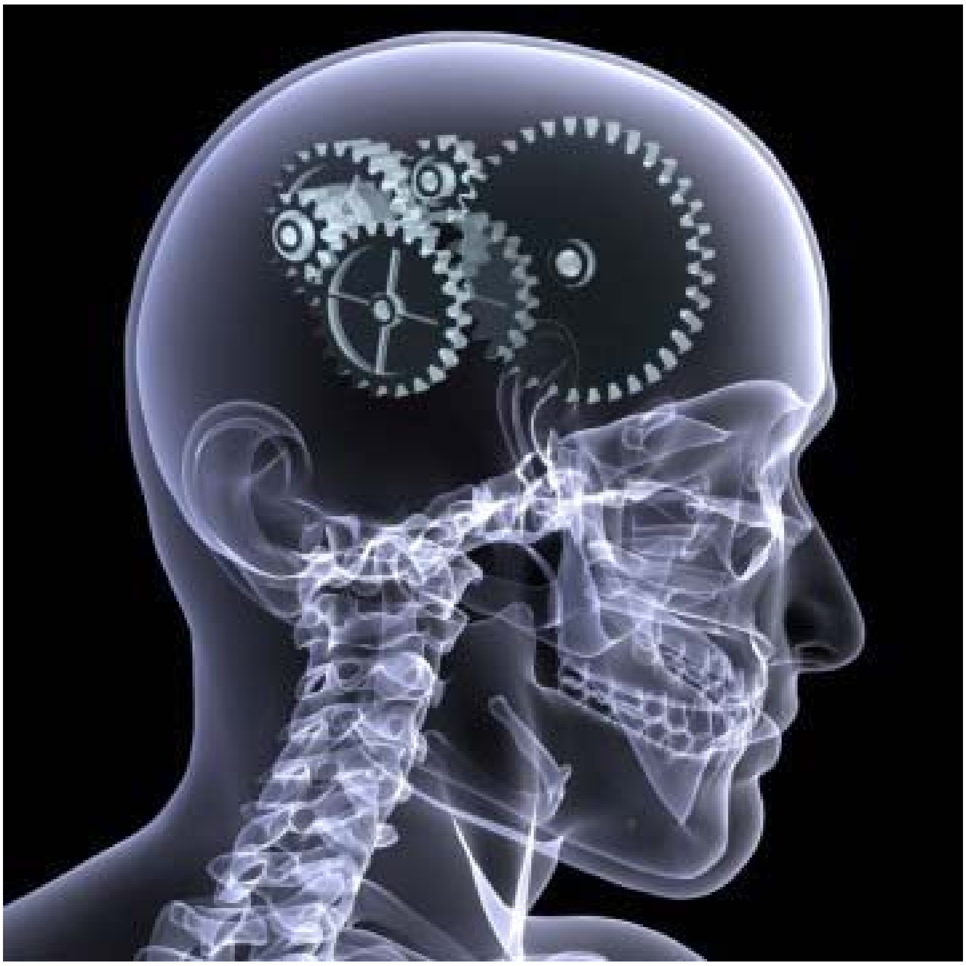
Limited working memory capacity and second language acquisition

Kasper Soeters



Supervisors: Elma Blom & Jan Don

August 2009

Table of contents

1. Introduction

Understanding and speaking a language seems to be one of the most complex things that human beings are capable of. Thousands of words and many subtle grammatical rules for modifying and combining these words must be learned, yet nearly all people manage to master their native language during childhood and are in fact already quite competent with language at the age of three or four years old. However, if people start learning a second language at a later age, the learning process advances with much more difficulty. Over the years, we seem to lose our ability to learn a language as efficiently and effortlessly as we do from birth on. The last few decades, several researchers have set out to try to explain this finding, yet the exact mechanism behind the decline in the ability to learn a language as people get older still remains unclear.

Recently, a number of second-language experimenters have suggested that working memory plays a key role in foreign language acquisition (e.g., Van den Noort, Bosch, & Hugdahl, 2006). Working memory is the part of cognition responsible for the active processing of information, including filtering sensory information from the environment and retrieving information from long-term memory. It is proposed that this working memory system is also dedicated to language processing and that it has a limited capacity (e.g., Just & Carpenter, 1992). Language processing draws on this limited capacity, and as soon as a shortage of capacity originates, language processing will slow down and language comprehension and production will deteriorate.

From this, it can be assumed that, since second language acquisition occurs more slowly than native language acquisition,apparently the working memory capacity is exceeded more often during second language processing than during native language processing. Based on the ideas of Just and Carpenter, that is exactly what the American psycholinguist Janet McDonald (2000) proposes in her ‘limited capacity theory’. She states that the processing of a second language poses a higher load on working memory than the processing of a native language. Therefore, the working memory capacity is exceeded sooner during the processing of a second language than during the processing of a native language, resulting in the breakdown of the second language acquisition process. If this is true, then working memory is a central component of second language aptitude, and examiningits influence on second language acquisition may therefore be a first step towards the unraveling of the mechanism behind the decline in the ability to learn a language fluently at a later age. The current study set out to test the assumed influence of working memory on second language acquisition.

In the first section of this paper I will give a theoretical background of this study. First of all I will describe the concept of working memory more thoroughly, after which I will mention the assumed influence of the working memory system on second language acquisition, with a description of McDonalds limited capacity theory deriving from that. I will also provide a linguistic framework in which McDonalds theory can be tested. At the end of this section, I will mention the predictions and the hypotheses of this study. In the second section I will describe the used materials and methods in this study. The results will be mentioned in the third section and these will be discussed in the fourth and final section of this paper.

2. Theoretical background

2.1 Working memory

The term working memory was first introduced by Baddeley and Hitch (1974), to emphasize that temporary storage is not just, as was previously thought, a station on the way to long-term memory; it is the way we store information while we are working with or attending to it. Baddeley and Hitch first distinguished three components of working memory: a phonological loop (which stores auditory information, including words), a visuospatial sketchpad (which stores visual information), and the central executive (which directs attention toward one stimulus or another and determines which items will be stored in working memory). A fourth subsystem, the episodic buffer, has recently been proposed (Baddeley, 2003). The episodic buffer is assumed to form a temporary storage system that allows information from the subsystems to be combined with that from long-term memory into integrated chunks. This system is assumed to form a basis for conscious awareness (Fig.1).

The phonological loop, which, in collaboration with the linguistic portions of the central executive and the episodic buffer, is concerned with language processing and therefore is of most importance here, was proposed to exist of two subcomponents: a temporary storage system and a subvocal rehearsal system. The temporary storage system was proposed to hold memory traces over a matter of seconds, during which these traces decay, unless the subvocal rehearsal system refreshes them. The rehearsal system maintains both auditory and visual linguistic information within the store in a comparable manner. The system registers visual information in a way that involves subvocalization of the visually presented items, provided these items can be named. So, these visually presented items are being registered as if they were acoustically presented. Hence, if a sequence of letters is visually being presented to a subject for immediate recall, then despite their visual presentation, subjects will subvocalize them, and hence their retention will depend crucially on their acoustic or phonological characteristics. Thus, while subjects can readily recall a sequence of letters such as *B, W, Y, K, R, X*, they are likely to have considerable difficulty in retaining a sequence of letters with similar sounding names, such as *T, C, V, D, B, G,* whether these items are being presented visually or acoustically (Baddeley, 2003).

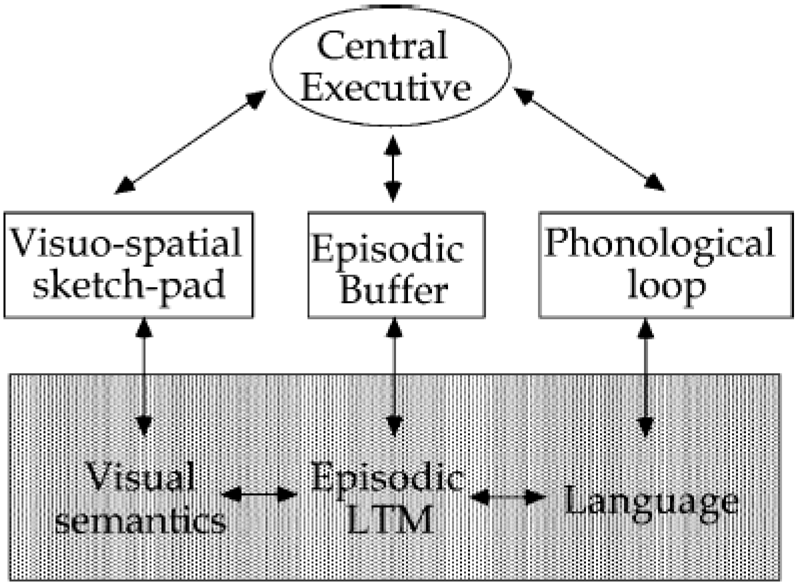


Fig.1 – Baddeley's four component model of working memory in which visual and verbal subsystems are controlled by a central attentional executive and stored in the episodic buffer. The shaded areas refer to long-term systems, which involve stored information that is capable of interacting with the working memory system.

Although nowadays Baddeley's model of working memory still is the most influential, other feasible models of working memory are available, such as the one put forth by Just and Carpenter (1992). In their theory, the working memory corresponds approximately to the part of the central executive in Baddeley's theory that deals with language comprehension. A contrast with Baddeley's model is that the working memory in Just and Carpenter's theory does not include modality-specific buffers, such as the phonological loop. However, in the current experiment, I focus on the interface between (second) language processing and working memory, and therefore the modularity discussion is not of importance here. More interesting here are the storage and processing functions of working memory. Just and

Carpenter's theory relies on the integration of these functions, and since McDonald used Just and Carpenter's model as a basis for her theory, I hereby adopt this view.[1]

Just and Carpenter present a model in which storage and processing are 'fueled' by activation. They state that during comprehension of a sentence, each element (such as a word, phrase, grammatical structure, or object in the external word) reaches a certain associated activation level. This process of activation occurs by virtue of being encoded from new language input or by retrieval from long-term memory. An element is considered part of working memory, as long as its activation level is above some minimum treshold value. Consequently, if the treshold value is reached, the element is available to be operated on by various processes. If, however, the system is being 'overloaded', e.g. if a comprehension task requires more activation than is available to the system, then some of the activation that is maintaining old elements will be used to help out the system in processing the new task, thereby producing a kind of forgetting by replacement. So, activation underlies both computation and the maintenance of information. Elements are added to or deleted from working memory by changing their activation level appropriately. It is important to note here, that although the storage and processing components are facilitating each other through activation of linked elements, these components are also in a way competing with each other. A trade-off takes place between storage and processing when the activation maximum of the system is about to be exceeded. That is, when the task demands are about to exceed the available resources, both storage and computational functions are degraded. As a result, the processing will breakdown and some elements may be forgotten.

In this framework, capacity can be expressed as the maximum amount of activation available in working memory to support the storage and processing of the representational elements. A crucial proposition in Just and Carpenter's theory is, that individuals vary in the amount of activation they have available for meeting the computational and storage demands of language processing. So, each individual has a different working memory capacity. Therefore, it is important in an experiment such as the current one, in which the relationship between language processing and working memory is being tested, to assess the participants' working memory capacities, to be able to find out whether differences between certain participants are actually due to the distribution of the experimental variables, or if they are perhaps due to differences in their working memory capacity.

[1] However, this does not mean that I am adopting it as the only accurate model available.

2.2 Limited capacity theory

Janet McDonald (2000) bases her hypothesis about the influence of working memory on second language acquisition on the idea of a limited working memory capacity that differs for each individual. McDonald tries to test the critical period hypothesis (Lenneberg, 1967), which claims that there is a specific and limited time period for language acquisition. According to Lenneberg, “the language acquisition device, like other biological functions, works successfully only when it is stimulated at the right time.” (Lenneberg, 1967, p.19) Although originally formulated to explain first language (L1) acquisition data, over the years, the critical period hypothesis (CPH) has also been applied to studies of second language (L2) acquisition. Nowadays however, claims for a critical period for second language acquisition are not without controversy. Critics of the CPH have argued that decreased L2 learning ability with age is not inevitable, and that the data supporting the CPH can be explained by other means (e.g., Juffs & Harrington, 1995).

McDonald is one of these critics and she tries to test the CPH by examining the grammar acquisition of native Spanish early and late acquirers of English as well as that of Vietnamese early and child acquirers of English (McDonald, 2000). One implication of the CPH is that all learners within the critical period should be capable of L2 mastery, while all learners outside of the critical period should fail to gain native-like mastery of the L2. Recently, however, a few studies found that the nature of the L1 might be an important factor that may influence L2 mastery of early as well as late learners, which may conflict with the implication of the CPH. The CPH is not able to explain the findings that some late L2 learners, with a L1 that does not differ much from their L2, are able to gain native-like competency levels, and that some early L2 learners are never able to reach native-like competency levels in their L2 (Yew 1995, as cited in Bialystok, 1997; Bialystok and Miller, 1998, as cited in Bialystok & Hakuta, 1999).

To test the influence of the L1 on L2 mastery, McDonald conducted an experiment in which native Spanish early and late acquirers of English as well as native Vietnamese early and child acquirers of English made grammaticality judgments of sentences in their second language. According to her findings, native Spanish early acquirers were not distinguishable from native English speakers, whereas native Spanish late acquirers had difficulty with almost all aspects of grammar tested. Native Vietnamese early acquirers had difficulty with those aspects of English that differ markedly from Vietnamese, while native Vietnamese child acquirers had more generalized problems, similar to those of native Spanish late acquirers.

Therefore, McDonald concludes that native language appears to make a difference for early acquirers, whereas a later age of acquisition causes a more general problem. Other studies indicate that the similarity of the L1 to the L2 plays a role even for late acquirers (Kellerman, 1995).

As has been mentioned, the CPH does not seem to be able to account for these findings, and therefore McDonald proposes an alternative to explain the recent data: a simple processing theory, which should capture both the effects of L1 on L2 as well as the general, age-related declines in syntax. In this *limited capacity theory*, McDonald claims that success in an L2 is dependant on individual working memory capacities. Language performance breaks down as soon as an individual's capacity is exceeded. This breakdown occurs more often for people with a smaller working memory capacity than for people with a larger working memory capacity. McDonald proposes that acquirers with a native language quite different from the target language would already be working under a higher working memory load than learners whose L1 differs less from the L2 would, and therefore their working memory capacity is exceeded earlier. Such an explanation would also account for the major differences reported between the data of the late L2 learners. According to McDonald, individual differences in memory capacity may also explain why some older L2 learners are able to achieve native levels of performance whereas others are not. Although some of the aspects of McDonald's research seem to be questionable (such as the comparability between her subject groups), the limited capacity theory as an alternative to the CPH seems to be interesting enough, and therefore it is unfortunate that McDonald fails to describe how exactly the trade-off between second language learning and working memory might work.

2.3 Omission of verbal inflection

To be able to find out how such a trade-off might work, it is necessary to find a feasible linguistic framework, in which second language acquirers can be tested. There is much evidence that second language learners frequently omit verbal inflection in their speech (Ionin & Wexler, 2002). McDonald (2000) only tested her participants on grammatical judgments and did not examine their speech production, but she does state that her limited capacity theory should also be able to capture the effects of a high processing load on speech production. She mentions that non-native language learners appear to have trouble decoding the surface form of the language, and that these problems "would add to the processing load of the learner and, depending on the level of the load, cause possible disruptions in their ability to *produce or judge* grammatical level information" (McDonald, 2000, p.416;

emphasis by KS). Therefore, following this train of thoughts, it seems plausible to assume that the often found omissions of verbal inflection occur as soon as the processing of the second language puts too high a load on working memory so that the working memory capacity is exceeded. Therefore, the present study focused on the omission of verbal inflection.

Previous research by Prévost and White (2000) and by Ionin and Wexler (2002) also focused on the omission of verbal inflection by second language learners with English as the second language. These researchers aimed at clarifying whether the optionality in the use of verbal inflection indicates a major impairment to the interlanguage grammar (ILG) in the domain of functional categories or whether it is the consequence of some more superficial problem. Based on spontaneous production data (used by Prévost and White as well as by Ionin and Wexler) and data from a grammaticality judgment task (used only by Ionin and Wexler) in English, they argue that omission of inflection is due to problems with the realization of surface morphology, rather than to feature impairment. This argument has been proposed independently by several researchers, of which some have begun to investigate the properties of missing surface inflection in some detail, such as Haznedar and Schwartz (1997).

Haznedar and Schwartz aimed at proving whether child L2 acquirers go through an optional infinitive state during their acquisition process or not. The Optional Infinitive (OI) (Wexler, 1994) or Root Infinitve (Rizzi, 1993/1994) phenomenon is a variation in the incidence of verbal inflection that is often been found in first language (L1) acquisition. Hereby, children's main clauses vary between having a finite or a non-finite main verb. However, this variation in the use of finiteness seems to be non-random. Typically, when the child uses non-finite forms in place of finite ones, these are actually found in structural positions typical of non-finite verbs, and therefore it is assumed that the non-finite forms used are indeed non-finite, and not bare forms exhibiting properties of finite verbs. Also, when the child uses finite forms, these are found in structural positions of finite verbs. So, following the OI account, the issue is not that the child does not know the difference between finite and non-finite forms, but rather that the child still lacks the knowledge that main clauses must be finite. After examining the data from a Turkish child learning English as L2, Haznedar and Schwartz argued against an OI phenomenon in L2 acquisition, and that there was no evidence for a syntactic deficit in their subject's English. Rather, the child's non-finite morphology was indicative of missing inflection; meaning that the child had difficulties with identifying the appropriate morphological realization of the existing functional categories. From this, Prévost and White propose the Missing Surface Inflection Hypothesis (MSIH), which states that the problems with verbal inflection during L2 acquisition lie in mapping from the abstract features to the corresponding surface morphology, while the grammar of L2 learners remains intact and actually does contain the relevant abstract categories and features.

If we take these findings into account, as well as the aforementioned claim by McDonald that both speech production and grammatical judging could be affected by a high processing load during the language acquisition process of L2 learners, it seems wise to keep in mind that in an experiment such as the current one (in which L2 learners are tested on verbal inflection under two conditions: with and without an extra working memory load) it is important to test both the comprehension and the production of verbal inflection. Since both Prévost and White (2000) and Ionin and Wexler (2002) state that the abstract categories and features in the grammar of the L2 learners are still intact, contrary to what McDonald claims, we should expect only few or perhaps even no mistakes in a comprehension task, even with an extra load on working memory during that task, while we should expect a clear increase in mistakes in a production task when working memory is loaded as compared to the same task without such a load, since the problems with verbal inflection are supposed to lie in the mapping from the abstract features to the surface morphology.

2.4 Predictions and hypothesis

With the previous theories in mind, it is possible to postulate some predictions about the current working memory experiment in which the participants were both tested on comprehension as well as on production of verbal inflection under two conditions: one with, and one without an extra working memory load. First of all, it may be assumed that L2 learners perform worse on all tasks and under every condition as compared to the native speakers, since the native speakers have already reached their ultimate attainment level in the language in which they are being tested and the L2 learners have not. However, the sentences in the current experiment were matched to the proficiency level of the L2 learners, and therefore, the differences in performance on both tasks between native speakers and L2 speakers are predicted to be very small, if not nonexistent, at least in the task without an extra working memory load.

Secondly, based on the limited capacity theory of McDonald (2000), differences between the performances of L1 and L2 speakers are expected to occur when working memory is loaded, since the working memory capacity of L2 learners is expected to be exceeded earlier than that of native speakers. This goes for both the comprehension and the production task. However, since it is proposed by Ionin and Wexler (2002) that L2 learners mostly have trouble with the mapping from the abstract features to the surface morphology, while the grammar is still intact, I expect to find larger differences in the production task when working memory is loaded than in the comprehension task. So, L1 and L2 speakers are expected to perform similarly on both tasks in the condition without an extra working memory load, but differences between the two groups are expected to occur in the condition with such a load, especially in the production task.

Lastly, I predict that both the native speakers and the L2 acquirers should perform better in the comprehension task as well as in the production task without an extra working memory load than in the task with such a load. So, I expected all the participants to make more mistakes and to have increased response times in the tasks with an extra working memory load as compared to the same tasks without such a load.

To summarize:

- All participants are expected to perform similarly on both tasks in the condition without an extra working memory load.
- All participants are expected to perform worse on both tasks in the condition with an extra working memory load as compared to the condition without an extra working memory load.
- L2 learners are expected to perform worse than native speakers on both tasks in the condition with an extra working memory.
- The differences in performance between native speakers and L2 learners are expected to be largest when working memory is loaded during the production task.

Table 1 - Predictions

	WM not loaded	**WM loaded**
Comprehension task	No significant difference between L1 and L2 performance	Performance decreases for both L1 and L2 speakers, but L2 learners are expected to perform worse than L1 speakers
Production task	No significant difference between L1 and L2 performance	Performance decreases for both L1 and L2 speakers, but L2 learners are expected to perform far worse than L1 speakers

3. Method

3.1 Participants

Nineteen native Dutch-speaking students (twelve females and seven males) and fifteen native English-speaking students (ten females, five males) participated in this study. The mean age of the Dutch participants was 17;9 (i.e., 17 years and 9 months), and that of the native English speaking subjects 17;3. Both the Dutch and native English speaking participants had no known history of any neurological or psychiatric impairment. All of them had normal or corrected-to-normal vision. None of them were dyslectic. They gave their written informed consent and received a small eatable reward for their participation.

The participants were divided into two groups on the basis of their native language. The experimental group consisted of the 19 Dutch participants. These students all started learning English in an academic atmosphere between the ages of nine and twelve years old and have continued to do so for the last five to eight years. The control group consisted of the fifteen native English-speaking participants, all of whom reported to have learned English as their first language and resided in a country with English as the main language for at least 90% of their life. All participants (both Dutch and English speaking) studied or had studied at (pre-) university level.

At the time of the study, all Dutch participants were able to speak and understand English, but were not entirely comfortable speaking English. These subjects were chosen on the basis of their proficiency level in English. On the one hand participants were needed who were able to understand the sentences used in both tasks in the current study, on the other it was to be prevented to test participants who had already reached their ultimate attainment level. As McDonald (2000) mentions, some second language learners are able to reach native-like competency levels. Following McDonald's train of thought, it may be so that the working memory load on language processing decreases as second language fluency increases. That is, at some point the acquirers might reach a competency level at which second language processing no longer inhibits working memory any more than first language processing would. Although this claim is purely hypothetical, it seemed best to select L2 participants, who had not yet reached their ultimate attainment level, so as to be able to avoid the potential aforementioned effect.

3.2 Materials

In the current study three tasks were used: a working memory task, a grammaticality judgment task, and a production task. The test that was used to assess working memory is the n-back task, which is consistently regarded as a valid indicator of working memory capacity (e.g., Owen et al., 2005). In this task, the participants viewed a serial presentation of stimuli and judged whether each is a repetition of the stimulus that appeared *n* trials previously. For example, in a three-back test, which was used in the present study, the fourth item of the series A B C A is a three-back match, whereas the fourth item of the series X B C A is not. A practice list of 14 letter items and an experimental list of 35 letter items were composed. Of the 35 letter items in the experimental list, 12 items were a three-back match. The Dutch participants received the instructions for this task in Dutch, and the native English-speaking participants received theirs in English.

The sentences in the grammaticality judgment task were based on those used by McDonald (2000). Of the 104 sentences used in the current study, 54 were directly taken from McDonald's list. McDonald divided her sentences into several categories. The 54 sentences used from McDonald's list, were taken from the categories in which the subjects in her study made the most mistakes, i.e. 'past tense', 'plurals', and 'third person subject-verb agreement'. These categories were also selected to fit the assumptions of the Missing Surface Inflection Hypothesis. The 50 other sentences used in the present experiment were based on the 54 sentences taken from McDonald and were equally spread out over the aforementioned categories. The 104 sentences in the present study were divided into four lists: A, B, C, and D (see Appendices). Each of these lists consisted of 26 sentences, of which 13 were grammatical, and 13 were not. Six of these 26 sentences were filler items. The sentences in list A were exactly the same as those in list C, except for their grammaticality: those that were grammatical in list A were altered to make them ungrammatical in list C, and vice versa. The same goes for list B and list D: the sentences that were grammatical in list B were altered to make them ungrammatical in list D, and vice versa. Therefore, the participants in the current study were only tested on a combination of list A and list B or on a combination of list C and list D. So, every participant judged 52 sentences on their grammaticality. The sentences were presented in a random order to every participant.

A gap task was used to evoke the production of words by the participants. Two lists (list 1 and list 2, see Appendices), each consisting of 22 sentences, were composed. In each of these sentences, three dots appeared instead of a target word, which was to be formulated on the basis of the sentence context and spoken out loud by the participants. The 44 sentences were divided into the same categories as the sentences in the grammaticality judgment task ('past tense', 'plurals', and 'third person subject-verb agreement'). These sentences were also comparable to the ones used in the grammaticality judgment task considering their difficulty level. I deliberately chose not to use the exact same sentences as were used in the grammaticality judgment task, to exclude a possible priming effect in the data. The sentences in all categories were equally divided over both lists.

Lastly, it is to be mentioned that the sentences in the grammaticality judgment task and in the production task were matched to the proficiency level of the Dutch participants. It was important to construct sentences that were not too difficult for these participants, since we wanted them to make few, if any, errors in the tasks without an extra working memory load. If they would already make many mistakes under this condition, the expected difference between the performance on the tasks with and without an extra working memory load could only be very small (from many errors to even more errors), and therefore it would be difficult to ascribe the assumed decrease in performance (when working memory was loaded as compared to the same tasks without an extra load) to the influence of the extra working memory load. However, the sentences could not be too easy as well, since we wanted them to load the working memory in such a way, that an extra working memory task could possibly lead to an overloaded working memory capacity.

3.3 Procedure

At the beginning of the experiment, the participants were told that the experiment consisted of three parts: a 3-back test, a grammaticality judgment task, and a production task. The purpose of the first test was to assess the participants' working memory capacity. The 3-back task involved presenting a black letter at the centre of a grey screen. The participants were asked to press a button as soon as they saw a letter, which they already had seen three screens previously. Each letter was presented in isolation in slide show presentation for exactly three seconds, after which a blank screen appeared automatically. This blank screen was presented for 250 milliseconds, after which the next letter appeared automatically. The time frame of three seconds allowed the participants to process the presented letter, but minimized the time to mentally rehearse the letters kept in working memory. All participants started of with a practice session, consisting of 14 letter items, after which the experimental session started, in which 70 letter items were presented. The number of correct responses, incorrect responses, and missed items all together formed the participants' score on this working memory test.

After the 3-back test, participants were told they would read English sentences, and it was their task to judge whether each sentence was a grammatical or an ungrammatical English sentence. They recorded their response by pressing either the YES/JA (grammatical) or NO/NEE (ungrammatical) button on the keyboard of a laptop. Response and response latencies from sentence onset were recorded by the computer for later analysis, using the program DmDX. All participants were notified that their response times were monitored, but it was stressed to them that making the correct judgment was more important than responding as fast as possible. Participants were asked to use only their preferred hand while pressing the keys. All participants were told the grammaticality judgment task consisted of two parts: one with and one without an extra working memory task. The meaning of the extra working memory task was to load working memory during the grammaticality judgment task. Both parts consisted of 26 sentences. The 26 sentences of the part without an extra working memory task were presented continuously. The 26 sentences of the part with an extra working memory task were divided into two sets of 13 sentences. Before each set, a list of three words was presented. To be able to load their working memory, participants were asked to remember these three words and to say one of these words out loud as soon as they saw that one in a sentence presented to them. Meanwhile, they continued to judge the sentences on their grammaticality. After each set of 13 sentences, the participants were asked to name the last 3 words they needed to remember beforehand, so that I was able to verify whether they had actually kept their working memory loaded during the whole experiment.

The third and final part of the experiment consisted of the gap task. This task also consisted of two parts: one with and one without an extra working memory task. In the gap task, two times 22 sentences were presented to the participants. During the first part, participants were asked to fill in the word left out of the sentence and to say that word out loud. Three dots showed the participants the place in the sentence where the word was meant. During the second part, the gap task needed to be fulfilled in the same way as before, and an extra working memory task that was comparable to the one used in the grammaticality judgment task was introduced. The second part of the gap task consisted of two sets of 11 sentences. Before each set, the subjects again were asked to remember 3 words presented to them. Also, they needed to say one of these words out loud as soon as they came across one in a sentence presented to them. After each set of 11 sentences, the participants again were asked to name the last 3 words they needed to remember.

4. Results

Eighteen Dutch subjects and fourteen native English-speaking participants were successfully tested on all three tasks. Due to technical problems with the minidisc player, which was used to record the participants' utterances, the production task data for two subjects turned out to be nonexistent. These two participants were equally divided over the experimental and the control group, so, the analysis of the results of the production task is based on the data of eighteen Dutch subjects and on the data of fourteen native English-speaking subjects. The analyses of the results of the 3-back task and of the grammatical judgment task are based on the data of all the nineteen Dutch participants and the fifteen native English-speaking participants.

Let's recall the predictions before analyzing the results:

- All participants are expected to perform similarly on both tasks in the condition without an extra working memory load.
- All participants are expected to perform worse on both tasks in the condition with an extra working memory load as compared to the condition without an extra working memory load.
- L2 learners are expected to perform worse than native speakers on both tasks in the condition with an extra working memory.
- The differences in performance between native speakers and L2 learners are expected to be largest when working memory is loaded during the production task.

4.1 Grammaticality judgment scores

Tables 2 and 3 show the mean amount of errors, the mean amount of both the errors and the missed items, and the mean response times for the Dutch and the native English-speaking participants respectively for the grammaticality judgment task. The standard deviations are also shown. In calculating the participants' mean response times the response time for the first sentence presented to them was always discarded, as well as any response times faster than 1000 ms.

	Dutch participants [N=19]		
	Mean amount of errors	*Mean amount of errors and missed items combined*	*Mean response times*
Without WM load	4.05 [2.20]	6.00 [2.94]	3.79 [0.59]
With WM load	5.47 [3.60]	8.11 [3.87]	4.11 [0.62]

Table 2 – The mean amount of errors [standard deviations], the mean amount of errorsand the missed items combined [standard deviations], and the mean response times for the Dutch participants [standard deviations (s)] in the task without an extra working memory load and the task with such a load.

	Native English participants [N=15]		
	Mean amount of errors	*Mean amount of errors and missed items combined*	*Mean response times*
Without WM load	2.33 [1.49]	2.73 [2.02]	2.85 [0.68]
With WM load	3.67 [3.13]	3.93 [3.08]	3.35 [0.55]

Table 3 – The mean amount of errors [standard deviations], the mean amount of errors and the missed items combined [standard deviations], and the mean response times [standard deviations (s)] for the native English speaking participants in the task without an extra working memory load and the task with such a load.

According to the first prediction, there should not be a significant difference between the scores of the Dutch and those of the English-speaking participants in the task without an extra working memory load. However, as can be seen in Tables 2 and 3, the Dutch do perform worse than the English-speaking participants under these conditions (e.g., 6.00 vs. 2.73 for the amount of errors and missed items combined). A Mann-Whitney Test shows that these results are significant, and this goes for all three categories: the mean amount of errors ($U = 78.000$, $N1 = 19$, $N2 = 15$, $p = 0.025$, two-tailed), the mean amount of errors and missed items combined ($U = 54.500$, $N1 = 19$, $N2 = 15$, $p = 0.002$, two-tailed), and the mean response times (U = 42.000, N1 = 19, N2 = 15, $p < 0.0005$, two-tailed). These results do not confirm the first prediction.

Table 2 and Table 3 also show that the mean amount of errors, the mean amount of errors and the missed items combined, and the mean response times all increase when working memory is loaded in both groups. If the scores of the grammaticality judgment task without an extra working memory task are compared to those of the task with such a load for all participants in a paired *t*-test, it is found that, again, these results are significant for all three categories (for the errors only: $t = 2.757$, df = 33, $p = 0.009$, one-tailed; for both the errors and the missed items: $t = 4.010$, df = 33, $p < 0.00025$, one-tailed; for the response times: $t = 5.460$, df = 33, $p < 0.00025$, one-tailed). This confirms the second prediction, which states that an extra working memory task during the grammaticality judgment task would lead the participants to make more mistakes and to react more slowly. It therefore may be concluded, that the used working memory task indeed loaded working memory.

If the scores of the Dutch and the English-speaking participants on the task with an extra working memory load are compared, it is clear that under these conditions, the Dutch also tend to make more mistakes and to be slower than the English-speaking participants. A Mann-Whitney Test again shows that these results indeed are significant (only errors: U = 92.500, N1 = 19, N2 = 15, $p = 0.083$, two-tailed; both errors and missed items: U = 49.500, N1 = 19, N2 = 15, $p = 0.001$, two-tailed; response times: U = 55.000, N1 = 19, N2 = 15, $p = 0.002$, two-tailed). These findings do confirm the third prediction.

	Increase in number of errors	*Increase in number of errors and missed items*	*Increase in response time*
Dutch	1.42	2.11	0.32
Native English speaking	1.34	1.20	0.50

Table 4 - The mean increase in the amount of errors, in the mean increase in the amount of both the errors and the missed items, and in the mean response times in the grammaticality judgment task with a working memory load as compared to the same task without such a load, for both the Dutch and the native English speaking participants.

We also expected a slightly more significant difference between the two participant groups when working memory was loaded than when it was not. Table 4 shows the mean increase in the amount of errors, in the amount of the errors and missed items combined, and in the response times in the task with a working memory load as compared to the task without such a load, for both the Dutch and the native English speaking participants. Although the performance of the Dutch seems to decrease more than that of the L1 participants (e.g., 2.11 vs. 1.20 for the amount of errors and missed items combined), the mentioned Mann-Whitney Tests however show that under both conditions (with and without an extra working memory task), the differences are highly significant (e.g., p = 0.001 and p = 0.002 for the number of errors and missed items combined). Therefore, relatively speaking, the difference between the performances of the two groups seems to stay about the same. This is not what was expected. Another remarkable finding is the result that the response times of the L1 participants actually increase faster when working memory is loaded than the response times of the Dutch participants (0.32 s for the Dutch vs. 0.50 for the L1).

4.2 Production task scores

Tables 5 and 6 show the mean amount of errors, and the mean amount of both the errors and the missed items combined for the Dutch and the native English-speaking participants respectively for the production task. The standard deviations are also shown.

If we take our first prediction into account, we should not expect a significant difference between the results of the L1 and the L2 participants in the production task without an extra working memory load. A Mann-Whitney test shows that this is indeed the case, both for the mean amount of errors (1.83 as compared to 1.64; U = 124.000, N1 = 18, N2 = 14, p = 0.955, two-tailed) and for the mean amount of errors and missed items combined (3.00 as compared to 1.86; U = 106.000, N1 = 18, N2 = 14, p = 0.464, two-tailed). These results do confirm our first prediction.

	Dutch participants [N=18]	
	Mean amount of errors	*Mean amount of errors and missed items combined*
Without WM load	1.83 [1.72]	3.00 [3.14]
With WM load	2.17 [2.04]	3.00 [3.22]

Table 5 – The mean amount of errors [standard deviations], and the mean amount of errors and the missed items combined [standard deviations] for the Dutch participants in the production task without an extra working memory load and the production task with such a load.

	Native English speaking participants [N=14]	
	Mean amount of errors	*Mean amount of errors and missed items combined*
Without WM load	1.64 [1.39]	1.86 [1.46]
With WM load	1.36 [1.45]	1.43 [1.50]

Table 6 – The mean amount of errors [standard deviations], and the mean amount of errors and the missed items combined [standard deviations] for the Dutch participants in the production task without an extra working memory load and the production task with such a load.

As was alsoexpected, the Dutch do perform worse in the production task with an extra working memory load as compared to the same task without such a load (2.17 as compared to 1.83). Remarkably enough however, a paired *t*-test shows that this difference is not significant (t = 1.304, df = 17, p = 0.105, one-tailed), and is even nonexistent when the missed items are taken into account (3.00 as compared to 3.00; t = 0.000, df = 17, p = 0.500, one-tailed). This might indicate that the task used to load working memory did not work thoroughly enough to actually cause a decrease in performance in the production task. This assumption can be confirmed by comparing the results of the native English-speaking participants. These participants even seem to perform better when working memory is loaded (1.36 as compared to 1.64 for the errors only, and 1.43 as compared to 1.86 for both the errors and the missed items combined), although these results are not significant (respectively: t = 0.844, df = 13, p = 0.207, one-tailed; t = 1.194, df = 13, p = 0.127, one-tailed). These results do not confirm the second prediction, which states that all participants are expected to perform worse when working memory was loaded as compared to the same task without such a load.

	Difference in number of errors (Dutch – L1)	*Difference in number of errors and missed items (Dutch – L1)*
Without WM load	0.19	1.14
With WM load	0.81	1.57

Table 7– The difference in the number of errors between the performance of the Dutch participants in the production task as compared to the L1 participants under the conditions with and without an extra working memory load.

Furthermore, Tables5, 6 and 7 show that, as we expected, the native English-speaking participants perform better under all conditions than the Dutch participants. Just as we predicted, these results were not significant for the task without an extra working memory load. However, a Mann-Whitney test shows that the results of the task with such an extra load are also non-significant, both for the mean amount of errors (2.17 as compared to 1.36; $U = 96.500$, $N1 = 18$, $N2 = 14$, $p = 0.232$, one-tailed) and for the mean amount of errors and the missed items combined (3.00 as compared to 1.43; $U = 86.000$, $N1 = 18$, $N2 = 14$, $p = 0.068$, one-tailed),which is *not* what we predicted. These results therefore do not confirm our third prediction, although the last result is almost significant.

Lastly, we expected the differences in performance between the native speakers and the L2 learners to be largest when working memory was loaded during the production task, as compared to their performances when working memory was loaded during the grammaticality judgment task. If we compare our results on both the grammaticality judgment task and the production task, we find that these do not confirm our fourth prediction. There is a significant difference between both participant groups for the grammaticality judgment task, while there is no such difference for the production task, which is not what was expected.

4.3 N-back scores

Table 8 shows the mean amount of mistakes and the mean amount of both the mistakes and the missed items in the N-back task for both the L1 and the L2 participants. A participant's score on the N-back task is compiled by both the mean amount of mistakes and the missed items.

	Dutch participants	**Native English speaking participants**
Mean amount of mistakes	1.58 [1.03]	2.36 [1.22]
Mean amount of mistakes and missed items combined (= participant's score)	3.42 [2.43]	5.27 [2.56]

Table 8 – The mean amount of mistakes [standard deviations] and the mean amount of the mistakes and missed items combined [standard deviations] in the N-back task for both the L1 and the L2 participants.

As can be seen in Table 8, the Dutch participants performed significantly better on the N-back task than the native English-speaking participants (t = 2.543, df = 29, p = 0.017, two-tailed). This is not what was expected, since both participant groups were matched according to age and educational level.

Based on the idea that there might as well be a within-group effect of the amount of working memory capacity, post-experimentally a median split on the N-back task was used to divide both participant groups (L1 and L2) into two, forming two groups of participants who had a low score on the N-back task (from which it may be concluded that, relatively speaking, they have a large working memory capacity, since a low score on the N-back task means that they made few mistakes and did not miss many items) and two groups of participants who had a high score on the N-back task (who are therefore thought to have a relatively small working memory capacity). According to McDonald's theory, the participants with a high score on the N-back task are expected to perform worse than the participants with a low score on the N-back task, since their working memory capacity is smallest. This goes for both the grammaticality judgment and the production task, especially for both tasks under the condition with an extra working memory task. No significant differences between the high and the low groups are expected in the task without an extra load on working memory.

Tables 9 and 10 show the mean amount of errors, the mean amount of errors and missed items combined and the mean response times on the grammaticality judgment task for the Dutch and the native English speaking participants respectively, divided into groups with a low and with a high N-back task score. The standard deviations are also shown.

	Dutch participants [N=19]					
	Mean amount of errors		*Mean amount of errors and missed items combined*		*Mean response times*	
	High [N=7]	**Low** [N=12]	**High** [N=7]	**Low** [N=12]	**High** [N=7]	**Low** [N=12]
Without WM load	4.29 [2.43]	3.92 [2.15]	6.29 [2.50]	5.83 [3.27]	3.82 [0.37]	3.78 [0.70]
With WM load	6.71 [4.82]	4.75 [2.63]	9.43 [4.12]	7.33 [3.68]	4.25 [0.58]	4.02 [0.65]

Table 9 – The mean amount of errors [standard deviations], the mean amount of errors and missed items combined [standard deviations], and the mean response times [standard deviations] on the grammaticality

judgment task with and without an extra working memory load for the Dutch participants, divided into a group with a low score on the N-back task and a group with a high score on the N-back task.

As was expected, the L2 participants with a high score on the N-back task, and therefore with a relatively small working memory capacity, perform worse on the grammaticality judgment task with an extra working memory load than the participants with a relatively large working memory capacity (near significant; t = 1.148, df = 17, p = 0.134, one-tailed) These differences also exists when the results of the test without an extra working memory task are being considered, but in this case, they are much smaller and also non-significant (t = 0.315, df = 17, p = 0.328, one-tailed).

	Native English speaking participants **[N=15]**					
	Mean amount of errors		*Mean amount of errors and missed items combined*		*Mean response times*	
	High **[N=9]**	**Low** **[N=6]**	**High** **[N=9]**	**Low** **[N=6]**	**High** **[N=9]**	**Low** **[N=6]**
Without WM load	2.44 [1.51]	2.17 [1.60]	2.89 [2.37]	2.50 [1.52]	3.02 [0.71]	2.59 [0.58]
With WM load	4.22 [3.90]	2.83 [1.33]	4.44 [3.81]	3.17 [1.47]	3.42 [0.59]	3.24 [0.51]

Table 10 – The mean amount of errors [standard deviations], the mean amount of errors and missed items combined [standard deviations], and the mean response times [standard deviations] on the grammaticality judgment task with and without an extra working memory load for the native English speaking participants, divided into a group with a low score on the N-back task and a group with a high score on the N-back task.

The results of the L1 participants are comparable to those of the L2 participants. Again as expected, the participants with a relatively small working memory capacity perform worse on all aspects of the grammaticality judgment test than the participants with a relatively large working memory capacity, both when working memory was loaded with an extra task and when it was not. The differences are larger when working memory was loaded with an extra task than when it was not, which was expected. However, in both cases, the results are non-significant (t = 0.775, df = 13, p = 0.226, one-tailed; t = 0.354, df = 13, p = 0.365 respectively for the mean amount of errors and the missed items combined), which may (in part) be due to the high standard deviations.

Tables 11 and 12 show the mean amount of errors and the mean amount of errors and the missed items combined in the production task of the Dutch and English speaking participants respectively, divided into groups with a high and with a low score on the N-back task. The standard deviations are also shown. Here, the data are quite straightforward: in all cases the participants with a high score on the N-back task, and therefore with a relatively small working memory capacity, perform worse than the participants with a low score on the N-back task and a relatively large working memory capacity. For the L1 speakers, these results are non-significant, and for the Dutch participants, the results are near significant (t = 1.265, df = 16, p = 0.112). Remarkably enough, the results of the Dutch participants show that the differences between both groups are just as big in the condition without an extra working memory task as in the condition with such a task, which is not what was expected.

	Dutch participants [N=18]			
	Mean amount of errors		*Mean amount of errors and missed items combined*	
	High [N=6]	**Low [N=12]**	**High [N=6]**	**Low [N=12]**
Without WM load	2.67 [1.97]	1.42 [1.51]	4.33 [3.67]	2.33 [2.77]
With WM load	2.83 [2.64]	1.83 [1.70]	4.33 [4.80]	2.33 [2.02]

Table 11 – The mean amount of errors [standard deviations] and the mean amount of errors and missed items combined [standard deviations] on the production task with and without an extra working memory load for the Dutch participants, divided into a group with a high score on the N-back task and a group with a low score on the N-back task.

	English speaking participants [N=14]			
	Mean amount of errors		*Mean amount of errors and missed items combined*	
	High **[N=9]**	**Low** **[N=5]**	**High** **[N=9]**	**Low** **[N=5]**
Without WM load	1.67 [1.41]	1.60 [1.52]	2.00 [1.50]	1.60 [1.52]
With WM load	1.44 [1.59]	1.20 [1.30]	1.56 [1.67]	1.20 [1.30]

Table 12 – The mean amount of errors [standard deviations] and the mean amount of errors and missed items combined [standard deviations] on the production task with and without an extra working memory load for the native English speaking participants, divided into a group with a high score on the N-back task and a group with a low score on the N-back task.

5. Discussion

The current experiment set out to clarify the link between language processing and working memory. Based on a working memory model by Just and Carpenter (1992), Janet McDonald (2000) proposed a theory in which language processing draws on a limited working memory capacity. Her idea is that language performance breaks down as soon as an individual's working memory capacity is exceeded. Furthermore, she proposes that second language acquirers may be working under a higher working memory load than native speakers, and therefore will have more trouble reaching a native-like competency level. Using basically the same grammaticality judgment task as McDonald did, as well as a production task, I however did not obtain sufficient data to fully support McDonald's theory.

5.1 Grammaticality judgment

First of all, it was expected that the L2 and the L1 speakers would perform similarly on the grammaticality judgment task under the condition without an extra working memory load. The obtained grammaticality judgment data however show that there is a significant difference between the scores of the L2 learners on this task and those of the native speakers. Therefore, despite the fact that it was tried to match the chosen items in the grammaticality judgment task to the competency level of the L2 participants, it may be argued that this task proved to be too hard for the L2 learners, even when their working memory was not loaded using an extra working memory task.

Furthermore, it was expected that the difference between the performance of the native speakers and that of the L2 learners would be significantly larger when working memory was loaded with an extra task than when it was not. However, no such increase in the difference between both groups could be found. The fact that there already was a highly significant difference between the scores of both groups in the condition without an extra working memory task may have had an influence on these results. As a result, it is difficult to assess to what extent the differences between the scores of the L1 and those of the L2 learners are attributable to the assumed overload of the participants' working memory capacities. The fact that the L2 learners performed worse than the native speakers may just as well be accounted for by assuming that the items that were to be judged were too difficult for the L2 learners. Therefore, it would have been better if the L2 participants were selected on the basis of their scores on the grammaticality judgment task without an extra working memory task, which ideally should have been exactly the same as the scores of the L1 speakers. In my case however, practical issues stood in the way, since it already was quite hard to find enough participants for this experiment. Also, the danger of selecting a group of L2 participants who perform just as well as native speakers under the condition without an extra task is, that they might have already reached a native-like competency level in their L2, so that they would perform similarly as the L1 speakers, even under the condition with an extra working memory load. So, it is very difficult to perfectly match the sentence items that are to be judged with the proficiency level of the L2 learners in an experiment such as the current one.

Nevertheless, some important conclusions can still be drawn from the obtained grammaticality judgment data. First of all, as was expected, the results of both the L1 and the L2 participants on this task got worse when working memory was loaded using an extra task. It therefore may be concluded that the used working memory task during the grammaticality judgment task indeed loaded working memory, and that this extra load on working memory caused the language system to make more mistakes, both for the L1 and for the L2 speakers. So, the exceeding of a limited working memory capacity seems to have at least some influence on the processing of a language.

Lastly, a remarkable finding was the fact that the response times of the L1 learners actually increased faster than those of the L2 learners when working memory was loaded, as compared to the response times on the task without such an extra load, which was contrary to what was expected. A possible explanation for this finding can be found in the trade-off that takes place between the amount of mistakes made and the mean response times. When a participant is confronted with a task that just became more difficult than before, such as when a participant in the current experiment was introduced to the extra working memory task combined with the familiar grammaticality judgment task, he has two options to deal with this new situation. First of all, he may decide to try to keep the amount of mistakes as low as possible. As a result, the response times will get slower, simply due to the fact that the task became more difficult. On the other hand, a participant may decide to try to react as quickly as possible, in which case he is likely to make more mistakes than before. It is possible that most of the L1 speakers chose to keep the amount of mistakes as low as possible, since the task at hand was quite a simple one for them, and they were afraid to lose face if they made too many mistakes.The Dutch participants however may have had less trouble with making a mistake, since the task items were not presented in their native language. It therefore seems plausible to assume that the L2 learners chose a different strategy than the native speakers, in which they mainly focused on keeping the response timesas low as possible, whereas the native speakers mainly focused on making as little mistakes as possible.

5.2 Language production

If we take the results of the production task into account, some noticeable differences between these results and those of the grammaticality judgment task are to be mentioned. First of all, the production task seems to have been somewhat easier for the participants than the grammaticality judgment task. As was expected and wished-for, there is no significant difference between the performance of the L2 learners and the native speakers on the production task in the condition without an extra working memory task. Participants from both groups tended to make very few mistakes, if any. The same can be said however for the production task combined with an extra working memory task. Again, there is no significant difference between the scores of both groups, although interestingly enough, the difference between both groups is near significant if we take both the errors and the missed items into account.

Furthermore, the extra working memory task added to the production task did not cause the participants to make significantly more mistakes, which is not what was expected. The L1 participants even tended to make fewer mistakes, although this difference is not significant either. Basically the same task was used to load working memory during the production task as the one that was used to load working memory during the grammaticality judgment task. Since this task proved to be enough to load working memory during the grammaticality task, it seems unlikely to assume that the used working memory task did not load working memory at all during the production task. In my opinion, it is more likely to assume that the production task was so simple, that even an extra working memory task did not cause the participants' working memory capacities to be exceeded, and therefore they did not start to make significantly more mistakes.

Although it is important to note that the production task could have been more daring for the participants, some important conclusions can still be drawn from the obtained production task data. As has been said, the native English-speaking participants tended to perform worse on the production task when working memory was not loaded with an extra task than when it was loaded with such a task, which was contrary to what was expected. This finding may have two explanations: First of all, the native English-speaking participants tended to make the same 'error' over and over again. This finding raises the question whether these utterances should indeed be viewed upon as errors, or whether the presented items should have been selected more carefully. However, the item lists were presented randomly to all participants, and therefore it is likely to assume that the influence of this presumed erroneous item should have been as big in the condition without an extra working memory task as in the condition with such a task. Another, more interesting explanation in my opinion can be found in the fact that all participants received both tests in the same order, and therefore, a learning effect may have taken place. To ensure that all participants could get used to the production task, they all started with the task without an extra working memory task, after which they did the same task (but with different items, of course) with the extra working memory load. Therefore, it may be so that all participants needed the first part of the production task experiment, in which working memory was not loaded, to get used to the test. So, the English-speaking participants may have made some mistakes in the condition without an extra working memory taskonly due to the fact that the task was new to them, and they did not make these mistakes when working memory was loaded, because at that time, they were used to the task. Now, if we follow this train of thought, there seems to be no reason to assume that this effect should not apply for the L2 participants. Despite the fact that the L2 participants also needed to get used to production task, and therefore probably made some mistakes just due to that, they still tended to make more mistakes in the task with an extra working memory load. So, the influence of the extra working memory task on the performance of the L2 learners may indeed have been bigger than the data actually show.

5.3 Individual's working memory capacity

The participants' scores on the N-back task may shed some new light on the overall results of this experiment as well. As can be seen in the results section of this paper, overall, the native-English speaking participants performed worse on the N-back task than the Dutch participants. From this, it may be concluded that the Dutch participants had slightly larger working memory capacities than the L1 participants. This seems logicalsince the Dutch participants were slightly older as well (the Dutch were aged 17;9, whereas the L1 participants were aged 17;3), and it is assumed that the working memory capacity is still growing at the participants' age (e.g. Wingfield et al., 1988). If the L2 learners indeed had larger working memory capacities in general than the L1 speakers, it seems plausible to assume that the found differences in the performances on the grammaticality judgment and the production tasks between both groups could have been even larger, if only there would have been no difference at all betweenthe mean working memory capacities of both participant groups.

The data of the N-back scores also gave us the opportunity to divide both participant groups into two, forming a group of participants with a relatively small working memory capacity and a group of participants with a relatively large working memory capacity. Interestingly enough, the results show that in all cases, both for the Dutch participants and for the native English-speaking participants and for the grammaticality judgment task as well as for the production task, participants with a relatively small working memory capacity perform worse than participants with a relatively large working memory capacity. Therefore, there seems to be a remarkable correlation between working memory capacity size and performance on both tasks. In my opinion, it would be interesting for future research to fully examine this correlation.

5.4 Consequences for the limited capacity theory

Although the current study did not find enough evidence to fully support the theory that second language acquirers are working under a higher working memory load than native speakers, McDonald's limited capacity theory might still prove to be a fruitful alternative to the critical period hypothesis. The L2 acquirers actually did perform worse on both the grammaticality judgment task and the production task than the L1 speakers did, and the differences between both groups indeed seemed to grow larger when working memory was loaded more thoroughly, although this effect could not be measured as accurately as was wished for, due to the fact that the grammaticality judgment task proved to be quite difficult for the Dutch participants, while the production task on the other hand proved to be fairly easy.

However, in my opinion, the results of this experiment do point towards an influence of a limited working memory capacity on language processing, which indeed may be larger for L2 acquirers than for L1 speakers. Biological maturation effects do still seem to be a factor in acquiring a (second) language, but evidence is mounting that other factors, such as an individual’s limited working memory capacity and the nature of the L1 as compared to the to be acquired L2, also play an important part in language acquisition. Much is still to be discovered about the actual influence of working memory capacity on second language acquisition.

6. References

- Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In: G. A. Bower (Ed.), *Recent advances in learning and motivation, 8*, 47-90. New York: Academic Press.
- Baddeley, A.D., & Hitch, G.J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*, 485-493.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second language research, 13*, 116-137.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161-181). Mahwah, NJ: Erlbaum.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.
- Haznedar, B., & Schwartz, B (1997). Are there optional infinitives in child L2 acquisition? In Hughes, E., Hughes, M., & Greenhill, A., editors, *Proceedings of the 21st Annual Boston University Conference on Language Development*, Somerville, MA: Cascadilla Press, 257-268.
- Ionin, T., & Wexler, K. (2002). Why is ‘is’ easier than ‘-s’?: acquisition of tense/agreement morphology by child second language learners of English. *Second language research, 18,2*, 95-136.
- Juffs, A., & Harrington, M. (1995). Parsing Effects in Second Language Sentence Processing: Subject and Object Asymmetries in "Wh"-Extraction. *Studies in second language acquisition, 17,4*, 483-516.
- Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review, 99*, 122-149.
- Kellerman, E. (1995). Age before beauty. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 219-231). Amsterdam: Benjamins.
- Lenneberg, E. (1967).*Biological foundations of language.* New York: Wiley.
- MacDonald, M. C., & Christiansen, M.H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109,* 35-54.

- McDonald, J.L. (2000).Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics, 21*, 395-423).
- Milner, A.D., Jeeves, M.A., Ratcliff, P.J., Cunnison, J. (1982). Interference effects of verbal and spatial tasks on simple visual reaction time. *Neuropsychologia*, *20*, 591-595.
- Noort, M. W. M. L. van den, Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European psychologist, 11*, 289-296.
- Owen, A.M., McMillan, K.M., Laird, A.R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25*, 46-59.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second language research, 16,2*, 103-133.
- Whitney, P., Arnett, P.A., & Driver, A. (2001). Measuring central executive functioning: What's in a reading span? *Brain and Cognition*, *45*, 1-14
- Wingfield, A., Stine, E.A.L., Lahar, C.J., & Aberdeen, J.S. (1988). Does the capacity of working memory change with age? *Experimental aging research*, *14*, *2*, 103-107.

Appendix A

List A of the grammaticality judgment task

1. Ten years ago the astronomer discovered a comet.
2. Last week Fred washes his clothes.
3. Yesterday the hunter shot a deer.
4. Last year the boxer fights his last match.
5. A bat flew into our attic last night.
6. Last Tuesday Peter felled from his bike.
7. The poet took his girlfriend to the recital last night.
8. The barber cutted my hair last week.
9. The little boy eats breakfast every morning.
10. Michael hate getting up every morning at 5 A.M.
11. The magician performs for charity functions.
12. The basketball player look really big from nearby.
13. The farmer bought two pigs at the market.
14. The girl walked her three brother to the park.
15. The soldier is hiding behind the wall.
16. The big, growling bear angry.
17. The boy kept three mice as pets.
18. The doctor visited the two womans.
19. A shoe salesman sees many feet throughout the day.
20. The mother took her two childrens to the zoo on Wednesday.
21. Did Bill dance at the party last night?
22. Closed Sally the door?
23. When will Sam fix the car?
24. When Jim did leave the train?
25. Larry went the home after the wedding.
26. Tom is reading a book in the bathtub.

Appendix B

List B of the grammaticality judgment task

1. The little boy running away from home.
2. The new car my father bought is green.
3. Last week the boy went on a fishing trip.
4. The day before yesterday Lucy finds a dollar.
5. Johnny brought a cake to the picnic last year.
6. Last night the police caughted a burglar.
7. He threw the football to his cousin yesterday afternoon.
8. We buyed a table last week.
9. The dog plays with his master in the afternoon.
10. The doctor work in the maternity ward.
11. The cyclist wants a drink as soon as possible.
12. The little girl need her teddy bear to feel safe.
13. I bought four new chairs for my living room.
14. The musician sold two of his guitar for charity.
15. Five people were at the party.
16. The reporter interviewed the three mens.
17. The little girls fed the three geese at the park.
18. The baby has two new tooths.
19. Yesterday Sally checked the books out of the library.
20. Last Wednesday the teacher yells at a student.
21. Did Sandra clean up her room?
22. Found he the marble he lost this morning?
23. When can you come to my house?
24. When you did know she broke the vase?
25. Death is one of life’s great sorrows.
26. Janice is following special recipe for the cake.

Appendix C

List C of the grammaticality judgment task

1. Ten years ago the astronomer discovers a comet.
2. Last week Fred washed his clothes.
3. Yesterday the hunter shoots a deer.
4. Last year the boxer fought his last match.
5. A bat flewed into our attic last night.
6. Last Tuesday Peter fell from his bike.
7. The poet taked his girlfriend to the recital last night.
8. The barber cut my hair last week.
9. The little boy eat breakfast every morning.
10. Michael hates getting up every morning at 5 A.M.
11. The magician perform for charity functions.
12. The basketball player looks really big from nearby.
13. The farmer bought two pig at the market.
14. The girl walked her three brothers to the park.
15. A shoe salesman sees many feets throughout the day.
16. The mother took her two children to the zoo on Wednesday.
17. The boy kept three mouses as pets.
18. The doctor visited the two women.
19. The soldier hiding behind the wall.
20. The big, growling bear is angry.
21. Did Bill dance at the party last night?
22. Closed Sally the door?
23. When will Sam fix the car?
24. When Jim did leave the train?
25. Larry went the home after the wedding.
26. Tom is reading a book in the bathtub.

Appendix D

List D of the grammaticality judgment task

1. Did Sandra clean up her room?
2. Found he the marble he lost this morning?
3. When can you come to my house?
4. When you did know she broke the vase?
5. Death is one of life's great sorrows.
6. Janice is following special recipe for the cake.
7. The little boy is running away from home.
8. The new car my father bought green.
9. The little girls fed the three gooses at the park.
10. The baby has two new teeth.
11. Five peoples were at the party.
12. The reporter interviewed the three men.
13. I bought four new chair for my living room.
14. The musician sold two of his guitars for charity.
15. The dog play with his master in the afternoon.
16. The doctor works in the maternity ward.
17. The cyclist want a drink as soon as possible.
18. The little girl needs her teddy bear to feel safe.
19. He throwed the football to his cousin yesterday afternoon.
20. We bought a table last week.
21. Johnny broughted a cake to the picnic last year.
22. Last night the police caught a burglar.
23. Last week the boy goes on a fishing trip.
24. The day before yesterday Lucy found a dollar.
25. Yesterday Sally checks the books out of the library.
26. Last Wednesday the teacher yelled at a student.

Appendix E

List 1 of the production task

1. Last summer, the film star … a role in a blockbuster movie.
2. After he had shouted at Kim, Mike felt sorry for her and ... to her.
3. This month, the car salesman only ... ten cars.
4. Yesterday, Freddy ... the dog for a walk.
5. Last year, we took a plane and ... to the Bahamas.
6. This morning, I ... my books and had to go back to get them.
7. The little boy ... very tall in just one year.
8. Last Saturday, Francine ... an earring and could not find it anymore.
9. Every day Linda ... her mother on the phone.
10. Every week, my mom ... the toilet with bleech.
11. On her birthday, Eric always ... a nice cake for his girlfriend.
12. Usually, Timmy … a glass of milk before going to school.
13. The magician pulled many ... out of his hat.
14. Monkeys like to eat ...
15. Mark could not sleep and therefore he decided to count ...
16. The man went to the river and caught many ... in his net.
17. The dog chased ... cat down the street.
18. At first, the water was cold, but then it got ...
19. In the autumn, leaves turn from green to ...
20. It was raining, so my clothes got all ...
21. I drank ... glass of wine.
22. I never put butter on my ...

Appendix F

List 2 of the production task

1. Yesterday, the writer ... his book and sent it to the publisher.
2. Kelly … on her strawberry flavoured chewing gum.
3. Many years ago, the pirate … a deep hole with his shovel to bury his treasure in."
4. This morning, Lisa stepped in her new car and … to work.
5. Yesterday, Dianne ... her rabbit an extra carrot for lunch.
6. Last year, I ... a new bike for my birthday.
7. Recently, Erik Hulzebosch ... the contest ‘So you wanna be a popstar’.
8. During hide and seek, Mike ... under the bed until Rudolf found him.
9. Every morning, Wendy ... her hair before going to school.
10. Sally never ... to what her mother tells her.
11. He hardly ever ... his books on time to the library.
12. Dennis got so fat that he now ... more than 100 kilos.
13. The war veteran in the wheelchair lost both his ... in Vietnam.
14. All ... have feathers.
15. The house was on fire, so my dad called the ... to come and put it out.
16. All ... have to go to school.
17. The baby cried and woke up ... parents.
18. Nathan was so tired that he fell asleep ... class.
19. The teacher was ... and yelled at her.
20. On her wedding, she wore a white ...
21. The jury said that the man was ...
22. He ate the whole ... of candy.