UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Radboud Universiteit

Master Thesis

Speech and Language Pathology

June 2019

THE ROLE OF AGING IN THE SUBJECT-VERB NUMBER AGREEMENT PRODUCTION PROCESS

Author: Rosemarije P. C. Weterings (S4565584)
Radboud University: Faculty of Arts, Master Speech and Language Pathology

Supervisor: Prof. dr. Robert J. Hartsuiker
Ghent University: Faculty of Psychology and Educational Sciences, Department of Experimental Psychology

Second reader: Dr. Marina B. Ruiter
Radboud University: Faculty of Arts, Centre for Language Studies / Department of Language & Communication

Acknowledgements

Before you lies my master thesis: *The role of aging in the subject-verb number agreement production process*. With this thesis, I will finish the specialisation program Language and Speech Pathology of the master Linguistics at the Radboud University in Nijmegen. I developed an interest in the effects of aging on language throughout my whole study, so it was a wonderful opportunity to specialize myself in this topic for half a year.

First, I want to thank my supervisor Prof. dr. Robert Hartsuiker for his guidance, help and enthusiasm. He taught me more about doing research and gave me new insights in the research field. The weekly lab meetings at Ghent University were very inspiring, and therefore I also want to thank the entire research group.

I would like to thank all my participants who joined the study and who were willingly to give their time and collaboration. In special, I want to thank my grandmother, who helped me recruiting many participants through all her social contacts. Additionally, I would like to thank the people who could not join my study, but nonetheless helped me with recruiting participants. Through all of you, I have found the participants I needed.

Last but not least, I want to thank all the people who helped me in the process of writing my thesis. Writing in English was not easy for me, and therefore all the help I got was very useful. With all good advice and tips, I managed to fulfil the writing part.

I am very proud to present my master thesis to you. Enjoy reading!

Rosemarije Weterings,

Nijmegen, 25 June 2019

Table of contents

Abstract

In order to investigate the effect of aging on the syntactic production process of subject-verb number agreement in Dutch, we compared the production of agreement errors between elderly people and young adults in a spoken sentence completion task. As in previous studies, effects of attraction (more agreement errors in sentences with a singular subject head noun and a plural local noun) and distributivity (more agreement errors when the conceptual number of the subject head noun mismatched the grammatical number) were found. No difference was found in the total amount of produced agreement errors. Aging, however, made the attraction effect stronger and the distributivity effect weaker. We presume therefore that the aging process changes the underlying mechanism of subject-verb number agreement. We suggest that the found limited working memory capacity of elderly people could be involved in this change.

1. Introduction

Aging, the process of growing older, happens to all of us. Aging in humans contains physiological, psychological, and social changes. During human life, a number of characteristic alterations could occur. Common health conditions associated with old age are for example, change of speech characteristics (Yorkson, Bourgeois, & Baylor, 2010), developing hearing loss (Gates, Feeney, & Mills, 2008), shrinking of brain volume (Peters, 2006), and affected language processes (Thornton & Light, 2006). Aging is also one of the main risk factors for the prevalence of diseases, such as cancer, cardiovascular disease, and neurodegeneration (Niccoli & Partridge, 2012). The amount of people getting age related diseases is expanding throughout the years. This goes together with the change of the demographic structure of the (Dutch) society (van Duin & Garssen, 2010). The change implies that the population of elderly people increases. The Dutch Central Bureau of Statistics (CBS) calls this process “double aging”: Improved living conditions, for instance enhanced medical care and technical interventions, support people to get older. Simultaneously, fertility rates decline.

As a consequence of these developments, gaining knowledge about the aging process and understanding how that process influences life in all aspects, is essential. Providing more insights in the aging process could lead to earlier recognition of age related decline. One of the main aims of aging research is supporting people to live healthy for more years. The current research contributes to this by investigating a small, but important aspect of the aging process: the course of the language production process and cognitive functioning.

Language is one of the most important aspects of human life. The production of speech, the use of language, and the possibility to communicate with other people is a special human characteristic and even essential for human physical (Hawkley, Thisted, Masi, & Cacioppo, 2010) and mental health (Wei, Russell, & Zakalik, 2005). The inability to communicate with others could, for instance, cause loneliness (Hawkley & Cacioppo, 2010). Being able to communicate is substantial for all age groups, in particular for elderly people. Loneliness and social isolation are potential risk factors for the increase of cognitive decline (Evans, Martyr, Collins, Brayne, & Clare, 2018; Gow & Mortensen, 2016). Moreover, healthy cognition plays a valuable role in the quality of life and independence of elderly people (Abrahamson, Clark, Perkins, & Arling, 2012). Thus, investigating the influence of aging on cognitive functions, especially language, is of crucial and social importance.

The current research contributes to the overall question what the influence of aging is on cognitive function and language of healthy people. It focusses on the effect of aging on the syntactic process of subject-verb number agreement in language production and the role of working memory in this process. A recent study using structural priming (Hardy, Messenger, & Maylor, 2017) suggested that syntactic representations and production processes do not suffer from aging. However, other studies (e.g. Hartsuiker and Barkhuysen, 2006; Lorimor, Jackson, & van Hell, 2019) have found indirect indications that aging does influence syntactic processes in sentence production. Those studies showed that subject-verb number agreement, a syntactic planning process which involves a dependency across a long distance in the sentence, relies heavily on working memory capacity. As it is conceivable that working memory capacity declines in normal aging (Craik, 1994), this syntactic process could be affected through this in elderly people. The study of Reifegerste, Hauer, and Felser (2017) showed that in a subject-verb number agreement comprehension task, elderly people indeed performed worse than young adults. This effect was modulated by working memory. As far as we know, no research has been done regarding the influence of aging and working memory capacity on the subject-verb number agreement production process. Our aim was to fill this hiatus in the literature. We expected elderly people to perform less in comparison with young adults on this syntactic language production process in which working memory is involved.

1.1 Cognitive functions and aging

Cognitive functions include all higher mental processes belonging to information registration and processing, and could be divided into different specific cognitive domains (Kolb & Whishaw, 2001). Cognitive functions like memory, executive function, and attention are examples of non-linguistic cognitive functions. Linguistic cognitive functions are the functions that are primary used by language processes such as phonology, morphology, syntax, semantics, and pragmatics. In normal aging, cognitive abilities often decline, although people are still healthy (Murman, 2015). Beside this normal cognitive decline in healthy people as a result of aging, brain diseases like neurodegeneration or brain traumas could also lead to cognitive decline. It is therefore very important to be able differentiating between the types of cognitive changes which occur in normal aging and those that point in the direction of brain diseases or traumas.

First, it should be mentioned that it is not easy to investigate cognitive changes occurring in the normal aging process, due to the limitations of aging research (Harada, Love, & Triebel, 2013; Murman, 2015). Biases could arise in subject selection and study design. It is, for example, possible that a bias occurs in recruiting subjects, because only the willingly and maybe consequently the healthiest and most advantaged participants join the study. There could also be a misclassification bias, which leads to classifying a subject as healthy and normal, although this is not the case. Two possible ways of studying the effects of aging are via a longitudinal design, following the same individuals over time, or via a comparison between two different age groups. By using a longitudinal design, it could happen that a practice effect appears of showing improvements on test results, because the same people do similar tests over time. By comparing two different groups, it could be that the difference found is not due to aging, but caused by other (unmeasured) distinctive features (Harada et al., 2013; Murman, 2015). These limitations should be kept in mind when we study aging effects. Despite these possible restrictions, adequate evidence appears of the existence of cognitive changes in normal aging.

The cognitive function memory has the main focus in the current research, because we believe it is involved in the syntactic process of subject-verb number agreement in language production. Memory plays a role in the encoding, storage and retrieval of intern and extern information (Kessels, Eling, Ponds, Spikman, & van Zandvoort, 2017). Memory could be divided into separate subsystems: working memory and long-term memory. In this paragraph we shortly explain what the cognitive function of memory enfolds, particularly the working memory system, and what is known about the influences of aging on this cognitive function.

1.1.1 Working memory

Working memory is a temporary place with a limited capacity for maintenance and active processing of information (Kessels et al., 2017). It plays an important role in the execution of complex cognitive operations. The functions of working memory are short-term retention and manipulation of information. A small amount of information is kept active for a short time period. This period is as long as attention is focused on the information. If attention turns to other information, the kept information in the working memory disappears.

The working memory model of Baddeley and Hitch (1974) and the adapted model of Baddeley (2000) are widely accepted and often used, although also criticised (Cowan, 2001). Baddeley's working memory model describes sub mechanisms that are involved in the working memory system: the central executive processor, the visuospatial scratchpads, the phonological loop, and the episodic buffer, which is added later by Baddeley. All these components have specific functions. See Figure 1 for a schematic representation of this adapted working memory model of Baddeley.

One of the sub mechanisms is the central executive processor, which is responsible for the control and regulation of higher-order cognitive functions and decides which information from the sub storage systems is brought under the attention (Baddeley, 2007). It is held responsible for information updating, modification and inhibition. This are executive functions, which are the higher control functions of the brain, and provide connections between different functions (Kessels et al., 2017). Awareness, planning and organisation, initialising and execution, regulation, and self-control are the different executive functions that contribute to consciously goal-orientated complex behaviour, which is also necessary in the working memory system.

The sub storage systems can be subdivided further on the basis of the type of information processing. The visuospatial scratchpad processes visual and spatial information (Baddeley, 2007). The phonological loop is responsible for the processing of auditory information. It is also called the auditory-verbal short-term memory (Kemmerer, 2014). It is the resource that is used to keep phonological information in an active state for a relatively short time period. The digit span task is often used to measure the capacity of the auditory-verbal short-term memory, which determines the longest correctly repeated sequence of arbitrary numbers (for a full review see Ramsay & Reynolds, 1995). The simple version of the digit span memory task is the forward digit span task. The given numbers have to be repeated in the same order in this task. Here, only the phonological loop is needed for the maintenance of information. A more complex version of the digit span memory task, is the digit span backward task. In this task, the given numbers have to be repeated in the reverse order. This is a working memory task, because also information processing is required. Both the phonological loop (maintenance of information) and the central executive processor (processing and manipulation of information) are necessary for executing this digit span backwards task (Baddeley & Logie, 1999). The two digit span tasks were used in the current research to measure the capacity of the working memory system.

The episodic buffer, added in the adapted model of Baddeley, links and associates visual, spatial, and verbal information together. It makes is possible to visualize an interaction between the working memory system and the long-term memory system. A transition from working memory to long-term memory is necessary to store information permanently. Via association and encoding in the episodic buffer, memory binding takes place to the long-term memory system. An episode is formed by the combination of visual, auditory, and perceptual information from the working memory system, and semantic and episodic information from the long-term memory system (Baddeley, 2007).

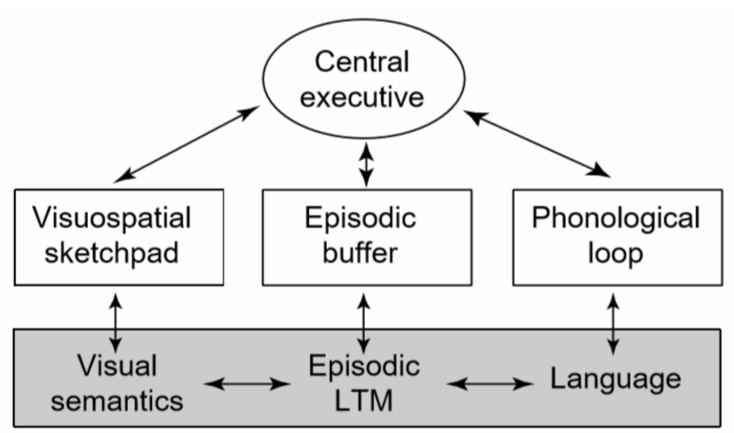


Figure 1. The adapted working memory model of Baddeley (2000).

1.1.2 Long-term memory

To retain information in memory, a transformation has to be made from working memory to long-term memory, which is done by forming episodes. Long-term memory refers to the storage of information over a longer period and could be divided into the declarative memory system and the non-declarative memory system (Kessels et al., 2017). In the non-declarative, also called implicit, memory system knowledge cannot be consciously recalled, although information is stored and behaviour is affected by the stored information. Processes such as priming, procedural learning, habituation and conditioning are all included in this implicit memory system. In the declarative, also called explicit, memory system stored knowledge consisting of facts could be consciously recalled. Subsystems of the explicit memory system are the episodic memory for storing events and the semantic memory for storing facts (Kessels et al., 2017).

1.1.3 Memory decline in aging

Memory capacity changes in normal aging, although not all aspects are equally affected. Some memory processes are more vulnerable to the effects of aging than other processes. Elderly people show a reduced capacity on simple and complex cognitive working memory tasks (Choi et al., 2014; Monaco, Costa, Caltagirone, & Carlesimo, 2013). Tasks that require primary storage capacity, like the digit span forward task, are done worse by elderly people in comparison with young adults. The same applies to tasks in which more working memory capacity is involved, like the digit span backward task. The capacity of the working memory system is important in the execution of complex cognitive tasks, because processed information must be hold (Cowan, 2010). In performing the digit span backward tasks, people not only have to store information, but also have to do active processes with this information. Research on the course of working memory founds that working memory improves with age in the period of infancy (Gathercole, Pickering, Ambridge, & Wearing, 2004; Pickering, 2001) and that it decreases in the period of later age (Park et al., 1996). So, aging affects working memory capacity in a way that it declines when adults are getting older (Kirova, Bays, & Lagalwar, 2015).

The question arises why elderly people lose working memory capacity. Possibly, there could be an influence of the reduction of the ability to inhibit distractors (Hasher & Zacks, 1988; McNab et al., 2015). As the executive processor of the working memory system decides which information is brought under the attention, the base lays here of the ability to inhibit distracting and non-relevant information to enter the working memory system. If more non-relevant information is held in the working memory system, less capacity remains available for relevant information. A brain activity study, for instance, has given evidence that a lower working memory capacity correlated with the storage of more information that belongs to distractors (Chadick, Zanto, & Gazzaley, 2014).

In addition to this aging effect on the working memory system, the episodic memory system declines over time (Fonseca, Zimmermenn, Scherer, Parente, & Ska, 2010; Irish, Lawlor, O'Mara, & Coen, 2011). Progressive declines in immediate and delayed recall, and recognition of stories, numbers, and words by elderly people over a period of four years were found by Fleischman, Wilson, Gabrieli, Bienias, and Bennett (2004). Memory systems that seem to be relatively spared from age related decline are the implicit memory system, which includes priming (Fleischman et al., 2004) and procedural memory (Chauvel et al., 2012; Smith et al., 2005), and the semantic memory system (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), which includes facts and general knowledge.

1.2 The language production process

The effects of aging on language and communication are widely studied. Studies have found that different language processes could be affected by aging (Thornton & Light, 2006). For example, studies about the lexical access process found that elderly people become slower and make more mistakes in picture naming tasks (Verhaegen & Poncelet, 2013), and have more difficulties in word finding (Meinzer et al., 2012). To understand more about the effects of aging on this process, it is necessary to comprehend the normal language production process. This process will be describe this in the next paragraph. Especially the syntactic process of language production is explained in more detail. This is one of the main subjects of this study, because little research is done about aging effects on this process.

1.2.1 Sentence production

As shown by most psycholinguistic models (Dell, 1986; Garret, 1988; Levelt, 1989) a number of processing levels are required for the formulation of a sentence in the language production process: the conceptual, syntactic, and phonological level (Schriefers & Vigliocco, 2015). However, psycholinguistic models differ in other areas: for example whether the levels are accessed parallel (Patterson & Shewell, 1987) or sequential (Levelt, 1989). For the explanation of language production process we follow the model of Levelt (1989) and Levelt, Roelofs, and Meyer (1999). Figure 2 visualized the levels in the different stages of the language production process, according to the adapted model from Levelt and Levelt et al. The stages are explained in more detail below.

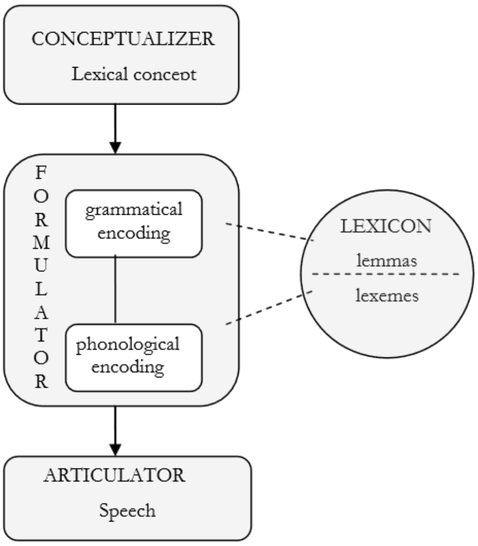


Figure 2. Schematic representation of a psycholinguistic language production model. The model is adapted from Levelt (1989; see also Levelt et al., 1999; figure from Mehotcheva, 2010, p. 33.).

The first step in the language production process is the forming of the intended abstract idea, without the involvement of words. This is the conceptualisation process, in which general world knowledge and information about the specific situation is used. A pre-verbal message is the outcome of this level.

The next step is to use language knowledge to express the intended idea. Here, words and a grammatical structure have to be selected. These processes happen in the formulation level, which involves grammatical and phonological encoding. Grammatical encoding is used to create a sentence structure for the conveyance of the intended message. This process could be divided into two processes: functional and positional processing. See Figure 3 for the detailed steps of grammatical encoding process in Garrett's Model (1975). The aim of the functional processing is to select the appropriate words (lexical selection) and give them their grammatical function (syntactical role) in the sentence in a way that it expresses the intended message (functional assignment). In the lexical selection process abstract forms of words, the lemmas, are gathered from the mental lexicon, which is the mental storage system for the concepts of all known words. The other component of grammatical encoding is the positional processing, which determines the order (constituent assembly) and inflection of the selected lemmas. After grammatical encoding, appropriate sound sequences are constructed in the phonological encoding process and form lexemes, the manifestation of the morphological and phonological version of words. At this stage the pre-verbal message has turned into structured language: a phonetic speech plan.

In order to speak the formulated message out loud and produce speech sounds, the articulation process has to be completed by converting the speech plan into speech movement. Following the described processes leads eventually to the production of language. The current research focusses on the syntactic processing in the formulation process of grammatical encoding: forming number agreement between subject and verb, which will be explained in more detail below.

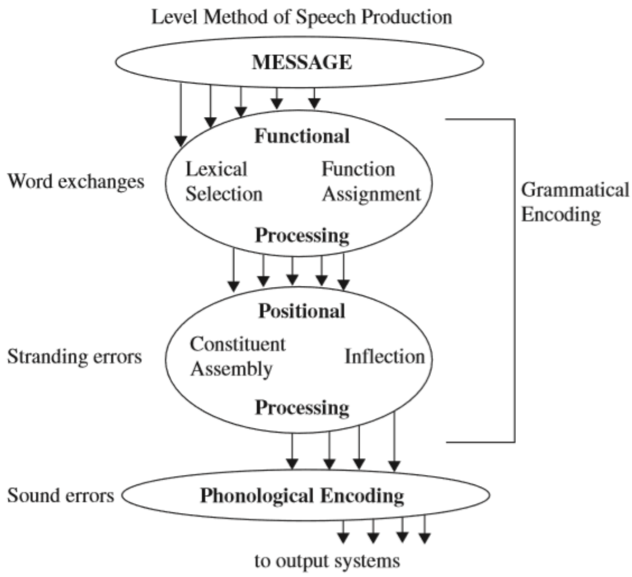


Figure 3. The components of the formulation process, including the subprocesses of syntactic encoding. This model is adapted from Garrett (1975; figure from Ferreira & Engelhardt, 2006, p. 63.).

1.2.2 The syntactic process of subject-verb number agreement

Syntax is used to bring mental concepts together on a linguistic manner. Syntax exists of a set of rules, principles, and processes that convey a sentence structure in a certain language. A part of the language production process must take care of the development of a syntactically well-formed sentence. The relationships of the participants in a conceptual representation (e.g. agent, patient, theme, etc.) are aligned onto the more functional syntactic relationship between the words in the sentence (e.g. subject, direct object, indirect object, etc.). This results in a hierarchically organized syntactic sentence frame of word order, that is organized in a linear frame in the next step (Vigliocco & Nicol, 1998). This linear frame of word order is the way the sentence will be spoken.

Part of this process is the syntactic processing of subject-verb number agreement. Subject-verb number agreement is the grammatical rule that determines that the subject and the verb should agree in number: A singular subject needs a singular verb and a plural subject needs a plural verb. This grammatical rule applies to many languages, including the language (Dutch) used for our study.

To investigate this process of subject-verb number agreement, Bock and Miller (1991) developed a paradigm that provokes subject-verb number agreement errors as a violation of the grammatical subject-verb number agreement rule. In this paradigm short sentence fragments without a verb (such as 1) are given to participants. They have to repeat these fragments out loud and complete them while forming a grammatically correct sentence. The short sentence fragments consist of a complex subject noun phrase, including a subject head noun, and a modifying prepositional phrase, including a local noun. The sentence fragments do not include a verb. In this paradigm, a verb is elicited on a natural way. According to the subject-verb number agreement rule, this verb should correspondent in number with the subject. The language process in the paradigm of Bock and Miller consists of both a comprehension part: understand what the given sentence fragment means, and a production part: repeat the heard sentence fragment, find the verb, and give the verb the right inflection, depending on the number of the subject. The outcome variable in this paradigm is the proportion of agreement errors, occurring in sentences with a verb that has an incorrect number according to the number of the subject.

(1) The baby on the photos

Agreement errors are more likely to be made if the local noun separates the subject head noun and the verb from each other. Therefore, the number of this local noun must be different from the number of the subject head noun. The occurrence of agreement errors appears particularly in the situation in which the subject head noun is singular and the local noun is plural (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Hartsuiker, Antón-Méndez, & van Zee, 2001). Thus, participants are more likely to make agreement errors in a sentence such as (1), where a mismatch is found between the number of the subject head noun (singular) and the number of the local noun (plural), than in a sentence such as (2), where the number of the subject head noun matches the number of the local noun (both singular). This phenomenon is called *attraction* and it is suggested that for the specification of the verb number a competition is going on between the number of the subject head noun and the number of the local noun (Bock & Miller, 1991).

(2) The baby on the photo

Additionally to syntactic features in the sentence, subject-verb number agreement is receptive for semantic features of the referents in the sentence as well (Thornton & MacDonald, 2003). The conceptual number of the subject is for example a feature that could influence the subject-verb number agreement process (e.g. Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Butterworth, & Garrett, 1996). The conceptual number of the subject could be singular like the grammatical number, such as in (3), or plural, such as in (4).

(3) The owner of the suitcases
(4) The collar of the coats

In (3) there is just one owner, who has multiple suitcases. In (4) there is a collar attached to each of several coats. Therefore, there are several collars. Thus, in comparison with (3), the subject in (4) has to refer to multiple collars (one of each coat) to be in line with our wold knowledge. Sentences with subjects in which the conceptual number is interpreted as plural, although the grammatical number is singular, are called *distributive*. Sentences in which the subject has a distributive interpretation such as (4) are found to elicit more agreement errors than sentences with a non-distributive interpretation such as (3) (Hartsuiker & Barkhuysen, 2006; Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Butterworth, & Garrett, 1996).

To sum up, the found attraction and distributivity effects show that a syntactic process such as subject-verb number agreement relies both on the syntactic information given by grammatical number of the subject and other nouns, and on the sematic information given by the conceptual number of the subject. Consequently, speakers use superfluous information that could lead to an inaccurate production of subject-verb number agreement.

1.2.3 Subject-verb number agreement and working memory

Non-linguistic cognitive functions like working memory, executive functions and attention are related with language functions. Language functions become active in the working memory system, if the executive control functions bring it to attention (Sallis, Kelly, & Code, 2015). There is consensus in the literature (Bock, 1982; Levelt, 1989) that the conceptualisation process in the language production process demands working memory capacity. Therefore, this process is relatively non-automatic. The views about the question whether working memory is required in the later stages of the language production process, however, are divided. Levelt (1989) for instance, stated that the syntactic and lexical processes are more automatic. Although, recently it is suggested that those processes are not fully automatic (Hartsuiker & Moors, 2016) and thus demands working memory.

The automatic view of Levelt asserts that the formulation level, which includes the syntactic planning process of subject-verb number agreement, is a largely automatic process. As it is an automatic process, no other cognitive functions are necessary to complete the process. Consequentially, working memory is also not involved in syntactic planning within this automatically view. Additional indications for this view were given by Bock and Cutting (1992), who did not found an obvious correlation between working memory and subject-verb number agreement.

On the other hand, more recent studies have provided evidence that subject-verb number agreement needs working memory. The resource-constrained hypothesis (Fayol, Largy, & Lemaire, 1994) suggest that verbal working memory limitations affect the ability to construct subject-verb number agreement. It makes sense to investigate the role of working memory in sentence production by testing subject-verb number agreement, because this demands a dependency across a long distance between the subject and the verb. Previous studies using the paradigm of Bock and Miller (1991) in combination with pathological or experimental driven working memory limitations have given evidence for the resource-constrained hypothesis.

Hartsuiker, Kolk and Huijnck (1999) demonstrated that their healthy elderly participants who joined the control group showed both the attraction effect (grammatical number affected the subject-verb number agreement) and the distributivity effect (conceptual number affected the subject-verb number agreement), though their participants with Broca's aphasia showed only the attraction effect. No effect of conceptual number was found in the production of agreement errors in the Broca's aphasia participants. The researchers argued that the participants with Broca's aphasia suffered from an impairment in verbal working memory capacity. Therefore, those participants could not take into account both the grammatical and conceptual information. Instead, they used the grammatical information only in the syntactic process of subject-verb number agreement. These results demonstrated that a severe capacity limitation (by a pathological cause such as aphasia) could change the interaction between the grammatical and conceptual information and makes it impossible to maintain the conceptual number information. It is reasonable to assume that reduction of working memory capacity plays a role in the processes of attraction and distributivity in subject-verb number agreement.

The studies of Fayol, Largy, and Lemaire (1994) and Hartsuiker and Barkhuysen (2006) with healthy young adults as participants demonstrated that the production of subject-verb number agreement is affected by the addition of an extrinsic working memory load. The number of agreement errors increased if there was a memory load condition added. However, Hartsuiker and Barkhuysen found no interaction effect between distributivity and memory load, which would suggest that working memory capacity plays no role in the distributivity effect. The study of Lorimor, Jackson, and van Hell (2019) investigated whether working memory capacity (measured by an OSpan task, in which math equations have to be made and simultaneous words have to be remembered) influenced the subject-verb number agreement process, especially the distributivity effect and a morphophonological effect (if the determiner of the subject was ambiguous for number (*de*) or not ambiguous for number (*het*)). No main effect of working memory capacity was found on production of agreement errors. However, it can be deduced from one of their findings that in the condition in which the subject noun had the ambiguous determiner *de*, participants with higher working memory scores made fewer agreement errors, especially in the non-distributive items. These findings could support that working memory capacity plays a role in subject-verb agreement, and affects the distributivity effect. Allen et al. (2015) found a way to test the effect of a (temporary) reduced working memory system on language processes, including the syntactic process of subject-verb number agreement. They did this by creating a situation of acute hypoglycemia, a low blood sugar, because this affects the cognitive domain of memory. The researchers found that during hypoglycemia fewer correct responses and more miscellaneous responses were given, although the amount of agreement errors was not significantly increased. This study gave partial evidence for the resource-constrained hypothesis.

To sum up, recent studies brought up evidence that working memory capacity is involved in the production of subject-verb number agreement, and that it is therefore a resource-constrained process. Also indications for the role of working memory on the distributivity effect are given, although no consensus is found for this. Following the resource-constrained view, as working memory capacity limitations develop, consequences are expected to occur for syntactic processes and the sub process of our interest: the subject-verb number agreement production process.

1.3 Aging, syntactic processes and working memory

As cognitive functions like memory, attention, and executive functions, which are involved in language production processes, decline with age, this could have consequences for the language processes. Indeed, research about the effects of aging on working memory and syntactic processes have found changes in those functions. However, not all studies about the effects of aging on working memory and syntactic processes have found aging effects.

1.3.1 Syntactic complexity and aging

Kemper, Greiner, Marquis, Prenovost, and Mitzner (2001) investigated language decline across life span and demonstrated aging effects on the use of grammatical complexity. They used the data of the longitudinal study about aging processes, called *the nun study*, which David Snowdon begun with American Roman Catholic sisters in 1986. Kemper et al. (2001) used already existing language production samples from the nuns' autobiographies, written by the nuns over different time periods, starting when the nuns had a mean age of 22 years and ending when the same nuns had a mean age of 83 years. The last ten sentences of each sample were code for grammatical complexity using the index D-level, which is based on a scale originally developed by Rosenberg and Abbeduto (1987) and modified by Cheung and Kemper (1992). On this scale, grammatical complexity ranges from a simple one-clause sentence to complex sentences. Moreover, it displays sentence relations. The complex sentences enfold, for example, different forms of embedding and subordination. By comparing the grammatical complexity of the samples written over a whole life span, the study of Kemper et al. (2001) showed that grammatical complexity of written language gradually declined across life span in the nuns.

Additionally to this longitudinally study, group comparison studies have been conducted. Kemper and Sumner (2001) investigated the structure of verbal abilities in young and elderly adults. Among other things they found an association between their index D-level and measures of working memory. Each complete sentence, elicited in an oral language production sample, was analysed for D-level. For working memory, the reading span test (read aloud unconnected sentences and remember the final word of each sentence), the digit span forward and backward tests were used. Results of Kemper and Sumner showed that young adults scored higher on the D-level, the reading span test, the digit span forward and digit span backward tests in comparison with the elderly people. Furthermore, the researchers showed that the D-level was related to the scores of the working memory measures in both groups. The capacity of the working memory demands how much information (in this case numbers and words) could be given. Working memory also influenced the limitations of the total of sentence relations given. The more complex a sentence is, the more relations it has in a form of embedded or subordinated clauses. This increases the load on the working memory. The study of Kemper and Sumner reveals that working memory plays a role in producing syntactic complex structures and that this production declines with age.

Corresponding results were found in the study of Kemper, Herman, and Lian (2003), in which young and elderly participants had to produce a sentence with given words (2, 3, or 4 words) that appears on a computer screen. The words disappeared when the participant began to speak. When given four words, elderly people produced shorter and less complex sentences in comparison with young adults. The elderly people also made more errors (non-fluent responses and memory errors) in the four words given condition. When given two or three words, the produced sentences of the elderly were comparable in length and grammatical complexity with those of the young adults. As the words disappeared when participants began to speak, this task also contained an effect of memory load. Elderly people, in contrast with young adults, were affected by this effect of memory load on their sentence production. Kemper et al. (2003) found a decline in the production of complex syntactic structures in elderly and thereby an effect of memory load.

In spite of the findings that with aging the production of complex syntactic structures declines, a recent study using structural priming (Hardy, Messenger, & Maylor, 2017) suggested that syntactic representations and processes do not suffer from aging. In the study of Hardy et al. (2017) young adults (18-23 years) and elderly people (69-80 years) described transitive verb pictures after they have heard an active or a passive sentence. The results showed that both groups produced more passive sentences after hearing a passive prime than after hearing an active prime. This is an indication for the occurrence of syntactic priming. When the priming sentence and the target picture were using the same verb, creating lexical overlap, the priming effect increased. The syntactic priming effect nor the lexical overlap effect differed significantly between young adults and elderly people. The researchers suggested therefore that aging does not affect syntactic representations underlying sentence production.

The different results in the production of complex syntactic structures in elderly people could be explained by the study paradigm. A syntactic priming task provides a cue that could support and facilitate the production of a sentence with the same syntactic structure. The independent production of complex syntactic structures is more challenging, as this is asking more cognitive capacity of the working memory system, which is likely to be more limited in elderly people.

Hardy et al. explained their contrasting findings by suggesting that the underlying syntactic representations do not change with age (no difference in the priming effect between young adults and elderly people), but gaining access to complex syntactic structures declines with age (less complex syntactic structures in elderly people). A comparison with the aphasia literature could be made, as this provides also evidence for the effect of syntactic priming in the maintenance of syntactic complex structures: Despite severe language production impairments, people with Broca's aphasia were able to use passive sentences after hearing a passive prime (Hartsuiker & Kolk, 1998). This indicates rather the loss of an easy access to syntactic representation than a complete loss of the representations.

To sum up the above described studies, an aging effect is found on working memory capacity and the independent production of syntactic structures: Elderly people use less complex syntactic structures and have a lower working memory capacity than young adults. Hereby, it is suggested that becoming older is related to a decline in working memory capacity. Moreover, it is argued that this effect could influence the access to syntactic complex structures, although it does not affect the underlying syntactic representations. These findings occurred by investigating the syntactic process of producing complexity in syntactic structures. As the subject-verb number agreement process is a part of syntactic processing too, it is interesting to explore whether aging effects occur in this particular process. Combining studies about this process gives indications for an age effect.

1.3.2 Subject-verb number agreement and aging

In addition to the research about syntactic complexity, suggestions are given that there may also be differences between young adults and elderly people in the syntactic process of subject-verb number agreement production. Those suggestions arise by looking at studies about this syntactic process (Hartsuiker & Barkhuysen, 2006; Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996) that use the paradigm of Bock and Miller (1991) and have materials derived from the same experimental items.

First, the elderly control group tested by Hartsuiker et al. (1999) made relatively more agreement errors (17%) than the young adults did (around the 7%) in the studies of Hartsuiker and Barkhuysen (2006), and Vigliocco et al. (1996). Furthermore, differences were found on the distributivity effect. The young participants of the studies of Hartsuiker and Barkhuysen, and Vigliocco et al. showed an effect of distributivity. Around three quarters of the agreement errors were made in the distributive condition in the study of Hartsuiker and Barkhuysen. Vigliocco et al. found that around 90% of the agreement errors were made in the distributive condition. The elderly control participants in the study of Hartsuiker et al. showed an effect of distributivity too. However, compared to the relative division of the agreement errors in the distributive and non-distributive conditions, these elderly control participants showed a difference with the young participants in the other studies. The elderly people made only two-thirds of all the agreement errors in the distributive condition, which is less than young adults did.

In comparing elderly people with young adults, these studies showed divergent reactions on the subject-verb number agreement production paradigm of Bock and Miller, with the same experimental material: Elderly people produced more agreement errors, and the effect of distributivity was lower in elderly people. However, the two age groups in those studies cannot directly be compared to each other, because it were separate investigations. Nevertheless, these results are very interesting and could be an indication of existing differences in the subject-verb number agreement production process between young adults and elderly people. Moreover, this could indicate an effect of aging in the subject-verb number agreement production process. In the current study, we investigated therefore whether young adults and elderly people differ in the subject-verb number agreement production process. To the best of our knowledge, no published studies exist focussing on the effects of aging on the subject-verb agreement production process measured by the paradigm of Bock and Miller.

1.4 Research questions and hypotheses

As described above, elderly people show a decline in the production of complex syntactic structures in comparison to young adults. Moreover, indications exist that elderly people differ from young adults on the subject-verb agreement production process, although the two age groups are not compared to each other in one investigation. In this current research, we wanted to know whether aging has an effect on the syntactic language production process of subject-verb number agreement. This leads to the following research question:

Is there a difference between healthy young adults (aged between 18 and 25 years) and healthy elderly people (aged above 70 years) in the subject-verb number agreement production process, measured by the paradigm of Bock and Miller (1991)?

Sub questions for this main research question were:

1) *Is there an effect of age (young adults compared to elderly people) on the amount of produced agreement errors?*
2) *Is there an effect of attraction on the amount of produced agreement errors?*
3) *Is there an effect of distributivity on the amount of produced agreement errors?*
4) *Are there interaction effects between the age, attraction, and distributivity effects on the amount of produced agreement errors?*

We expected an age effect in the way that elderly people were overall more likely to produce subject-verb number agreement errors than young adults (*Hypothesis 1*). We predicted an attraction effect for both groups (*Hypothesis 2*). We also predicted that there was an effect of distributivity in both groups (*Hypothesis 3)*, but that this effect was lower in elderly people than in young adults. We expected elderly people to make less agreement errors in the distributive condition than young adults (*Hypothesis 4)*.

If we find, as predicted, this effect of aging on the subject-verb number agreement production process and the extend of the influence of the distributivity effect, the question arises: what lies beyond this aging effect? Could this aging effect be linked to working memory capacity since previous research suggested that working memory capacity could influence syntactic processes? To investigate this, we added the additional sub questions:

5) *Is there a difference between healthy young adults (aged between 18 and 25 years) and healthy elderly people (aged above 70 years) on working memory capacity, measured by the digit span tests?*
6) *Is there an effect of working memory score, measured by the digit span backward test, on the amount of produced agreement errors?*
7) *Are there interaction effects between the age, distributivity, and working memory capacity effects on the amount of produced agreement errors?*

We expected elderly people to have lower working memory scores than young adults (*Hypothesis 5*). We predicted an effect of working memory capacity and expected to find more subject-verb number agreement errors made by people with a low digit span score and less subject-verb number agreement errors made by people with a high digit span score (*Hypothesis 6*). Because we hypothesized that aging affected working memory, we predicted that elderly people had lower digit span scores and produced more agreement errors in comparison to young adults (*Hypothesis 7*).

2. Methods

2.1 Participants

Participants for this study were recruited from our social networks. Eventually, we recruited 69 young adults aged between 18 and 25 years, and 70 elderly people aged above 70 years. Conform to our exclusion criteria, the data of 63 young adults and 59 elderly people were used in the current study. All included participants were healthy and did not report speech or language disorders, nor neurological or psychological problems. Additionally, the participants did not have any cognitive problems, which was measured by the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). All participants were native Dutch speakers and lived in the surroundings of Roosendaal or Nijmegen. The participants joined the experiment on a voluntary basis and signed an informed consent form before the start of the tests. Table 1 shows the group characteristics and background information of the participants, divided into the two age groups. There was no significant difference in gender and the highest completed level of education between the two age groups. We also compared the groups on reading behaviour, which we defined as reading out of amusement, for instance voluntarily reading a book or the newspaper. Reading behaviour was significant higher for elderly people group than in for young adults.

The group of young adults consisted of 22 men and 41 women; The mean age was 21.27 years (SD=1.89), with a range of 18 to 25 years; The highest completed level of education varied from low education (MULO) to high education (WO), level 5 until level 7 on the seven point scale of Verhagen (1964; see also Hendriks, Kessels, Gorissen, Schmand, & Duits, 2014) with a mean level of 6.00 (SD=0.72), where 25.4% of the young adults had an education level of 5 (MULO), 49.2% had an education level of 6 (VHMO) and 25.4% had an education level of 7 (WO). Seven of the young adults were left handed, the others were right handed. The reading behaviour scores of the young adults had a mean of 3.11 (SD=1.48) on a 5 point scale (1=never, 2=once a month, 3=once a week, 4=several times a week, 5=daily).

The group of elderly people consisted of 21 men and 38 women; The mean age was 75.75 years (SD=5.44), with a range of 70 to 93 years; The highest completed level of education varied from very low education (a not competed secondary education) to high education (WO), level 3 until level 7 on the seven point scale of Verhagen, with a mean level of 5.76 (SD=0.82), where 1.7% of the elderly people had an education level of 3 (a not completed secondary education), 3.4% had an education level of 4 (an education level lower than MULO), 27.1% had an education level of 5 (MULO), 52.5% had an education level of 6 (VHMO) and 15.3% had an education level of 7 (WO). Eight of the elderly people were left handed, the others were right handed. The reading behaviour scores of the elderly people group had a mean of 4.95 (SD=0.22) on the 5 point scale.

Table 1. Mean (M) and standard deviation (SD) of the characteristics and background information of the group of young adults and the group of elderly people, and the results of the comparison between the two age groups (Independent samples t-test).

	Young (N=63, 22 men)		**Old** (N=59, 21 men)		**Comparison**	
	M	**SD**	**M**	**SD**	**t**(120)	***p***
Age (years)	21.27	1.89	75.75	5.44	x	x
Education level[a]	6.00	0.72	5.76	0.82	-1.71	0.09
Reading behaviour[b]	3.11	1.48	4.95	0.22	9.42	<0.001***

[a]Education level was measured on the 7 point scale of Verhagen (1964)

[b]Reading behaviour was measured with a self-assessment question on a 5 point scale (I read e.g. a book or the newspaper out of amusement …. Never(1), once a month(2), once a week(3), several times a week(4), daily(5))

*** Difference is significant on the 0.001 alfa level

2.2 Materials

This study consisted of three different tasks: A sentence completion task following the paradigm of Bock and Miller (1991), which investigated subject-verb number agreement, a forward and backward digit span task, which measured working memory capacity, and the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), which examined cognitive function.

2.2.1 Sentence completion task

In the sentence completion task, following the paradigm of Bock and Miller (1991), participants had to listen to short sentence fragments. Then, they had to repeat those fragments and complete them with a verb (the correct version of the Dutch verb *to be* for that fragment) and a given adjective. In this way a grammatically correct sentence should be made. The materials used for the sentence completion task were the same as the experimental items of Vigliocco, Hartsuiker, Jarema, and Kolk (1996), Hartsuiker, Kolk, and Huinck (1999), and Hartsuiker and Barkhuysen (2006). The sentence completion task existed of experimental and filler items. Table 2 gives an overview of all the conditions of the experimental items and the filler items, and containing examples for each condition. All experimental items are included in Appendix A and Appendix B provides information about the filler items.

Experimental items

We used 24 experimental sentence fragment items consisting of a subject noun phrase with an embedded prepositional phrase. The experimental sentence fragment items had a singular subject head noun in the noun phrase, causing the participants to complete the fragment with the singular form of the verb to be: *is* (*is* in Dutch). The number of the local noun in the prepositional phrase varied. Therefore, each experimental sentence fragment item had two versions, creating the *matching* variable, which was divided into the *match* and *mismatch* condition. The match condition was the version with a singular local noun, in which the number of the subject head noun and the local noun were the same such as (1). The mismatch condition was the version with a plural local noun, in which the number of the subject head noun and the local noun differed such as (2). The examples (1) and (2) belong to the same experimental sentence fragment item. Each participant received only one of the two versions of these experimental sentence fragment items.

(1) The cup for the winner (*match* condition)
(2) The cup for the winners (*mismatch* condition)

Beside the matching variable, the *distributivity* variable was included. Half of the 24 experimental items had a distributive reading in the mismatch condition, in which the local noun was plural. In these cases the subject head noun was grammatically singular, but had a plural reading, such as (3). Therefore, it had a distributive interpretation (there is a back on each of several chairs and therefore there are several backs). This is the *multiple token* condition. These 12 multiple token items had the distributive reading exclusively in the mismatch condition. In the match condition they just had a singular reading. The subject head noun of the other 12 items always had a singular reading, such as (4). These items belong to the *single token* condition. For the distributivity variable, only the multiple token items in the mismatch condition had a discrepancy between grammatical and conceptual number of the subject head noun.

(3) The back of the chairs (*multiple token* condition)
(4) The director of the films (*single token* condition)

Only non-neuter nouns were used for the subject head and local noun. All those nouns needed the determiner *de* in Dutch and thus were unmarked for number. As a result, all the nouns in the sentence fragments, singular or plural, had the same determiner. Each sentence fragment consisted of five words. The number of syllables of the sentence fragment items did not differ significantly between the match condition (M=6.79, SD=1.10) and the mismatch condition (M=7.33, SD=1.05), t(46)=1.74, *p*=0.09. Although the number of syllables of the sentence fragment items did differ significantly between the single token condition (M=7.5, SD=0.89) and the multiple token condition (M=6.63, SD=1.14), t(46)=-2.98, *p*=0.05, the number of syllables of the sentence fragment items in the mismatch condition did not differ significantly between the single token condition (M=7.67, SD=0.99) and the multiple token condition (M=7.00, SD=1.04), t(22)=1.61, *p*=0.12.

Table 2. Overview of the conditions and the amount of items of the experimental and filler items used in the sentence completion task. An example is given for each condition together with the corresponding adjective. The original Dutch items are translated into English.

Condition	Number of items	Sentence fragment e.g.	Adjective
Experimental items	**24**		
Single Token	12		
Match	6	De diefstal van de diamant	SUCCESVOL
		The theft of the diamond	*SUCCESSFUL*
Mismatch	6	De diefstal van de diamanten	SUCCESVOL
		The theft of the diamonds	*SUCCESSFUL*
Multiple Token	12		
Match	6	De sleutel van de kast	KLEIN
		The key of the cupboard	*SMALL*
Mismatch	6	De sleutel van de kasten	KLEIN
		The key of the cupboards	*SMALL*
Filler items	**36**		
Simple NP singular	6	De eikenhouten tafel	ZWAAR
		The oak table	*HEAVY*
Simple NP plural	6	De mislukte grappen	VERVELEND
		The unsuccessful jokes	*ANNOYING*
Plural head noun + singular local noun	12	De appels in de mand	ROT
		The apples in the basket	*ROTTEN*
Plural head noun + plural local noun	12	De geluiden uit de klassen	HARD
		The sounds from the classes	*LOUD*

Fillers

In addition to the experimental items, 36 filler sentence fragment items were included. Of these fillers, 24 had the same syntactic structure as the experimental items, but these fillers had a plural instead of a singular subject head noun. As a result, participants needed the plural form of to be, *are* (*zijn* in Dutch), to complete these filler sentence fragment items. Furthermore, half of these 24 filler items had a singular local noun, the other half had a plural local noun. The remaining 12 filler sentence fragment items were simple noun phrases, which consist of a determiner, one or more adjectives and a noun. Half of these simple noun phrases were singular, the other half those phrases were plural. Here, we also used non-neuter nouns for the subject head and local noun. As a result of the inclusion and distribution of the fillers, the complete item list was balanced for number of subject head and local noun. So, half of the time participants had to use the singular form of the verb to be, *is,* and the other half of the time they had to use the plural form of the verb to be, *zijn*.

Adjectives
An adjective was coupled to each experimental and filler sentence fragment item. These adjectives were selected by Hartsuiker et al. (1999) and also used by Hartsuiker and Barkhuysen (2006). The participants were instructed to use that given adjective to complete the sentence. In this way, they only had to give the correct form of the verb by themselves, which restricted the variation in the answers that could be given. Therefore, variation in complexity of conceptualisation and duration in the sentence production process was limited. The adjectives gave no information about the number of the noun. In the match and mismatch condition the same adjectives were used for each sentence fragment item. The adjectives in the single and multiple token condition were dissimilar, but did not differ significantly in length of syllables and letters, frequency (Keuleers, Brysbaert, & New, 2010; Uit den Boogaart, 1975), concreteness and age of acquisition (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014; van Loon-Vervoorn, 1985).

Item-lists
We created two lists, which consisted of 60 items each. Each list contained 24 different experimental sentence fragment items (six single token match items, six single token mismatch items, six multiple token match items, and six multiple token mismatch items) and the 36 filler items. In each list, the items were organised in a pseudo-random order, with the constraint that the list started with four fillers and that no more than two experimental items followed each other. The items had the same order in the two lists, but the match and mismatch conditions of the experimental items changed over the lists. Thus, the items in the mismatch condition in list 1 were the items in the match condition in list 2 and vice-versa. As a result, the experimental items occurred once in the match condition and once in the mismatch condition. Appendix B shows the two lists with the distribution of all the items. The lists were divided over the participants in the way that every list appeared equally in both age groups.

Recordings
All items were recorded in a sound proof room by a female native speaker of Dutch. The speaker, who was one of the researchers, was instructed to speak as normal as possible. In each recording, the adjective was named first followed by a natural pause and after that the sentence fragment was named. The pause lengths between the adjective and the sentence fragment did not differ significantly from each other per condition. Measurements with phonetic software (Praat, version 6.0.50; Boersma, Paul, & Weenink, 2019) showed that the mean speech rate of the experimental sentence fragments was 336 ms per syllable (SD=40.89) in the single token condition and 350 ms per syllable (SD=54.15) in the multiple token condition, and did not differ significantly from each other ($t(46)=0.97$, $p=0.34$). The mean speech rate of the experimental sentence fragments in the match condition (M=355 ms per syllable, SD= 41.51) and mismatch condition (M=331 ms per syllable, SD= 51.85) did not differ significantly from each other, $t(46)=1.73$, $p=0.09$. A pilot containing two people of the intended groups was taken to ensure all the recordings were clear.

2.2.2 Digit span tests
The digit span test measures the cognitive function working memory (Baddeley & Hitch, 1974) and is used in different psychological test batteries (Ramsay & Reynolds, 1995), for example the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008). We used the forward as well as the backward digit span test. At first, the forward digit span test was orally performed. During the test, the participants had to repeat the numbers in the same order as the test leader said them. If the participant repeated the numbers correctly, a new sequence of numbers was given with the addition of an extra number. This pattern continued until the moment the participant could

not repeat the correct order of numbers anymore. When participants made a mistake, they had one more chance to try repeating another number set of the same length. If the participant repeated that correctly, the test continued as before. If the participant failed again, the test ended and the number of each sequence correctly repeated was counted. The same protocol was used for the backward digit span test. In this version, however, the participants were instructed to repeat the given numbers in the reverse order.

2.2.3 Montreal Cognitive Assessment

The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) is designed as a short screening instrument for detecting mild cognitive impairment. The MoCA is available in 27 languages and freely available at http://www.mocatest.org/. In this study we used the version of the MoCA translated into Dutch by Dautzenberg and de Jonghe in 2005. The MoCA assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuo-constructive skills, conceptual thinking, calculating and orientation. The test takes about 10 minutes. The maximum number of points is 30. Originally, a score of 26 points or higher is considered as a normal cognitive functioning. However, it could be argued that this cut-off point is too high for the population of our research.

Nasreddine et al. (2005) used a cut-off score of 26, and found a sensitivity of 87%. Different results came from an applicability and validity study of the Dutch version of the MoCA. Thissen, van Bergen, de Jonghe, Kessels, and Dautzenberg (2010) found that the Dutch version of the MoCA could distinguish between healthy elderly people, MCI patients and dementia patients. Nevertheless, an insufficient sensitivity (88%) and poor specificity (43%) for the cut-off score of 26 came was found in their study. The researchers found a mean score of 26.4 (SD=2.2) for healthy people at an age of 75.8 years. The standard deviation of 2.2. in this case is big in our opinion. These findings suggest that a cut-off score of 26 could be too high for our intended age groups. To compare cognitive function scores of healthy people and people with dementia, Thissen et al. found a cut-off score of 23 as an optimal cut-off point, because at that point the sensitivity (94%) and specificity (90%) were both balanced and good. As a result of these findings, we used a cut-off score of 23 or higher. We did this, in order to avoid including people with serious cognitive problems, and to avoid excluding people who were not likely to have serious cognitive problems, although their scores were not that high. Participants who scored below this point were excluded. The Dutch version of the MoCA was carried out manually and scores were calculated afterwards by the test leader.

2.3 Design of study

Participants of this research were divided into two groups by the independent between-subject variable *age* (young adults and elderly people). Different tasks were used in this research. First, the sentence completion task was taken. In this task the syntactic and semantical processes belonging to subject-verb number agreement in sentence production were measured. In the sentence completion task, the within-subject variable *matching* divided the experimental items in two conditions (match or mismatch). Furthermore, the within-subject variable *distributivity* divided the experimental items also in two conditions (single token or multiple token). The outcome variable in the sentence completion task was the categorical variable agreement error or not (*accuracy*) and was divided into the options correct response, agreement error, or miscellaneous error. The second assignment was the digit span task, the forward version as well as the backward version. These tests appealed to working memory. The more sequences could be repeated correctly, the better the function of the working memory should be. In the forward and backward digit span tasks the outcome variable was the amount of number sequences that were repeated correctly. The third task was the MoCA, a screening form for cognitive impairment. The outcome variable in the MoCA was the total achieved score. After these tasks a general questionnaire was taken to gain more insight in the background of the participants, such as information about age, education level and potential neurological or psychological problems.

2.3.1 Procedure

We recruited participants from our social networks and via a snowball effect, in which participants or others involved asked people in their networks to join this study. The experiment was performed at the homes of the participants or at other quiet and easy accessible locations, such as the (university) library. The test leader travelled to all these places. Sessions took about 30 minutes and consisted of four parts. The first part was the sentence completion task. The second part contained the forward and backward digit span test. The third part was the MoCA and the fourth part was the general questionnaire.

Sentence completion task

The sentence completion task was taken through a computer program, PsychoPy Builder version 3.0.6 (Peirce et al., 2019). Participants sat at a table in a quiet room in front of a laptop screen from whereon the program was played. Each session consisted of two practical trials and two experimental blocks of 30 trials. Between those two blocks participants had the opportunity to take a break. The entire session was recorded on an extern audio tape. At the beginning of the experiment participants got the instructions. They also had the opportunity to ask questions about the task. After that, the experiment began. Each trial started with the participant pressing a button. After 1000 milliseconds a warning sound of 500 ms was played. Then, after a silence interval of 500 ms, a recorded sentence fragment item containing an adjective and sentence fragment was played. The playing of the item and a fixation cross appeared simultaneously on the screen. After the item was played, the fixation cross still remained on the screen. When the audio was not playing anymore, the participants were instructed to repeat the sentence fragment and complete the sentence with a form of *to be* (*zijn* in Dutch) and the given adjective. There was no time limit given, but if the test leader had the impression that a participant was too slow in an amount of trials, she encouraged the participant to answer more quickly. If the participants asked for a repetition of a trial, one extra chance was given per trial.

Digit span tests

Participants sat at a table and the test leader sat near them. The test leader read series of numbers in a clear and loud tone, with a second of silence between each number. The participant was instructed to repeat the series of numbers in the same order for the forward digit span test and in the reversed order for the backward digit span test. If the participant performed the repetition of the numbers correctly, a new series was given with one more number than the previous one. If the participant did not perform the repetition of the numbers correctly, one other change was given with the same amount of numbers. If the participant succeeded in repeating this new number sequence, the test continued as normal. If the participant gave a wrong number again, then the test was finished.

MoCA

The participants were seated at a table and the test leader was sitting close by. The MoCA was taken according the administration instructions of that assessment (Nasreddine et al., 2005).

2.4 Analysis

2.4.1 Data treatment

For every participant, the recorded experimental items of the sentence completion task were transcribed and the responses were classified within the following coding categories. Appendix C provides an overview with examples of the different coding categories. The first answer was coded. If participants corrected themselves, this correction was additional marked and coded but this second coding was not valid for the eventual data analysis.

A *0* was given if the participant had given a *correct response*. This meant that the participants repeated the beginning of the given sentence correctly, produced the correct inflected verb form appropriate to the subject and completed the sentences with an adjective. We allowed the use of a different adjective or another kind of finish of the sentence after the verb, as long as the sentence remain syntactically the same. A correct inflected verb other than a form of *to be* was also allowed, as well as the past tense of the correct form of *to be.* The sentence had to be spoken fluently without any hesitation.

A *1* was given if the participants made an *agreement error*. This code was given if the response met all the criteria of the correct response, with one exception: the number of the verb form was wrongly inflected and did not match the number of the subject. This error has priority over all the others. For example, even if there was a non-fluency in the sentence beside an incorrect inflection, the response still got the code for an agreement error.

A *2* was given if the participant repeated the number of the subject noun wrong. This *number subject noun repetition error* was given if the participants repeated the number of the subject incorrectly, but produced a correct corresponding form of the verb with that incorrectly repeated subject. A *22* was given if the participants made a *number local noun repetition error*, which included that the participants repeated the number of the local noun wrong.

A *3* was given if the participants made a *repetition plus agreement error*. This code was given if the participant incorrectly repeated the number of subject noun and if the produced number of the verb did not match that incorrectly repeated number of the subject noun.

A *4* was given as a more general code for *miscellaneous* responses. Within this category, a distinction was made between different subcategories of errors: The beginning utterance had to be repeated twice or more by the test leader (A); The subject or local noun was omitted (B); An incomplete sentence was produced (C); The preposition was changed into a different preposition (D); The local noun was changed into a different local noun (E); The subject noun was changed into a different subject noun (F); In the sentence were hesitations, repetitions, interruptions, or other non-fluencies (G); Long silences in the sentence, around 300 ms (H).

Apart from this, we marked and coded the corrections made by the participants by giving a *5*. A correction was defined as the repetition of a whole word or more words in a different form than the first response. Within this category, a distinction was made between different subcategories of correction: The adjective was corrected right (Q); The subject noun was corrected right (R); The local noun was corrected right (S); The preposition was corrected right (T); An incomplete sentence was corrected into a complete sentence (U); Participants did not correct what they said, but mentioned that they had a strange feeling about it after finishing the sentence, like saying something as " this is weird, something sounds wrong, I think" (V); The verb was corrected wrong (X); The verb was corrected right (Y); The subject noun was corrected wrong (Z).

For our analysis, the codes of *correct response* (*0*), *agreement error* (*1*), and *miscellaneous response* (*99*) were used. In doing so, the categories *2*, *22*, *3*, and *4* felt under the category of *miscellaneous response*. This was done to provide getting to many categories. Nevertheless, it was useful to categorize all the different errors to get some insight in the kind of errors that were made.

2.4.2 Statistical analysis

For the sentence completion task, it was known per participant per item in what category the response could be classified. Beforehand, descriptive statistics were calculated to see the distribution of the correct responses, agreement errors and miscellaneous responses per group per conditions. With the categorical data of the sentence completion task, statistical analysis were executed to examine the hypotheses. Therefore, we used the programme RStudio version 1.2.1335 (RStudio Team, 2018). It has been argued that using ANOVA for the statistical analysis of categorical outcome data causes serious problems and could lead to false results (Jaeger, 2008). A solution for the problems which arise by using ANOVA with categorical data, is using a logit mixed effect analysis. An important benefit from logit mixed effect models is that they take into account that items and participants are sampled from a potential infinite population. Thus, the model includes the effects of random factors such as subjects and items.

In short, a logit mixed effect model tries to find the best fit of the categorical data to a model. It considers several different predictor variables, which are defined on forehand, and give them a certain weight. Distinctive combinations of the predictor variables are made and for these different models it is estimated what the likelihood of each model is compared to the fundamental model. Comparisons between all the different models tell whether there are significantly differences between the fits of the models. Doing this leads to a final model that has the most optimal fit and the least number of predictor variables (Baayen, Davidson, & Bates, 2008). Another option of finding the best fitting-model is to make the model as complete as possible first, including all fixed factors as predictor, intercepts for random effects, and random slopes of the fixed effects. A reversion of the model is necessary if the model does not converge, what is done in a stepwise reduction (Barr, Levy, Scheepers, & Tilt, 2013). This latest option was used for our statistical analysis.

Our categorical data were submitted to a logit mixed effects analysis by using the lme4 software (Bates, Mächler, Bolker, & Walker, 2015). The dependent variable was the categorical outcome variable accuracy (correct response or agreement error). Fixed factors were the independent between-subject factor age (young adults or elderly people), the independent within-subject factors matching (match or mismatch) and distributivity (single token or multiple token), and their interactions. Subjects and items were included as crossed random factors. The fixed effects were also included in the random slopes.

Furthermore, correlations were measured (with SPSS version 21.0) between the dependent variable accuracy in the sentence completion task and the scores on the forward and backward digit span task, and the MoCA scores. Additionally, correlations were also measured between the dependent variable accuracy in the sentence completion task and the more general factors gender, education level, reading behaviour, list type, and list number of the items. We also measured correlation between the factors themselves. All correlations were executed for the separate age groups. Because a lot of comparisons were made, the alfa level was lowered to 0.01 to reduce the chance on a false positive effect.

3. Results

3.1 Overview results sentence completion task

3.1.1 Item analysis

First, an item analysis was carried out to make sure no item yielded a much larger percentage of correct responses, agreement errors or miscellaneous responses than the rest of the items. An item consisting of more than 50% miscellaneous responses was considered an unsuitable item, that should be removed. None of the items had to be excluded as a result of this analysis. Figure 4 shows the distribution of the response types per item in ascending order of the number of correct responses. The items are included in Appendix A, which shows the sentence fragment belonging to each item name. The correct responses in the items ranged from 54.1% till 92.6% (M=77.77, SD=10.17). The agreement errors in the items ranged from 1.6% till 29.5% (M=11.17, SD=8.55). The miscellaneous responses in the items ranged from 4.1% till 20.5% (M=11.07, SD=4.67). Thus, all the items could be included in the further analysis.

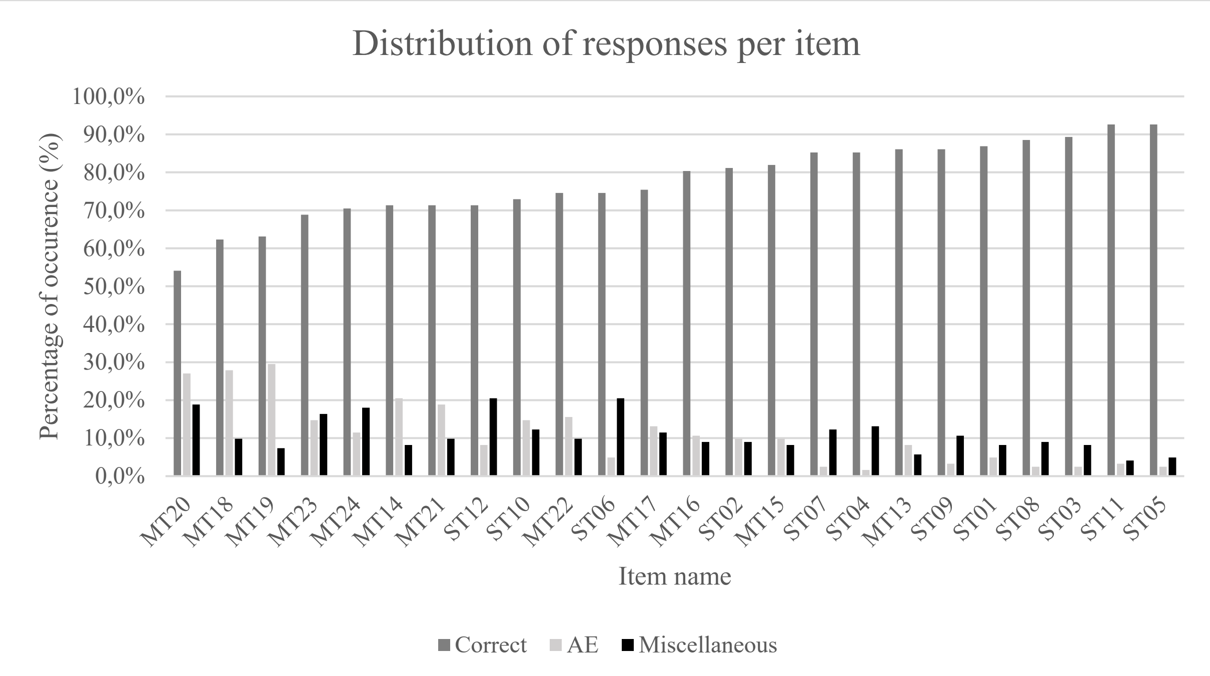


Figure 4. Bar chart of the distribution of response types (Correct Response, Agreement Error (AE), Miscellaneous Response) in percentages (%) per Multiple Token (MT) item and Single Token (ST) item.

3.1.2 Distribution of the responses

In total 2928 experimental responses were given, of which 2277 (77.8%) were correct responses, 327 (11.2%) were agreement errors, and 324 (11.0%) were miscellaneous responses (which were non fluencies in 134 (41.4%) cases). Table 3 provides an overview of the absolute and relative number of responses in the different scoring categories per group per condition. In the group of young adults 1512 experimental responses were given, of which 1249 (82.6%) were correct responses, 153 (10.1%) were agreement errors, and 110 (7.3%) were miscellaneous errors (which existed of non-fluencies in 60 (54.5%) cases). In the group of elderly people 1416 experimental responses were given, of which 1028 (72.6%) were correct responses, 174 (12.3%) were agreement errors and 214 (15.1%) were miscellaneous errors (which existed of non-fluencies in 74 (35.0%) cases).

Figure 5 shows a graph of the agreement errors as proportion of the sum of correct responses and agreement errors produced in each condition. Miscellaneous responses were not included in Figure 5, nor in the calculations of the proportions. Appendix D shows a table with the absolute and relative numbers of the responses with the correct responses and agreement errors per group per condition only.

Table 3. Distribution (in absolute numbers and percentages) of responses by scoring category (Correct Responses, Agreement Errors, Miscellaneous Responses) per age group (Young, Old) per condition (Single Token, Multiple Token, Match, Mismatch).

Condition	**Total Responses**		**Correct Responses**		**Agreement Errors**		**Miscellaneous Responses**	
	Young	Old	Young	Old	Young	Old	Young	Old
Single Token								
Match	378	354	366	309	1	5	11	40
			96.8%	87.3%	0.3%	1.4%	2.9%	11.3%
Mismatch	378	354	321	232	23	45	34	77
			84.9%	65.5%	6.1%	12.7%	9.0%	21.8%
Multiple Token								
Match	378	354	357	303	1	7	20	44
			94.4%	85.6%	0.3%	2.0%	5.29%	12.4%
Mismatch	378	354	205	184	128	117	45	53
			54.2%	52.0%	33.9%	33.0%	11.9%	15.0%
Total								
	1512	1416	1249	1028	153	174	110	214
			82.6%	72.6%	10.1%	12.3%	7.3%	15.1%
	2928		2277		327		324	
			77.8%		11.2%		11.0%	

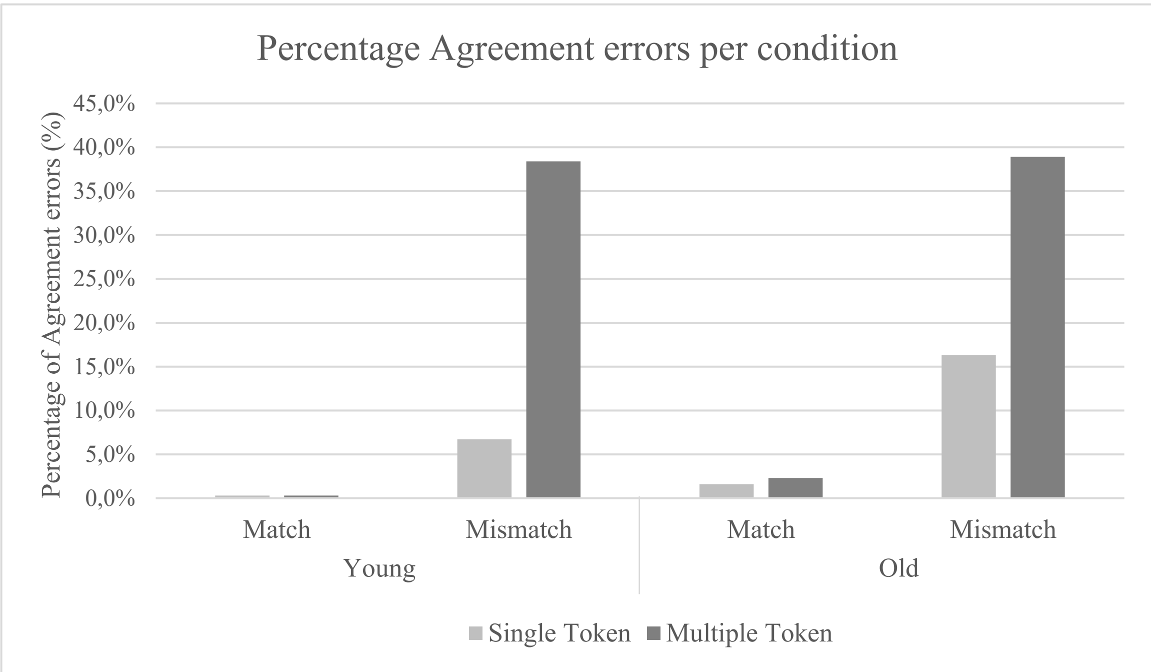


Figure 5. Bar chard of the percentage (%) produced Agreement errors per age group (Young and Old) in the matching variable (Match, Mismatch) and distributivity variable (Single Token, Multiple Token).

Table 4 shows an overview of the descriptive statistics for the young adults and the elderly people per scorings category separately. An independent samples t-test showed that young adults (M=19.83, SD=2.62) gave significantly more correct responses than elderly people (M=17.42, SD=3.84), t(101)=4.01, p<0.001. The production of agreement errors by the young adults (M=2.43, SD=2.33) did not significantly differ from the production of agreement errors by the elderly people (M=2.90, SD=3.10), t(108)=-0.95, p=0.35. Young adults (M=1.75, SD=1.47) gave significantly less miscellaneous responses than elderly people (M=3.63, SD=2.14), t(102)=-5.62, p<0.001 (Independent samples t-test).

Table 4. Descriptive statistics (Mean, Standard deviation (SD), Minimum (Min) and Maximum (Max) of the responses per scorings category (Correct Responses, Agreement Errors, Miscellaneous Responses) per age group (Young or Old), and the results of the comparison between the two age groups (Independent samples t-test).

Age group		**Correct Responses**	**Agreement Errors**	**Miscellaneous Responses**
Young	Mean	19.83	2.43	1.75
	SD	2.62	2.33	1.47
	Min	11	0	0
	Max	24	11	5
Old	Mean	17.42	2.90	3.63
	SD	3.84	3.10	2.14
	Min	7	0	0
	Max	24	11	9
Comparisons	t	4.01	-0.95	-5.62
	df	101	108	113
	p	<0.001***	0.35	<0.001***

*** Difference is significant on the 0.001 alfa level

3.2 Statistical analysis sentence completion task

In the first place it has to be said that the miscellaneous responses were not included in the following statistical analysis[1]. In the second place, we have to take into account that the number of agreement errors that occurred in the match condition was very low in both groups. The matching variable had a big effect, as a result of the few errors in the match condition (14 of the 2604 responses, 0,5%) in comparison with the mismatch condition (313 of the 2604 responses, 12,0%). So, there were too little observations for agreement errors in the match condition, which could possibly disturb the effects of the other variables. Moreover, the distributivity effect is only possible in the mismatch condition and not in the match condition. Consequently, the distributivity effect appears as a main effect of item type (multiple token vs. single token) in the mismatch condition only. In addition, as the attraction effect of the matching variable is clearly evident in earlier research (e.g. Bock & Miller, 1991), and other researchers did not even include the matching variable in their study (Lorimor, Jackson, & van Hell, 2019), it could be argued to drop this variable. To see whether problems arise with the inclusion of the matching variable, we rapport the statistical analysis with both the items in the match and mismatch conditions first. Second, we report the statistical analysis without the match condition. Therefore, we eliminated the items of the match condition and restricted the statistical analysis on the items in the mismatch condition only[2], just like Lorimor et al. (2019).

[1] See Appendix E for the statistic main findings of the miscellaneous responses

[2] See Appendix F for a table with absolute and relative numbers of mismatch items only

3.2.1 Sentence completion task with all conditions

The data of the sentence completion task were analysed in RStudio version 1.2.1335 (RStudio Team, 2018) using the lme4 software (Bates, Mächler, Bolker, & Walker, 2015). The R-script is available in Appendix G. Accuracy responses were coded as 0 for a correct response and 1 for an agreement error. A logit mixed effects analysis was performed on the data, as we had a categorical outcome variable and this is the most suitable way to analyse them without a separate item and participant analysis (Jaeger, 2008). The model predicted the probability of an agreement error in each condition. Accuracy (correct response vs. agreement error) was the dependent variable. The variables age group (young vs. old), distributivity (single token vs. multiple token), matching (match vs. mismatch) and their interactions were fitted into the model as fixed effects. We included intercepts for subjects and items as random effects, as well as random slopes of the fixed effects. All factors were categorical, but we transformed them into numeric values by sum-coding them to the contrast -0.5 and 0.5, which gave them a mean of 0 and a range of 1 (Davis, 2010). As the full model (Model 1) did not converge, we performed a stepwise reduction of the model to get a best-fitting model of the data, as outlined by Barr, Levy, Scheepers, and Tily (2013). We did this by eliminating the interactions in the random effects first. Second, we removed the slopes one by one. At this point the model converged. The final model (Model 2) included the main and interaction effects of the fixed factors, the random effect intercepts for subject and item, the random by-subject slope for the effects of distributivity and matching, and the random by-item slope for the effect of matching. The log-likelihood (the fit) of the final model was -560.4. A Wald Chi-Square Test was taken of the model to investigate main and interaction effects, by using the command "summery" of the R Base Package (R Core Team, 2019). The descriptions of the parameters of Model 2 are reported in table 5.

Model 1) Accuracy ~ agegroup*distributivity *matching + (1|subject) + (1|item) + (0+distributivity|subject) + (0+matching|subject) + (distributivity:matching|subject) + (0+matching|item) + (0+agegroup|item) + (matching:agegroup|item)

Model 2) Accuracy ~ agegroup*distributivity*matching + (1|subject) + (1|item) + (0+distributivity|subject) + (0+matching|subject) + (0+matching|item)

Table 5. Coefficients and probability estimates (Coefficient, Standard Error, Wald Chi-Squared Z-score, p-value) of the fixed effects in the final version of the logit mixed effect model (Model 2)

Predictor	Coefficient	Std. Error	Wald Z	*p*
(Intercept)	-46.86	0.41	-11.56	< 2e-16***
Age group	13.02	0.61	2.15	0.03*
Distributivity	14.53	0.56	2.60	0.01**
Matching	49.36	0.60	8.18	2.8e-16***
Age group:Distributivity	0.64	0.86	0.74	0.46
Age group:Matching	-13.77	10.20	-1.35	0.18
Distributivity:Matching	23.63	0.99	2.38	0.02*
Age group: Distributivity:Matching	-17.26	17.14	-1.01	0.31

* Effect is significant on the 0.05 alfa level
** Effect is significant on the 0.01 alfa level
*** Effect is significant on the 0.001 alfa level

There was a significant main effect of distributivity, whereby participants produced more agreement errors in the multiple token condition (77.4%) than in the single token condition (22.6%). There was a significant main effect of the matching variable, whereby participants produced more agreement errors in the mismatch condition (95.7%) than in the match condition (4.3%). There was a significant main effect for age group, whereby the amount of produced agreement errors was lower for young adults (10.9%) than for elderly people (14.5%). There was also a significant interaction effect between distributivity and matching, because there were significantly more agreement errors for multiple token items in the mismatch condition than for the three other conditions.

3.2.2 Sentence completion task without the match condition

For the second analysis, the same steps were taken as in 3.2.1 with the exception that in this analysis the items of the match condition were excluded. Accuracy (correct response vs. agreement error) was the dependent variable. The variables age group (young vs. old), distributivity (single token vs. multiple token) and their interactions were fitted into the model as fixed effects. We included intercepts for subjects and items as random effects, as well as a random by-subject slope for distributivity and a random by-item slope for age group. As the full model (Model 3) did not converge, we performed a stepwise reduction of the model to get a fitting model of the data. We did this by eliminating the slopes one by one. At this point the model converged. The final model (Model 4) included the main and interaction effects of the fixed factors, the random effect intercepts for subject and item, and the random by-subject slope for distributivity. The log-likelihood (the fit) of the model was -504.8. A Wald Chi-Squared Test was taken of the model to investigate main and interaction effects. The descriptions of the parameters of Model 4 are reported in table 6.

Model 3) Accuracy ~ agegroup*distributivity + (1|subject) + (1|item) + (0+token|subject) + (0+agegroup|item)

Model 4) Accuracy ~ agegroup* distributivity + (1|subject) + (1|item) + (0+distributivity|subject)

Table 6. Coefficients and probability estimates (Coefficient, Standard Error, Wald Chi-Squared Z-score, p-value) of the fixed effects in the final version of the logit mixed effect model (Model 4).

Predictor	Coefficient	Std. Error	Wald Z	*p*
(Intercept)	-2.18	0.32	-6.81	9.78e-12***
Age group	0.62	0.43	1.46	0.14
Distributivity	2.60	0.50	5.25	1.50e-07***
Age group:Distributivity	-1.49	0.46	-3.23	0.001**

** Effect is significant on the 0.01 alfa level
*** Effect is significant on the 0.001 alfa level

There was a main effect of distributivity, whereby participants produced more agreement errors in the multiple token condition (78.3%) than in the single token condition (21.7%). There was no significant main effect for age group. Thus, there was no difference in the amount of produced agreement errors between young adults (22.3%) and elderly people (28.0%). However, a significant interaction effect was found for age group and distributivity. The difference in produced agreement errors (44.5%) between the single token condition (27.8%) and multiple token condition (72.2%) for elderly people deviated significantly from the difference (69.6%) between the produced agreement errors in the single token condition (15.2%) and multiple token condition (84.8%) for young adults, in such a way that the difference was smaller for elderly people. This effect was driven by the single token condition, as in that condition showed the biggest difference between the two groups. These results indicated that the distributivity effect (more agreement errors in the multiple token condition than in the single token condition) for elderly people is smaller than the distributivity effect for young adults. As argued before, we will use these statistic results in the sequel, obtained without the match condition.

3.2.3 Sentence completion task with additional factors

We also investigated other factors (gender, education level, reading behaviour, MoCA and digit span scores, list type and number) that probably might have influenced the occurrence of agreement errors. This was done by adding these factors individually as an extra fixed main effect on Model 4. The outcomes of these added fixed factors showed no differences between this model (Model 4) and the model with the match condition (Model 2). Before running the models, the outcomes of the ordinal factors were centred at the sample mean. In this way, a zero value arose, as a result of changing the values of the scores whereby the scale remained the same. By centring the scores at the mean, the data became more interpretable. A comparison between centring the scores at the mean and standardizing the scores showed that this did not changes the outcomes. Wald Chi-Square Tests were taken of each of the supplemented models to investigate the effects of the added factors.

Results demonstrated that education level ($p<0.001$), MoCA score ($p<0.001$), digit span forward score ($p<0.05$), digit span backward score ($p<0.001$), and list number ($p<0.05$) had a significant effect, which is further reported in Appendix H. None of the extra added factors significantly changed the original outcomes of Model 4, described in Table 6. This means that the main effect of distributivity remained together with the interaction effect of age group and distributivity, even by adding those extra factors.

The factors involving working memory were of particular interest, because previous research had shown that working memory could influence the production of agreement errors (Hartsuiker & Barkhuysen, 2006). Therefore, we created a new model (Model 5) to investigate the effect of the working memory by adding the digit span backward score as a full fixed effect with interactions and random effect structures.

Model 5) accuracy ~ agegroup*distributivity*backward.digit.span + (1|subject) + (1|item) + (0+distributivity|subject) + (0+backward.digit.span |item)

Beside the already in Table 6 reported effect for distributivity and the interaction between distributivity and age group, results showed a significant effect of the digit span backward score ($p<0.001$). Participants with high digit span backward scores produced less agreement errors than participants with low scores. The interaction between distributivity and digit span backward score was marginal significant ($p=0.065$), but it is interesting to look at which direction this interaction went. The difference in producing agreement errors between the single token condition and multiple token condition was bigger for participants with low digit span scores than for participants with high digit span scores. Low digit span participants had a bigger distributivity effect, and high digit span participants had a smaller distributivity effect. A full overview of the report of the model outcomes is added in appendix I.

3.3 Correlation analysis

Beside the logit mixed effects analysis, a correlation analysis was carried out to investigate whether the occurrence of agreement errors and correct responses correlated with the included other factors (gender, education level, reading behaviour, MoCA and digit span scores, list type and list number). Table 7 shows the Pearson's Correlation Coefficient r. For each factor with an effect on the outcome variables a brief summarization is given. Appendix J shows scatter plots for the significant correlations. No correlation was found with gender (man vs. woman), list type (list1 vs. list 2) or reading behaviour (never until daily) on the amount of agreement errors and correct responses. Also correlations between the extra factors were calculated.

Table 7. Correlations (Pearsons r) with p-values for the Agreement errors and Correct responses for the total group, and the divided age groups (Young, Old) with the added factors (list, gender, list number, education level, reading activity, MoCA scores, digit span forward scores (DSFW), and digit span backward scores (DSBW)).

Correlations

		List	Gender	List nr.	Edu. level	Read. Act.	MoCA	DSFW	DSBW
Agreement Errors									
Total	r	-0.114	-0.057	-0.533	-0.425	0.005	-0.436	-0.221	-0.382
	p	0.212	0.533	0.007**	0.000**	0.959	0.000**	0.014*	0.000**
Young	r	-0.061	-0.026	-0.599	-0.5	-0.057	-0.398	-0.136	-0.334
	p	0.633	0.841	0.002**	0.000**	0.657	0.001**	0.288	0.000**
Old	r	-0.158	-0.085	-0.406	-0.361	-0.161	-0.48	-0.278	-0.334
	p	0.321	0.52	0.049*	0.005**	0.223	0.000**	0.033*	0.01**
Correct Responses									
Total	r	0.193	0.025	0.487	0.409	-0.072	0.498	0.302	0.434
	p	0.033*	0.787	0.016*	0.000**	0.433	0.000**	0.001**	0.000**
Young	r	0.167	-0.111	0.545	0.492	0.157	0.273	0.188	0.414
	p	0.19	0.388	0.006**	0.000**	0.22	0.030*	0.14	0.001**
Old	r	0.219	0.139	0.324	0.315	0.131	0.45	0.307	0.499
	p	0.096	0.293	0.123	0.015*	0.322	0.000**	0.018*	0.002**

* Correlation is significant on the 0.05 alfa level
**Correlation is significant on the 0.01 alfa level

3.3.1 List number

The results of the correlation analysis showed that there was a significant strong negative correlation between list number and the amount of agreement errors for young adults ($r=-0.60$, $p<0.01$): The lower the list number, the more agreement errors were given. Moreover, there was a significant medium negative correlation for elderly people ($r=-0.41$, $p=0.05$), although, by adding a Bonferroni correction, this correlation could no longer be called significant. In addition, the results of the correlation analysis showed that there was a significant strong positive correlation between list number and the amount of correct responses for young adults ($r=0.55$, $p<0.01$): The higher the list number, the more correct responses were given. For elderly people no significant correlation with the amount of correct responses was found. Appendix J.I shows the scatterplots of the significant correlations with list number.

3.3.2 Education level

Education level did not differ significant between the two age groups, because we controlled that both groups had the same distribution of education level (see Table 1). The results of the correlation analysis showed that education level had a significant medium negative correlation with the amount of agreement errors within both young adults ($r=-0.50$, $p<0.001$) and elderly people ($r=-0.36$, $p<0.01$): The higher the educational level, the less agreement errors were given. The results of the correlation analysis showed that education level had a significant medium positive correlation with the amount of correct responses within both young adults (r=0.49, *p*<0.001) and elderly people (r=0.32, *p*=0.02), although, by adding a Bonferroni correction, this last correlation could no longer be called significant. Appendix J.II shows the scatterplots of the significant correlations with education level.

3.3.3 MoCA

The descriptive scores of the MoCA are reported in table 8. An independent samples t-test showed that the MoCA scores of young adults (M=27.98, SD=1.93) were significantly higher than the scores of elderly people (M=26.49, SD=1.93), t(112)=4.65, *p*<0.001. The results of the correlation analysis showed that MoCA score had a significant medium negative correlation with the amount of agreement errors in both young adults (r=-0.40, *p*=0.001) and elderly people (r=-0.48, *p*<0.001): The higher the MoCA scores, the less agreement errors were given. The results of the correlation analysis showed that MoCA scores had a significant medium positive correlation with the amount of correct responses in both young adults (r=0.27, *p*<0.05) and elderly people (r=0.50, *p*<0.001), although, by adding a Bonferroni correction, the first correlation could no longer be called significant. Appendix J.III shows the scatterplots of the significant correlations with MoCA score.

Table 8. Descriptive statistics (Mean, Standard deviation (SD), Minimum (Min) and Maximum (Max) of the scores on the different test (MoCA, Digit Span Forward (DSFW), Digit Span Backward (DSBW)) per group (Young, Old), and the results of the comparison between the two age groups (Independent samples t-test).

Age group		**MoCA scores**	**DSFW scores**	**DSBW scores**
Young	Mean	27.98	6.59	4.78
	SD	1.58	1.14	1.07
	Min	24	4	2
	Max	30	9	7,00
Old	Mean	26.49	5.90	4.29
	SD	193	1.12	1.15
	Min	23	4	2
	Max	30	8	7
Comparisons	t	4.65	3.35	2.44
	df	112	120	120
	p	<0.001***	0.001***	<0.05*

*Difference is significant on the 0.05 alfa level
** Difference is significant on the 0.001 alfa level

3.3.4 Digit span forward

The descriptive scores of the digit span forward are reported in table 8. An independent samples t-test showed that the scores of the forward digit span test of young adults (M=6.59, SD=1.15) were significantly higher than the scores of elderly people (M=5.90, SD=1.13), t(120)=3.35, *p*=0.001. The results of the correlation analysis showed that there was no significant correlation between the digit span forward score and the amount of agreement errors for young adults. Still, there was a significant marginally negative correlation with the amount of agreement errors for elderly people (r=-0.28, *p*=0.03), although, by adding a Bonferroni correction, this correlation could no longer be called significant. The results of the correlation analysis showed that there was no significant correlation between the digit span forward score and the amount of correct responses for young adults. For elderly people a significant medium positive correlation with the amount of correct responses was found (r=0.31, *p*=0.02), although here too, by adding a Bonferroni correction, this correlation could no longer be called significant.

3.3.5 Digit span backward

The descriptive scores of the digit span backward are reported in table 8. An independent samples t-test showed that the scores of the backward digit span test of young adults (M=4.78, SD=1.07) were significantly higher than the scores of elderly people (M=4.29, SD=1.15), t(120)=2.44, $p<0.05$. The results of the correlation analysis showed that digit span backward scores had a significant medium negative correlation with the amount of agreement errors in both young adults (r=-0.33, $p<0.001$) and elderly people (r=-0.33, $p=0.01$): The higher the digit span backward scores, the less agreement errors were given. The results of the correlation analysis showed that digit span backward scores had a significant medium positive correlation with the amount of correct responses in both young adults (r=0.41, $p=0.001$) and elderly people (r=0.50, $p=0.002$): The higher the digit span backward scores, the more correct responses were given. Appendix J.IV shows the scatterplots of the significant correlations with the digit span backward score.

3.3.6 Mutual correlations

Beside the correlations between the added factors and both agreement errors and correct responses, correlations between the added factors were calculated. Appendix K.I shows a table with an overview of the correlations between education level, MoCA scores, digit span forward scores, and digit span backward scores.

As the effect of education level is found to be a strong predictor variable of digit span scores (Ostrosky-Solís & Lozano, 2006; Zimmermann, Cardoso, Trentini, Grassi-Oliveira, & Fonseca, 2015) we performed a correlation analysis between those two factors. The results of the correlation analysis showed that education level had a marginally significant effect on the scores of the digit span forward score for both groups together, (r=0.325, $p<0.01$), although separately, no correlation was found. On the contrary, education level had a significant medium positive correlation with the scores of the digit span backward test in both young adults (r=0.38, $p=0.002$) and elderly people (r=0.43, $p=0.001$): The higher the educational level, the higher the digit span scores. Appendix K.II shows the scatterplots of the significant correlations with education level and the digit span backward score.

The results of the correlation analysis show that the digit span forward and backward scores had a positive high correlation with each other in both young adults (r=0.43, $p=0.01$) and elderly people (r=0.65, $p<0.001$). Appendix K.III shows the scatterplots of the significant correlations with the digit span backward score and the digit span forward score.

The results of the correlation analysis show that educational level had a marginally significant effect on the MoCA scores for both groups together, (r=0.26, $p<0.01$), although separately, no correlation was found. Results of the correlation analysis showed that MoCA score had no effect on the digit span forward score. The results of the correlation analysis show that MoCA scores had a marginally significant positive correlation with the digit span backward score for both groups together, (r=0.25, $p<0.01$). However, only a marginally significant correlation was found for young adults (r=0.85, $p<0.05$), which was not significant anymore by adding a Bonferroni correction.

3.4 Explorative post hoc analysis

Our elderly people ranged from an age of 70 until 93 years, which is a broad range. To investigate whether age in the elderly people correlated with the production of agreement errors and correct responses, and working memory scores, we executed a explorative post hoc correlation analysis. Within the elderly people, a mediate negative correlation was found between age and the amount of correct responses (r=-0.30, $p=0.023$) and a mediate positive correlation was found between age and the amount of agreement errors (r=0.32, $p=0.013$). No correlations between age and digit span or MoCA scores were found.

4. Discussion

The current research is executed to gain more insight into the process of aging in healthy people by investigating the influence of aging on the syntactic process of subject-verb number agreement in language production. This was done by comparing young adults and elderly people on several tests: A sentence completion task following the paradigm of Bock and Miller (1991) to measure the subject-verb number agreement process; The digit span forward and backward tasks to measure working memory capacity; The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) to measure cognitive functioning. With those tasks we aimed to be able to gain further insight about the role of aging and working memory capacity in the subject-verb number agreement production process.

Based on a combination of existing research literature, we expected elderly people to produce more agreement errors than young adults in the subject-verb number agreement task and that elderly people showed a lower distributivity effect (Hartsuiker & Barkhuysen, 2006; Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Butterworth, & Garrett, 1996). Thereby, we expected that elderly people had a reduced working memory capacity in comparison to young adults (Kirova, Bays, & Lagalwar, 2015) and that working memory capacity effected the subject-verb number agreement process (Allen et al., 2015; Hartsuiker & Barkhuysen, 2006; Kemper, Herman, & Lian, 2003; Kemper & Sumner, 2001; Lorimor, Jackson, & van Hell, 2019). Combining all those predictions, we expected that the by age reduced working memory capacity of elderly people could (partly) explain the foreseen different results on the subject-verb number agreement task.

4.1 Interpretation results

4.1.1 Sentence completion task

The results of the sentence completion task showed that elderly people did not make more agreement errors than young adults. However, young adults produced more correct responses and less miscellaneous responses than elderly people. The comparable number of agreement errors in the two age groups was against our prediction. It could be seen as a positive finding: No indications were found for the influence of aging on the production of agreement errors in total. Nevertheless, the fewer correct answers and the more miscellaneous responses suggest that elderly people are less precise in their language production than young adults in this task. These results strongly suggest that aging causes difficulties in a relatively seen easy linguistic task: A task in which a large part exist of hearing, storing, and reproducing a sentence fragment and then complete the sentence fragment. This indicates age related influences on a language production task in general. It would be interesting to investigate the miscellaneous errors in further detail of even include a part of them as an additional factor in the analysis. Lorimor, Jackson, and van Hell (2019) included the non-fluent utterances in their study as additional factor, instead of coded them as miscellaneous error like we did. They found an effect for fluency, since participants produced more agreement errors in non-fluent utterances than in the fluent ones. Thus, fluency could also be an indication for the operationalisation of the subject-verb number agreement process. Unfortunately, our coding was not sufficient to include this analysis, but it is highly recommended to include it in the future.

In the sentence completion task, we found robust evidence for the attraction and distributivity effect, just like earlier research found (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Hartsuiker, Antón-Méndez, & van Zee, 2001; Hartsuiker & Barkhusen, 2006; Hartsuiker et al., 1999; Vigliocco et al., 1996). Participants were more likely to produce agreement errors when the subject noun and local noun differed in number from each other (the mismatch condition) than when they had both the same number (the match condition), which indices the attraction effect. Furthermore, participants produced more agreement errors when the number of the subject noun was grammatically singular but had a plural conceptual number (the multiple token condition), than when both the grammatical and conceptual number were singular (the single token condition). The found distributivity effect in both groups provides further evidence for the hypothesis that our participants consider even more information than actually needed for subject-verb number agreement.

Moreover, an interesting interaction effect was found for the variables distributivity and age group, just as we expected. The distributivity effect was smaller for elderly people than for young adults, as the difference between the amount of agreement errors in the single token and multiple token condition was smaller for elderly people than for young adults. However, this smaller difference in the elderly people was driven by the single token condition, as elderly people produced relatively more agreement errors in this condition. This was not expected, because we thought that elderly people would make less agreement errors in the multiple token condition and as a result of that the distributivity effects would be lower. An explanation for these results could be that elderly people show a stronger attraction effect. Given the fact that the subject head noun differs in number from the local noun, this could be enough for elderly people to produce agreement errors. The situation that the subject head noun is grammatical singular but conceptual plural, has a relatively weaker effect when compared to the attraction effect. The question arises what could cause this result.

According to these results, the elderly people in our study showed more similarities with the Broca's aphasia participants in the study of Hartsuiker et al. (1999) than our young adults. In their study, the Broca's aphasia participants did not seem to be sensitive for a mismatch between grammatical and conceptual number, and did not take semantic information into account when constructing subject-verb number agreement. Suggested is that for Broca's aphasia participants it is too resource consuming considering the conceptual number in subject-verb number agreement as well. As a result, they only use grammatical information. Of course, our elderly people are not nearly comparable with the Broca's aphasia participants of Hartsuiker et al. The elderly people did show the distributivity effect after all, but it was weaker than for the young adults. Thus, it could be that a limited resource capacity plays a role in the subject-verb number agreement process in elderly people too.

4.1.2 Sentence completion task and working memory tasks

To see if resource capacity is involved in subject-verb number agreement, we included working memory capacity tasks. In the working memory tasks, we found evidence for the effect of age on working memory capacity, just like earlier research found (Kirova, Bays, & Lagalwar, 2015): Elderly people had lower scores than young adults on the digit span forward and backward task, and also on the MoCA.

Combining the digit span backward score as strongest measure for working memory capacity with the sentence completion task, we found very interesting results of the effect of working memory capacity on the production of agreement errors (just as found by e.g. Allen et al., 2015; Hartsuiker & Barkhuysen, 2006; Lorimor et al., 2019). Participants with low digit span scores produced more agreement errors than participants with high digit span scores. Elderly people were found to have lower working memory scores, thus we found indications that age is affecting the subject-verb number agreement process. It supports the role of working memory capacity in the process of subject-verb number agreement in language production and is evidence for the resource-constrained hypothesis (Fayol, Largy, & Lemaire, 1994). The availability of working memory capacity is a determining variable for the accurate production of subject-verb number agreement.

An interesting question is where in the subject-verb number agreement process working memory is demanded, and which step in this process is affected by a reduced working memory capacity. The assumption of Hartsuiker and Barkhuysen (2006) is that working memory in the used subject-verb number agreement paradigm is required in the step that both the subject head noun and the local noun are integrated in the complex subject noun. Both nouns carry their own independent number specification and if these numbers are not the same, they compete with each other to specify the number of the complex subject noun. This process could be carried out automatically, at the point where the number of the nouns moves up higher in the syntactic tree representation to the place of the complex subject noun. The first one that reaches that place wins the competition. The one that wins the competition gives the number to the representation of the complex subject noun. The competition between number specification of the head and local noun could provoke number agreement errors. However, if working memory capacity is available, it controls this competition to avoid number agreement errors. If working memory capacity is limited, there is no control and the number agreement is just determined by the competition winner. If the local noun wins this, an agreement error is made. Following this assumption of Hartsuiker and Barkhuysen, it is difficult to control the subject-verb number agreement process for people with a lower working memory capacity or people who have to do a double task. As a consequence, more agreement errors occur in these situations, which happened also in our task.

Furthermore, we found a marginally significant effect for the interaction between distributivity and working memory capacity. Because it was not significant, we cannot draw conclusions from it. Nevertheless, it is interesting to look at the direction of the interaction effect. The effect of distributivity was smaller for participants with a high working memory capacity than for participants with a low working memory capacity, as the difference between the amount of agreement errors in the single token and multiple token condition was smaller for participants with a high working memory capacity than participants with a low working memory capacity. This smaller difference in participants with a high working memory capacity was driven by the multiple token condition, as participants with a high working memory capacity produced relatively seen less agreement errors in this condition. At a certain point for people with very high digit span scores, no more agreement errors were made in single token items and very few in multiple token items. As the digit span backward scores correlated with education level, this effect could be explained from that. Participants with a high education level, had high digit span backward scores and made almost no agreement errors. Our finding is opposite to the findings of Lorimor et al. (2019): they found in the same items used by us no effect of their measure for working memory capacity (OSpan) in the multiple token items, but only in the single token items. This could be a result from education level as well, since their participants seemed to be all students in higher education and our study consisted of participants with varying education levels.

4.1.3 The correlations

The before mentioned correlation we found between backward digit span scores and education level was in line with earlier research (Ostrosky-Solís & Lozano, 2006; Zimmermann, Cardoso, Trentini, Grassi-Oliveira, & Fonseca, 2015). Because our groups did not differ significantly from each other in the level of education and for both groups the correlation with digit span backward scores occurred, this reduced the risk on incorrect interpretations. Nevertheless, by making conclusions concerning digit span backward score it should be taken into account that those scores correlated with education level. Education level correlated with the amount of produced agreement errors too in both groups. This shows that we have made a right decision controlling for education level between the two age groups. It reveals also that education level is an important factor in investigating language production and that this effect must be kept in mind.

Another interesting finding is the correlation between the place in the list of the items and the production of correct responses and agreement errors for young adults only. Instead of using a randomized design in the distribution of the items, in which every participant had a complete random item order, we used a fixed order for the items. We found an effect of this fixed order, since young adults became better in the task (more correct answers and less agreement errors) as the task progressed. This task learning effect is in principle something we do not want to have, since this influences the results. Nevertheless, we found it very interesting that this task learning effect was shown by young adults only. This suggests that young adults are better in increasing cognitive control during the experiment in comparison with elderly people.

4.2 Answering the main research question

To sum up our most important findings to answer the research question, elderly people did not make more agreement errors than young adults, although their repetition of the sentence fragments was more affected. Elderly people showed a reduced distributivity effect in comparison with young adults and a stringer attraction effect. Elderly people scored lower on the digit span tasks and the MoCA than young adults, which indicates a lower working memory capacity for elderly people. Working memory capacity was a predictor for the production of agreement errors.

We could answer our research question: *Is there a difference between healthy young adults (aged between 18 and 25 years) and healthy elderly people (aged above 70 years) in the subject-verb number agreement production process?*. We found that the distributivity effect was weaker in elderly people than in young adults. Thus, we could presume that the process of subject-verb number agreement differs between young adults and elderly people and that aging influences this process.

An explanation for this difference could be the low working memory capacity of elderly people, as we found that working memory capacity influenced the subject-verb number agreement process. Less control on the competition between the head noun and the local noun appears if working memory capacity is reduced. As no more agreement errors were made by elderly people, this control mechanism is still working well, although it seems like something is going on in this mechanism. It could be that in elderly people, who have a reduced working memory capacity, less information could be processed. This could explain the weaker distributivity effect and the stronger attraction effect in elderly people: Less influence came from conceptual information in the competition process and more influence came from the interference local noun. It costs too much capacity to take into account conceptual information as there already is a lot of information from both the head and local noun. Here could be suggested that also other cognitive functions, which are affected by age-related changes, are involved in the subject-verb number agreement process. The strong attraction effect in elderly people might be explained on the basis of reduced inhibition abilities, as it could be difficult to ignore the intervening local noun attraction without enough control capacity.

In this research we combined the subject verb number agreement process with the role of working memory capacity by using the digit span backward test. According to the working memory model of Baddely (2001), working memory exists of multiple processes, including information storage, capacity, the executive functions of attention control and inhibition, and information retrieval. It would be very interesting to investigate more specifically which sub processes of executive function including working memory play a role in the subject-verb number agreement process and what functions or process deteriorate with aging. Vandierendonck, Loncke, Hartsuiker, and Desmet (2018) already showed a link between the executive function control and agreement errors in sentence comprehension. Reifegerste, Hauer and Felser (2017) found indications in their comprehension study that elderly people had difficulties in blocking intervening information in the subject-verb number agreement and that working memory could modulate this. These studies investigated the role of executive control and information blocking in the comprehension process of subject-verb number agreement for elderly people, but to our knowledge no further research is done in specifying this to the production process. To continue and expand this research field it is important to gain more insight in the sub mechanisms underlying working memory and in what way these mechanisms are related to language production processes. It is thereby relevant to combine this with investigating the role of aging on these processes. A useful contribution to the current research might be the inclusion of a measurements in which inhibition is determined, such as the Stroop test.

4.3 Limitations and recommendations

4.3.1 Participants

For the current research we used a comparison between two age groups (young adults and elderly people) to investigate the effect of aging. This was for us the most practical and feasible way, although it involves some unpredictable factors, because (unknown) variation between the two age groups cannot be excluded. For example, education level can still differ between the two groups due to changes in the school systems and teaching ways over the past decades. These factors are inevitable for the research design and only changing the design could solve these problems. It would be interesting to follow elderly people longitudinally, to see if changes also occur over the years individually. A short exploratory post hoc correlation analysis showed that age of elderly people correlated with the number of agreement errors and correct responses: the older the person, the more agreement errors and the less correct responses were produced.

Another limitation of this study was that the included participants were all very interested in scientific research and language (they had according to their own saying “a feeling for the Dutch language”), which could influence the results. Therefore, we probably had not a good representation of the whole Dutch population. For further research, it is recommended to keep this in mind. Despite the mentioned limitations, the power of this research lays in the relatively big participant groups (±60), which increases the statistic power.

4.3.2 Experiment

One of the problems in the current research was in the statistical analysis with the matching variable, that measures the attraction effect. Because both groups produced a really small amount of agreement errors in the match condition (both in the single and multiple token condition), a logit mixed effect model that includes those cells would be inappropriate. Because we defined the matching variable on forehand and also predicted an attraction effect (and found a very strong effect), we decided to report the statistics with this variable. After that, we performed a second statistical analysis, excluding the match condition and only including the age, distributivity and (the additional) working memory variables. Those two analyses differed from each other in the outcome effects: in the second analysis age group had no effect, but there was an interaction effect between age group and distributivity. Our results and conclusion are based on the second analysis, without the match condition. We believe we could defend our decision by making a comparison with the study design of Lorimor et al. (2019). They do mention the attraction effect, but do not include it in their design and only investigate the distributivity effect and two other variables, including working memory capacity.

In this research we used materials from earlier research: the items of Vigliocco, Hartsuiker, Jarema, and Kolk (1996), Hartsuiker, Kolk, and Huinck (1999), and Hartsuiker and Barkhuysen (2006). These items are thus more than 20 years old, which is not necessarily bad, but language use changes over time and perhaps these items are no longer entirely up to date. The short item analysis to see how many correct responses, agreement errors and miscellaneous responses are given per item revealed the interesting result that one of the single token items (ST10: *de discussie over de wetten*) provoked a lot of agreement errors, even more than some of the multiple token items. The fact that this happened says something about the included material as well. We suggest that further studies, which would like to use this material, first check how the intended participant group assesses the items in the different conditions (multiple token or single token).

4.3.3 Clinical implications

Our study contributed to the clinical field of speech and language pathology, because we provide more insight in the effect of the healthy aging process on language processing. Investigating the syntactic process of subject-verb number agreement in elderly people may help to gain a better understanding of this syntactic process in clinical populations. For studies in these clinical populations, healthy age-and-education matched elderly people represent the bases line to whom clinical populations are compared. Furthermore, our aging research contributes to the differentiation between age related changes in language production and changes that are caused by other impairments. If we know what happens in the aging process, these age effects could be set against the effects of the impairment, so we know what causes what.

We found in our study that the syntactic process of subject-verb number agreement is affected by age and that a reduced working memory plays a role in this. On the long term, when we as scientists have more insights in the exact mechanism underlying this, our findings should find a way to the clinical field, in a way that people actually could benefit from it. A better understanding of language processes could help providing good communication as long as possible for people. Eventually, we will all benefit from this.

5. Conclusion

Investigating aging and the effects of the aging process on cognitive functions and language becomes more and more important in society. The current research has contributed to this by examining a small part of the language production process, the syntactic process of subject-verb number agreement. By comparing elderly people to young adults on a subject-verb number agreement task, together with collecting data on working memory capacity of the participants, we gained insight in the way aging and working memory affect subject-verb number agreement in Dutch.

The findings support that the attraction and distributivity effect are distinctive, and related mechanism in subject-verb number agreement in both age groups. Aging, however, changes the underlying processes of distributivity and attraction: the distributivity effect is weaker and the attraction effect is stronger in elderly people. We provided evidence for the effect of a limited working memory capacity in the subject-verb number agreement process. In doing so, we demonstrated that the subject-verb number agreement process is not fully an automatic process, because working memory capacity plays a role in this. We suggested that the lower working memory capacity in elderly people could be used to explain the found difference between young adults and elderly people in the subject-verb number agreement process.

To conclude, aging affects the underlying mechanism in the syntactic language production process of subject-verb number agreement. A reduced working memory capacity could be underlying in this process. We proposed for follow-up research to investigate which specific processes in the working memory system and executive functions are involved in subject-verb number agreement. As the population of elderly people is expanding the need for research about this topic increases. We will all becoming old eventually.

6. References

Abrahamson, K., Clark, D., Perkins, A., & Arling, G. (2012). Does cognitive impairment influence quality of life among nursing home residents?. *The Gerontologist, 52*(5), 632-640. https://doi.org/10.1093/geront/gnr137

Allen, K. V., Pickering, M. J., Zammitt, N. N., Hartsuiker, R. J., Traxler, M. J., Frier, B. M., & Deary, I. J. (2015). Effects of acute hypoglycemia on working memory and language processing in adults with and without type 1 diabetes. *Diabetes Care, 38*(6), 1108-1115. https://doi.org/10.2337/dc14-1657

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences, 4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174909.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistic Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bock, K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1-47. http://dx.doi.org/10.1037/0033-295X.89.1.1

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*(1), 99-127. https://doi.org/10.1016/0749-596X(92)90007-K

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes, 8*(1), 57-99. https://doi.org/10.1080/01690969308406949

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*(1), 45-93. https://doi.org/10.1016/0010-0285(91)90003-7

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.05. http://www.praat.org/

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica, 150*, 80-84. https://doi.org/10.1016/j.actpsy.2014.04.010

Chadick, J. Z., Zanto, T. P., & Gazzaley, A. (2014). Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing. *Nature Communications, 5*, 4223. https://doi.org/10.1038/ncomms5223

Chauvel, G., Maquestiaux, F., Hartley, A. A., Joubert, S., Didierjean, A., & Masters, R. S. (2012). Age effects shrink when motor learning is predominantly supported by

nondeclarative, automatic memory processes: Evidence from golf putting. *The Quarterly Journal of Experimental Psychology, 65*(1), 25-38. https://doi.org/10.1080/17470218.2011.588714
Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics, 13*(1), 53-76. https://doi.org/10.1017/S0142716400005427
Choi, H. J., Lee, D. Y., Seo, E. H., Jo, M. K., Sohn, B. K., Choe, Y. M., . . . & Jhoo, J. H. (2014). A normative study of the digit span in an educationally diverse elderly population. *Psychiatry Investigation, 11*(1), 39-43. http://dx.doi.org/10.4306/pi.2014.11.1.39
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current Directions in Psychological Science, 19*(1), 51-57. https://doi.org/10.1177/0963721409359277
Craik, F. I. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*(5), 155-158. Retrieved from http://www.jstor.org/stable/20182295
Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of Data Science, 8*(1), 61–73. http://dx.doi.org/10.6339/JDS.2010.08(1).563
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321. http://dx.doi.org/10.1037/0033-295X.93.3.283
Evans, I. E., Martyr, A., Collins, R., Brayne, C., & Clare, L. (2018). Social isolation and cognitive function in later life: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*. Advance online publication. https://doi.org/10.3233/JAD-180501
Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology, 47*(2), 437-464. https://doi.org/10.1080/14640749408401119
Ferreira, F., & Engelhardt, P. E. (2006). Syntax and Production. In M. J. Traxler and M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics,* (2nd ed., pp. 61-92). San Diego, CA, US: Academic Press. https://doi.org/10.1016/B978-012369374-7/50004-3
Fleischman, D. A., Wilson, R. S., Gabrieli, J. D., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19*(4), 617-625. http://dx.doi.org/10.1037/0882-7974.19.4.617
Fonseca, R. P., Zimmermann, N., Scherer, L. C., Parente, M. A. D. M. P., & Ska, B. (2010). Episodic memory, concentrated attention and processing speed in aging: A comparative study of Brazilian age groups. *Dementia & Neuropsychologia, 4*(2), 91-97. http://dx.doi.org/10.1590/S1980-57642010DN40200003
Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 133-175). San Diego, CA, US: Academic Press. https://doi.org/10.1016/S0079-7421(08)60270-4
Garrett, M. F. (1988). Processes in language production. In F. J. Nieumeyer (Ed.), *Linguistics: The Cambridge Survey, III. Language: Psychological and Biological Aspects* (pp. 69-96). Cambridge, UK: Cambridge University Press.
Gates, G. A., Feeney, M. P., & Mills, D. (2008). Cross-sectional age-changes of hearing in the elderly. *Ear and Hearing, 29*(6), 865-874. https://doi.org/10.1097/AUD.0b013e318181adb5

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology, 40*(2), 177-190. http://dx.doi.org/10.1037/0012-1649.40.2.177

Gow, A. J., & Mortensen, E. L. (2016). Social resources and cognitive ageing across 30 years: the Glostrup 1914 Cohort. *Age and Ageing, 45*(4), 480-486. https://doi.org/10.1093/ageing/afw070

Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine, 29*(4), 737-752. https://doi.org/10.1016/j.cger.2013.07.002

Hardy, S. M., Messenger, K., & Maylor, E. A. (2017). Aging and syntactic representations: Evidence of preserved syntactic priming and lexical boost. *Psychology and Aging, 32*(6), 588-596. http://dx.doi.org/10.1037/pag0000180

Hartsuiker, R. J., Antón-Méndez, I., & van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language, 45*(4), 546-572. https://doi.org/10.1006/jmla.2000.2787

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21*(1-3), 181-204. https://doi.org/10.1080/01690960400002117

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic facilitation in agrammatic sentence production. *Brain and Language, 62*(2), 221-254. https://doi.org/10.1006/brln.1997.1905

Hartsuiker, R. J., Kolk, H. H. J., & Huinck, W. J. (1999). Agrammatic production of subject–verb agreement: The effect of conceptual number. *Brain and Language, 69*(2), 119-160. https://doi.org/10.1006/brln.1999.2059

Hartsuiker, R. J., & Moors, A. (2017). On the automaticity of language processing. In H.-J. Smith (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 201-225). Berlin, Germany: De Gruyter Mouton.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation, 22*, 193-225. https://doi.org/10.1016/S0079-7421(08)60041-9

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine, 40*(2), 218-227. https://doi.org/10.1007/s12160-010-9210-8

Hawkley, L. C., Thisted, R. A., Masi, C. M., & Cacioppo, J. T. (2010). Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. *Psychology and Aging, 25*(1), 132-141. https://doi.org/10.1037/a0017805

Hendriks, M. P., Kessels, R. P. C., Gorissen, M. E. E., Schmand, B. A., & Duits, A. A. (2014). *Neuropsychologische diagnostiek: De klinische praktijk*. Amsterdam, The Netherlands: Uitgeverij Boom.

Irish, M., Lawlor, B. A., O'Mara, S. M., & Coen, R. F. (2011). Impaired capacity for autonoetic reliving during autobiographical event recall in mild Alzheimer's disease. *Cortex, 47*(2), 236-249. https://doi.org/10.1016/j.cortex.2010.01.002

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Kemmerer, D. (2014). *Cognitive neuroscience of language*. New York, NY, US: Psychology Press.

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: Findings from the nun study. *Psychology and Aging, 16*(2), 227-239. https://doi.org/10.1037/0882-7974.16.2.227

Kemper, S., Herman, R., & Lian, C. (2003). Age differences in sentence production. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*(5), 260-268. https://doi.org/10.1093/geronb/58.5.P260

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16*(2), 312-322. https://doi.org/10.1037//0882-7974.16.2.312

Kessels, R., Eling, P., Ponds, R., Spikman, J., van Zandvoort, M. (2017). *Clinical Neuropsychology.* Amsterdam, The Netherlands: Boom Publishers

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643-650. https://doi.org/10.3758/BRM.42.3.643

Kirova, A. M., Bays, R. B., & Lagalwar, S. (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Research International, 2015*, 748212. http://dx.doi.org/10.1155/2015/748212

Kolb, B., & Whishaw, I. Q. (2001). *An introduction to brain and behavior*. New York, NY, US: Worth Publishers.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, UK: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences, 22*, 1-75. https://doi.org/10.1017/S0140525X99001776

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychology and Aging, 17*(4), 677-689. http://dx.doi.org/10.1037/0882-7974.17.4.677

Lorimor, H., Jackson, C. N., & van Hell, J. G. (2019). The interaction of notional number and morphophonology in subject–verb agreement: A role for working memory. *Quarterly Journal of Experimental Psychology, 72*(4), 890-900. http://dx.doi.org/10.1177/1747021818771887

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences, 112*(20), 6515-6518. https://doi.org/10.1073/pnas.1504162112

Mehotcheva, T. H. (2010). *After the fiesta is over: Foreign language attrition of Spanish in Dutch and German Erasmus students*. Barcelona, Spain: Universitat Pompeu Fabra.

Meinzer, M., Seeds, L., Flaisch, T., Harnish, S., Cohen, M. L., McGregor, K., ... Crosson, B. (2012b). Impact of changed positive and negative task-related brain activity on word-retrieval in aging. *Neurobiology of Aging, 33*(4), 656–669. https://doi.org/10.1016/j.neurobiolaging.2010.06.020

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurological Sciences, 34*(5), 749-754. http://dx.doi.org/10.1007/s10072-014-2019-7

Murman, D. L. (2015). The impact of age on cognition. *Seminars in Hearing, 36*(3), 111-121. https://doi.org/10.1055/s-0035-1555115

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Niccoli, T., & Partridge, L. (2012). Ageing as a risk factor for disease. *Current Biology, 22*(17), R741-R752. https://doi.org/10.1016/j.cub.2012.07.024

Ostrosky-Solís, F., & Lozano, A. (2006). Digit span: Effect of education and culture. *International Journal of Psychology, 41*(5), 333-341. https://doi.org/10.1080/00207590500345724

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*(4), 621-637. http://dx.doi.org/10.1037/0882-7974.11.4.621

Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori and R. Job (Eds.), *The Cognitive Neuropsychology of Language* (pp. 273-294). London, UK: Lawrence Erlbaum Associates.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., & . . . Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01193-y

Peters, R. (2006). Ageing and the brain. *Postgraduate Medical Journal*, *82*(964), 84-88. http://dx.doi.org/10.1136/pgmj.2005.036665

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory, 9*(4-6), 423-432. https://doi.org/10.1080/09658210143000182

R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramsay, M. C., & Reynolds, C. R. (1995). Separate digits tests: A brief history, a literature review, and a reexamination of the factor structure of the Test of Memory and Learning (TOMAL). *Neuropsychology Review, 5*(3), 151-171. https://doi.org/10.1007/BF02214760

Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: Evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition, 24*(6), 672-702. https://doi.org/10.1080/13825585.2016.1251550

Rosenberg, S., & Abbeduto, L. (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults 1. *Applied Psycholinguistics, 8*(1), 19-32. https://doi.org/10.1017/S0142716400000047

RStudio Team (2018). RStudio: Integrated Development for R. Boston, MA, US: RStudio, Inc. http://www.rstudio.com/

Salis, C., Kelly, H., & Code, C. (2015). Assessment and treatment of short-term and working memory impairments in stroke aphasia: a practical tutorial. *International Journal of Language & Communication Disorders, 50*(6), 721-736. https://doi.org/10.1111/1460-6984.12172

Schriefers, H., & Vigliocco, G. (2015). Speech production, psychology of [repr.]. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., vol. 23, pp. 255–258). New York, NY, US: Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.52022-4

Smith, C. D., Walton, A., Loveland, A. D., Umberger, G. H., Kryscio, R. J., & Gash, D. M. (2005). Memories that last in old age: motor skill learning and memory preservation. *Neurobiology of Aging, 26*(6), 883-890. https://doi.org/10.1016/j.neurobiolaging.2004.08.014

Thissen, A. J. A. M., van Bergen, F., de Jonghe, J. F. M., Kessels, R. P. C., & Dautzenberg, P. L. J. (2010). Bruikbaarheid en validiteit van de Nederlandse versie van de Montreal Cognitive Assessment (MoCA-D) bij het diagnosticeren van Mild Cognitive Impairment. *Tijdschrift voor Gerontologie en Geriatrie, 41*(6), 231-240. https://doi.org/10.1007/s12439-010-0218-0

Thornton, R., & Light, L. L. (2006). Language comprehension and production in normal aging. In J. E. Birren & K. W. Schaire (Eds.), *Handbook of the psychology of Aging*

(6th ed., pp. 261-287). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/B978-012101264-9/50015-X

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language, 48*(4), 740-759. https://doi.org/10.1016/S0749-596X(03)00003-2

Uit den Boogaart, P. C. (1975). *Woordfrequenties in geschreven en gesproken Nederlands* [Word frequencies in written and spoken Dutch]. Utrecht, The Netherlands: Oosthoek, Scheltema & Holkema.

van Duin, C., & Garssen, J. (2010). *Bevolkingsprognose 2010–2060: sterkere vergrijzing, langere levensduur.* Retrieved from https://www.cbs.nl/nl-nl/achtergrond/2010/50/bevolkingsprognose-2010-2060-sterkere-vergrijzing-langere-levensduur

van Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden: 4600 substantieven, 1000 verba en 500 adjectieven*. [Imageability ratings of Dutch words: 4600 nouns, 1000 verbs, and 500 adjectives]. Lisse, The Netherlands: Swets & Zeitlinger.

Vandierendonck, A., Loncke, M., Hartsuiker, R. J., & Desmet, T. (2017). The role of executive control in resolving grammatical number conflict in sentence comprehension. *The Quarterly Journal of Experimental Psychology, 71*(3), 759-778. https://doi.org/10.1080/17470218.2016.1276610

Verhaegen, C., & Poncelet, M. (2013). Changes in naming and semantic abilities with aging from 50 to 90 years. *Journal of the International Neuropsychological Society, 19*(2), 119-126. https://doi.org/10.1017/S1355617712001178

Verhage, F. (1964*). Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar*. Assen, The Netherlands: van Gorcum.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*(3), 261-298. https://doi.org/10.1016/S0010-0277(96)00713-5

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*(4), 407-442. https://doi.org/10.1080/016909696387169

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear?. *Cognition, 68*(1), B13-B29. https://doi.org/10.1016/S0010-0277(98)00041-9

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, TX, US: NCS Pearson.

Wei, M., Russell, D. W., & Zakalik, R. A. (2005). Adult attachment, social self-efficacy, self-disclosure, loneliness, and subsequent depression for freshman college students: A longitudinal study. *Journal of Counseling Psychology, 52*(4), 602-614. http://dx.doi.org/10.1037/0022-0167.52.4.602

Yorkston, K. M., Bourgeois, M. S., & Baylor, C. R. (2010). Communication and aging. *Physical Medicine and Rehabilitation Clinics, 21*(2), 309-319. https://doi.org/10.1016/j.pmr.2009.12.011

Zimmermann, N., Cardoso, C. D. O., Trentini, C. M., Grassi-Oliveira, R., & Fonseca, R. P. (2015). Brazilian preliminary norms and investigation of age and education effects on the Modified Wisconsin Card Sorting Test, Stroop Color and Word test and Digit Span test in adults. *Dementia & Neuropsychologia, 9*(2), 120-127. http://dx.doi.org/10.1590/1980-57642015DN92000006

7. Appendices

A. Appendix: Experimental items

Experimental items used in the sentence completion task. The items are the same as used by Vigliocco, Hartsuiker, Jarema, and Kolk (1996), Hartsuiker, Kolk, and Huinck (1999), and Hartsuiker and Barkhuysen (2006). Each experimental item consist of the adjective and sentence fragment; the item name is included; below each item, is the English translation.

SINGLE TOKEN ITEMS

Item name	***Sentence***	***Adjective***
ST01	de regisseur van de film/s	bekend
	the director of the film/s	*well-known*
ST02	de aanslag op de minister/s	brutaal
	the strike on the minister/s	*brutal*
ST03	de beker voor de winnaar/s	groot
	the cup for the winner/s	*large*
ST04	de boodschap voor de reiziger/s	kort
	the message for the traveller/s	*brief*
ST05	de eis van de ontvoerder/s	onredelijk
	the demand of the kidnapper/s	*unreasonable*
ST06	de klacht van de scholier/en	terecht
	the complaint of the student/s	*appropriate*
ST07	de monteur van de motorfiets/en	handig
	the mechanic of the motorcycle/s	*handy*
ST08	de arts van de zieke/n	knap
	the doctor of the sick person/s	*clever*
ST09	de baby op de foto/'s	lief
	the baby on the photo/s	*sweet*
ST10	de discussie over de wet/ten	saai
	the discussion about the law/s	*boring*
ST11	de eigenaar van de koffer/s	spoorloos
	the owner of the suitcase/s	*traceless*
ST12	de diefstal van de diamant/en	succesvol
	the theft of the diamond/s	*successful*

MULTIPLE TOKEN ITEMS

Item name	***Sentence***	***Adjective***
MT13	de stop op de fles	groen
	the stopper on the bottle/s	*green*
MT14	de beschrijving in de gids	onduidelijk
	the description in the guide/s	*unclear*
MT15	de datum op de munt	onleesbaar
	the date on the coin/s	*illegible*
MT16	de reclame op de bus	opvallend
	the advertisement on the bus/ses	*catchy*
MT17	de ingang van de flat	small
	the entrance of the apartment/s	*narrow*
MT18	de kraag van de jas	vuil
	the collar of the coat/s	*dirty*
MT19	de leuning van de stoel	gammel
	the back of the chair/s	*dodgy*
MT20	de sleutel van de kast	klein
	the key of the cupboard/s	*small*
MT21	de bel op de fiets	luid
	the bell on the bicycle/s	*loud*
MT22	de afbeelding op de mok	mooi
	the picture on the mug/s	*pretty*
MT23	de bon in de folder	ongeldig
	the coupon in the flyer/s	*invalid*
MT24	de paraaf op de declaratie	vals
	the initials [singular] on the declaration/s	*forged*

B. Appendix: List 1 and 2 of the sentence completion task

I: List 1 includes experimental items and filler items (with item number, item name, adjective, and sentence fragment). For the experimental items, the conditions of the distributivity variable (Single Token, Multiple Token) and Matching variable (Match, Mismatch) are given.

Item Nummer	Name	Adjective	Sentence	Distributivity	Matching
1	F19	ROT	De appels in de mand		
2	F06	ZWAAR	De eikenhouten tafel		
3	F22	STOFFIG	De boeken op de plank		
4	F31	HARD	De geluiden uit de klassen		
5	MT20	KLEIN	De sleutel van de kasten	Multiple Token	Mismatch
6	F29	FRAAI	De foto's van de gorilla’s		
7	ST06	TERECHT	De klacht van de scholier	Single Token	Match
8	F21	SMERIG	De ramen van het huis		
9	ST12	SUCCESVOL	De diefstal van de diamanten	Single Token	Mismatch
10	F36	VIES	De varkens bij de stallen		
11	MT23	ONGELDIG	De bon in de folders	Multiple Token	Mismatch
12	F12	VERVELEND	De mislukte grappen		
13	F03	JONG	De pas afgestudeerde dokter		
14	F24	WIT	De lichten van de auto		
15	MT22	MOOI	De afbeelding op de mokken	Multiple Token	Mismatch
16	F10	ONTERECHT	De hoge rekeningen		
17	MT18	VUIL	De kraag van de jas	Multiple Token	Match
18	F13	BANG	De schapen in de wei		
19	F27	DUUR	De geschenken voor de meisjes		
20	F23	VERS	De garnalen voor de cocktail		
21	MT14	ONDUIDELIJK	De beschrijving in de gids	Multiple Token	Match
22	F14	DROOG	De oases in de woestijn		
23	ST02	BRUTAAL	De aanslag op de minister	Single Token	Match
24	F01	ENG	De gevaarlijke haai		
25	ST04	KORT	De boodschap voor de reiziger	Single Token	Match
26	F02	GEZELLIG	De lange en mooie wandeling		
27	F26	DIK	De benen van de voetballers		
28	ST09	LIEF	De baby op de foto's	Single Token	Mismatch
29	MT19	GAMMEL	De leuning van de stoelen	Multiple Token	Mismatch
30	F16	KORT	De mouwen van het overhemd		
31	ST05	ONREDELIJK	De eis van de ontvoerder	Single Token	Match
32	F34	NIEUW	De munten in haar zakken		
33	MT15	ONLEESBAAR	De datum op de munt	Multiple Token	Match
34	ST01	BEKEND	De regisseur van de film	Single Token	Match
35	F28	FOUT	De diagnoses van de dokters		
36	F08	ENG	De spannende verhalen		
37	F09	LANGDRADIG	De saaie vergaderingen		
38	MT24	VALS	De paraaf op de declaraties	Multiple Token	Mismatch
39	MT13	GROEN	De stop op de fles	Multiple Token	Match
40	F18	LUIDRUCHTIG	De soldaten in de trein		
41	ST03	GROOT	De beker voor de winnaar	Single Token	Match
42	F32	HEERLIJK	De taarten van de koks		
43	MT16	OPVALLEND	De reclame op de bus	Multiple Token	Match
44	F35	PIJNLIJK	De operaties aan de voeten		
45	ST10	SAAI	De discussie over de wetten	Single Token	Mismatch
46	F15	GIFTIG	De pillen in de fles		
47	F25	CORRUPT	De vrienden van de ministers		
48	MT21	LUID	De bel op de fietsen	Multiple Token	Mismatch
49	ST08	KNAP	De arts van de zieken	Single Token	Mismatch
50	F30	GOED	De platen van de zangers		
51	F07	AARDIG	De beleefde studenten		
52	F05	SAAI	De plaatselijke krant		
53	ST07	HANDIG	De monteur van de motorfietsen	Single Token	Mismatch
54	F04	LEKKER	De vegetarische maaltijd		
55	F11	VALS	De nieuwe paspoorten		
56	F17	LEKKER	De geuren uit de keuken		
57	MT17	SMAL	De ingang van de flat	Multiple Token	Match
58	F33	LANG	De files op de wegen		
59	ST11	SPOORLOOS	De eigenaar van de koffers	Single Token	Mismatch
60	F20	SCHATTIG	De baby’s op de deken		

II: List 2 includes experimental items (STX/MTX) and filler items (FX) (with item number, item name, adjective, and sentence fragment). For the experimental items, the conditions of the distributivity variable (Single Token, Multiple Token) and Matching variable (Match, Mismatch) are given.

Item Nummer	Name	Adjective	Sentence	Distributivity	Matching
1	F19	ROT	De appels in de mand		
2	F06	ZWAAR	De eikenhouten tafel		
3	F22	STOFFIG	De boeken op de plank		
4	F31	HARD	De geluiden uit de klassen		
5	MT20	KLEIN	De sleutel van de kast	Multiple Token	Match
6	F29	FRAAI	De foto's van de gorilla’s		
7	ST06	TERECHT	De klacht van de scholieren	Single Token	Mismatch
8	F21	SMERIG	De ramen van het huis		
9	ST12	SUCCESVOL	De diefstal van de diamant	Single Token	Match
10	F36	VIES	De varkens bij de stallen		
11	MT23	ONGELDIG	De bon in de folder	Multiple Token	Match
12	F12	VERVELEND	De mislukte grappen		
13	F03	JONG	De pas afgestudeerde dokter		
14	F24	WIT	De lichten van de auto		
15	MT22	MOOI	De afbeelding op de mok	Multiple Token	Match
16	F10	ONTERECHT	De hoge rekeningen		
17	MT18	VUIL	De kraag van de jassen	Multiple Token	Mismatch
18	F13	BANG	De schapen in de wei		
19	F27	DUUR	De geschenken voor de meisjes		
20	F23	VERS	De garnalen voor de cocktail		
21	MT14	ONDUIDELIJK	De beschrijving in de gidsen	Multiple Token	Mismatch
22	F14	DROOG	De oases in de woestijn		
23	ST02	BRUTAAL	De aanslag op de ministers	Single Token	Mismatch
24	F01	ENG	De gevaarlijke haai		
25	ST04	KORT	De boodschap voor de reizigers	Single Token	Mismatch
26	F02	GEZELLIG	De lange en mooie wandeling		
27	F26	DIK	De benen van de voetballers		
28	ST09	LIEF	De baby op de foto	Single Token	Match
29	MT19	GAMMEL	De leuning van de stoel	Multiple Token	Match
30	F16	KORT	De mouwen van het overhemd		
31	ST05	ONREDELIJK	De eis van de ontvoerders	Single Token	Mismatch
32	F34	NIEUW	De munten in haar zakken		
33	MT15	ONLEESBAAR	De datum op de munten	Multiple Token	Mismatch
34	ST01	BEKEND	De regisseur van de films	Single Token	Mismatch
35	F28	FOUT	De diagnoses van de dokters		
36	F08	ENG	De spannende verhalen		
37	F09	LANGDRADIG	De saaie vergaderingen		
38	MT24	VALS	De paraaf op de declaratie	Multiple Token	Match
39	MT13	GROEN	De stop op de flessen	Multiple Token	Mismatch
40	F18	LUIDRUCHTIG	De soldaten in de trein		
41	ST03	GROOT	De beker voor de winnaars	Single Token	Mismatch
42	F32	HEERLIJK	De taarten van de koks		
43	MT16	OPVALLEND	De reclame op de bussen	Multiple Token	Mismatch
44	F35	PIJNLIJK	De operaties aan de voeten		
45	ST10	SAAI	De discussie over de wet	Single Token	Match
46	F15	GIFTIG	De pillen in de fles		
47	F25	CORRUPT	De vrienden van de ministers		
48	MT21	LUID	De bel op de fiets	Multiple Token	Match
49	ST08	KNAP	De arts van de zieke	Single Token	Match
50	F30	GOED	De platen van de zangers		
51	F07	AARDIG	De beleefde studenten		
52	F05	SAAI	De plaatselijke krant		
53	ST07	HANDIG	De monteur van de motorfiets	Single Token	Match
54	F04	LEKKER	De vegetarische maaltijd		
55	F11	VALS	De nieuwe paspoorten		
56	F17	LEKKER	De geuren uit de keuken		
57	MT17	SMAL	De ingang van de flats	Multiple Token	Mismatch
58	F33	LANG	De files op de wegen		
59	ST11	SPOORLOOS	De eigenaar van de koffer	Single Token	Match
60	F20	SCHATTIG	De baby’s op de deken		

C. Appendix: Coding categories

Table with an overview of the different coding categories for the given answers in the subject-verb agreement task.

Code			Meaning	Example
0			Correct response	De sleutel van de kasten is klein
1			Agreement error	De sleutel van de kasten zijn klein
99			Miscellaneous response	
	2		Number subject noun repetition error	De sleutels van de kasten zijn klein
	22		Number local noun repetition error	De sleutel van de kast is klein
	3		Repetition plus agreement error	De sleutels van de kasten is klein
	4		Miscellaneous	
		A	The beginning utterance had to be repeated twice or more by the test leader	
		B	The subject or local noun was omitted	De sleutel is klein/De kasten zijn klein
		C	An incomplete sentence was produced	Is klein
		D	The preposition was changed in a different preposition	De sleutel in de kasten is klein
		E	The local noun was changed in a different local noun	De sleutel van de klassen is klein
		F	The subject noun was changed in a different subject noun	Het slot van de kasten is klein
		G	Hesitations, repetitions, interruptions and other non-fluencies	De sleutel van de k-kasten euh is klein
		H	Long silence around 300 ms	De sleutel van de kasten [ ] is klein

5	Q	The adjective is corrected right
	R	The subject noun is corrected right
	S	The local noun is corrected right
	T	The preposition is corrected right
	U	An incomplete sentence is corrected into a complete sentence
	V	Not a real correction but mentioned that there was "something strange"
	X	The verb is corrected wrong
	Y	The verb is corrected right
	Z	The subject noun is corrected wrong

D. Appendix: Table of distribution without miscellaneous responses

Table of the distribution (in absolute numbers and percentages) of responses by scoring category (Correct Responses, Agreement Errors) per group (Young, Old) per condition (Single Token, Multiple Token, Match, Mismatch)

Condition	Total Responses		Correct Responses		Agreement Errors	
	Young	Old	Young	Old	Young	Old
Single Token						
Match	367	314	366	309	1	5
			99.7%	98.4%	0.3%	1.6%
Mismatch	344	277	321	232	23	45
			93.3%	83.7%	6.7%	16.3%
Multiple Token						
Match	358	310	357	303	1	7
			99.7%	97.7%	0.3%	2.3%
Mismatch	333	301	205	184	128	117
			61.6%	61.1%	38.4%	38.9%
Total						
	1402	1202	1249	1028	153	174
			89.1%	85.5%	10.9%	14.5%
	2604		2277		327	
			87.4%		12.6%	

E. Appendix: Report statistic main findings miscellaneous responses

Because the miscellaneous responses were not included in the main analysis, we report some general main findings concerning this category here. Using a Wald Chi-Square Test, results showed that the production of miscellaneous responses was not significantly affected by the distributivity variable (p>0.1), although a significant interaction effect between age group and distributivity was found. The difference in produced miscellaneous responses (9.4%) between the single token condition (54.7%) and multiple token condition (45.3%) for elderly people deviated significantly from the differences (18.2%) between the production of miscellaneous responses in the single token condition (40.9%) and multiple token condition (59.1%) for younger people, in the way that the difference was smaller for the elderly people, $\chi^2(1)=5.03$, $p<0.05$. Furthermore, a significant main effect of the matching variable was found. In the mismatch condition (7,1%) significantly more miscellaneous errors were produced than in the match condition (3,9%), $\chi^2(1)=39.82$, $p<0.001$, but no interaction was found between the age group and the matching variable.

F. Appendix: table of distributions without items of the match condition

Table of the distribution (in absolute numbers and percentages) of responses by scoring category (Correct Responses. Agreement Errors) per age group (Young. Old) per condition (Single Token, Multiple Token) for the mismatch items only

Condition	Total Responses		Correct Responses		Agreement Errors	
	Young	Old	Young	Old	Young	Old
Mismatch						
Single Token	344	277	321	232	23	45
			93.3%	83.7%	6.7%	16.3%
Multiple Token	333	301	205	184	128	117
			61.6%	61.1%	38.4%	38.9%
Total						
	677	578	526	416	151	162
			77.7%	72.0%	22.3%	28.0%
	1255		942		313	
			75.1%		24.9%	

G. Appendix: R script for statistical analysis

```
# Read the required packages
library(ggplot2)
library(lme4)
library(Matrix)
library(car)
library(carData)

# Read the table with all the required data for the analysis (without the miscellaneous data)
Data_alles = read.csv("EIND_data_alles.csv")

# Make factors of accuracy subjectname en itemname
Data_alles$accuracy = factor(Data_alles$accuracy)
Data_alles$subjectname = factor(Data_alles$subjectname)
Data_alles$itemname = factor(Data_alles$itemname)

# Sum code the independent variables that are binominal and change them in numeric values,
to gain a mean of 0 with a range of 1. R likes it that way. Young = -0.5; Old = 0.5. Single
token = -0.5; Multiple token = 0.5. Match = -0.5; Mismatch = 0.5. list 1 = -0.5; list 2 = 0.5.
Man = -0.5; Woman = 0.5.
Data_alles$agegroup <- ifelse(Data_alles$agegroup=="jong",-.5,.5)
Data_alles$token <- ifelse(Data_alles$token=="S", -.5,.5)
Data_alles$match <- ifelse(Data_alles$match=="match", -.5,.5)
Data_alles$list <- ifelse(Data_alles$list=="1", -.5,.5)
Data_alles$gender <- ifelse(Data_alles$gender=="M", -.5,.5)

# Make the full model (Model 1)
Full_Model_Data_alles = glmer(accuracy ~ agegroup*token*match + (1|subjectname) +
(1|itemname) + (0+token|subjectname) + (0+match|subjectname) +
(token:match|subjectname) + (0+match|itemname) + (0+agegroup|itemname) +
(match:agegroup|itemname), data=Data_alles, family=binomial, control =
glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

# Make different models via reduction to gain a fitting model that converge. Remove the
interactions in the random effects first, remove the slopes apiece to see if removing one
influences the model. Model2 converged (= without the interactions in the by subject en item
slopes and without the by subject slope for item)
Model2Data_alles = glmer(accuracy ~ agegroup*token*match + (1|subjectname) +
(1|itemname) + (0+token|subjectname) + (0+match|subjectname) + (0+match|itemname),
data=Data_alles, family=binomial, control = glmerControl(optimizer="bobyqa",
optCtrl=list(maxfun=1e6)))

######################################################################

# Read the table with all the required data for the analysis (without the miscellaneous data and
without the match condition)
Data_zonderMatch = read.csv("EINDdata_alles_zonderMatch.csv")

# Make factors of accuracy subjectname en itemname
```

```
Data_zonderMatch$accuracy = factor(Data_zonderMatch$accuracy)
Data_zonderMatch$subjectname = factor(Data_zonderMatch$subjectname)
Data_zonderMatch$itemname = factor(Data_zonderMatch$itemname)

# Sum code the independent variables that are binominal and change them in numeric values, to gain a mean of 0 with a range of 1. R likes it that way. Young = -0.5; Old = 0.5. Single token = -0.5; Multiple token = 0.5. list 1 = -0.5; list 2 = 0.5. Man = -0.5; Woman = 0.5.
Data_zonderMatch$agegroup <- ifelse(Data_zonderMatch$agegroup=="jong",-.5,.5)
Data_zonderMatch$token <- ifelse(Data_zonderMatch$token=="S", -.5,.5)
Data_zonderMatch$list <- ifelse(Data_zonderMatch$list=="1", -.5,.5)
Data_zonderMatch$gender <- ifelse(Data_zonderMatch$gender=="M", -.5,.5)

# Make the full model (Model 3)
Full_Model_Data_zonderMatch = glmer(accuracy ~ agegroup*token + (1|subjectname) + (1|itemname) + (0+token|subjectname) + (0+agegroup|itemname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

# Make a reduced converging model (Model 4)
Reduced_Model_Data_zonderMatch = glmer(accuracy ~ agegroup*token + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

# Get the output
summary(Full_Model_Data_zonderMatch)

# Scale to mean (educaltion level, reading code, MoCA score, digit span forward score, digit span backward sore)
Data_zonderMatch$c.eductaional.level.code = scale(Data_zonderMatch$eductaional.level.code, scale = FALSE)
Data_zonderMatch$c.reading.code = scale(Data_zonderMatch$reading.code, scale = FALSE)
Data_zonderMatch$c.MoCA.score = scale(Data_zonderMatch$MoCA.score, scale = FALSE)
Data_zonderMatch$c.forward.digit.span = scale(Data_zonderMatch$forward.digit.span, scale = FALSE)
Data_zonderMatch$c.backward.digit.span = scale(Data_zonderMatch$backward.digit.span, scale = FALSE)

# Make new models with the extra added other factors. Include these factors as a main effect
Reduced_Model_ZM_gender = glmer(accuracy ~ agegroup*token + gender + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_list = glmer(accuracy ~ agegroup*token + list + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_listnr = glmer(accuracy ~ agegroup*token + listnumber + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))
```

```
Reduced_Model_ZM_cRead = glmer(accuracy ~ agegroup*token + c.reading.code + (1|subjectname) + (1|itemname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_cEdu = glmer(accuracy ~ agegroup*token + c.eductaional.level.code + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_cMoCA = glmer(accuracy ~ agegroup*token + c.MoCA.score + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_cFWDS = glmer(accuracy ~ agegroup*token +c.forward.digit.span + (1|subjectname) + (1|itemname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

Reduced_Model_ZM_cBWDS = glmer(accuracy ~ agegroup*token + c.backward.digit.span + (1|subjectname) + (1|itemname) + (0+token|subjectname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

# Make a reduced converging model with digit span backward as a real factor (Model 5)
Model_ZM_cBWDS_fullPRED = glmer(accuracy ~ agegroup * token * c.backward.digit.span + (1 | subjectname) + (1 | itemname) + (0 + token | subjectname) + (0 + c.backward.digit.span | itemname), data=Data_zonderMatch, family=binomial, control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

# Get the output
summary(Model_ZM_cBWDS_fullPRED)
```

H. Appendix: Table of models with individual added predictors

This table shows the coefficients and probability estimates (Coefficient, Standard Error, Wald Chi-Squared Z-score, p-value) of the added fixed effects in Model 4 (Accuracy ~ agegroup* distributivity + X + (1|subject) + (1|item) + (0+distributivity|subject)), in which X stands for an as main factor added fixed effect, without any interaction. This is done to see which extra factors predicted the production of agreement errors.

Predictor	**Coefficient**	**Std. Error**	**Wald Z**	***p***
Gender	0.17	0.44	0.39	0.69
Education Level	-1.44	0.27	-5.41	2.27e-08***
Reading Behaviour	-0.93	0.62	-1.51	0.13
MoCA score	-0.58	0.11	-5.20	2.00e-07***
Forward Digit Span	-0.45	0.18	-2.43	0.015*
Backward Digit Span	-0.96	0.18	-5.29	1.21e-07***
List type	-0.68	0.59	-1.14	0.25
List number	-0.03	0.01	-2.30	0.02*

* Effect is significant on the 0.05 alfa level
*** Effect is significant on the 0.001 alfa level

There was a significant main effect of education level, because high educated participant produced less agreement errors than low educated participants. There was a significant effect for MoCA scores, because participants with high MoCA scores produced less agreement errors than participants with low MoCA scores. There was a significant effect for digit span forward scores, because participants with high digit span forward scores produced less agreement errors than participants with low digit span forward scores. There was a significant effect for digit span backward scores, because participants with high digit span backward scores produced less agreement errors than participants with low digit span backward scores. There was a significant effect for the place of the item in the list, because participants produced less agreement errors by items later in the list than by items earlier in the list.

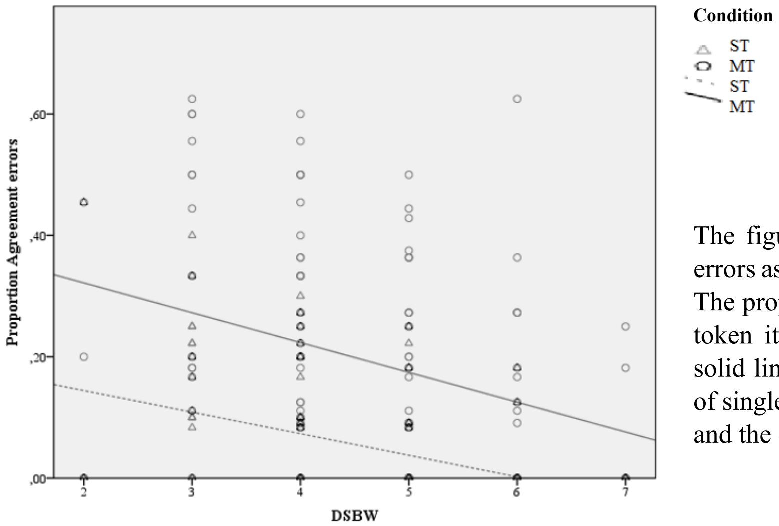
I. Appendix: Table of Model 5 with digit span backward scores

Table with coefficients and probability estimates (Coefficient, Standard Error, Wald Chi-Squared Z-score, p-value) of the fixed effects in the extra version of the logit mixed effect model with the digit span backward scores (Model 5: accuracy ~ agegroup*distributivity*backward.digit.span + (1|subject) + (1|item) + (0+distributivity|subject) + (0+backward.digit.span |item))

Predictor	Coeff.	Std. Error	Wald Z	*p*
(Intercept)	-2.29	0.32	-7.07	1.51e-12***
Age group	0.20	0.42	0.49	0.62
Distributivity	2.81	0.52	5.37	7.85e-08***
Backward digit span	-0.98	0.19	-5.17	2.39e-07***
Age group: Distributivity	-1.50	0.52	-2.87	0.004**
Age group: Backward digit span	0.21	0.36	0.57	0.57
Distributivity: Backward digit span	0.44	0.24	1.85	0.065~
Age group: Distributivity: Backward digit span	-0.18	0.46	-0.40	0.69

~ Effect is marginal significant (almost 0.05)
** Effect is significant on the 0.01 alfa level
*** Effect is significant on the 0.001 alfa level

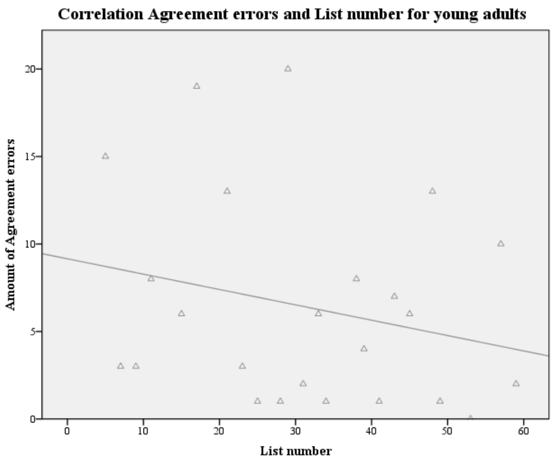
The log-likelihood (the fit) of the model was -489.8. A Wald Chi-Squared Test was taken of the model to investigate main and interaction effects. There was no significant main effect for age group. There was a significant main effect of distributivity, whereby participants produced more agreement errors in the multiple token condition than in the single token condition. There was a significant main effect for digit span backward scores, whereby participants with high scores produced less agreement errors than participants with low scores. There was a significant interaction effect for age group and distributivity, whereby the difference in produced agreement errors between the single token condition and multiple token condition for elderly people was significantly smaller than the difference between the produced agreement errors in the single token condition and multiple token condition for young adults. The interaction between distributivity and digit span backward score was marginal significant (*p*=0.065), but it is interesting to look at which direction this interaction went. The difference in producing agreement errors between the single token condition and multiple token condition was bigger for participants with low digit span scores than for participants with high digit span scores. Low digit span participants had a bigger distributivity effect, and high digit span participants had a smaller distributivity effect.

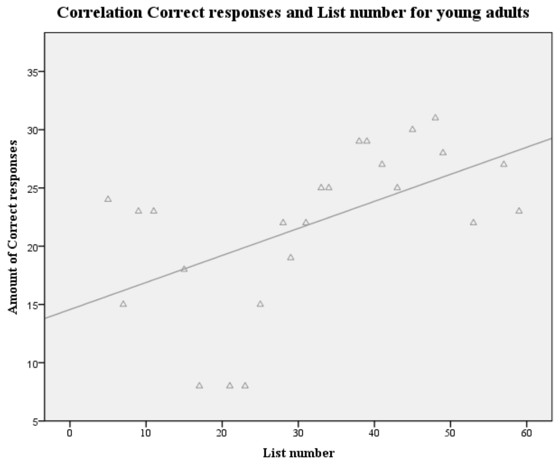


The figure gives the proportion of agreement errors as a function of digit span forward scores. The proportion of agreement errors of multiple token items is represented by circles and the solid line. The proportion of agreement errors of single token items is represented by triangles and the dotted line.

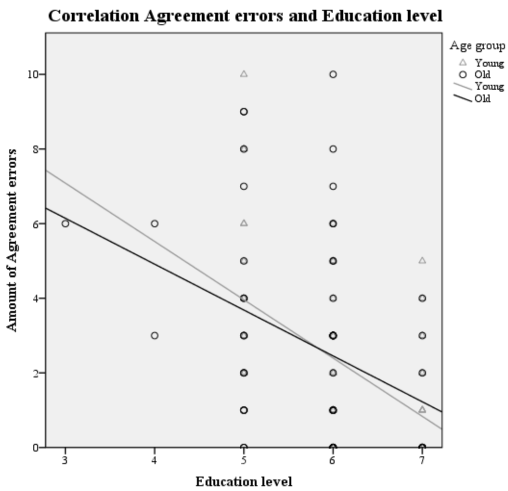
J. Appendix: Correlation analysis figures

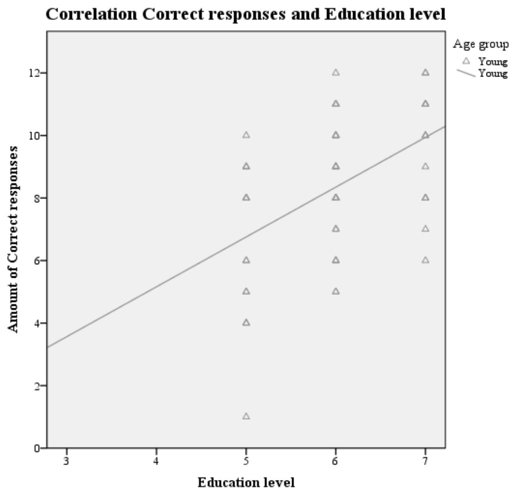
I: Scatter plots of the significant correlations between list number and the amount agreement errors (left) and correct responses (right) of the young adults.



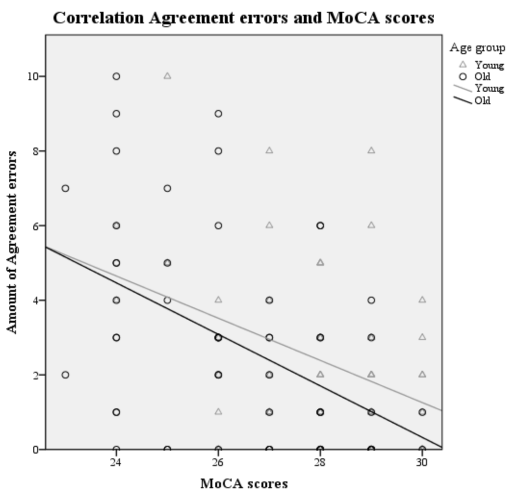


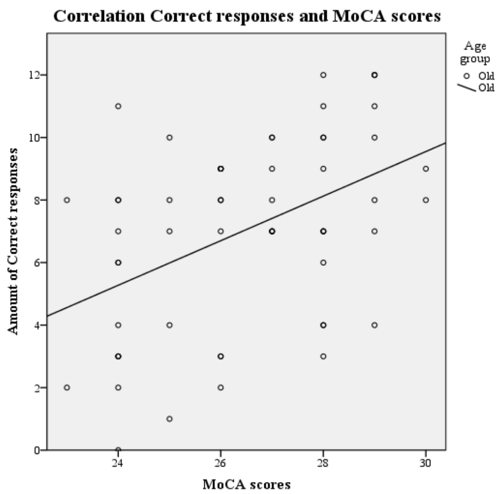
II: Scatter plots of the significant correlations between the education level and the amount of agreement errors (left, both young adults and elderly people) and the amount of correct responses (right, only young adults).



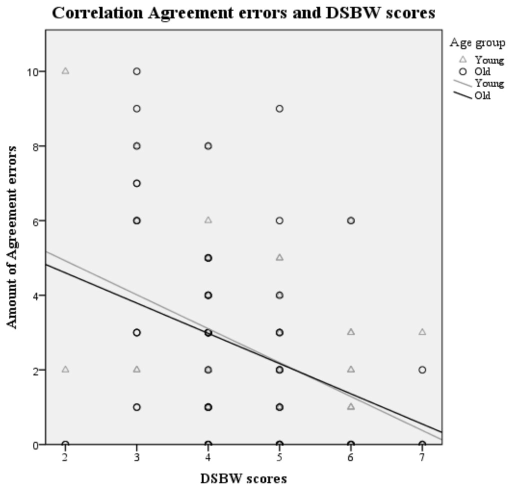


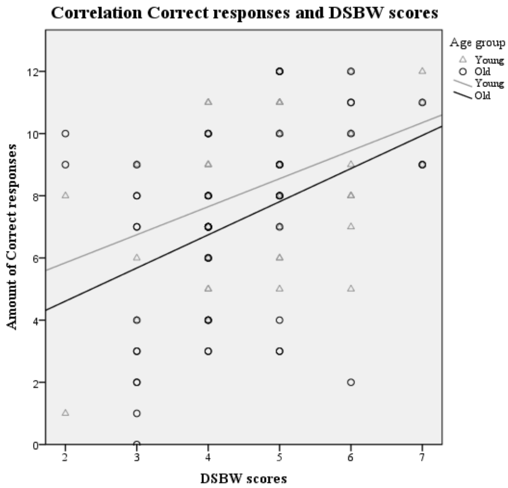
III: Scatter plots of the significant correlations between the MoCA scores and the amount of agreement errors (left, both young adults and elderly people) and the amount of correct responses (right, only elderly people).





IV: Scatter plots of the correlations between the Digit Span Backward scores of the young adults and elderly people and the amount of agreement errors (left) and the amount of correct responses (right).





K. Appendix: Mutual correlations

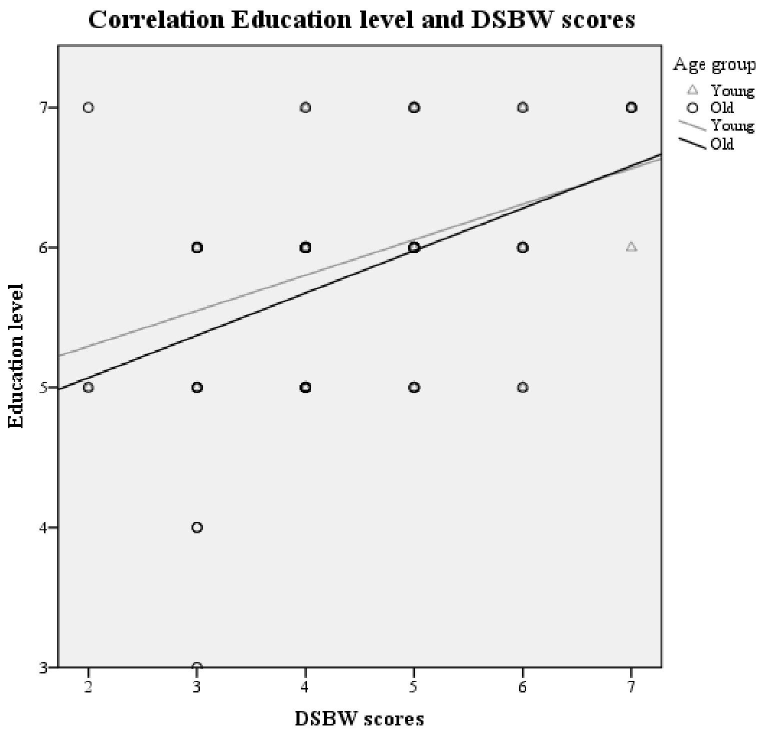
I: Table with correlations (Pearsons r) with p-values of educational level, MoCA scores, digit span forward scores (DSFW), and digit span backward scores (DSBW) for the total group, and the divided age groups (Young, Old).

Correlations

		Edu. level	MoCA	DSFW	DSBW
Educational level					
Total	r	1	0,256	0,249	0,422
	p		0,004**	0,006**	0,000**
Young	r	1	0,185	0,176	0,378
	p		0,148	0,166	0,002**
Old	r	1	0,239	0,255	0,425
	p		0,068	0,051	0,001**
MoCA scores					
Total	r	0,256	1	0,176	0,252
	p	0,004**		0,053	0,005**
Young	r	0,185	1	0,041	0,275
	p	0,148		0,750	0,029*
Old	r	0,239	1	0,095	0,114
	p	0,068		0,475	0,390
DSFW					
Total	r	0,249	0,176	1	0,565
	p	0,006**	0,053		0,000**
Young	r	0,176	0,041	1	0,425
	p	0,166	0,750		0,001**
Old	r	0,255	0,095	1	0,652
	p	0,051	0,475		0,000**
DWBW					
Total	r	0,422	0,252	0,565	1
	p	0,000**	0,005**	0,000**	
Young	r	0,378	0,275	0,425	1
	p	0,002**	0,029*	0,001**	
Old	r	0,425	0,114	0,652	1
	p	0,001**	0,390	0,000**	

* Correlation is significant on the 0.05 alfa level
**Correlation is significant on the 0.01 alfa level

II: Scatter plot of the correlations between the digit span backward scores (DSBW) and education level of the young adults and elderly people.



III: Scatter plot of the correlations between the digit span backward scores (DSBW) and digit span forward scores (DSFW) of the young adults and elderly people.