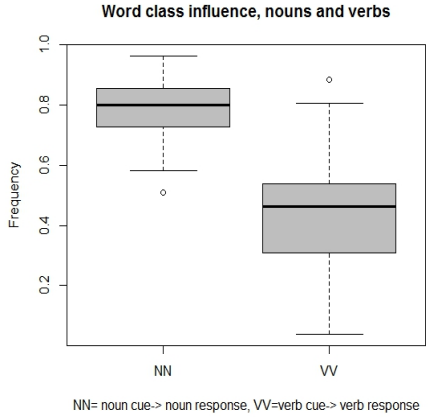
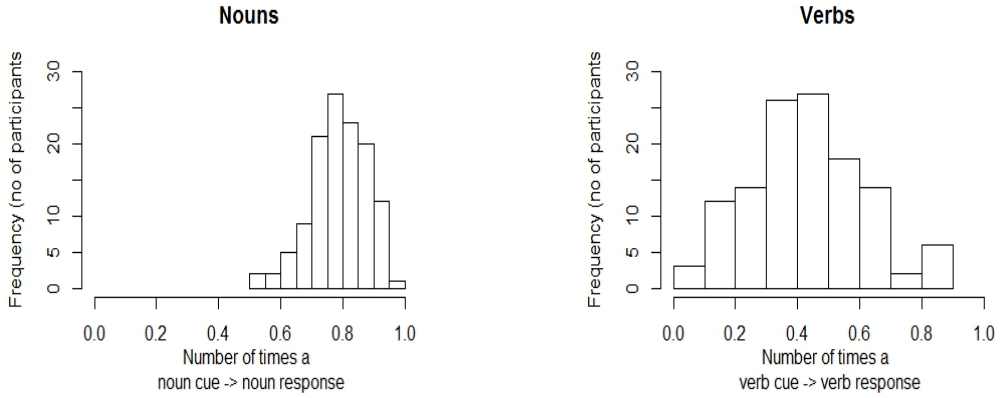


Figure 5.9: Cue word class influence compared. Data based on the proportion of responses in the same word class as the cue for each participant

Figures 5.10a and 5.10b show how many of the noun cues and verb cues that elicited noun responses and verb responses, respectively, by the participants in the younger cohort. The noun cues exhibited a higher frequency of responses from the same word class as themselves, and were also more consistent in this behaviour (median=0.8, range=0.5091 - 0.9636), compared to verbs (median=0.46154, range=0.03846 - 0.8846).[56] A note should be made in relation to all these results that due to the lower number of verb cues (n=26) than noun cues (n=55), the results for verbs are less strong than the noun results. However, the Wilcoxon test does take the different sample sizes into consideration, and thus the p-value at any rate reflects the probability of the results.

[56]As the test did not have equal amounts of noun cues and verb cues, proportional values are given here. A value of 1.0 means that the cue only elicited responses from its own word class.



(a) Proportion of noun cue words resulting in noun responses for each participant. 1.0 equals that all noun cues elicited a noun response for the given participant

(b) Proportion of verb cue words resulting in verb responses for each participant. 1.0 equals that all verb cues elicited a verb response for the given participant

Figure 5.10: Cue word class influence

The response word class distribution for all verb cues can be seen in Appendix K. I initially planned to test the relationship between the transitivity of the verbs and the word class distribution of the responses the verbs triggered. This was however not done, for two reasons. Firstly, a high number of verbs have alternating transitivity. In the word association test they are given without any context, and as such it seems any results on this would still have to be presented with some caution.

Secondly, from looking at the data, the response word class distribution fairly clearly does not seem to follow the transitivity of the verbs. For instance, the verb cue *ansette* (eng: *employ*) triggered, in line with what might be expected based on its transitivity, mainly noun responses (n=98, from 122 participants). However, a verb cue such as *formidle* (Eng: *convey/impart/arrange*), triggered mainly verb responses (n=83). The variation between the response patterns of the verb cues that could be observed just by looking at the data appeared to have very little relation between transitivity and response word class. Thus no further statistical tests were carried out; however, it should be an aim of further research to explore which factors, if any, that influence the response word class patterns for verb cues. The observations from the current study strongly suggest that transitivity on its own is not the decisive factor in this matter.

5.6 Response type categorisation

Following the aims of the first set of research questions, on whether the word associations are different in the two age cohorts (see 1.2), and the previous research on word associations and ageing, it was predicted that one might see age-related differences in response

category patterns. The results for this prediction are presented in 5.6.1. The response categories will also be used to further explore the way word association data might shed light on the differences in representations of nouns and verbs in the mental lexicon. These are presented in 5.6.2. See also 3.1.1 for theoretical background and predictions, and 5.5 for the results concerning cue word class influence on the word class of the responses.

For both the cohort comparison and the cue word class comparison, two sets of comparisons were carried out using both the five overarching response categories, and the fourteen subcategories presented in table 5.15. See also table 4.4 in Chapter **??** for examples of all fourteen subcategories.

Category	**Subcategory**
Meaning-based association	Synonym
	Lexical set
	Other conceptual
Position-based association	Cue-response collocation
	Response-cue collocation
	Cue-response & response-cue collocation
Form-based association	Affix manipulation
	Similar in form only
	Two-step association
Dual association	Lexical set *and* cue-response collocation
	Lexical set *and* response-cue collocation
	Synonym *and* cue-response collocation
	Synonym *and* response-cue collocation
Erratic association	Erratic

Table 5.15: Overarching categories and subcategories for response classification, based on Fitzpatrick et al. (2015)

The response category data also provides an overview of what response types are most common in general. As can be seen in figure 5.11, and tables 5.16 and 5.18, for both cohorts there is a preference for meaning-based responses. Position-based responses are the second most common for both cohorts. In fact, these two overarching categories to a large extent dominate the category distribution, as will be seen in the next sections.

5.6.1 Response categories: cohort differences

To prepare the data for the analysis based on response category differences between the two age cohorts, the instances of each response category were counted per participant.

This was done for both the five overarching categories, and for the full list of fourteen subcategories. The median number of responses given for the two cohorts within each of the overarching categories can be seen in table 5.16. The median number of responses given for the two cohorts for all of the fourteen response categories can be seen in table 5.18. Boxplots for all fourteen subcategories can be found in Appendix L.

Category	**20-30 y/o median (range)**	**Over 60 y/o median (range)**	**Group with higher median**
Meaning-based	70 (26 - 84)	80 (48 - 95)	O
Position-based	18 (4 - 34)	8 (1 - 23)	Y
Form-based	2 (0 - 17)	1 (0 - 3)	Y
Dual	3 (0 - 9)	3 (0 - 9)	Equal
Erratic	3 (0 - 22)	3 (0 - 33)	Equal

Table 5.16: Median proportions of responses for the overarching response categories for both cohorts

Shapiro-Wilk tests gave $p<0.05$ for all samples except the position-based responses for the younger cohort ($p=0.143$, $W=0.98351$), see table 5.17. Wilcoxon rank sum tests were then used to compare the response category distributions for the two age cohorts (table 5.17). The distributions can also be seen in figure 5.11.

Category	**20-30 y/o Shapiro-Wilk**	**Over 60 y/o Shapiro-Wilk**	**Wilcoxon rank sum test p-value**	**W**	**Cohen's *d***
Meaning-based	*** p<0.001, W=0.90496	** p=0.007398, W=0.93444	*** <0.001	1137	-1.165
Position-based	p=0.143, W=0.98351	*** p<0.001, W=0.89926	*** <0.001	5509.5	1.652
Form-based	*** p<0.001, W=0.80018	*** p<0.001, W=0.78881	*** <0.001	4920	0.917
Dual	** p=0.001269, W=0.96062	*** p<0.001, W=0.90578	0.3309	2823.5	-0.181
Erratic	*** p<0.001, W=0.82486	*** p<0.001, W=0.60606	0.9451	3090	0.002

Table 5.17: Shapiro-Wilk tests, Wilcoxon rank sum tests, and Cohen's d effect size values for the overarching response categories compared across the two cohorts

When comparing the response category patterns of the two cohorts based on the five overarching categories, there were clear, significant differences and pronounced effect sizes between the cohorts for the number of meaning-based, position-based, and form-based responses (see table 5.17).

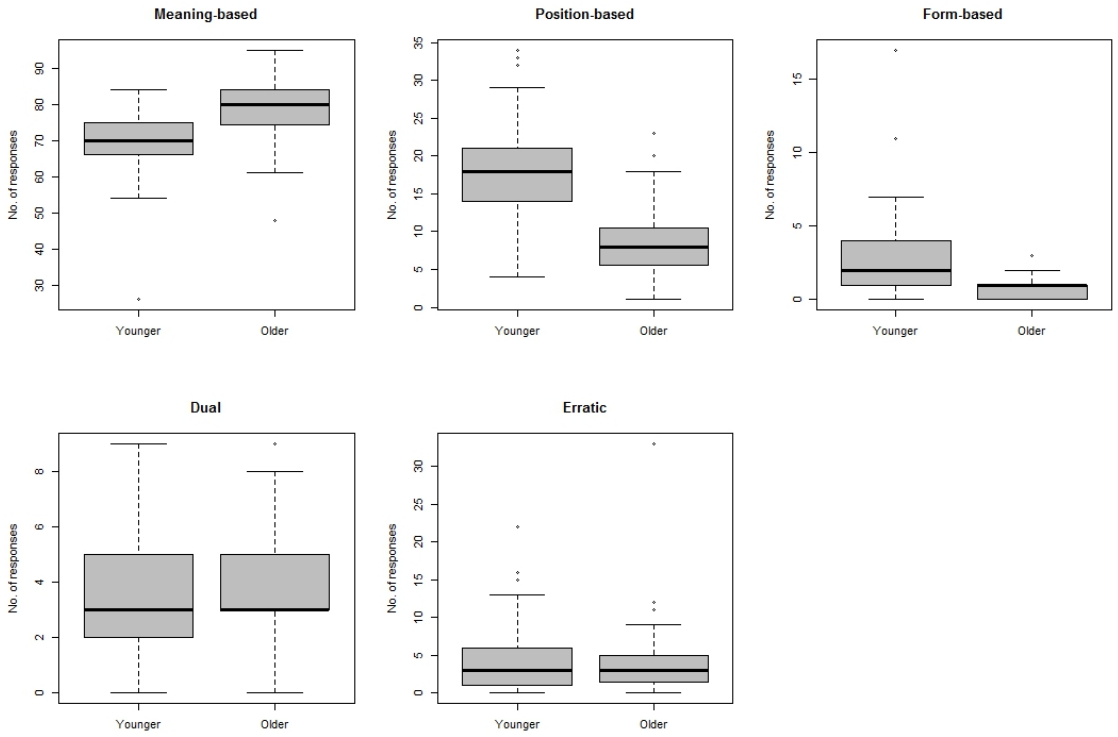


Figure 5.11: Overarching response categories, both age cohorts

However, when looking at the patterns based on all fourteen categories instead of the five overarching ones, the results showed that the differences between the cohorts are not equal across all the subcategories within an overarching category (see table 5.18).

Category	**20-30 y/o median (range)**	**Over 60 y/o median (range)**	**Wilcoxon rank sum test p (adjusted)**	**W**	**Cohen's *d***	**Group with higher median**
Synonym	19 (2 - 44)	44 (6 - 83)	*** <0.001	938	-1.786	O
Lexical set	12 (4 - 23)	11 (4 - 23)	1	3581	0.186	Y
Other concep.	38 (5 - 56)	26 (1 - 52)	*** <0.001	4793.5	1.122	Y
Cue-response	9 (1 - 21)	4 (1 - 13)	*** <0.001	5356.5	1.354	Y
Response-cue	6 (1 - 15)	2 (0 - 10)	*** <0.001	5027.5	1.202	Y
CR-RC	2 (0 - 7)	1 (0 - 4)	0.2856	3752.5	0.428	Y
Affix mani.	1 (0 - 7)	0 (0 - 3)	* 0.0132	4004	0.575	Y
Form only	0 (0 - 10)	0 (0 - 2)	0.3864	3625.5	0.335	Equal
Two-step	1 (0 - 4)	0 (0 - 1)	*** <0.001	4864.5	1.037	Y
Erratic	3 (0 - 22)	3 (0 - 33)	1	3090	0.002	Equal
Lex + CR	1 (0 - 6)	1 (0 - 7)	1	3092.5	-0.013	Equal
Lex + RC	1 (0 - 3)	1 (0 - 3)	0.7776	3482.5	0.198	Equal
Syn + CR	0 (0 - 3)	1 (0 - 3)	** 0.0026	2090	-0.637	O
Syn + RC	0 (0 - 1)	0 (0 - 1)	1	3167.5	0.082	Equal

Table 5.18: Median proportions of responses for all 14 response categories, compared across the two cohorts

Shapiro-Wilk showed that at least one of the cohorts did not have a normally distributed sample ($p<0.05$) for each of the fourteen categories, and thus Wilcoxon rank sum tests

were used to discover potential significant differences between the response category distributions of the two cohorts (table 5.18). For instance, as illustrated in table 5.16, for the total of meaning-based responses, the younger cohort had a median of 70 responses within the category, whereas the older cohort had a median of 80 responses. However, when the meaning-based response category was divided into synonyms, lexical sets, and 'other conceptual' responses, the results showed that the older cohort only had a higher median for synonyms (44, compared to 19 for the younger cohort). For both lexical sets and 'other conceptual' responses, the younger cohort had the higher median (see table 5.18). In fact, the younger cohort had a higher median for 'other conceptual' responses (38) than for synonyms (19). For lexical sets, there was very few difference between the cohorts..

As presented in 4.5.7, the p-values need to be adjusted to avoid Type I errors when performing tests on all fourteen subcategories. The p-values from the Wilcoxon tests were multiplied by the formula given in 4.5.7, $(1\text{-}0.95^{14})/0.05 \approx 10.25$. The adjusted p-values can be seen in table 5.18. With the new, adjusted p-values, most of the significant results still held; however, the p-values for *cue-response + response-cue collocations* and *form only* responses were not significant after they had been adjusted.

5.6.2 Response categories: cue word class

As for the statistical tests on the influence of cue word class on the word class of the responses (see Section 5.5), the calculations in the current section were also based on the data from the younger cohort only (n=122, see 5.5). As in 5.5, tests to discover potential statistical differences were performed on noun cues and verb cues only; however, the median and range values for the number of responses in each category are also provided for adjective cues (see tables 5.19 and 5.20).

Category	**Noun cues median (range)**	**Verb cues median (range)**	**Adj. Cues median (range)**	**Word class w/ higher median**
Meaning-based	83 (22-118)	93.5 (47-113)	76 (41-112)	V
Position-based	17 (0-73)	11 (1-63)	24 (1-52)	A
Form-based	1 (0-18)	3 (0-47)	1 (0-47)	V
Dual	0 (0-73)	0 (0-15)	0 (0-19)	Equal
Erratic	3 (0-15)	5 (1-12)	6 (1-12)	A

Table 5.19: Median number of responses in each overarching category, for noun, verb and adjective cues

Category	Noun cues median (range)	Verb cues median (range)	Adj. Cues median (range)	Word class w/ higher median
Synonym	6 (0-89)	28 (1-91)	21 (0-66)	V
Lexical set	15 (0-96))	3 (0-28)	5 (0-38)	N
Other conceptual	33 (9-98)	42 (9-112)	48 (7-76)	A
Cue-response	5 (0-46)	7.5 (1-63)	20 (0-50)	A
Response-cue	7 (0-52)	0 (0-23)	0 (0-28)	N
CR-RC	0 (0-33)	0 (0-5)	0 (0-28)	Equal
Affix mani.	0 (0-16)	1 (0-8)	0 (0-8)	V
Form only	0 (0-5)	0 (0-4)	0 (0-4)	Equal
Two-step	0 (0-4)	0 (0-47)	0 (0-47)	Equal
Erratic	3 (0-15)	5 (1-12)	6 (1-12)	A
Lex + CR	0 (0-35)	0 (0-2)	0 (0-19)	Equal
Lex + RC	0 (0-52)	0 (0-7)	0 (0-8)	Equal
Syn + CR	0 (0-48)	0 (0-7)	0 (0-0)	Equal
Syn + RC	0 (0-7)	0 (0-0)	0 (0-0)	Equal

Table 5.20: Median number of responses in all fourteen categories, for noun, verb and adjective cues

Shapiro-Wilk tests showed that none of the samples for the noun cues for the overarching categories were normally distributed ($p<0.05$). For the verb cues, only the meaning-based category gave a $p>0.05$ ($p=0.05046$, $W=0.92213$). The other four categories for verbs also gave $p<0.05$. Wilcoxon rank sum tests were performed to check for potential statistically significant differences between the response categories given to noun cues compared to verb cues (see table 5.21).

Category	Wilcoxon rank sum test			Cohen's *d*
Meaning-based		p=0.4299	W=636.5	-0.2464
Position-based		p=0.2442	W=830.5	0.2872
Form-based	***	p<0.001	W=366.5	-0.5453
Dual	*	p=0.01455	W=922.5	0.4761
Erratic	*	p=0.01841	W=485	-0.4348

Table 5.21: Significance test results and effect sizes for the effect of cue word class on the five overarching response categories (younger cohort only)

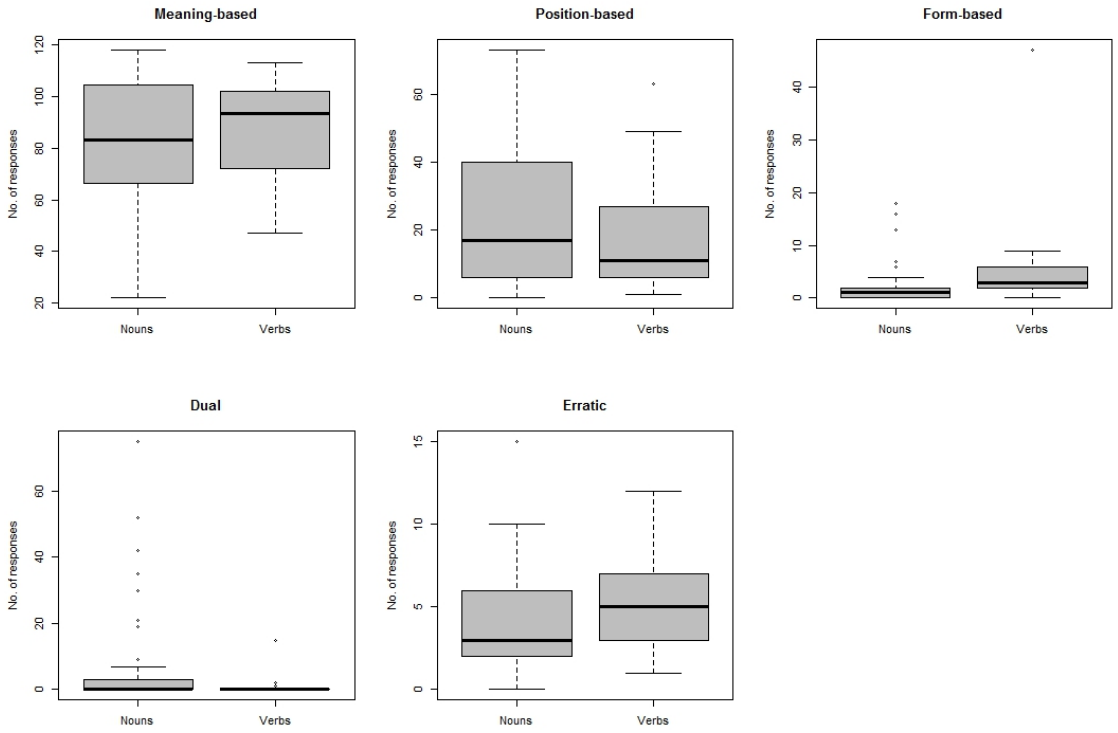


Figure 5.12: Overarching response categories, noun cues and verb cues. Based on data from the younger cohort (n=122)

The results from the Wilcoxon rank sum tests showed that there were statistically significant differences between noun cues and verb cues for form-based responses, dual category responses, and erratic responses. However, these are as previously mentioned, much less frequent than the meaning-based and position-based categories.

For the fourteen overarching categories, Shapiro-Wilk tests showed that none of the noun samples were normally distributed ($p<0.05$). For the verb samples, 11 categories were not normally distributed ($p<0.05$). The three categories that returned a value of $p>0.05$ for the Shapiro-Wilk tests were synonyms ($p=0.1372$), 'other conceptual' ($p=0.09274$), and erratic responses ($p=0.06707$). Wilcoxon rank sum tests were performed to check whether there were significant differences between the response category patterns of noun cues and verb cues (table 5.22, boxplots can be found in Appendix L).

Category		p (adjusted)	W	Cohen's d
Synonym	**	0.004	364.5	-0.6975
Lexical set	*	0.01001	1040.5	0.8218
Other conceptual		1	665.5	-0.2062
Cue-response		0.389	510	-0.5743
Response-cue	***	<0.001	1210	0.943
CR-RC		0.1655	912.5	0.5042
Affix mani.		0.183	516	-0.2343
Form only		1	614	-0.3733
Two-step		0.6355	580.5	-0.4078
Erratic		0.1887	485	-0.4348
Lex + CR		0.3319	866	0.4192
Lex + RC		1	761	0.324
Syn + CR		1	713.5	0.1198
Syn + RC		1	741	0.2357

Table 5.22: Significance of differences (from Wilcoxon rank sum tests) between the response categories for the responses made to noun cues and verb cues, all categories (for the younger cohort)

As for the significance tests using all fourteen subcategories in 5.6.1, the p-values were also adjusted for these calculations to avoid Type I errors. The formula presented in 4.5.7 was used for these adjustments as well. With the new, adjusted p-values, significant differences were found for synonyms, lexical sets, and response-cue collocations.

5.6.3 Response categories: imageability

It was predicted in 3.1.3 that the imageability ratings of the cue words might affect response category patterns, as imageability might affect storage and retrieval in the mental lexicon. The differences between the response category patterns of noun cues and verb cues were only found to be significant for some categories (see tables 5.21 and 5.22). However, other results, such as for cue word class influence on response word classes (5.5), did show significant differences between the responses for nouns and verbs. There were only two verbs with a high imageability rating in the available imageability data, and this sample was thus too small for tests on the difference between verbs with high and with low imageability ratings to be fruitful. Testing was done for the potential difference between nouns with high and with low imageability ratings, and for the total of all cue words with available imageability ratings.[57] Median values and ranges for the different samples can be seen in table 5.23.

[57]The *total* calculations include all 71 nouns, verbs, and adjectives imageability ratings were available for.

	Meaning-based		Position-based		Form-based	
	median	range	median	range	median	range
Noun-low	80	22 - 118	17	0 - 73	1	0 - 16
Noun-high	69.5	32 - 114	23.5	3 - 59	0	0 - 2
Total low	83	22 - 118	20	0 - 73	1	0 - 47
Total high	75.5	32 - 114	29	3 - 59	0.5	0 - 6
	Dual		Erratic			
	median	range	median	range		
Noun-low	0	0 - 75	3	0 - 15		
Noun-high	5	0 - 42	2	0 - 10		
Total low	0	0 - 75	4	0 - 15		
Total high	2	0 - 42	2	0 - 10		

Table 5.23: Median values for cue word imageability and response categories

Statistical testing was done only for the five overarching response categories, to get more robust results. Shapiro-Wilk tests were performed to see whether the samples were normally distributed. The results of these tests can be seen in table 5.24.

	Meaning-based		Position-based		Form-based	
	p	W	p	W	p	W
Noun-low	* 0.0317	0.92479	* 0.0146	0.91204	*** <0.001	0.53119
Noun-high	0.2146	0.89919	* 0.01341	0.7972	** 0.002088	0.73087
Total low	* 0.01078	0.94579	*** <0.001	0.9155	*** <0.001	0.47408
Total high	0.2787	0.91912	* 0.04468	0.85687	*** <0.001	0.62327
	Dual		Erratic			
	p	W	p	W		
Noun-low	*** <0.001	0.38275	*** <0.001	0.82602		
Noun-high	* 0.01974	0.81103	*** <0.001	0.64132		
Total low	*** <0.001	0.31544	*** <0.001	0.91708		
Total high	** 0.003066	0.75569	** 0.001349	0.72088		

Table 5.24: Shapiro-Wilk test results for cue word imageability ratings and response category patterns

Some samples returned a value of p>0.05, however, there were no low imageability/high imageability pairs for any of the categories that had >0.05 for both samples in the pair.[58] Thus, Wilcoxon rank sum tests were used for all comparisons (see table 5.25).

[58]A low imageability/high imageability pair is here used to denote for instance the meaning-based responses for nouns with low imageability and the meaning-based responses for nouns with high iamgeability, i.e. the pairs that would be compared in a test for statictically significant difference between the two samples.

	Meaning-based		Position-based		Form-based	
	p-value	**W (Cohen's d)**	**p-value**	**W (Cohen's d)**	**p-value**	**W (Cohen's d)**
Nouns	0.3947	183.5 (0.3430)	0.6161	138 (-0.2012	0.3469	184 (0.3954)
Total	0.3732	412.5 (0.3596)	0.394	298 (-0.2917)	0.1768	439 (0.3613)
	Dual		**Erratic**			
	p-value	**W (Cohen's d)**	**p-value**	**W (Cohen's d)**		
Nouns	0.124	108 (-0.4999)	0.1579	200.5 (0.3851)		
Total	* 0.03156	234 (-0.6210)	0.06567	473 (0.4929)		

Table 5.25: Wilcoxon rank sum results (and Cohen's d) for imageability influence on response categories

No statistically significant effects were found for the effect of imageability on responses for either of the five response categories for the noun cue words. For all cue words considered together, significant effects were only found for the 'dual' category.[59]

[59]Medians and ranges for the distribution of response categories for nouns and verbs, without imageability being considered, can be seen in Section 5.6.2.

Chapter 6

Discussion

The results of the word association study will be discussed in accordance with the research questions for this project, as presented in 1.2. The first section (6.1) will discuss findings related to the first set of research questions: Are the word associations different in the two cohorts? How do these compare to the results found in Fitzpatrick et al. (2015), and in other previous studies using word association tests across age cohorts?

The second section (6.2) is linked to the second set of research questions: How does the data shed light on theories on the mental lexicon? What might the data tell us about the organisation of the mental lexicon, and of storage and access in particular? The data will also be compared to relevant findings from other previous word association studies. Lastly, the third section (6.3) discusses the findings relevant for the third set of research questions for the project: How does the data shed light on theories on language and ageing? What might they tell us about potential changes in linguistic abilities with old age?

6.1 Cohort differences in word association responses

In Chapter 5, several sets of results were presented that explored what cohort differences could be seen in the word association responses from the current study. Specifically, the focus was on the implications of the difference in age between the two cohorts. Age-related differences were found in the dominant responses for the cue words. Differences were also found in the rest of the norms lists for the two cohorts (5.1), in terms of the semantic properties of the responses to some of the cue words (5.2), in the amount, and types of foreign language responses (5.3), and for the distribution of at least some response categories (5.6.1). There were also differences between the cohorts in the number of multi-word and blank responses. These will be discussed in Section 6.3.

6.1.1 Norms lists

Both Hirsch and Tree (2001) and Fitzpatrick et al. (2015) found age related differences in the dominant responses of the norms lists for the age cohorts in their studies (see 2.1.4). Hirsch and Tree found only a moderate level of overlap between the dominant responses for the two cohorts, where the most dominant response was shared for 36 out of the 90 cue words used in their study (40%). When comparing the top three responses for each cue word, the two cohorts had an overlap of 57% (Hirsch and Tree, 2001, p. 7). In Fitzpatrick et al. (2015) the amount of overlaps for the dominant responses were 47%. The data from the current study thus coincides with these two other studies, with an overlap of 41% for the dominant responses, and 45% when comparing the top three responses (5.1). Note should again be made in this section that Hirsch and Tree used only noun cues (Hirsch and Tree, 2001, p. 3); however, the results from all three studies show that there is age-related variation in word association responses.

As was seen in 5.1.1, both cohorts got a higher stereotypy score when being scored against a norms list that was based on their own cohort rather than a list that was based on the other cohort. In Fitzpatrick et al. (2015), the study found that participants overall gained an average of 6.46 stereotypy points when being compared to a norms list for their own cohort, instead of norms from the other age cohort (Fitzpatrick et al., 2015, p. 36). In the present study, the corresponding overall average gain was 8.49 points, somewhat larger than in Fitzpatrick et al. (2015). As mentioned in 5.1.1, it appears that the older cohort gained more than the younger cohort from being compared to a norms list matched for age. However, as also mentioned in 5.1, this might be caused by the relative low number of participants in the older cohort (n=51), which led to more split dominant responses (i.e. two or more responses sharing the first place to a given cue word). However, despite this potential effect of the ties for more dominant responses in the older than in the younger cohort's norms list, there is still a clear, significant difference between scores to *own list* and *other list* for both cohorts, and for all participants taken together. The differences between the norms lists for the two cohorts clearly show that there is a need for norms lists to be matched for age. In potential clinical uses, or further studies on age related changes in linguistic abilities, to use a standardised norms list based on another age cohort will be likely to seriously skew the results.

6.1.2 Comparison with data from Håseth (1968)

In 2.1.4, it was predicted that the comparison of the responses from the five overlapping cue words from the two cohorts in the present study and the study in Håseth (1968), might be able to shed some light on the reason for the observed cohort differences in norms (see

5.1.2 for more on the results of the comparison). Clearly, the very small number of cue words, and the limited data to base the analysis on might not be enough to reveal any differences. It might be that other cue words would have revealed differences; however, for these five cues, there were few longitudinal and/or cross-cohort differences, and the results were not really able to shed any further light on the observed age cohort differences in the norms lists.

The cue words used in Håseth (1968) were Norwegian translations of the Kent-Rosanoff list (Kent and Rosanoff, 1910). The cue words from Kent and Rosanoff (1910) were compiled to be used for psychological diagnosis (see also 2.1.1), and their continued use in linguistic research has been criticised by for instance Fitzpatrick, who notes that "researchers seem surprisingly untroubled by the fact that the lists were originally compiled in such an unprincipled way" (Fitzpatrick, 2007, p. 321). The Kent-Rosanoff list has also been criticised for almost only containing nouns and adjectives, which means that one cannot generalise outside of these two word classes if using the list (Deese, 1962, p. 79). The data from the five cue words from Håseth (1968) might thus not be the best basis for discovering longitudinal changes in word associations. Little research has been done on longitudinal word association studies, and it should be an aim of further research to study age-related changes not just cross-sectionally as in the current study, but also longitudinally. It would have been beneficial to be able to retest the same participants after a longer period of time, but this was alas not possible for the current study, as the participant data from Håseth (1968) had been destroyed (see 4.5.2).

The high level of overlap in the comparison with the data from Håseth (1968) suggests that one might be able to, to some extent, control the degree of overlap by the choice of cue words for the test. Compare this for instance to the generally moderate level of overlap between the norms lists for the two cohorts in the current study (6.1.1), where the cue words were chosen in a very different manner (4.1.3 than for the Kent-Rosanoff list). This observation supports Fitzpatrick's notion that cue word lists need to be compiled in a principled manner (Fitzpatrick et al., 2015, p. 124). For one this means that one should take the way the cue words were compiled into consideration when analysing the responses they elicited. Secondly, it might also be beneficial to choose cue words with a specific set of attributes if one wishes to study a particular phenomenon, or the effects of a certain factor on the responses.

6.1.3 Meaning differences in responses

Although the comparison with the data for five cue words from Håseth (1968) did not reveal any clear longitudinal or generational differences, other data from the current study showed how the meanings and usage of words may change over time. For a more thorough

presentation of the findings of meaning differences across the two cohorts, see 5.2.

In 2.1.5, three hypotheses from Fitzpatrick et al. (2015) on what may cause age-related differences in word association behaviour were presented. Firstly, the differences might be due to a change in word association strategies as a function of ageing. A second possible reason is that different generations have different preferences for vocabulary and/or associations. A third hypothesis is that these two factors interact, and that associations are affected by both ageing and generation (Fitzpatrick et al., 2015, pp. 36-7). The findings discussed here appears to match with the third of these hypotheses. As was seen in the results, there appears to be changes in meaning and/or usage across the cohorts for some cue words, but that these are also difficult to predict. Some differences were rather clearly linked to generational differences, such as for instance the many responses among the younger cohort related to *russefeiring* (celebrations of leaving upper secondary school in Norway) for the cue word *konsept* (Eng: *concept* (N)). This celebration would have taken place when the older cohort were at that age as well, however, it has in more recent times become a massive celebration, with new adjacent traditions, which are then reflected in the younger cohort's responses. The celebrations are also a more recent memory for the younger cohort. Additionally, regardless of the changes in the celebrations themselves, the word *konsept* also appears to have changed semantically, to include, for the younger cohort, a very specific meaning of *konsept* as a theme for an event, or similar uses. Other differences were more difficult to explain, such as for the cue word *psykisk* (Eng: *psychic*, where the top response for the younger cohort did not occur at all in the responses for the older cohort. For other cue words again, the expected differences did not occur, such as for the cue word *innvandrer* (Eng: *immigrant*) (see 5.2 for more on the differences in meaning and usage across the cohorts).

It is also possible that the fact that all participants resided in Oslo might have influenced the responses made. For instance, if participants rather had been recruited from more rural areas, the responses might have differed in other ways than they did in the current study. It has for example been reported that attitudes to immigration differ based on whether informant reside in sparsely populated areas or urban settlements (Blom, 2015, p. 43), which might have led to a change in the responses given to the cue words *innvandrer* (Eng: *immigrant* (N)), one of the cue words which produced surprisingly few differences between cohorts in the current study (5.2), or to *mangfold* (Eng: *diversity* (N)). Likewise, if one had had a larger group of participants with more diversity in terms of education level, one might also have seen differences in responses based on this. In the data collected from the younger cohort, who had all undertaken some amount of higher education, their academic background was mirrored in the responses they produced. For instance, the cue word *analyse* (Eng:*analysis* (N)), triggered responses such as *metode, kvalitativ, SPSS* (statistics software) and *forske* (Eng: *methodology (N), qualitative (A), SPSS (N)* and

research(V)) for the younger, but not for the older cohort. These responses seem likely to be linked to the participants' experience with higher education, and the responses given to *analyse* could have been different if the younger cohort had rather been recruited from for example vocational courses. The difficulty in predicting where changes or differences might be observed suggests that norms lists should perhaps be matched not just for age, but also potentially for other factors relevant for the study at hand, such as education level, or place of residence, or socio-economical factors.

6.1.4 Foreign language responses

Another set of differences between the two cohorts was observed in the surprising amount of, and distribution of, foreign language responses. As presented in 5.3, the younger cohort produced more English responses than the older cohort (1.27% compared to 0.41%), whereas the older cohort produced more other foreign language responses than the younger cohort (0.55% compared to 0.05%). This reflects a development in language education in Norwegian schools. The older cohort was likely to have learnt some German, French and English in school over the last decades. The younger cohort would have learnt English at school from an earlier age than the older cohort, and have had more English instruction. They would also have been able to choose between more languages for a second foreign languages, such as for instance Spanish in addition to German and French, and thus their knowledge of other foreign languages than English would be likely to be more diverse than for the older cohort. In general, the younger cohort would also have been exposed to English in many more domains of life, and more often than the older cohort.

This increased exposure to English might explain the observed variation in the English responses among the younger cohort, compared to the responses in other foreign languages (for both cohorts, see 5.3 for examples). English appears to be readily available for activation for the younger cohort. For some cases, such as the cue-response pair *offer-tilbud* (Eng: *victim/sacrifice- offer supply*) (see 5.3), it even appears some participants read the Norwegian cue word *offer* as if it was an English word, and responded to it as such. As the response, *tilbud*, was given in Norwegian, this kind of responses suggest that at some point in the word association process, English either was briefly activated, or that Norwegian and English were both activated at the same time for the given participant. One might also, in terms of storage in and access from the mental lexicon (which will be further discussed in 6.2), suggest that these results show that there is little separation between the storage of Norwegian and English items in the lexicons of these participants.

The apparent ease with which the younger participants activate and use English, even in a monolingual Norwegian test setting, may be taken as a sign of a very high knowledge of English as a second language among younger Norwegians, at least among those in

higher education. These results are in line with previous research on the role of English in Norway today,and particularly in academia. A study by Ljosland (2007) concluded that English is favoured over Norwegian in academia, and that in the area of scientific publishing, "[d]omain loss is thus very near complete" (p. 409). Ljosland sees the situation in academia as a "diglossic-like state", which might in turn affect the linguistic situation in Norway as a whole, with other areas of society as well moving towards diglossia (Ljosland, 2007, p. 409). The data from the current study are not adequate to say anything directly about the prediction made by Ljosland; however, for Norwegian society in general, there is a tendency, like in many other countries, for English to gain influence in more and more areas. These changes have happened continuously over a long period of time, but appear to have accelerated the last decades (Mæhlum, 2007, p. 160). The difference in the proportion of English responses between the cohorts thus reflects this rapid increase in the use of English in Norway during the last decades. Additionally, due to the younger cohort all having some level of higher education, the amount of English in the current group of participants may have been higher than it would have been for another population with a different level of education.

6.1.5 Response categories

In 2.1.5 it was predicted, based on previous studies such as Fitzpatrick et al. (2015), that a more detailed classification system than what has been traditionally used in word association research might reveal age-related differences in response category patterns that might be concealed by the use of fewer, broader categories. The results presented in 5.6.1 found cohort differences both for the calculations made for the five overarching categories, and for the full set of fourteen categories. However, the tests for all fourteen categories revealed additional differences, that had not been visible from the analysis using the five overarching categories.

The comparison of the response category data from the two cohorts showed significant differences for four of the five overarching categories (including the two most common ones, *meaning-based* and *position-based* responses). For the full list of fourteen categories, significant differences were found for seven categories. This pattern is similar to what was seen for the overarching categories, in that six of the nine categories that make up the overarching meaning-based, position-based, and form-based categories showed significant differences between the cohorts. However, the use of fourteen rather than five categories also revealed a new set of differences between the cohorts. For example, there was no significant difference for lexical sets, and a very large difference in terms of the frequency of synonyms, with the older cohort producing a median number of 44 (range: 6 - 83) synonyms, compared to the younger cohort's median of 19 (range: 2 - 44) (see 5.6.1 for a

full overview of the results of cohort differences in response category patterns).

The differences in the three meaning-based categories (*synonyms*, *lexical sets* and *other conceptual*) reflect an observed tendency in attitude to the word association test experienced for some of the participants in the older cohort. In some of the test sessions (see 4.4), the older participants expressed worries that they were making "wrong" responses, or stated that they were not able to find a word with the same meaning as the cue word. Some also talked about[60] their "crossword gene" (i.e. that they usually solve a lot of crosswords) post-testing as a reason for why they had wanted to write a lot of synonyms. Although only some expressed this, it is possible that a larger part of the older cohort also responded to the test in this manner, which would potentially explain (some of) the larger frequency of synonyms for the older cohort. Based on this it also seems that some of the older participants had a different approach to the word association test than the younger cohort, and felt more like they were being tested, in the sense of their skills in association being measured. Again, this is based on notions expressed by some of the participants during or after the test session. All the participants were informed, both orally, and in the written instructions on the front page of the list of cue words, that they were to write down the first word that came to mind as quickly as possible, and that no responses were right or wrong. In the cases where participants expressed difficulties with finding a "correct" response word, they were also reminded of this. It thus appears that the observed effect might be difficult to avoid, and potential solutions to this should be a focus of future word association studies. In retrospect, one such change could have been to use a different term than *test* to describe the task at hand to the participants, in order to not trigger any additional concerns about making "wrong" responses.

A different attitude to the test situation may however not explain all the observed differences in response category patterns between the two cohorts. For example is the significantly greater amount of cue-response and response-cue collocations for the other cohort more difficult to explain (see 5.6.1). It is possible that these other differences reflect some sort of age-related difference in association strategies, although the data from the current study is not on its own enough to propose a specific change that might have caused the observed differences.

In 2.1.5 it was predicted that the use of fourteen, rather than three, classification categories, might reveal differences between age cohorts which had been hidden in previous studies. Tendencies of this has already been presented above in realtion to the statistical testing of response category distributions. One previous study on age differences in word association behaviour is Hirsch and Tree (2001) who found few differences in the

[60]This was done self-prompted by the participants, as no formal talk about the test was conducted after the test sessions. However, many still commented on one or several aspects of the test when they had finished it, before leaving the testing room.

distribution of response categories between their two age cohorts. Two things however need to be noted in relation to a comparison to Hirsch and Tree (2001). Firstly, Hirsch and Tree only had noun cues. In 5.5, the results based on the data from the current study showed that nouns were more consistent in the way they elicit responses from their own word class than verbs were. However, when comparing the effect of word classes on response category patterns (5.6.2), a significant effect between nouns and verbs was only found for the form-based, dual, and erratic overarching response categories, and only for three of the fourteen responses categories. Nouns, verbs, and adjectives all had *other conceptual* as their most frequent response category. As a rather low amount of cue word class differences was found for the response categories, despite the findings of the strong effect of noun cues on the word class of their responses, a comparison with the data from Hirsch and Tree (2001) might still be interesting, even if they only used noun cues. A second note to make is that even the overarching categories in the current study were more detailed than the binary classification system used in Hirsch and Tree (2001), and thus the differences might have been smaller if such a simpler division had been used for the Norwegian data as well.

The results presented in 5.6.1 go against the findings of Hirsch and Tree (2001), as there were significant differences between the cohorts for a large number of the response categories, including substantial significant differences and effect sizes for the overarching meaning-based and position-based categories, which correspond approximately to the *hierarchical-categorical* and the *propositional-relational* categories in Hirsch and Tree (2001).[61] Additional differences were observed when using the full list of fourteen categories for analysis. As seen in 5.6.1, the results for the subcategories within the meaning-based category (*synonyms, lexical sets* and *other conceptual*) revealed that the differences between the cohorts did not apply for all of the meaning-based categories. The older cohort produced a much higher amount of synonyms than the younger cohort, and the younger cohort produced a much larger amount of 'other conceptual' responses than the older cohort. For lexical sets, the difference between the cohorts was not statistically significant. These results thus show that the use of more narrow response categories can help determine more specifically how response category behaviour may differ between two cohorts. The results from the current study appear to support a hypothesis that predicts changes in selection strategies with age. This is the prediction made in the first, and to some extent the third, hypothesis on age-related changes in word associations presented in Fitzpatrick et al. (2015, p. 36) (these hypotheses are also presented in 2.1.4).

Seen in connection with the other cohort differences observed for the norms lists and the meaning differences for the cue words discussed above, a combination of age-related factors

[61] These two categories again largely overlap with the paradigmatic-syntagmatic distinction (Hirsch and Tree, 2001, p. 6). See also 4.5.4 for more on the difference between a binary syntagmatic-paradigmatic classification system and a more complex set of response categories.

and generational factors might be the best fitting hypothesis for the cohort differences observed in response category patterns as well. As mentioned, a different attitude to the word association test, and a wish to or habit of producing synonyms might explain some of the age-related differences. However, other differences, such as the significant difference in the frequency of *other conceptual* responses, or the lack of significant difference for the frequency of lexical sets, are more difficult to explain, and might be due to changes in strategies with age.

6.1.6 Summary

The results discussed above show that word association tests are useful for discovering a set of different age-related linguistic differences. The word association results from the current study revealed differences in the frequency of use of words, in changes in the meaning of words over time, and in the categories of responses that were produced. Based on the results from the current word association study, it appears that a combination of age-related changes, and generational effects may best explain the observed differences between the two cohorts.

Secondly, it also seems clear from these results, in line with other previous studies, that one has to use age specific norms in word association research, as there might be large cohort differences in responses, and it is hard to anticipate where these differences might occur. It was also hypothesised that other factors such as level of education, area of residence, or socioeconomic background might influence association behaviour. Future research might further explore the potential effects of these and other factors, and whether such factors should also be taken into account when compiling norms lists and studying response data.

In the usage-based model presented in Lamb (2000), a relational network in the lexicon is described as "a model of the end-result of a long learning process" (p.104). Based on the usage-based theory of language presented in 3.1, the differences between the age cohorts discussed above appear to follow naturally from differences in, and amount of experience with language in the two cohorts. The results from the current word association study will be further discussed below in terms of their ability to shed light on theories of the mental lexicon.

6.2 Word associations and the mental lexicon

In 3.1.3, it was predicted that we would see individual differences within the cohorts due to the role of experience in usage-based theory. As was seen in the results from the norms

list calculations (5.1), and for the meaning differences (5.2), although some cue words produced strong dominant responses, there was also a vast number of idiosyncratic, or uncommon responses, as can be seen in the full norms lists (available at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/sprakmateriale.html). This fits with Schmitt's observation that almost any cue tends to have a small number of relatively frequent responses, in addition to a larger number of relatively infrequent responses (Schmitt, 2010). The variation in responses also reflects the role of personal experience in usage-based theory; this will be further discussed in 6.2.4.

Further, it was predicted that the responses from the word association test would support the role of frequency in usage-based theory, and in particular in the structure of the network. Other predictions concerning the structure of the network were that associations would happen on many or all of the representational levels in the network and that meaning-based responses should be most common if associations reflect categorical clusters. Network structure should also be reflected in differences between the responses given to noun cue words compared to those given to verbs, and other factors that might affect storage and retrieval, such as imageability, should also be visible in analysis of the collected data. These predictions will be discussed below following the relevant sets of results (presented in Chapter 5).

6.2.1 Cue word class influence

The basic contrasting properties of nouns and verbs in cognitive grammar were presented in 3.1.1. To survey how word association data might shed light on the differences between the two word classes, several sets of comparisons were carried out (cf. the results presented in 5.5). The first of these was related to the effect of the word class of the cue word on the word class of the responses it elicited. The results presented in 5.5 showed a significant difference between the amount of noun responses given to noun cues compared to the number of verb responses given to verb cues.[62] As mentioned in 3.1.1, in cognitive grammar, one of the features that distinguishes archetypical nouns from archetypical verbs is how nouns are conceptually autonomous, and may be conceptualised on their own. Verbs, on the other hand, are bound in time, and are dependent on other participants to be conceptualised. It was predicted in 3.1.3 that this would lead nouns to elicit more nouns than verbs elicited verbs, which is what the results also show.

The results also showed that the noun cues vary less in their behaviour, as the nouns had a smaller range for the proportion of noun responses they elicited than the verb cues had for verb responses. It is in line with the proposed word class differences in cognitive

[62]The results are based on data from the younger cohort only (see 5.5).

grammar that verbs would differ more with regard to which word class they mainly elicit, as they are dependent on more external factors than nouns are. As verbs can only be conceptualised with the participants who interact to constitute an event, one is more likely to first think of one of those participants, which would (in most cases) belong to another word class.

It was hypothesised that the response word class elicited by verb cues might be related to the transitivity of the verb. However, as seen in 5.5, there does not seem to be any clear correspondence between cue transitivity and response word class. Rather, it appears from the current results that access and retrieval of responses for verb cues is dependent on other factors as well. Firstly, it appears that the property which mainly affects the response word class pattern might differ between verbs. Some verb cues seem to elicit responses mainly based on co-occurence and/or transitivity, such as *nyte* (Eng: *enjoy*), which mainly triggered nouns,[63] i.e. *nyte* elicited a number of position-based responses (*å nyte X* (Eng: *to enjoy X*)). Other verb cues, such as *beholde* (Eng: *keep*) mainly elicited verb responses, i.e. likely elicited a number of meaning-based responses, such as synonyms. One might, based on the current results, hypothesise that frequency of co-occurrence is a strong factor in association patterns of verbs. However, it might be that the design of the word association test facilitates meaning-based responses which for verbs would mean to primarily produce verb responses. Although none of the younger participants expressed any desire to find a synonym to the cue (as many of the older participants did), it is possible that also the younger cohort inadvertently to some extent searched for synonyms when associating to the cue words. The response category patterns for the verb cues (see 5.6.2 for the results, which will be further discussed below in 6.2.3), showed that verbs mainly elicit meaning-based responses.

The observed variation in response behaviour to verb cues might also be explained in terms of the models of language processing and production presented in 3.2. In the relational network model (Lamb, 2000), the repeated co-occurrence between a verb and another item, such as a noun in the mentioned co-occurrence of *to enjoy NOUN*, would lead to the connection between these two items becoming strong. In the model, information is stored in the connectivity of the network, and strength affects activation and spread in the network (Lamb, 2000, p. 101). One might thus hypothesise that in the cases where the verb cues mainly elicited other word classes, this reflects high frequencies of co-occurrence between the units. On the other hand, if verb cues mainly elicit other verbs, this might reflect that other verbs with an overlapping, or opposite, meaning are frequent in Norwegian. In usage-based theory (see 3.1), exemplars may potentially contain information on all aspects of the linguistic experience related to the exemplar, including

[63]See Appendix K for an overview of response word class distributions for all verb cues in the word association test.

phonetic detail, meaning, pragmatic dimensions, and context of use (Bybee, 2010, p. 14). Thus, if two (or more) verbs are experienced in similar contexts, and/or co-occurring with similar other items, they might end up being strongly connected in the network based on their shared properties.

6.2.2 Cue word class: Multi-word and blank responses

In addition to revealing age related differences (6.3.1), it was hypothesised that the number of blank responses given would correlate with the imageability rating of the cue word. Significance tests showed that there were significant differences between the number of blank responses and the imageability rating of the cue word for both age cohorts, and for the total number of blanks for all participants (5.4.4). The imageability rating of the cue words thus appears to affect the participants' ability to produce a response. This is supportive of previous research that has found that imageability affects the storage and processing of words (see references in Simonsen et al., 2013). The current findings suggest that a low imageability rating for a cue word lower the chances of successfully associating to it. These results are not on their own enough to explain exactly how the imageability rating affects the association process. However, one may hypothesise that a lower imageability rating either affects the comprehension of the cue word, and/or that there are fewer available connections to other units in the network for words of low imageability, and that this makes it more difficult to associate.

The current study found a correlation between imageability ratings and the number of blank responses given; however, no such correlation was found for word class and the number of blanks (6.3.1). This supports the notion from other previous studies (e.g. Kausche and Stenneken (2008)) that one should look into more properties of the words than just word class alone when researching differences between nouns and verbs.

6.2.3 Response categories and cue word class

A further prediction for the influence of cue word class was that word class might also affect response category patterns. It was predicted that nouns would elicit more meaning-based responses, and verbs more position-based responses (see 3.1.3). As seen in 5.6.2, this hypothesis only held for parts of the data. When looking at the overarching categories, form-based, dual, and erratic responses showed a significant difference between the response categories triggered by nouns compared to by verbs. When using all fourteen categories to look at differences, significant differences between the responses given to nouns compared to verbs were only found for three of the categories; synonyms, lexical sets and response-cue collocations. For synonyms, the prediction held that verbs would

elicit more synonym responses than nouns would, and nouns also elicited significantly more lexical sets, which are part of the meaning-based category. However, for the only position-based category where a significant difference was found between nouns and verbs, response-cue collocations, nouns had a higher median number of responses than verbs did. This does not fit with the predictions made in 3.1.3.

One other study that has looked at the influence of cue word class on response categories is Nissen and Henriksen (2006) (see also 2.1.3).[64] The results from the current study will be compared to Nissen and Henriksen's findings. The test in Nissen and Henriksen (2006) consisted of equal amounts of verbs, nouns, and adjectives. Nissen and Henriksen (2006) found that, for the participants' L1, nouns triggered more paradigmatic responses than the other two word classes, whereas verbs and adjectives elicited more syntagmatic responses than nouns (Nissen and Henriksen, 2006, p. 397). This is in line with the predictions for the current study (3.1.3), if one uses the overarching *position-based* and *meaning-based* categories as comparable to the syntagmatic-paradigmatic categories used in Nissen and Henriksen (2006). No statistically significant differences were found in the current study for differences between noun cues and verb cues for either meaning-based nor position-based responses; however, statistically significant differences were not found for the opposite pattern either.

When all fourteen response categories were used, the current data appears to support the predictions that were made somewhat more. Nouns produced significantly more lexical sets than verbs did, which supports the idea of nouns as mainly organised in hierarchies, and the prediction that noun cues would elicit more meaning-based responses. The prediction made that verbs might elicit more synonyms than nouns also held. However, within the three categories that make up the overarching position-based category, only one had a higher frequency for verbs than nouns, and thus the prediction that verbs would elicit significantly more position-based responses than nouns was not supported by either set of comparison.

Due to the number of factors that might influence association behaviour, it may be difficult to explain why some of the results from the current study do not fit with neither the predictions made in 3.1.3, nor with the findings of Nissen and Henriksen (2006). The current study seems to match Nissen and Henriksen's (2006) study on several background factors. The participants in Nissen and Henriksen (2006) were 17-19 years old, and the calculations for cue class influence were made for the data from the younger cohort in the present study, so the differences should not be due to age-related factors. Also, the frequencies for the cue words were the same for both the current study, Nissen and Hen-

[64]The participants in Nissen and Henriksen were tested in their L1 Danish and in their L2 English. The participants were aged 17-19 years old, and had all had 7-8 years of classroom instruction in English (Nissen and Henriksen, 2006, p. 392).

riksen (2006) and Fitzpatrick et al. (2015), and this factor should thus also be compatible across the studies.

One reason why the current results neither fit with the predictions made for the current study (see 3.1.3), nor with the findings of Nissen and Henriksen (2006), might be that there is less overlap between the paradigmatic/syntagmatic categories and the meaning-based/position-based categories than what has been assumed in the current study (see 4.5.5 for more on the differences between the two classification sets). It is possible that for example the category *other conceptual*, which is part of the overarching meaning-based category, could contain responses that might have been considered syntagmatic in a paradigmatic-syntagmatic classification system. However, in the current study they were thought to be connected to the cue word based more on meaning than position, even if they might not have been able to fill the criteria of paradigmatic responses, i.e. that they may replace the cue without changing the grammaticality of the sentence they occur in. The *other conceptual* category was by far the most popular one for all three cue word classes, and it is thus possible that the sheer amount of responses in this category might have ended up concealing other differences.

It was also predicted in 3.1.3 that imageability might affect storage of words, and that this might again affect response category patterns. However, as seen in 5.6.2, very limited amounts of differences could be found in the distributions of the response categories based on the imageability ratings of the cue words. This was true for both the within word class comparisons, and for the total sum of response categories for the 58 verbs and nouns for which imageability ratings were available. This is not to say that imageability does not influence storage and retrieval in the mental lexicon. As already discussed in 6.2.2, the imageability of the cue word appears to affect the ability to produce a response, as there was a large, significant difference between the number of blank responses for the cue words with low imageability ratings compared to those with high imageability ratings. It is of course possible that significant differences might have been found if the imageability ratings and response categories had been compared based on the full list of the fourteen response categories rather than the five overarching categories. However, due to the size of the current data set, this would not have produced very robust results. It is nonetheless an issue that should be considered in future word association studies.

The response category patterns might also tell us something about the role of frequency in usage-based theory. Position-based responses are likely to reflect frequencies of co-occurrence between the cue word and other units in the lexicon. The findings of the current study support the role of frequency in usage-based theory for constructions and for the effect of common co-occurence of units (see 3.1.1). For position-based responses based on formulaic or fixed expressions, such as the cue-response pair *gård-grunn* (Eng:

farm-grounds,[65]) it seems likely that the formulaic sequence is stored as a whole in the lexicon. When one word in the sequence is presented as a cue, the whole sequence is activated, and the main lexical item in the sequence which is not identical to the cue, is used as the response. In other cases, for such cue-response pairs as *hull-tenner* (Eng: *hole (N)-teeth (N)*, tooth decay), one may hypothesise that *tenner* was given as a response to *hull* because the participants have experienced these two words occur together several times, and that the connection between the two would thus have been strengthened in the lexicon, potentially affecting selection and retrieval of *tenner* at the expense of other potential candidates.

6.2.4 Word association and units of storage in the mental lexicon

As already pointed out in some of the above sections, the data from the current word association test might tell us something about the units of storage in the mental lexicon, as they are postulated in usage-based theory, i.e. exemplars and constructions. Each exemplar is thought to contain all aspects of linguistic knowledge known to a person, and this fits with the predictions and findings that associations happen on different levels of representation, and along different dimensions of the cue word (see 6.2.5).

For constructions, several aspects of the collected data support the role of larger units in the lexicon.[66] For one, the amount of multi-word responses appear to show that sometimes, a multi-word construction is the first unit of activation, and one would thus assume that unit to be stored as a whole.

Secondly, the fact that participants make such response pairs as *gøy-landet* (Eng: *fun (A)- country (N)*,), from *gøy på landet* (Norwegian equivalent to *fun in the sun*, literally *fun in the country*), might be seen as support for the postulated idea in usage-based grammar that the units that make up a construction might both be accessed on their own, and as a whole (i.e. the construction). In cases such as the cue-response pair *gøy-landet*, it appears that the activation of one part of the construction (*gøy*) leads to the rest of the construction getting activated as well. Although some researchers define a construction as a pattern that is "not strictly predictable from its component parts or from other constructions recognized to exist" (Goldberg, 2003, p. 219) many researchers also consider fully predictable patterns to be constructions, as long as they are frequent

[65] 'gå fra gård og grunn', literally *go from/leave farm and grounds*, is a Norwegian expression meaning to lose everything.

[66] Constructions might be form and function pairings of many different sizes, from morphemes to words to "fully general linguistic patterns" (Goldberg, 2003, p. 219). However, as smaller units are covered in other parts of the discussion, the focus here will be on more complex constructions.

enough (Goldberg, 2003). If ones decides to follow the latter definition,[67] the constructions represented in the data from the current study may inform us about the frequency of co-occurrence for words in larger multi-word units.

6.2.5 Associations on all levels of representation?

In 3.1.3 it was predicted that associations would be made on all levels of representation in the network, that is, one would see associations based on the phonetic, semantic, and pragmatic dimensions in the network proposed by Bybee (2010). The results presented in 5.6 showed that responses were made within all the fourteen response categories. Although the distributions within the different responses differed substantially, responses were made both based on the meaning of the cue (meaning-based responses), on their position in schema-instance relations (such as for many lexical set responses),[68] and on their typical patterns of co-occurence, as in the position-based response categories. Smaller numbers of responses were also made based on what affixes the cue might take, or based on the phonetic properties of the cue.

Another set of responses, those in the two-step association category, appear to either be triggered by the participants mis-reading the cue word, or by an additional "mediator" association having been activated in between the comprehension of the cue and the production of the response. For the first case, examples such as *støy-hybelkanin* (Eng: *noise (N)- dust bunny*) appears to have been triggered by the participant reading *støy* as *støv* (Eng: *dust* (N)). In the other case, examples such as *tolke-Hobbit* or *tolke- Ringenes herre* (Eng: *interpret (V)- Hobbit* and *interpret (V)- Lord of the Rings*) makes for a more complicated discussion. It seems unlikely that several participants (n=3 in the younger cohort) have read *tolke* as *Tolkien*, the author of The Hobbit and Lord of the Rings. Rather, perhaps *Tolkien* first was activated, and then quickly another association was made, which is the one the participant wrote down.

The observation that responses did appear on all levels of representation also supports bi-directional, cascading models of language production and comprehension, such as Dell (1986) or Lamb (2000), presented in 3.2. In Lamb's relational network model (Lamb, 2000), activations would spread based on which connections were the strongest from the nection activated from the cue word, and this could thus lead to different types of responses, based on the participant's previous experience with language. The variation between participants can also be seen in the ranges of the amount of responses within

[67] This latter definition of constructions is very similar to the definition presented in Bybee (2010), see 3.1.

[68] These types of responses also support the schema-instance relations of storage in cognitive grammar, see 3.1.2.

each response category (see for example table 5.18). One might also hypothesise that the individual differences in response category patterns might not only purely affect language experience and frequencies, but also differences in semantic strategies, or storage in the mental lexicon, too.

6.2.6 Summary

The discussion above has pointed at many aspects of theories of the mental lexicon that word association data might shed light on. In general, the results from the current study seem to support the predictions of usage-based theory and cognitive grammar. Firstly, the fact that association responses are often inflected forms of a word, or even multi-word units, supports a non-modular view of language, which allows for storage of units of different sizes, such as usage-based theory. Secondly, the role of experience in usage-based theory is visible through for instance the individual variation in response behaviour. Thirdly, the data clearly supports some of the suggested fundamental differences between nouns and verbs, as postulated in cognitive grammar. Additionally, the fact that associations happen on all levels of representation supports both a network model with different types of clusters (i.e. meaning clusters or phonological clusters), and also cascading models of language processing and production, such as the model assumed in relational network theory.[69] The variation in types of associations also supports previous findings that word associations do not only reflect co-occurrence, but also other mental properties (De Deyne and Storms, 2015, p. 471).

At the beginning of Section 3.1, the dual nature of the mental lexicon was presented. The mental lexicon is often discussed both in terms of being the dictionary represented in the mind, and in terms of the lexical activity that goes on in the lexicon (Jarema and Libben, 2007, p. 1). Data from word association tests may touch upon both of these sides of the mental lexicon. The norms lists might tell us something about the contents of the dictionary, and analysis of the response category patterns and the influence of the cue word class on responses, might shed light on lexical processing.

Although word association data has been shown in the current discussion to be able to shed light on theories of the mental lexicon, and on storage and retrieval from the lexicon, much is still unknown about the processes that happen when a participant is asked to associate to a list of cue words. In terms of the logogen model (see 3.2), there is still a lot to discover about what goes on in the *cognitive system* component. However, future word association studies might reveal more about what happens in the lexicon when we associate between words.

[69] However, see also the discussion of the data in relation of the transmission deficit hypothesis in 6.3.1.

6.3 Word association and theories on language and ageing

In 3.3.2 it was predicted, based on theories on language and ageing, that changes in linguistic abilities with age would cause more blank responses, and more multi-word responses in the older cohort.

6.3.1 Blank responses

As presented in 5.4.3, there was a significant difference in the number of blank responses produced by the two cohorts, with more blank responses in the older cohort. This is in line with the prediction in 3.3.2. Due to the lack of context for the presented cue words, it is likely that in word association tasks naming problems will show up more clearly than in confronting naming tasks, where prompts might be given to the participants. As mentioned in Chapter 3, according to the *transmission deficit hypothesis* (TDH), there is an increase in problems with naming in older age due to a weakening of the connections between the nodes in the lexicon (see 3.3.1 for more on the hypothesis). Naming is particularly vulnerable, due to the small amount of connections between the lexical nodes and the phonological representation(s); however, the greater number of connections within the rest of the semantic system leaves the activation of the semantic nodes intact. See 6.3.2 below on how the cohort differences in multi-word responses may also support the TDH.

Based on the predictions made in 3.3.2, it was assumed that, as comprehension is not affected in ageing in the TDH, the word class of the cue word should not affect the ability to successfully produce a response. This should be seen in the data through a lack of difference in the number of blank and multi-word responses to noun cues compared to verb cues. Nouns have been shown to be more affected by naming difficulties (see 3.3), but this should not affect comprehension. However, after analysing the effect of cue word class on the word class of the responses the cue triggers, it was found that noun cues triggered a very high proportion of noun responses(see 5.5). It is then possible that noun cues would have more blank and multi-word responses than verbs, as they elicit mainly nouns, and nouns are more prone to naming difficulties than the other word classes. The results showed that for the number of blank responses there was no statistical difference between noun cues and verb cues (see 5.5), which is in line with the predictions. For the number of multi-word responses however, the difference between noun cues and verb cues was significant for both age cohorts. Surprisingly though, verb cues caused substantially more multi-word responses than nouns. This was true for both age cohorts (5.5). It thus seems that this difference is not due to noun cues triggering noun responses that are

difficult to produce.

In Section 6.2.6 the ways in which the data from the current study supports cascading models of language processing and production were presented. However, this section has shown that parts of the data also support the TDH, which postulates a more serial processing of language. Although the current data are not comprehensive enough draw firm conclusions, it appears that the processing and/or production process is neither strictly cascading nor stricktly discrete, and word association data should perhaps be further studied to see if for instance a limited cascading model (see 3.2) might be a better fit.

6.3.2 Multi-word responses

The calculations of cohort differences in multi-word responses (5.4.1) showed that there was a significant difference between the cohorts also in terms of these responses. The higher number of multi-word responses for the older cohort fits with the TDH, and the other theories on language and ageing outlined in 3.3. In fact, the effect size for the difference between the cohorts was larger for the multi-word responses than for the blank responses (see 5.4). In 3.3 it was mentioned that when asked to give definitions for words, there are few differences between younger and older participants (Obler and Pekkala, 2008, p. 351). Combined with the ideas in the TDH that semantic information is easier to retrieve than phonological information in older age, the large amount of multi-word responses suggests that if the associated word cannot be retrieved, the participants may in some cases instead access different propositional nodes related to the associated word. Based on this, rather than to produce the associated word itself, a multi-word definition may be produced. However, as both the effect size for the difference between the cohorts, and the frequency of multi-word responses were higher than the corresponding numbers for blank responses, it is likely that multi-word responses are also caused by other factors. As mentioned in 3.3, older persons tend to be more verbose, especially in discourse production. It has been suggested that this might be a mechanism to mask linguistic difficulties, for instance related to retrieval or selection of individual words (Mortensen, Meyer, and Humphreys, 2006). This would again support the previous suggestion that multi-word responses are given when a wanted word cannot be retrieved. However, other factors, such as problems with inhibiting irrelevant information, have also been suggested as explanations of the more wordy utterances observed among older persons. Thus, the large number of multi-word responses observed among the responses from the older cohort could stem from a combination of several potential age-related linguistic changes.

Multi-word responses also occurred in some of the responses in the younger cohort, and it is thus not only age-related changes in linguistic skills that might cause this type of

responses. Some of the multi-word responses, for both cohorts, may be explained based on the notion of *constructions* in usage-based theory (see 6.2.4).

In the calculations on the influence of cue word class for the amount of blank and multi-word responses, it was shown that the number of multi-word responses produced significantly differed depending on the word class of the cue (5.4.2). Verbs elicited more multi-word responses than nouns for both age cohorts. However, as presented above, multi-word responses are potentially made based on one of several different reasons, and these results might not be related to the comprehension of the cues, but rather just the types of responses made to cues from different word classes.

6.3.3 Summary

The data from the word association test revealed age-related differences in both the number of blank responses and in the number of multi-word responses, in line with the predictions made in 3.3.2. These results support theories on naming issues in healthy ageing, such as the transmission deficit hypothesis, and they also support the findings of other studies that these problems are mainly in production, and not in comprehension. The data on multi-word responses also support the notion that older persons tend to be more verbose, and, for both cohorts, the multi-word responses supports the idea of storage of larger chunks in the mental lexicon.

Chapter 7

Conclusions and ideas for further research

The current study aimed to answer three sets of research questions (see 1.2): to see if the word associations were different in the two age cohort, to explore how the data collected in the current study sheds light on theories of the mental lexicon, and lastly, to explore how the data sheds light on theories on language and ageing.

The results of the current study clearly showed the potential of word association tests as a research tool. The data from the word association test gave insights related to all three sets of research questions for the thesis. In terms of age-related differences in word association behaviour, the results of the study showed that there are both generational factors, and factors related to changes in linguistic strategies involved in producing these differences. The age-related differences found here support the findings in previous studies, such as Fitzpatrick et al. (2015), that norms lists should be tailored for each study using word associations.

The analysis and discussion of the data from the current study also showed that word associations might provide valuable input towards theories on the mental lexicon and on language and ageing. The results related to the organisation of the mental lexicon, such as the observed differences between nouns and verbs, show the possibilities to also use word association data for studies of topics that go outside of the the response words themselves.

However, as has been voiced by for instance Schmitt (2010, p. 248), there is a clear potential in word association data, but there is also a need for further methodological improvements in order to explore its full potential. One methodological change that was done for the current study was to make sure all the cue words were unambiguous in terms of word class (see 4.1.3). This change meant that the study was able to get several sets

of results that would not have been possible with ambiguous cue words. For instance, the statistical analysis of the data showed clear differences between the responses given to noun cues compared to verb cues, which again reflect differences in the storage and retrieval of different word classes in the mental lexicon. To have unambiguous cue words also eases response classification. With an ambiguous cue, it is difficult to know for sure which word class the participants associated from, and thus some responses might have several potential classification categories.

The current study also replicated the findings of the recent methodological developments in Fitzpatrick et al. (2015) in terms of both age-related differences in the norms lists, and in terms of the additional findings when using a set of fourteen response categories compared to a smaller set of broader categories. The findings of these methodological changes, and the omission of ambiguous cue words, suggest that the current strive in the field for tailored word association tests for each study, with smaller, specific categories, and as many factors as possible controlled for is a clear progression to the field of word association research.

7.1 Afterthoughts and ideas for future research

7.1.1 Methodological issues in the current study

Some methodological issues specific to the current study need to be addressed. It is problematic that data on how long each participant took to complete the word association test was only collected for 22 of the 122 younger participants (for the older cohort, the time was measured for all except one participant). The lack of time measures for the younger cohort happened during the classroom sessions, where many participants were tested at the same time. Ideally, a strategy for measuring, at least approximately, how long time the participants took to complete the test should have been developed for these sessions as well. For instance, a method such as the one used in Håseth (1968), where a letter was written on a blackboard in the test room every minute could have been used. Once a participant finished the test, (s)he wrote the current letter from the blackboard on his or her test. This way at least one would have had a fairly close time measure for the participants in the younger cohort as well.

It would also probably have been beneficial to operate with some kind of upper time limit for the tests. In Fitzpatrick et al. (2015) the participants were only given ten minutes to complete the test. Based on the spread in age for the participants in the current study, some of them being in their 80s or 90s, ten minutes would not have been sufficient. It is however possible that to not have any upper limit might have led to some of the

participants giving themselves a little more time to think, even if they were told both orally and in writing that they should do the test as quickly as possible.

Another issue that emerged retrospectively, was the selection of participants for the pilot study. It is possible that the participants for the pilot study should have been recruited from the two cohorts that would be used in the study itself. The participants in the pilot study included four persons over 60 years, and three persons between 20 and 30, in addition to seven people between 30 and 60 years. As the results of the study proper showed large differences between the age cohorts, it might have been more accurate to also only use participants from the two cohorts for the pilot study. It is of course impossible to tell how a different age distribution for the participants in the pilot study might have affected its outcome on the final cue word list. However, in retrospect, it might have been a stronger, methodological decision to include equal amounts of participants from each of the two age groups that were to be represented in the study proper in the pilot. This way one would be more certain that cue words that were removed after the pilot study represented the association behaviour of both age groups.

In terms of the questionnaire given to the participants, one revision that should be done to the form is the inclusion of a question asking specifically about the mother tongue of the participants. Although it was stated in the information letter given to all participant that the study was only recruiting native speakers of Norwegian, there were still cases of participants that had been bilingual from birth, or who had only learned Norwegian later in life. As the questionnaire asked about other languages known by the participant, and when these were acquired, I am fairly certain that most of these cases were discovered. However, it is possible that some participants who had either been bilingual from birth, or had acquired Norwegian later in life, ended up as part of the participants used in the study.

It is also a weakness of the current study that there were relatively few participants in the older cohort. The low number of older participants (n=51) meant that the results concerning the older cohort were not as strong as those for the younger cohort. There were also some aspects that it would have been interesting to conduct analysis on, such as for instance potential gender differences in association behaviour, or age differences within the older cohort (i.e. for instance to compare those below 75 years to those above 75 years); however, to divide the older cohort any further would have resulted in samples too small to make any conclusions.

7.1.2 Suggestions for future research

This section will present some thoughts on the methodology in general, and on changes in the methodology that I believe future word association research might benefit from.

Firstly, there is the discussion on whether the test should be paper based or be digitalised. For the purposes of the current study, a paper-based version was chosen for several reasons. For one, it was an aim to follow the methodology of Fitzpatrick et al. (2015) as closely as possible, and in their study the test was paper-based. Further, it would probably would have been difficult to recruit many of the older participants, in particular those in the top end of the age span, if the test had been digitalised.[70] Additionally, it would have been unlikely to be able to develop a digital version of the test within the time limits of a master's thesis. However, there are also clear advantages of a digitalised version. It would likely have been easier to recruit participants, in particular for the younger cohort, from more different backgrounds[71] if a digitalised test was used. In terms of measuring the time used by each participant, this would also have been much easier in a digitalised test. Not only could one achieve precise timings for each participants, but one could also have timed each response on its own.[72] To have time measurements for each cue-response pair would allow for a new set of analyses. For example, previous research on word association and imageability (e.g. de Groot (1989)), found correlations between imageability ratings for the cue and the time used to respond. One may also envision future studies for instance looking at the response time effects of such properties of the cue word as frequency of use or the number of phonological neighbours of the cue word.

Concerning the test sessions, some of the older participants appeared to find the list of 100 cue words somewhat long and tiring to complete. At the same time, a larger number of cue words, with adjacent data, provides the basis for better results concerning properties of the cues. It might thus be an idea for further, larger studies to have more participants, who each only respond to a subset of the total list of cue words. This would also be likely to reduce the amount of tests being discarded due to a high amount of blank responses.

Although it would not have been feasible within the time frame of this thesis, I believe it would have been beneficial to conduct post-test interviews with the participants. For one, this would aid the classification process. It might help solve issues of ambiguity in responses, and it might correct some responses that are likely to be mis-classified. Although a big part of the responses fell naturally into one of the categories, there were also

[70] See Appendix F for age distribution within the cohorts.

[71] In other words, to not just recruit students enrolled in higher education for the younger cohort.

[72] This section will not go into detail as to the specifics of how a digitalised version would be designed. The focus here will be on its implications for the results and analysis of word association data.

those that were harder to place, some due to ambiguity, and others due to problems with fully understanding the link between the cue and the response. By interviewing the participants after they have finished the test one would probably end up with less ambiguous responses, and less erratic responses as there might be links behind an association that is not obvious to the researcher. Additionally, post-test interviews might also be able to provide some reflections from the participants on the association process, and perhaps also help decide on new directions in research on what word associations might tell about the mental lexicon.

For the analysis of the word association data, some aspects of the response categories should also be mentioned here. In the process of categorising the responses, and in the subsequent statistical analysis, the *other conceptual* response category in particular emerged as somewhat troublesome. Firstly, as was seen in 5.6, this category has a tendency to become very voluminous. This is unfortunate, as a main reason for using a set of fourteen categories rather than the traditional syntagmatic-paradigmatic(-clang) division is that this will reveal differences and nuances of association behaviour that a set with fewer, broader categories might conceal. As mentioned in 6.2.3, the *other conceptual* category might end up being filled with very different response types. Examples of responses classified as *other conceptual* in the current study include responses that are semantically linked to the cue without being a synonym or part of a lexical set, such as in the cue-response pair *analyse-forsker* (Eng: *analysis (N)-researcher(N)*). Other responses classified as *other conceptual* were more personal, and although it is understood that the response is linked to the cue for the specific participant, there is no general semantic link between the two. Such cue-response pairs were *pappa-skjegg* (Eng: *dad (N)- beard (N)*), or *litteratur-gøy* (Eng: *literature (N)- fun (A)*). I believe it would be fruitful, and provide a more precise classification system, with clearer results, if the *other conceptual* category was only used for the first of these categories. In order to be able to classify the other mentioned types of responses, one suggestion might be to establish a fifteenth category for 'personal responses', that would cover responses that relate to assumed personal experiences, or subjective attitudes to the cue item. As mentioned in 4.5.6, changes in the response category set might also be beneficial for validation of response classification. Two raters might evaluate a cue-response pair the same way, but end up classifying it differently as there are several possible categories. As a solution, attempts should perhaps be made with both the suggested division of the *other conceptual* category, and secondly with allowing for more dual response categories, along the lines of the existing lexical set *and* cue-response collocation type of categories. However, there is of course also such a thing as too many categories, and the analysis will become more complicated, with perhaps weaker results, if one ends up with too many small categories instead. This is a trade-off that future research should explore.

Bibliography

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Barcroft, Joe, Gretchen Sunderman, and Norbert Schmitt (2011). "Lexis". In: *The Routledge Handbook of Applied Linguistics*. Ed. by James Simpson. London and New York: Routledge, pp. 571–584.

Bauer, Laurie and Paul Nation (1993). "Word Families". In: *International Journal of Lexicography* 6.4, pp. 253–279.

Bertram, R., R.H. Baayen, and R. Schreuder (2000). "Effects of family size for complex words". In: *Journal of Memory and Language* 42, pp. 390–405.

Blom, Svein (2015). *Holdninger til innvandrere og innvandring 2015*. Report 52. Statistisk Sentralbyrå (*Statistics Norway*).

Bonin, Patrick et al. (2012). "Evidence for a Limited-Cascading Account of Written Word Naming". In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 38:6, pp. 1741–1758.

Burke, D. M., D. G. MacKay, and L. E. James (2000). "Theoretical approaches to language and aging". In: *Models of Cognitive Aging*. Ed. by T. J. Perfect and E. A. Maylor. Debates in Psychology. Oxford: Oxford University Press, pp. 204–237.

Burke, Deborah M. and Meredith A. Shafto (2004). "Aging and language production". In: *Current Directions in Psychological Science* 13(1), pp. 21–24.

Bybee, Joan (2008). "Usage-Based Grammar and Second Language Acquisition". In: *Handbook of Cognitive Linguistics and Second Language Acquisition*. Ed. by P. Robinson and N.C. Ellis. New York/London: Routledge, pp. 216–236.

– (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan L. and Clay Beckner (2010). "Usage-based theory". In: *The Oxford Handbook of Linguistic Analysis*. Ed. by Bernd Heine and Heiko Narrog. Oxford Handbooks in Linguistics. Oxford: Oxford University Press, pp. 827–856.

Carroll, David W. (2008). *The Psychology of Language*. 5th ed. Belmont, California: Thomson Wadsworth.

Clark-Cotton, M. R. et al. (2007). "Language and Communication in Aging". In: *Encyclopedia of gerontology: Age, aging and the aged*. Ed. by J. E. Birren. 2nd edition. London: Elsevier, pp. 1–8.

De Deyne, Simon and Gert Storms (2008). "Word associations: Network and semantic properties". In: *Behaviour Research Methods* 40 (1), pp. 213–231.

De Deyne, Simon and Gert Storms (2015). "Word associations". In: *The Oxford Handbook of the Word*. Ed. by John R. Taylor. Oxford Handbooks in Linguistics. Oxford: Oxford University Press, pp. 465–480.

de Groot, Annette M. B. (1989). "Representational Aspects of Word Imageability and Word Frequency as Assessed Through Word Association". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:5, pp. 824–845.

Deese, James (1962). "Form Class and the Determinants of Association". In: *Journal of Verbal Learning and Verbal Behavior* 1, pp. 79–84.

Dell, Gary S. (1986). "A Spreading-Activation Theory of Retrieval in Sentence Production". In: *Psychological Review* 93:3, pp. 283–321.

Eastman, Frederic and A.J. Rosanoff (1912). "Association in Feeble-Minded and Delinquent Children". In: *American Journal of Insanity* LXIX, pp. 125–141.

Elevar i grunnskolen, 1. oktober 2015. http://www.ssb.no/utgrs/. [Online, accessed 19.01.2016]. Statistisk Sentralbyrå (Statistics Norway).

Entwisle, D.R. (1966). *The Word Associations of Young Children*. Baltimore, Md.: John Hopkins University Press.

Entwisle, D.R., D.F. Forsyth, and R. Muuss (1964). "The Syntactic-Paradigmatic Shift in Children's Word Associations". In: *Journal of Verbal Learning and Verbal Behavior* 3, pp. 19–29.

Ervin, Susan M. (1961). "Changes with Age in the Verbal Determinants of Word-Association". In: *The American Journal of Psychology* 74.3, pp. 361–372.

Eustache, F. et al. (1990). "Word-association responses and severity of dementia in Alzheimer Disease". In: *Psychological Reports* 66, pp. 1315–1322.

Fitzpatrick, T. et al. (2015). "Establishing the Reliability of Word Association Data for Investigating Individual and Group Differences". In: *Applied Linguistics* 36.1, pp. 23–50.

Fitzpatrick, Tess (2006). "Habits and rabbits. Word association and the L2 lexicon". In: *EUROSLA Yearbook*. Ed. by Susan H. Foster-Cohen, Marta Medved Krajnovic, and Jelena Mihaljević Djigunović. Vol. 6. John Benjamins Publishing Company, pp. 121–145.

– (2007). "Word association patterns: unpacking the assumptions". In: *International Journal of Applied Linguistics* 17.3, pp. 319–331.

– (2009). "Word Association Profiles in a First and Second Language: Puzzles and Problems". In: *Lexical Processing in Second Language Learners: Papers and Perspectives in Honour of Paul Meara*. Ed. by Tess Fitzpatrick and Andy Barfield. Bristol/ Buffalo/ Toronto: Multilingual Matters, pp. 38–52.

Fitzpatrick, Tess and Cristina Izura (2011). "Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation". In: *Studies in Second Language Acquisition* 33, pp. 373–398.

Folstein, Marshal F., Susan E. Folstein, and Paul R. McHugh (1975). "'Mini-mental state': A practical method for grading the cognitive state of patients for the clinician". In: *Journal of Psychiatric Research* 12:3, pp. 189–198.

Fromkin, Victoria A. and Nan Bernstein Ratner (1998). "Speech production". In: *Psycholinguistics*. Ed. by Jean Berko Gleason and Nan Bernstein Ratner. 2nd edition. Belmont, California: Wadsworth/Thomson Learning, pp. 309–340.

Gewirth, L.R., A.G. Shindler, and D.B. Hier (1984). "Altered Patterns of Word Associations in Dementia and Aphasia". In: *Brain and Language* 21.2, pp. 307–317.

Goldberg, Adele E. (2003). "Constructions: a new theoretical approach to language". In: *TRENDS in Cognitive Science* 7:5, pp. 219–224.

Gollan, T.H., D.P Salmon, and J.L. Paxton (2006). "Word Association in Early Alzheimer's Disease". In: *Brain and Language* 99, pp. 289–303.

Goral, Mira et al. (2007). "Change in lexical retrieval skills in adulthood". In: *The Mental Lexicon* 2:2, pp. 215–240.

Guevara, Emiliano Raul (2010). "NoWaC: a large web-based corpus for Norwegian". In: *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*. Association for Computational Linguistics, pp. 1–7.

Gundersen, Dag (2000). *Norske synonymer blå ordbok*. 3rd edition. Accessed online through http://www.ordnett.no. Oslo: Kunnskapsforlaget.

Haastrup, Kirsten and Birgit Henriksen (2000). "Vocabulary acquisistion: acquiring depth of knowledge through network building". In: *International Journal of Applied Linguistics* 10:2, pp. 221–240.

Higginbotham, George (2010). "Individual learner profiles from word association tests: The effect of word frequency". In: *System* 38, pp. 379–390.

Hirsch, Katherine W. and Jeremy J. Tree (2001). "Word association norms for two cohorts of British adults". In: *Journal of Neurolinguistics* 14, pp. 1–44.

Håseth, Kjell Johan (1968). *Norske ord-assosiasjonsnormer*. Oslo: Universitetsforlaget.

Jarema, Gonia and Gary Libben (2007). "Introduction: Matters of Definition and Core Perspectives". In: ed. by Gonia Jarema and Gary Libben. Oxford/Amsterdam: Elsevier, pp. 1–6.

Jung, Carl G. (1910). "The Association Method". In: *The American Journal of Psychology* 21.2, pp. 219–269.

Kausche, Christina and Prisca Stenneken (2008). "Differences in Noun and Verb Processing in Lexical Decision Cannot be Attributed to Word Form and Morphological Complexity Alone". In: *Journal of Psycholinguistic Research* 37, pp. 443–452.

Kent, G.H. and A.J. Rosanoff (1910). "A Study of Association in Insanity". In: *The American Journal of Insanity* 67, pp. 37–96.

Lamb, Sydney M. (2000). "Bidirectional Processing in Language". In: *Usage-based Models of Language*. Ed. by Michael Barlow and Suzanne Kemmer. Stanford, California: CSLI Publications, pp. 87–121.

Langacker, Ronald W. (2000). "A Dynamic Usage-Based Model". In: *Usage Based Models of Language*. Ed. by Michael Barlow and Suzanne Kemmer. Center for the Study of Language and Information, Stanford, California, pp. 1–63.

– (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Levelt, Willem J. M. (1989). *Speaking. From Intention to Articulation*. ACL-MIT Press Series in Natural-Language Processing. Cambridge, Massachusetts: The MIT Press.

Lind, Marianne et al. (2013). "'Ordforrådet' – en leksikalsk database over et utvalg norske ord". In: *Norsk tidsskrift for logopedi* 59:1, pp. 18–26.

Ljosland, Ragnhild (2007). "English in Norwegian academia: a step towards diglossia?" In: *World Englishes* 24:4, pp. 395–410.

Malt, Ulrik (2009). *Sinnslidelser. In Store medisinske leksikon.* https://sml.snl.no/sinnslidelser. [Online, accessed 28.05.2016].

Meumann, E. (1905). "Intelligenzprüfungen an Kindern der Volkschule". In: *Die experimentelle Pädagogik* 1, pp. 86–101.

Mortensen, Linda, Antje S. Meyer, and Glyn W. Humphreys (2006). "Age-related effects on speech production: A review". In: *Language and Cognitive Processes* 21:1-3, pp. 238–290.

Mæhlum, Brit (2007). *Konfrontasjoner: Når språk møtes.* Oslo: Novus forlag.

Namei, Shidrokh (2004). "Bilingual lexical development: a Persian-Swedish word association study". In: *International Journal of Applied Linguistics* 14:3, pp. 363–388.

Nicholas, Marjorie et al. (1997). "Naming in Normal Aging and Dementia of the Alzheimer's Type". In: *Anomia: The neuroanatomical and cognitive correlates of name-finding failure following focal brain lesion.* Ed. by H. Goodglass and A. Wingfield. Orlando, FL: Academic Press, pp. 166–188.

Nissen, Henriette Bagger and Birgit Henriksen (2006). "Word class influence on word association test results". In: *International Journal of Applied Linguistics* 16.3, pp. 389–408.

Obler, Loraine K. and Seija Pekkala (2008). "Language and Communication in Aging". In: *Handbook of the Neuroscience of Language.* Ed. by B. Stemmer and H. Whitaker. Oxford: Elsevier, pp. 351–359.

Paivio, Allan, John C. Yuille, and Stephen A. Madigan (1968). "Concreteness, Imagery, and Meaningfullness Values for 925 Nouns". In: *Journal of Experimental Psychology* 76:1, pp. 1–25.

Palermo, D.S. and J.J. Jenkins (1964). *Word Association Norms: Grade school through college.* MN: University of Minnesota Press.

Patterson, Karalyn and Christina Shewell (1987). "Speak and Spell: Dissociations and Word-Class Effects". In: *The Cognitive Neuropsychology of Language.* Ed. by Max Coltheart, Giuseppe Sartori, and Remo Job. London/ Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 273–294.

Politzer, Robert L. (1978). "Paradigmatic and Syntagmatic Associations of First Year French Students". In: *Papers on Linguistics and Child Languages. Ruth Hirsch Memorial Volume.* Ed. by V. Honsa and M.J. Hardman de Bautista. The Hague: Mouton Publishers, p. XX.

Prado, Elizabeth L. and Michael T. Ullman (2009). "Can Imageability Help Us Draw the Line Between Storage and Composition?" In: *Journal of experimental psychology : Learning, memory, and cognition* 35:4, 849–866.

Roelofs, Ardi (2008). "Tracing Attention and the Activation Flow in Spoken Word Planning Using Eye Movements". In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 34:2, pp. 353–368.

rope, n.1. http://www.oed.com/view/Entry/167402?rskey=OkVtUR&result=1&isAdvanced=false. [Online, accessed 24.04.2016]. OED Online, Oxford University Press.

Rosanoff, Isabel R. and A.J. Rosanoff (1913). "A Study of Association in Children". In: *Psychological Review* XX, pp. 43–89.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA. URL: http://www.rstudio.com/.

Russell, W.A. and J.J. Jenkins (1954). *Language in Behavior: The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test*. Report No. 11, N80, nr 66216. University of Minnesota.

Schiller, Niels O. and Rinus G. Verdonschot (2015). "Accessing words from the mental lexicon". In: *The Oxford Handbook of the Word*. Ed. by John R. Taylor. Oxford Handbooks in Linguistics. Oxford: Oxford University Press, pp. 481–492.

Schmitt, Norbert (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Research and Practice in Applied Linguistics. Basingstoke: Palgrave Macmillan.

Schmitt, Norbert and Cheryl Boyd Zimmerman (2002). "Derivative Word Forms: What Do Learners Know?" In: *TESOL Quarterly* 36.2, pp. 145–171.

Simonsen, Hanne Gram et al. (2013). "Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective". In: *Clinical Linguistics Phonetics* 27:6-7, pp. 435–446.

Steyvers, Mark and Joshua B. Tenenbaum (2005). "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth". In: *Cognitive Science* 29:1, pp. 41–78.

Taylor, John R. (2002). *Cognitive Grammar*. Oxford Textbooks in Linguistics. Oxford: Oxford University Press.

Ullman, Michael T. (2005). "A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model". In: *Mind and context in adult second language acquisition: Methods, theory and practice*. Ed. by C. Sanz. Washington DC: Georgetown University Press, pp. 141–178.

Walker, Ian (2007). *Null hypothesis testing and effect sizes*. http://staff.bath.ac.uk/pssiw/stats2/page2/page14/page14.html. [Online, accessed 11.05.2016].

Whitworth, Anne, Janet Webster, and David Howard (2014). *A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia. A clinician's guide*. 2nd edition. London and New York: Psychology Press.

Wilks, Clarissa and Paul Meara (2002). "Untangling word webs: graph theory and the notion of density in second language word association networks". In: *Second Language Research* 18:4, pp. 303–324.

Wingfield, Arthur and Elizabeth A. L. Stine-Morrow (2000). "Language and speak". In: *The Handbook of Aging and Cognition*. Ed. by Fergus I. M. Craik and Timothy A. Salthouse. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 359–417.

Wolter, Brent (2002). "Assessing proficiency through word associations: is there still hope?" In: *System* 30, pp. 315–329.

Woodrow, Herbert and Frances Lowell (1916). "Children's association frequency tables". In: *Psychological Monographs* 22.5, pp. i–110.

Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge, UK: Cambridge University Press.

Wray, Alison and Kazuhiko Namba (2003). "Use of Formulaic Language by a Japanese-English Bilingual Child: A Practical Approach to Data Analysis". In: *Japan Journal of Multilingualism and Multiculturalism* 9:1, pp. 24–52.

Appendix

Appendix A: Information letter to participants

The next two pages show the information letter that was given to all potential participants for the study.

Forespørsel om deltakelse i forskningsprosjektet

En ny assosiasjonstest for norsk

Hva tenker du på når du hører ordet *påske*? Og hva assosierer du med ordet *dugnad*?

Ordassosiasjoner er et nyttig verktøy innenfor mange områder av språkforskning. Data fra ordassosiasjonstester blir for eksempel brukt for å studere hva som skjer når barn og voksne lærer språk, eller hvordan språket kan rammes ved sykdom og skade. Data fra ordassosiasjonstester blir også brukt til å studere hvordan vi lagrer og henter frem språk i hjernen.

Masteroppgaven min går ut på å utvikle en ny ordassosiasjonstest for norsk, med tilhørende lister over de vanligste assosiasjonene for ulike aldersgrupper. Testen og normlistene vil være et nyttig verktøy i videre forskning.

For å kunne kartlegge assosiasjonene til morsmålsbrukere av norsk trenger jeg hjelp fra mange språkbrukere, og vil derfor invitere deg til å delta i studien. Studien vil samle inn data fra personer i to aldergrupper: 20-30 år, og over 60 år. Om du er innenfor en av disse aldergruppene, og har norsk som morsmål, setter jeg stor pris på om du vil bidra til studien.

Hva innebærer deltakelse i studien?

Undersøkelsen går ut på at du først svarer på noen spørsmål knyttet til alder, utdanning og språklig bakgrunn. Siden vil du gjennomføre selve ordassosiasjontesten. Du får en liste med 100 ord, og skal skrive ned det første ordet du kommer på for hvert av de 100 ordene. Det er ikke meningen at du skal tenke deg om lenge, du skal bare skrive ned det første som faller deg inn. Det tar ca. 15 minutter å gjennomføre undersøkelsen. Jeg er til stede når du svarer på undersøkelsen, og der det er mulig, vil den bli gjennomført gruppevis.

Alle data fra undersøkelsen vil bli behandlet konfidensielt. Det vil ikke være mulig å spore opplysninger tilbake til enkeltpersoner i den ferdige oppgaven. Prosjektslutt er 1.9.2016. Opplysningene vil da lagres frem til 2026 med personidentifikasjon. Dersom opplysningene skal

benyttes til nye prosjekter, vil du motta oppdatert informasjon om dette. Bare personer tilknyttet prosjektet har adgang til opplysningene, og de lagres konfidensielt ved Universitetet i Oslo.

Det er frivillig å delta i studien, og du kan når som helst trekke deg fra undersøkelsen uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli slettet.

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Om du ønsker å delta, eller har spørsmål til studien kan du kontakte meg på telefon 95 87 21 16 eller epost maleneb@student.iln.uio.no.

Du kan også kontakte en av veilederne mine:
Marianne Lind (marianne.lind@statped.no)
eller
Hanne Gram Simonsen (h.g.simonsen@iln.uio.no)

Vennlig hilsen,
Malene Bøyum
Masterstudent i lingvistikk
Institutt for nordiske og lingvistiske studier
Universitetet i Oslo

Appendix B: Consent form

The following consent form was given to, and signed by, all participants prior to administering the word association test.

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og er villig til å delta

Navn (med blokkbokstaver):_____

Signatur:_____ Dato: _____

Jeg samtykker til jeg kan bli kontaktet ved et senere tidspunkt for å bli spurt om å delta i eventuelle oppfølgingsstudier (sett kryss):

Ja ☐ Nei ☐

Appendix C: Questionnaire for participants

Prior to taking the word association test, all participants filled out this questionnaire. Its design and contents are further presented in section 4.2.

SPØRRESKJEMA FOR DELTAKERE

Navn:_____

Telefonnummer:_____

E-post:_____

Hvilket år er du født?_____

Kjønn: _____

Hva er din høyeste fullførte utdannelse? (sett kryss)

Folkeskole (7 år)	◯
Ungdomsskole (9 eller 10 år)	◯
Yrkesfaglig skole	◯
Videregående skole	◯
Universitet/høgskole inntil 4 år	◯
Universitet/høgskole 4 år eller mer	◯

Hva er ditt/dine nåværende eller tidligere yrke(r)? _____

Har du eller har du hatt noen kjente nevrologiske skader eller sykdommer? (sett kryss)

Ja	◯
Nei	◯
Ønsker ikke å svare	◯

Språklig bakgrunn:

Hvilken dialekt vil du si at du snakker?_____

Skriver du hovedsakelig bokmål eller nynorsk? (sett ring) *bokmål* *nynorsk*

Har du lært andre språk enn norsk? (sett ring) *ja* *nei*

Hvis ja, vennligst svar på spørsmålene på neste side om det/de andre språkene (fyll ut for så mange språk som du har lært og la resten stå tomt).

Fremmedspråk 1:

Hvilket annet språk enn norsk har du lært?_____

Hvor gammel var du da du lærte språket?_____

Er du i stand til å gjennomføre de følgende tingene på språket? (sett kryss)

- Stille enkle spørsmål og forstå svarene (for eksempel bestille mat på ferie) ◯
- Holde en hverdagslig samtale ◯
- Lese aviser ◯

Har du noen gang studert språket i høgere utdanning eller brukt det i jobbsammenheng?

- Ja, har studert det ◯
- Ja, har studert på det ◯
- Ja, i jobbsammenheng ◯
- Nei ◯

Fremmedspråk 2:

Hvilket annet språk enn norsk har du lært?_____

Hvor gammel var du da du lærte språket?_____

Er du i stand til å gjennomføre de følgende tingene på språket? (sett kryss)

- Stille enkle spørsmål og forstå svarene (for eksempel bestille mat på ferie) ◯
- Holde en hverdagslig samtale ◯
- Lese aviser ◯

Har du noen gang studert språket i høgere utdanning eller brukt det i jobbsammenheng?

- Ja, har studert det ◯
- Ja, har studert på det ◯
- Ja, i jobbsammenheng ◯
- Nei ◯

Fremmedspråk 3:

Hvilket annet språk enn norsk har du lært?_____

Hvor gammel var du da du lærte språket?_____

Er du i stand til å gjennomføre de følgende tingene på språket? (sett kryss)

- Stille enkle spørsmål og forstå svarene (for eksempel bestille mat på ferie) ◯
- Holde en hverdagslig samtale ◯
- Lese aviser ◯

Har du noen gang studert språket i høgere utdanning eller brukt det i jobbsammenheng?

- Ja, har studert det ◯
- Ja, har studert på det ◯
- Ja, i jobbsammenheng ◯
- Nei ◯

Fremmedspråk 4:

Hvilket annet språk enn norsk har du lært?_____

Hvor gammel var du da du lærte språket?_____

Er du i stand til å gjennomføre de følgende tingene på språket? (sett kryss)

- Stille enkle spørsmål og forstå svarene (for eksempel bestille mat på ferie) ◯
- Holde en hverdagslig samtale ◯
- Lese aviser ◯

Har du noen gang studert språket i høgere utdanning eller brukt det i jobbsammenheng?

- Ja, har studert det ◯
- Ja, har studert på det ◯
- Ja, i jobbsammenheng ◯
- Nei ◯

Fremmedspråk 5:

Hvilket annet språk enn norsk har du lært?_____

Hvor gammel var du da du lærte språket?_____

Er du i stand til å gjennomføre de følgende tingene på språket? (sett kryss)

- Stille enkle spørsmål og forstå svarene (for eksempel bestille mat på ferie) ◯
- Holde en hverdagslig samtale ◯
- Lese aviser ◯

Har du noen gang studert språket i høgere utdanning eller brukt det i jobbsammenheng?

- Ja, har studert det ◯
- Ja, har studert på det ◯
- Ja, i jobbsammenheng ◯
- Nei ◯

Appendix D: Word association test

The word association test, as it was given to the participants in the study. The front page provided test instructions. English translations of the instructions, and for all the cue words can be found in Appendix E.

Alder: Dato: Total tid brukt:

Vennligst skriv ned det første ordet som faller deg inn når du leser hvert av ordene på de neste to sidene.

Skriv bare ett ord per svar.

Husk at det er ingen rette eller gale svar.

Kode:

andel_____

anledning_____

ansette_____

avis_____

beholde_____

direktør_____

eksistere_____

formidle_____

forventning_____

frykte_____

gøy_____

gård_____

hverdag_____

hyggelig_____

høgskole_____

innføre_____

intervju_____

kald_____

kjærlighet_____

konsert_____

kose_____

litteratur_____

materiale_____

middel_____

nyte_____

oppfordre_____

opprinnelig_____

overføre_____

planlegge_____

posisjon_____

prege_____

psykisk_____

religion_____

savne_____

stolt_____

størrelse_____

sunn_____

sykehus_____

søt_____

telle_____

trolig_____

understreke_____

unik_____

usikker_____

utvalg_____

vedtak_____

vindu_____

virksomhet_____

øl_____

årsak_____

analyse_____

annonse_____

avansere_____

bevare_____

drømme _____

fellesskap_____

forbund_____

fremtidig_____

fødsel_____

gidde_____

gulv_____

hemmelig_____

hull_____

håndtere_____

industri_____

informere_____

initiativ_____

innvandrer_____

komité_____

konsentrere _____

konsept_____

kulturell_____

kvinnelig_____

landbruk_____

leverandør_____

likhet_____

mangfold_____

midlertidig_____

motstander_____

nabo_____

offer_____

pappa_____

potensiell_____

regelverk_____

rektor_____

revisor_____

rope_____

servere_____

struktur_____

støy_____

sur_____

svekke_____

tempo _____

tolke_____

transport_____

uheldig _____

vert_____

variant _____

vind_____

økning_____

Appendix E: English translation of the word association test

Below are English translations of the instructions given to the participants, and of the cue words in the Norwegian word association test. In some cases, the Norwegian cue word could have a rather large number of potential English translations. In these cases, a selection has been made while also trying to show all potential meanings of the word.

Age: Date: Total time used:

Please write down the first word that comes to mind when you read each of the words on the next two pages.

Write only <u>one</u> word per response

Remember that there are no right or wrong responses.

Code:

Age: Date: Total time used:

andel -share/part/portion (N)

anledning – occasion/opportunity(N)

ansette- employ (V)

avis- newspaper (N)

beholde- keep (V)

direktør – director / manager (N)

eksistere – exist (V)

formidle – convey/ impart/arrange (V)

forventning -expectation/anticipation(N)

frykte – fear (V)

gøy – fun (A)

gård – farm/ estate (N)

hverdag – everyday (N)

hyggelig – nice / pleasant (A)

høgskole- academy/ university college (N)

innføre- introduce/ import (V)

intervju- interview (N)

kald – cold (A)

kjærlighet- love (N)

konsert – concert (N)

kose - make pleasant for /cuddle (V)

litteratur- litterature (N)

materiale – material (N)

middel – means/ resource (N)

nyte – enjoy (V)

oppfordre – encourage (V)

opprinnelig –original/ initial/natural(A)

overføre – transfer/transmit (V)

planlegge – plan (V)

posisjon – position (N)

prege – characterize/mark/stamp (V)

psykisk – mental/ psychic (A)

religion- religion (N)

savne – miss (V)

stolt- proud (A)

størrelse – size (N)

sunn – healthy (A)

sykehus – hospital (N)

søt –sweet/cute (A)

telle – count (V)

trolig – likely/probable (A)

understreke –emphasize/underline (V)

unik – unique (A)

usikker – unsure/insecure/risky (A)

utvalg- selection/sample/panel (N)

vedtak- decision/ resolution (N)

vindu- window (N)

virksomhet-firm/operation/activity(N)

øl- beer (N)

årsak- cause/ reason (N)

Code:

analyse – analysis (N)

annonse – advertisement (N)

avansere- advance (V)

bevare- keep / preserve (V)

drømme- dream (V)

fellesskap- community/fellowship (N)

forbund- association/ alliance (N)

fremtidig- future/ prospective (A)

fødsel- birth (N)

gidde- be bothered (V)

gulv- floor (N)

hemmelig- secret (A)

hull- hole/gap (N)

håndtere- handle/ manage (V)

industri- industry (N)

informere – inform (V)

initiativ- initiative (N)

innvandrer- immigrant (N)

komité- committee (N)

konsentrere – concentrate (V)

konsept- concept (N)

kulturell- cultural (A)

kvinnelig- female/feminine (A)

landbruk- agriculture/ farming (N)

leverandør – supplier/ contractor (N)

likhet- similarity/ equality (N)

mangfold- diversity/ variety (N)

midlertidig- temporary (A)

motstander- opponent (N)

nabo- neighbour (N)

offer- victim/ sacrifice (N)

pappa- dad (N)

potensiell- potential (A)

regelverk- regulations (N)

rektor- headmaster (N)

revisor- accountant (N)

rope- shout /scream (V)

servere- serve/wait/deliver (V)

struktur- structure (N)

støy- noise (N)

sur- grumpy/ sour (A)

svekke- weaken (V)

tempo – speed / pace (N)

tolke- interpret (V)

transport- transport (N)

uheldig – unfortunate/ unlucky (A)

vert- host/landlord (N)

variant – variant/ version (N)

vind- wind (N)

økning- increase (N)

Appendix F: Age distribution within the cohorts

The current section presents the age distributions within both cohorts in the study.

20-30 y/o					
Age	**No.**	%	**Age**	**No.**	%
20	2	1,64%	**26**	11	9,02%
21	12	9,84%	**27**	3	2,46%
22	33	27,05%	**28**	2	1,64%
23	25	20,49%	**29**	0	0,00%
24	17	13,93%	**30**	5	4,10%
25	12	9,84%			
				122	%

Over 60 y/o								
Age	**No.**	%	**Age**	**No.**	%	**Age**	**No.**	%
60	0	0,00%	**72**	2	3,92%	**84**	2	3,92%
61	0	0,00%	**73**	3	5,88%	**85**	0	0,00%
62	0	0,00%	**74**	2	3,92%	**86**	1	1,96%
63	4	7,84%	**75**	4	7,84%	**87**	0	0,00%
64	2	3,92%	**76**	3	5,88%	**88**	0	0,00%
65	0	0,00%	**77**	1	1,96%	**89**	1	1,96%
66	2	3,92%	**78**	3	5,88%	**90**	3	5,88%
67	0	0,00%	**79**	3	5,88%	**91**	0	0,00%
68	4	7,84%	**80**	1	1,96%	**92**	0	0,00%
69	2	3,92%	**81**	3	5,88%	**93**	0	0,00%
70	1	1,96%	**82**	0	0,00%	**94**	1	1,96%
71	2	3,92%	**83**	1	1,96%	**95**	0	0,00%
							51	**100,00%**

Appendix G: Comparison with data from Håseth (1968)

Below is an overview of the top ten responses for the five cue words that overlapped between the word association test used in Håseth (1968), and the current study. See 2.1.5 for more on the comparison.

kald (A)					
20-30 y/o		Over 60 y/o		Håseth	
varm (A)	21,3%	fryse(r) (V)	17,6%	varm(t)	42,8%
vinter(en) (N)	18%	varm (A)	17,6%	is (N)	12,5%
fryse/fryst (V)	13,1%	kjølig (A)	11,8%	vinter (N)	7,4%
is (N)	7,4%	is (N)	9,8%	kjølig (A)	4,3%
snø (N/V)	4,9%	frossen (A)	5,9%	fryse (V)	3,9%
blå(tt) (A)	4,1%	vinter (N)	5,9%	sur(t) (A)	2,6%
frossen (A)	4,1%	avvisende (A)	3,9%	hard(t) (A)	2,1%
våt (A)	2,5%	snø (N/V)	3,9%	våt (A)	1,4%
kjølig (A)	1,6%			frossen (A)	1,2%

religion (N)					
20-30 y/o		Over 60 y/o		Håseth	
kristendom(men) (N)	21,3%	tro (N)	51%	tro (N)	16,8%
tro (N)	13,1%	kristendom (N)	9,8%	kristendom(men) (N)	13%
gud (N)	7,4%	kristen (A)	5,9%	bibel(en) (N)	10,8%
islam (N)	6,6%	bibel (N)	3,9%	gud(er) (N)	6,1%
kors (N)	4,1%	krig (N)	3,9%	prest (N)	6%
kirke (N)	3,3%			kirke(n) (N)	4,1%
kristen (A)	3,3%			lære (N)	2,5%
muslim (A)	3,3%			protestant (N)	1,8%
Jesus (N)	2,5%			buddhisme (N)	1,3%
krig (N)	2,5%			katolikk(er) (N)	1,1%

sur (A)					
20-30 y/o		Over 60 y/o		Håseth	
sint (A)	17,2%	gretten (A)	17,6%	søt(t) (A)	20,3%
gretten (A)	7,4%	grinete (A)	15,7%	sinna/sint (A)	17,2%
grinete (A)	4,9%	sinna/sint (A)	7,8%	gretten (A)	14,7%
godteri (N)	4,1%	søt (A)	5,9%	grinet(e) (A)	9,9%
sitron (N)	4,1%	tverr (A)	5,9%	sitron (N)	2,7%
blid (A)	3,3%	lei (A)	3,9%	blid (A)	2,5%
kjip (A)	3,3%			tverr (A)	2,5%
lei (A)	3,3%			ekkel(t) (A)	1,6%
glad (A)	2,5%			melk (N)	1,4%
melk (N)	2,5%			bitter (A)	1,3%

Comparision of responses to cue words that were also part of Håseth (1968). Idiosyncratic responses are excluded from the comparison.

søt (A)					
20-30 y/o		Over 60 y/o		Håseth	
sukker (N)————	12,3%	pen (A)———	15,7%	sur(t) (A)————	22,6%
snill (A)	6,6%	nydelig (A)	9,8%	pen (A)	22,5%
godteri (N)	5,7%	sukker (N)	9,8%	god(t) (A)	7,8%
fin (A)	4,9%	snill (A)	7,8%	sukker (N)	7,8%
pen (A)	4,9%	god(t) (A)	5,9%	pike (N)	3,8%
jente (N)	4,1%	yndig (A)	5,9%	nydelig (A)	2,8%
jordbær (N)	4,1 %	deilig (A)	3,9%	deilig (A)	2,7%
dyr (N)	2,3%	fin (A)	3,9%	snill (A)	2,7%
rosa (A)	2,5%	sur (A)	3,9%	yndig (A)	2,3%
sjokolade (N)	2,5%			jente(r) (N)	2,1%

vindu (A)					
20-30 y/o		Over 60 y/o		Håseth	
glass (N)	22,1%	utsikt (N)	15,7%	glass (N)	33,3%
hus (N)	7,4%	åpning (N)	11,8%	dør (N)	8,4%
utsikt (N)	7,4%	glass (N)	9,8%	utsikt (N)	5,9%
lys (N)	6,6%	se (V)	7,8%	rute(r) (N)	5,6%
dør (N)	4,9%	utsyn (N)	7,8%	gardin(er) (N)	5,2%
gardin(er) (N)	4,9%	glugg(e) (N)	3,9%	lys (N)	5,2%
karm (N)	4,1%	lys (N)	3,9%	hus (N)	3,2%
se (V)	3,3%	åpen(t) (A)	3,9%	åpning (N)	3%
ut (Adv)	3,3%	åpne (V)	3,9%	se(r) (V)	2,9%
himmel(en) (N)	2,5%			åpen(t) (A)	2,6%

Appendix H: Multi-word responses, blank responses, and unreadable responses

The following pages provides numbers of blank responses, multi-word responses, and unreadable responses for all cue words for both age cohorts in the study.

20-30 y/o						
	Blanks		**Multi-word**		**Unreadables**	
andel (N)	0	0,00 %	3	2,46 %	0	0,00 %
anledning (N)	0	0,00 %	2	1,64 %	0	0,00 %
ansette (V)	0	0,00 %	2	1,64 %	0	0,00 %
avis (N)	0	0,00 %	0	0,00 %	0	0,00 %
beholde (V)	0	0,00 %	7	5,74 %	1	0,82 %
direktør (N)	0	0,00 %	0	0,00 %	0	0,00 %
eksistere (V)	0	0,00 %	9	7,38 %	1	0,82 %
formidle (V)	0	0,00 %	2	1,64 %	0	0,00 %
forventning (N)	4	3,28 %	4	3,28 %	0	0,00 %
frykte (V)	1	0,82 %	2	1,64 %	0	0,00 %
gøy (A)	0	0,00 %	1	0,82 %	0	0,00 %
gård (N)	0	0,00 %	0	0,00 %	0	0,00 %
hverdag (N)	1	0,82 %	0	0,00 %	0	0,00 %
hyggelig (A)	0	0,00 %	0	0,00 %	0	0,00 %
høgskole (N)	1	0,82 %	2	1,64 %	1	0,82 %
innføre (V)	2	1,64 %	4	3,28 %	0	0,00 %
intervju (N)	0	0,00 %	1	0,82 %	0	0,00 %
kald (A)	0	0,00 %	0	0,00 %	0	0,00 %
kjærlighet (N)	0	0,00 %	1	0,82 %	0	0,00 %
konsert (N)	0	0,00 %	1	0,82 %	0	0,00 %
kose (V)	2	1,64 %	2	1,64 %	0	0,00 %
litteratur (N)	1	0,82 %	0	0,00 %	0	0,00 %
materiale (N)	1	0,82 %	1	0,82 %	1	0,82 %
middel(N)	2	1,64 %	0	0,00 %	1	0,82 %
nyte (V)	1	0,82 %	4	3,28 %	1	0,82 %
oppfordre (V)	1	0,82 %	7	5,74 %	1	0,82 %
opprinnelig (A)	1	0,82 %	3	2,46 %	0	0,00 %
overføre (V)	0	0,00 %	1	0,82 %	0	0,00 %
planlegge (V)	0	0,00 %	4	3,28 %	0	0,00 %
posisjon (N)	0	0,00 %	2	1,64 %	0	0,00 %
prege (V)	3	2,46 %	4	3,28 %	0	0,00 %
psykisk (A)	0	0,00 %	2	1,64 %	0	0,00 %
religion (N)	0	0,00 %	0	0,00 %	0	0,00 %
savne (V)	0	0,00 %	1	0,82 %	0	0,00 %
stolt (A)	2	1,64 %	1	0,82 %	1	0,82 %
størrelse (N)	4	3,28 %	2	1,64 %	0	0,00 %
sunn (A)	0	0,00 %	0	0,00 %	0	0,00 %
sykehus (N)	0	0,00 %	4	3,28 %	1	0,82 %
søt (A)	2	1,64 %	3	2,46 %	0	0,00 %
telle (V)	1	0,82 %	1	0,82 %	0	0,00 %
trolig (A)	4	3,28 %	3	2,46 %	1	0,82 %
understreke (V)	2	1,64 %	1	0,82 %	0	0,00 %
unik (A)	0	0,00 %	2	1,64 %	0	0,00 %
usikker (A)	2	1,64 %	3	2,46 %	0	0,00 %
utvalg (N)	1	0,82 %	3	2,46 %	0	0,00 %
vedtak (N)	1	0,82 %	2	1,64 %	1	0,82 %
vindu (N)	0	0,00 %	1	0,82 %	1	0,82 %
virksomhet (N)	2	1,64 %	1	0,82 %	0	0,00 %
øl (N)	0	0,00 %	0	0,00 %	0	0,00 %
årsak (N)	0	0,00 %	1	0,82 %	0	0,00 %

	Blanks		Multi-word		Unreadables	
analyse (N)	0	0,00 %	0	0,00 %	0	0,00 %
annonse (N)	0	0,00 %	1	0,82 %	0	0,00 %
avansere (V)	0	0,00 %	7	5,74 %	0	0,00 %
bevare (V)	0	0,00 %	19	15,57 %	1	0,82 %
drømme (V)	0	0,00 %	2	1,64 %	0	0,00 %
fellesskap (N)	0	0,00 %	1	0,82 %	2	1,64 %
forbund (N)	2	1,64 %	1	0,82 %	0	0,00 %
fremtidig (A)	0	0,00 %	6	4,92 %	1	0,82 %
fødsel (N)	0	0,00 %	3	2,46 %	0	0,00 %
gidde (V)	1	0,82 %	1	0,82 %	1	0,82 %
gulv (N)	0	0,00 %	2	1,64 %	0	0,00 %
hemmelig (A)	0	0,00 %	2	1,64 %	0	0,00 %
hull (N)	0	0,00 %	1	0,82 %	0	0,00 %
håndtere (V)	2	1,64 %	6	4,92 %	0	0,00 %
industri (N)	0	0,00 %	3	2,46 %	1	0,82 %
informere (V)	0	0,00 %	5	4,10 %	0	0,00 %
initiativ (N)	2	1,64 %	5	4,10 %	0	0,00 %
innvandrer (N)	1	0,82 %	1	0,82 %	0	0,00 %
komité (N)	2	1,64 %	2	1,64 %	1	0,82 %
konsentrere (V)	0	0,00 %	0	0,00 %	0	0,00 %
konsept (N)	1	0,82 %	4	3,28 %	0	0,00 %
kulturell (A)	0	0,00 %	1	0,82 %	1	0,82 %
kvinnelig (A)	1	0,82 %	0	0,00 %	0	0,00 %
landbruk (N)	0	0,00 %	0	0,00 %	1	0,82 %
leverandør (N)	2	1,64 %	0	0,00 %	0	0,00 %
likhet (N)	0	0,00 %	3	2,46 %	0	0,00 %
mangfold (N)	0	0,00 %	1	0,82 %	1	0,82 %
midlertidig (A)	2	1,64 %	9	7,38 %	0	0,00 %
motstander (N)	0	0,00 %	0	0,00 %	0	0,00 %
nabo (N)	1	0,82 %	5	4,10 %	0	0,00 %
offer (N)	0	0,00 %	1	0,82 %	0	0,00 %
pappa (N)	0	0,00 %	1	0,82 %	1	0,82 %
potensiell (A)	1	0,82 %	1	0,82 %	0	0,00 %
regelverk (N)	1	0,82 %	0	0,00 %	0	0,00 %
rektor (N)	0	0,00 %	3	2,46 %	0	0,00 %
revisor (N)	1	0,82 %	1	0,82 %	0	0,00 %
rope (V)	0	0,00 %	0	0,00 %	1	0,82 %
servere (V)	0	0,00 %	1	0,82 %	0	0,00 %
struktur (N)	1	0,82 %	0	0,00 %	0	0,00 %
støy (N)	0	0,00 %	0	0,00 %	0	0,00 %
sur (A)	0	0,00 %	4	3,28 %	2	1,64 %
svekke (V)	1	0,82 %	2	1,64 %	1	0,82 %
tempo (N)	0	0,00 %	1	0,82 %	1	0,82 %
tolke (V)	0	0,00 %	1	0,82 %	0	0,00 %
transport (N)	0	0,00 %	0	0,00 %	0	0,00 %
uheldig (A)	2	1,64 %	2	1,64 %	1	0,82 %
vert (N)	0	0,00 %	0	0,00 %	1	0,82 %
variant (N)	2	1,64 %	2	1,64 %	1	0,82 %
vind (N)	0	0,00 %	1	0,82 %	0	0,00 %
økning (N)	1	0,82 %	1	0,82 %	0	0,00 %
	69	**0,57 %**	**213**	**1,75 %**	**31**	**0,25 %**

Multi-word respones, blank responses, and unreadable responses for the younger cohort

Over 60 y/o						
	Blanks		**Multi-word**		**Unreadables**	
andel (N)	0	0,00 %	3	5,88 %	0	0,00 %
anledning (N)	0	0,00 %	1	1,96 %	0	0,00 %
ansette (V)	4	7,84 %	7	13,73 %	1	1,96 %
avis (N)	0	0,00 %	1	1,96 %	0	0,00 %
beholde (V)	1	1,96 %	7	13,73 %	0	0,00 %
direktør (N)	0	0,00 %	1	1,96 %	0	0,00 %
eksistere (V)	0	0,00 %	3	5,88 %	0	0,00 %
formidle (V)	1	1,96 %	4	7,84 %	0	0,00 %
forventning (N)	0	0,00 %	4	7,84 %	0	0,00 %
frykte (V)	0	0,00 %	9	17,65 %	0	0,00 %
gøy (A)	1	1,96 %	1	1,96 %	0	0,00 %
gård (N)	0	0,00 %	0	0,00 %	0	0,00 %
hverdag (N)	1	1,96 %	3	5,88 %	0	0,00 %
hyggelig (A)	0	0,00 %	1	1,96 %	2	3,92 %
høgskole (N)	2	3,92 %	2	3,92 %	1	1,96 %
innføre (V)	4	7,84 %	8	15,69 %	3	5,88 %
intervju (N)	1	1,96 %	2	3,92 %	0	0,00 %
kald (A)	0	0,00 %	1	1,96 %	1	1,96 %
kjærlighet (N)	0	0,00 %	7	13,73 %	1	1,96 %
konsert (N)	0	0,00 %	0	0,00 %	0	0,00 %
kose (V)	0	0,00 %	9	17,65 %	0	0,00 %
litteratur (N)	0	0,00 %	2	3,92 %	0	0,00 %
materiale (N)	1	1,96 %	5	9,80 %	1	1,96 %
middel (N)	6	11,76 %	3	5,88 %	1	1,96 %
nyte (V)	0	0,00 %	5	9,80 %	2	3,92 %
oppfordre (V)	2	3,92 %	11	21,57 %	0	0,00 %
opprinnelig (A)	2	3,92 %	5	9,80 %	0	0,00 %
overføre (V)	1	1,96 %	8	15,69 %	2	3,92 %
planlegge (V)	3	5,88 %	14	27,45 %	0	0,00 %
posisjon (N)	2	3,92 %	2	3,92 %	0	0,00 %
prege (V)	5	9,80 %	7	13,73 %	1	1,96 %
psykisk (A)	1	1,96 %	2	3,92 %	0	0,00 %
religion (N)	0	0,00 %	2	3,92 %	0	0,00 %
savne (V)	3	5,88 %	5	9,80 %	0	0,00 %
stolt (A)	1	1,96 %	4	7,84 %	1	1,96 %
størrelse (N)	3	5,88 %	4	7,84 %	0	0,00 %
sunn (A)	0	0,00 %	4	7,84 %	1	1,96 %
sykehus (N)	1	1,96 %	4	7,84 %	1	1,96 %
søt (A)	0	0,00 %	3	5,88 %	0	0,00 %
telle (V)	1	1,96 %	5	9,80 %	0	0,00 %
trolig (A)	2	3,92 %	3	5,88 %	0	0,00 %
understreke (V)	2	3,92 %	8	15,69 %	0	0,00 %
unik (A)	1	1,96 %	7	13,73 %	1	1,96 %
usikker (A)	3	5,88 %	9	17,65 %	0	0,00 %
utvalg (N)	2	3,92 %	9	17,65 %	3	5,88 %
vedtak (N)	4	7,84 %	3	5,88 %	0	0,00 %
vindu (N)	0	0,00 %	3	5,88 %	0	0,00 %
virksomhet (N)	4	7,84 %	0	0,00 %	1	1,96 %
øl (N)	0	0,00 %	1	1,96 %	0	0,00 %
årsak (N)	3	5,88 %	3	5,88 %	0	0,00 %

	Blanks		Multi-word		Unreadables	
analyse	4	7,84 %	6	11,76 %	0	0,00 %
annonse (N)	0	0,00 %	2	3,92 %	0	0,00 %
avansere (V)	3	5,88 %	20	39,22 %	1	1,96 %
bevare (V)	1	1,96 %	17	33,33 %	0	0,00 %
drømme (V)	1	1,96 %	6	11,76 %	1	1,96 %
fellesskap (N)	0	0,00 %	4	7,84 %	0	0,00 %
forbund (N)	0	0,00 %	2	3,92 %	0	0,00 %
fremtidig (A)	3	5,88 %	11	21,57 %	0	0,00 %
fødsel (N)	1	1,96 %	9	17,65 %	0	0,00 %
gidde (V)	0	0,00 %	5	9,80 %	0	0,00 %
gulv (N)	0	0,00 %	5	9,80 %	0	0,00 %
hemmelig (A)	5	9,80 %	9	17,65 %	1	1,96 %
hull (N)	0	0,00 %	10	19,61 %	1	1,96 %
håndtere (V)	0	0,00 %	3	5,88 %	0	0,00 %
industri (N)	0	0,00 %	2	3,92 %	0	0,00 %
informere (V)	0	0,00 %	8	15,69 %	0	0,00 %
initiativ (N)	2	3,92 %	8	15,69 %	1	1,96 %
innvandrer (N)	0	0,00 %	7	13,73 %	0	0,00 %
komité (N)	2	3,92 %	3	5,88 %	1	1,96 %
konsentrere (V)	3	5,88 %	5	9,80 %	0	0,00 %
konsept (N)	8	15,69 %	1	1,96 %	2	3,92 %
kulturell (A)	4	7,84 %	8	15,69 %	1	1,96 %
kvinnelig (A)	1	1,96 %	2	3,92 %	1	1,96 %
landbruk (N)	0	0,00 %	3	5,88 %	0	0,00 %
leverandør (N)	3	5,88 %	6	11,76 %	1	1,96 %
likhet (N)	4	7,84 %	6	11,76 %	0	0,00 %
mangfold (N)	2	3,92 %	6	11,76 %	1	1,96 %
midlertidig (A)	3	5,88 %	16	31,37 %	0	0,00 %
motstander (N)	1	1,96 %	5	9,80 %	1	1,96 %
nabo (N)	2	3,92 %	8	15,69 %	1	1,96 %
offer (N)	3	5,88 %	8	15,69 %	0	0,00 %
pappa (N)	0	0,00 %	3	5,88 %	0	0,00 %
potensiell (A)	6	11,76 %	3	5,88 %	1	1,96 %
regelverk (N)	1	1,96 %	3	5,88 %	0	0,00 %
rektor (N)	0	0,00 %	2	3,92 %	1	1,96 %
revisor (N)	1	1,96 %	2	3,92 %	1	1,96 %
rope (V)	0	0,00 %	4	7,84 %	0	0,00 %
servere (V)	1	1,96 %	5	9,80 %	0	0,00 %
struktur (N)	4	7,84 %	4	7,84 %	1	1,96 %
støy (N)	0	0,00 %	3	5,88 %	0	0,00 %
sur (A)	0	0,00 %	4	7,84 %	2	3,92 %
svekke (V)	3	5,88 %	5	9,80 %	1	1,96 %
tempo (N)	0	0,00 %	1	1,96 %	0	0,00 %
tolke (V)	1	1,96 %	4	7,84 %	0	0,00 %
transport (N)	0	0,00 %	4	7,84 %	0	0,00 %
uheldig (A)	4	7,84 %	4	7,84 %	1	1,96 %
vert (N)	3	5,88 %	4	7,84 %	1	1,96 %
variant (N)	3	5,88 %	5	9,80 %	1	1,96 %
vind (N)	0	0,00 %	4	7,84 %	0	0,00 %
økning (N)	1	1,96 %	4	7,84 %	1	1,96 %
	152	**2,98 %**	**494**	**9,69 %**	**49**	**0,96 %**

Multi-word respones, blank responses, and unreadable responses for the older cohort

Appendix I: Cue word imageability

The table below shows the imageability ratings for the 71 of the cue words in the word association test such ratings were available for. See also 4.5.8.

Cue	Class	Imageability
analyse	N	low
anledning	N	low
annonse	N	high
ansette	V	low
avansere	V	low
avis	N	high
beholde	V	low
drømme	V	low
eksistere	V	low
fellesskap	N	low
forbund	N	low
forventning	N	low
fremtidig	A	low
frykte	V	low
fødsel	N	high
gidde	V	low
gøy	A	low
gård	N	high
hull	N	low
hverdag	N	low
hyggelig	A	low
industri	N	low
initiativ	N	low
innføre	V	low
innvandrer	N	low
intervju	N	low
kald	A	low
kjærlighet	N	low
komité	N	low
konsentrere	V	low
konsert	N	high
kose	V	high
kulturell	A	low
landbruk	N	low
litteratur	N	low
mangfold	N	low
materiale	N	low
nabo	N	low
offer	N	low
pappa	N	high
planlegge	V	low
posisjon	N	low
prege	V	low
psykisk	A	low
religion	N	low
revisor	N	low
rope	V	high
savne	V	low
servere	V	low
stolt	A	low
struktur	N	low
størrelse	N	low
sur	A	low
svekke	V	low
sykehus	N	high
søt	A	low
telle	V	low
tempo	N	low
tolke	V	low
transport	N	low
trolig	A	low
uheldig	A	low
usikker	A	low
variant	N	low
vedtak	N	low
vind	N	high
vindu	N	high
virksomhet	N	low
økning	N	low
øl	N	high
årsak	N	low

Imageability ratings and number of blanks for 71 of the cue words in the word association test

Appendix J: Norms lists

The full norms lists from the study can be accessed online at http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/sprakmateriale.html.

The dominant responses for each age cohort can be found in 5.1.

Appendix K: Response word class distribution for verb cues

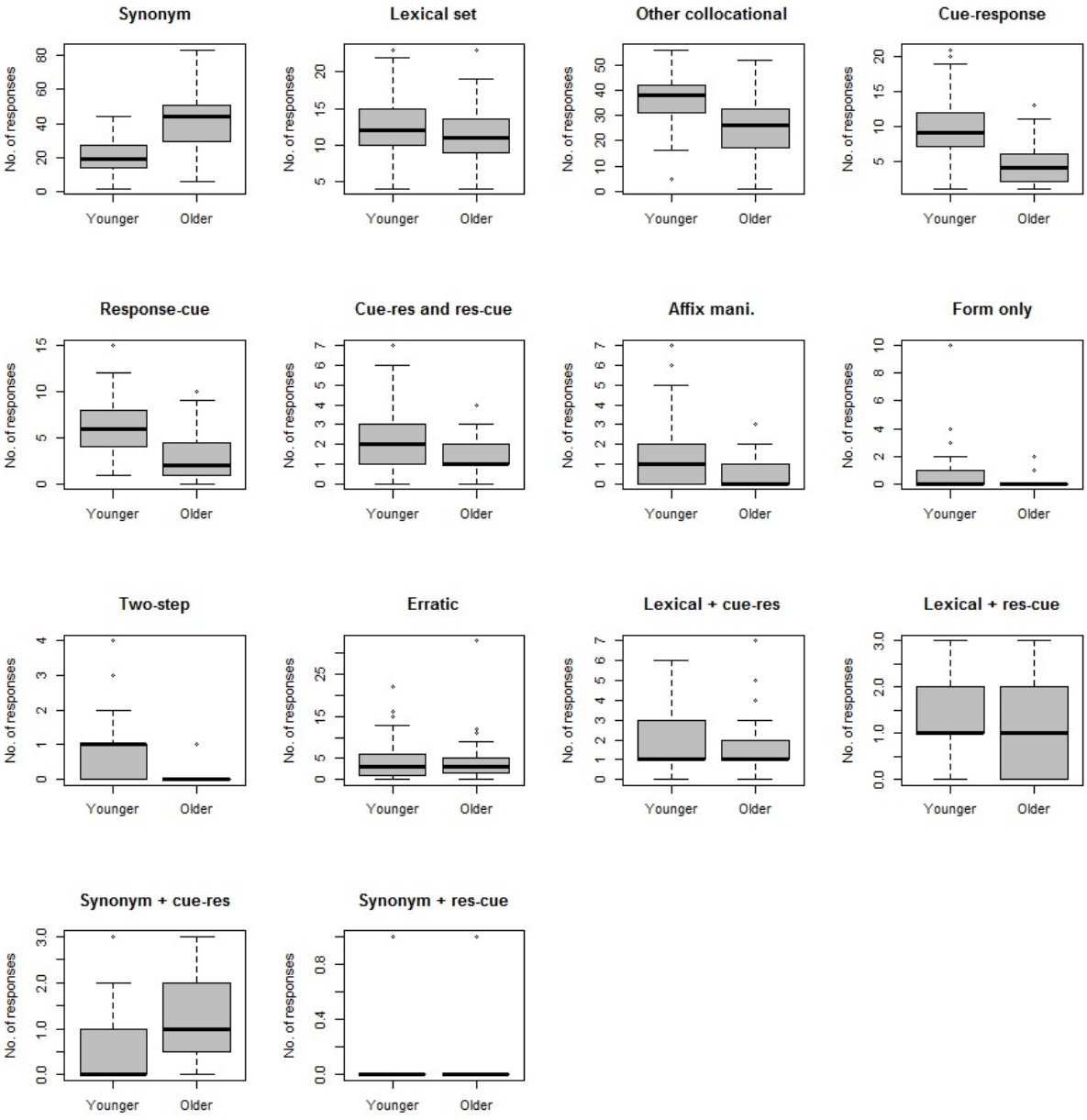
The table below shows the distribution of word classes for the responses given to the verb cues in the word association test. The table is based on the responses from the younger cohort (n=122). *Adj.=adjective*, *Det.=determiner*, *Prep.=preposition*, *Inter.=interjection*m *Subord.=subordinator*.

	Noun	Verb	Adj.	Adverb	Det.	Prep.	Pronoun	Inter.	Subord.
ansette	98	19	2	0	0	0	0	0	0
beholde	29	75	5	0	7	0	0	0	0
eksistere	21	97	1	0	1	0	0	0	0
formidle	33	83	2	0	0	0	0	0	0
frykte	35	11	75	0	0	0	0	0	0
innføre	49	52	14	0	0	0	0	0	0
kose	80	24	12	0	0	1	0	0	0
nyte	65	33	15	0	0	0	0	1	0
oppfordre	17	93	4	0	0	1	0	0	0
overføre	85	38	0	0	0	0	0	0	0
planlegge	89	24	3	0	0	0	0	0	0
prege	45	59	9	0	0	1	0	0	0
savne	62	30	25	0	4	1	0	0	0
telle	94	11	2	0	10	0	0	0	0
understreke	31	51	35	0	0	0	1	0	0
avansere	29	40	36	5	1	2	0	0	0
bevare	37	60	6	0	0	0	0	0	0
drømme	56	58	4	2	0	2	0	0	0
gidde	19	87	12	0	0	0	0	2	0
håndtere	29	78	6	0	2	0	0	0	0
informere	47	65	5	0	0	0	0	0	1
konsentrere	45	62	14	0	0	0	1	0	0
rope	27	75	17	0	0	0	0	1	0
servere	94	26	2	0	0	0	0	0	0
svekke	18	45	52	1	2	0	0	0	0
tolke	41	75	5	0	0	0	0	0	0

Response word class distributions for verb cues. Data based on the younger cohort only, n=122.

Appendix L: Additional boxplots for response category distributions

Cohort differences, all fourteen response categories



Difference between noun cues and verb cues, all response categories