

The Nun Shot the Monk

An investigation into incidental acquisition of German cases and vowel change rules among Dutch adults

Nora Kennis

s1006973

July 30 2020

BA Linguistics

Primary supervisor: dr. Kristin Lemhöfer

Secondary supervisor: dr. Sharon Unsworth

Word count: approx. 14900

Radboud Universiteit



Preface

Before you lies my thesis, entitled “The Nun Shot the Monk: an investigation into incidental acquisition of German cases and vowel change rules among Dutch adults”. It was written as the conclusion to my bachelor’s degree in linguistics at Radboud University. As I have been interested in the cognition of multilingualism since before I started the programme, I am very glad to have been able to explore this topic. The research and writing for this project were conducted between March and June of 2020.

The process of creating my thesis did not go entirely as planned. Because of the university’s closure due to COVID-19 I was unable to conduct my experiment in the labs of the Donders Centre for Cognition, to which I had really been looking forward. While replacing lab hours with Zoom calls and inspiring institute offices and engaging colleagues with my own bedroom with only myself for company was not ideal, I am very satisfied with the result of my work.

I would like to thank my supervisor, dr. Kristin Lemhöfer, for her excellent guidance during the process, her enthusiasm for the project and her sizable time investment during a pandemic. I am also very grateful for the help I received from the other staff at the DCC when setting up and conducting my experiment and in particular from Miriam Kos, who always responded promptly and patiently to my many emails about participant payment and online experimentation. A large thanks goes to my mother, who occasionally reminded me to sleep before my brain fell out, as well as my father who repeatedly suggested I reach my required word count by filling the pages with “waf waf waf”. Finally, I would like to thank my friends, who helped me with ideas and provided wonderful suggestions after listening to my lengthy ramblings about the problem repeatedly.

And thank you to you, the reader. I hope you enjoy my work.

Nora Kennis
Nijmegen, June 30, 2020

Contents

Preface	I
Contents	II
Abstract	II
Chapter 1: Introduction	1
1.1 Previous research	1
Incidental acquisition of second language grammars	1
Learning styles	7
1.2 The current study	8
Chapter 2: Methods	11
2.1 Participants	11
2.2 Materials	11
2.3 Procedure	13
2.4 Data-analysis	14
Chapter 3: Results	16
3.1 Learning effect GJT	16
3.2 Comparison of old and new incorrect items	17
3.3 Analysis of mistake types	18
3.4 Comparison of language modalities	19
3.5 Effects of awareness	20
Chapter 4: Discussion	21
4.1 Incidental learning	21
4.2 Data-based or rule-based learning?	23
4.3 Grammatical rule learning or lexical learning?	24
4.4 Explorative questions	25
Chapter 5: Conclusion	27
References	28
Appendices	30
Appendix 1: Images of characters and objects	32
Appendix 2a: Number of occurrences per character per experiment phase	33
Appendix 2b: Specification of character occurrences in the GJT	33
Appendix 3: Occurrences per rule per experiment phase	34
Appendix 4: All stimuli sentences	34
Appendix 5: Examples of task layouts	37

Abstract

This thesis explores incidental second language grammar acquisition by investigating learning of German case-gender rules and vowel change rules among Dutch adults. The main research questions are whether incidental learning occurs in this context, whether learning is either strongly data-driven or rule-driven (or if this differs between learners) and how levels of difficulty of the different grammatical rules compare to one another. This was tested using an experiment during which participants were exposed to grammatical sentences while performing a task examining semantic understanding. Participants then performed a grammaticality judgement task (GJT) and a sentence production task.

Results show participants were unable to incidentally acquire the case-gender and vowel change rules. This indicates the difficulty of acquiring grammatical elements that do not exist in a learner's native language. Nominative case is more easily acquired than other cases, which seems to be mostly based on interference of the native language. While evidence for either data-based learning or rule-based learning is inconclusive, large differences between learners have been observed with respect to old-new effects. Future research is needed to further explore learning styles in grammar acquisition.

Chapter 1: Introduction

As the world becomes an increasingly smaller place due to globalisation, it becomes more and more common for adults to learn a second language, to benefit them in their career or their personal life. There are many different methods of acquiring that second language, only one of which is traditional classroom instruction. With this development comes a surge in research on second language acquisition (SLA) amongst linguists, since the second half of the last century. Not only can these studies tell us about the human language system in general and what differentiates adult learning from the way children learn their native language, they could also serve as information to determine which linguistic concepts learners are and are not able to acquire naturally through immersion, which should influence which concepts are (more frequently) addressed in classroom teaching plans. What are the best approaches to second language learning and which aspects of the process are particularly difficult?

In this thesis I will explore a number of facets of this topic by investigating incidental acquisition of German grammar in Dutch adult learners. Incidental learning is fascinating, because learning in the absence of explicit instruction shows the human mind's ability for statistical learning and pattern recognition. Learning a grammar "by accident", while paying attention to understanding the material or communicating with speakers is a very natural process in most methods of language learning. However, research on incidental second language grammar learning has been reasonably limited, often using artificial grammars rather than natural languages. For this reason, a study that details the occurrence of and factors influencing incidental grammar learning of a natural language will be a valuable addition to current knowledge.

1.1 Previous research

This section endeavours to review some of the research done in the subfields of linguistics that are relevant to the topics of this thesis in which I will focus specifically on incidental acquisition of the German case system and vowel change rules. Because the field of second language acquisition (SLA) of syntax is vast, this discussion will be limited to a small number of topics that have a direct link to this experiment. Firstly, I will discuss studies related to the question whether any incidental second language grammar learning occurs in adults literature with respect to both artificial grammar learning and natural languages. These will be subdivided based on the applicable questions they endeavour to answer, such as the role of awareness in the success of incidental learning. This section will be followed by a section discussing previous research on individual differences between learners of a second language, with a focus on individual learning styles in grammar acquisition.

1.1.1 Incidental Second Language Grammar Acquisition

To start this review, a general overview of what makes second language (L2) grammar acquisition difficult will be an important basis upon which the rest of the literature review builds. In a review article on this topic, DeKeyser (2005) explores the characteristics of the second language that complicate acquisition. He defines "difficulty" as being determined by three factors: complexity of form, complexity of meaning and the complexity of the form-meaning mapping. Additionally, the transparency of the form-meaning relationship determines the level of difficulty.

The first of DeKeyser's (2005) three factors is the meaning of a grammatical element, which can constitute a source of difficulty if it is either novel or abstract to a learner. This is the main factor that interferes with the process of acquiring the German case system for Dutch learners, as Dutch does not have a distinction between a feminine and a masculine gender for nouns the way German does. Dutch also does not inflect its articles for case. German expresses syntactic function using a combination of constituent order and case, while Dutch is mainly reliant on constituent order. This novelty of case use in combination with a known structure may hinder acquisition of German. The fact that German articles do inflect for both gender and case also poses a problem of form, as DeKeyser would call his second factor, because it raises the number of choices speakers have to make to pick the appropriate morphemes to express the meanings.

The third factor to determine difficulty is the complexity of form-meaning mappings. In his exploration of this topic, DeKeyser (2005) names opacity as an important determiner for this, one form of which is when a single grammatical morpheme has multiple meanings. This is the case for some inflected German articles that are used for multiple gender-case combinations. Furthermore, focus-motivated word order changes and their frequency make the relationship between NP inflection and word order more difficult to understand.

These L2-specific characteristics also interact with the learning context, as described by DeKeyser (2005). For example, Williams (1999) differentiated between grammatical elements with meaningful form-meaning mappings, which were acquired best through conceptual, explicit learning, and semantically redundant elements, which were more easily acquired through "data-driven, implicit learning". On the other hand, research by Robinson (1996) in which he compares instructed to incidental learning on complex and simple grammatical rules, shows that, while incidental learning is successful for complex rules, instructed learning leads to better performances on all sentence types. Still, Taraban (2004), found that learning gender categories when learning time was restricted was easiest when patterns were made explicit, while implicit learning would increase in cases where learners had more time and exposure. These factors are instrumental in justifying which aspects of German case use may be more or less difficult for Dutch learners and could explain their successes and failures in incidental learning.

As DeKeyser (2005) described in his work on difficulties in L2 learning, there are many ways of learning a language. For this thesis, I am specifically interested in incidental acquisition of L2 grammar. All other works to be discussed in this section will focus on one central question: does incidental grammar learning occur in adults? In addition, these studies address a number of other relevant aspects of incidental L2 grammar learning.

In a review of literature on incidental and intentional learning with a focus on language learning, Hulstijn (2003) derived the following definition of incidental learning from Schmidt (1994, p.16): "learning of one thing (e.g., grammar) when the learner's primary objective is to do something else (e.g. communicate)". Learning grammar whilst paying attention to learning and understanding semantics would therefore be an example of incidental learning.

Hulstijn (2003) discusses a number of empirical studies on incidental grammar learning and their operationalisations of "incidental learning". In all of these studies, participants were exposed to L2 data without being told the data are examples of some rule in the L2 grammar. Participants are also not informed that the goal of the study is learning how successfully they acquire the feature. However, not all of these studies explicitly use the term "incidental learning" to describe this.

Notably, not all studies to be discussed in the next sections in relation to incidental learning meet the requirement set by Schmidt (1994) that incidental learning involves the

learner's focus being on something other than grammar. While McAndrews and Moscovitch (1985), Denhovska and Serratrice (2017) and Andringa (2020) distract their readers, making them pay attention to other aspects of the stimuli, other authors' definition of incidental learning seems to be "learning by mere exposure". Likewise, the studies to be discussed reported by Williams (2010), Franck, Rotondi and Frauenfelder (2016) and Mueller, Oberecker and Friederici (2009) do all inform their participant that their goal is to "learn the language". Although the instruction is more general than the specific study goal of acquiring grammar, this does not entirely fit Schmidt's definition of incidental learning.

Example-based vs. rule-based learning

One essential outstanding question in incidental grammar learning is to what extent learning, if it happens, is rule-based or example-based. Two studies to deal with this question are authored by McAndrews and Moscovitch (1985) and Williams (2010).

Most works on incidental L2 grammar learning have involved artificial grammars. These grammars are constructed and presented in a meaningless context (often consonant strings), ensuring controllability of the input and the variation within them. When used in incidental learning studies, participants are not asked to detect the patterns within the strings, but often have to engage with them in another way. An example of such an experiment was reported by McAndrews and Moscovitch (1985), who explore two different categorisation models in their paper. The abstraction model states speakers deduce and store abstract rules based on examples, while an exemplar-based model claims decisions about which category a novel item belongs to are solely based on analogy to previous examples. Therefore, if participants perform better on stimuli that are similar to ones previously encountered, this would be evidence for an exemplar-based model. They presented the training phase as a consumer marketing survey in which participants had to rate letter strings generated by a finite-state grammar for how well they would work as computer names. Afterwards, participants completed a grammaticality judgement task (GJT).

Results show there is an effect of the similarity to previously seen examples on performance in the classification task. However, the authors also found large differences between performances of different participants. A subset of participants clearly uses different classification strategies besides similarity to other examples. Some individuals seem to base their grammaticality judgements on the stored examples, while others are able to non-consciously induce the rules. Clearly learning is neither purely example-based nor rule-based across the whole participant sample.

Another study pertaining to this example-based/rule-based split was reported by Williams (2010). This paper focusses on word order regularities and inquires whether they are just sequence learning (based on previously seen data) or actual acquisition of the abstract rules. According to Williams, this question cannot be answered using artificial grammar literature, which provides little evidence for learning grammar as an abstract system because of its dissimilarity to natural language.

This study is based on an earlier study by Williams and Kuribara (2008) which investigated incidental learning of generalisable rules, specifically using Japanese word order of regular and scrambled sentences. A GJT shows the canonical patterns were clearly incidentally acquired, even when new lexical material was presented during the task. Evidence for the acquisition of scrambled word order is less conclusive. Learning seems to take place at the level of grammatical categories and abstract structures, because participants could generalise to new sentences.

However, Williams (2010) emphasises this does not necessarily mean participants learned rules per se, as participants did not show learning in every tested syntactic category.

Sentences with scrambled word order were not acquired at all. Participants would only have been able to perform well on these sentences if they had fully acquired the rules. Thus, he concludes associative sequence learning mechanisms are mostly used in early incidental learning.

Between McAndrews and Moscovitch's (1985) evidence for a combination of data-based and rule-based learning and Williams' (2010) conclusion that learning must be based on an in-between form in which learners extract abstract structures from previously seen data while being unable to generalise to different word orders, current evidence seems to neither fully support pure data-based learning nor pure rule-based learning.

Number of stimuli required for learning

A second question Williams (2010) focusses on in his study is the amount of input needed in order for incidental learning to be successful. In the 2010 study, Williams increased the number of exposure sentences from 194 sentences to 388 and compared the results to Williams and Kuribara's (2008) 194-sentence exposure group and control group (zero exposure) to find if increased exposure shows more robust learning effects. The GJT results show a clear difference between the control group and both exposure groups (194 vs. 388 sentences), indicating incidental learning of new word order patterns after exposure. Increased exposure to 388 sentences failed to show a significant increase in performance in the GJT performance for grammatical (simple and complex) and ungrammatical sentences. These structures are therefore quite rapidly learnt, although 194 sentences is still a sizable training phase.

Where Williams researches the difference in performance between a two quite large numbers of exposure sentences, Franck et al. (2016) tested adults' ability to generalise rules about hierarchical structure and structure dependency from very limited exposure to an artificial grammar. The artificial language they created had more similarities to natural languages than the language used in McAndrews and Moscovitch (1985), as it incorporated phonological and prosodic cues that are associated with grammar in most natural languages, as well as a distinction between nouns and verbs using suffixes.

Franck et al. (2016) studied whether participants could generalise the rules about long-distance subject-verb number agreement from the examples they had seen to new material. Participants were trained using 16 two- to three-word sentences presented three times made up of twelve different verbs or nouns and tested on two- to four-word sentences. Their results show participants could successfully generalise the constructions from the training phases to new words for both adjacent and non-adjacent dependencies, indicating they were able to induce agreement constraints despite very limited input and minimal training. Furthermore, a subset of the group was able to extend the hierarchical agreement rule to new four-word grammatical structures.

Therefore, learning may occur using very limited input, as shown in the Franck et al. study, or may need a larger number of exposure sentences like in Williams' (2010) study. While there probably is a minimum of input required in order to facilitate learning, the exact number will be largely dependent on the complexity of the rule to be learnt and the difficulty for the particular group of learners in question.

Measuring incidental learning

A third interesting question in incidental L2 grammar learning is which methods to use to measure the success of incidental learning. When performing a behavioural study, a relevant possible distinction is the use of a receptive or a productive task. Denhovska and Serratrice (2017) aim to explore incidental learning of gender agreement in adult learners based on their

success in both receptive and productive tasks. According to some previous studies, dependency relations like gender agreement cannot be acquired incidentally after the critical period ends, especially if the learner's L1 does not have this grammatical feature (e.g. Bley-Vroman, 2009; Lew-Williams & Fernald, 2010). Denhovska and Serratrice use a natural language, to increase ecological validity, that is completely new to their participants, to ensure controlled input. To this end, they use Russian as a test language. Russian adjectives agree with nouns in gender, case and number, to signal relations between linguistic units in an utterance.

Denhovska and Serratrice (2017) compare between-subject effects of the learning condition (incidental vs. explicit) on accuracy and RTs in the production and recognition of gender agreement patterns in Russian. Participants in the incidental condition were asked to focus on meaning and were not notified they would be tested, while participants in the explicit condition were informed about testing and provided with a metalinguistic explanation of the rule. It should be noted that it is appropriate to interpret the results of this study with some caution, as there are two differences between the incidental and explicit conditions. In the training phase, the words were presented in simple sentences with the English translation and a picture of the object. Participants were assessed using a GJT and a fill-in-the-blank task.

Results show participants in both conditions obtained high levels of receptive knowledge. There was a significant effect of learning condition and an interaction effect between incidental learning and old items. Participants in the explicit condition performed better than participants in the incidental condition (if only slightly) and all participants scored better on old items than new items in the GJT. For the productive task, results show a large significant difference between conditions, as participants in the incidental condition performed a lot more poorly than explicit condition participants (Denhovska & Serratrice, 2017). This is in accordance with earlier findings by Robinson (1996), who states instructed participants outperform participants in incidental learning conditions on all structure types. Their results shows speakers of a language without grammatical gender can accurately judge the grammaticality of agreement in a new language without receiving instruction. Because performances were above chance, this grammatical element seems to be learnable via procedural learning mechanisms without intention. The production of gender agreement, however, does seem to require explicit knowledge about grammatical gender.

While receptive and productive tasks are both used in behavioural testing, neuroimaging methods are also able to determine whether incidental learning was successful. Mueller et al. (2009) tested learning of distant grammatical dependency relations by mere exposure using Event Related Potential (ERP) techniques. The ability to process these non-adjacent dependencies (like in sentences such as “the baby who is in the bed is laughing”) is a prerequisite for understanding syntax. Mueller et al. compared the processing of learner vs. native language processing. Participants were presented with simple Italian sentences that contained a non-adjacent dependency relation between the auxiliary verb and the main verb's suffix in four separate learning phases and performed four GJTs.

Results show that the L2 speakers (who had previously never learned Italian) performed significantly better in later blocks than earlier blocks, indicating that they successfully learned the non-adjacent dependency. ERP data show that adults can incidentally learn to show similar behavioural and EEG signatures of non-adjacent dependency processing in a novel language when compared to native speakers. One slight difference is that native speakers have a P600 effect, which suggests they experienced syntactic integration difficulties. The fact that this remains absent in learners suggests they do not apply abstract rules in the same way native speakers do, which supports statistical learning mechanisms rather than rule-based extraction.

Awareness in incidental L2 grammar acquisition

A final topic on incidental L2 grammar acquisition that is not as intrinsically linked as the ones previously discussed but still strongly associated is awareness. How necessary is awareness of either the goal of learning or the specific rules in question for successful incidental acquisition? A work focussing on awareness in L2 grammar acquisition in general is a book chapter by Leow and Bowles (2005), wherein they discuss major theories on attention and awareness in SLA .

The most widely accepted view in the field is based on the noticing hypothesis by Schmidt (1990), according to which learning is impossible without awareness, because linguistic input can only be processed in the short-term memory if the learner is unaware of the rules to be acquired. While noticing is essential for learning to occur, Schmidt (2001) also poses learners have to reach a deeper level of awareness called understanding to be able to systematically acquire the new material.

A number of empirical research papers discussed by Leow and Bowles (2005) support Schmidt's noticing hypothesis, showing the effects of different levels of awareness on the success of SLA. According to Leow and Bowles, these studies provide evidence that awareness facilitates foreign language learning, both in case of mere noticing of patterns and of deeper understanding. Furthermore, understanding is found to be more beneficial than noticing alone (e.g. Leow, 2001). For the present study, this could result in differences between scores of participants who report awareness of the rules in comparison to those who do not.

However, a paper by Williams (2004) found that language learning can occur without noticing or awareness as well. Even participants who did not report awareness of the regularities in the artificial language they were learning performed better than chance level on the tests. These results should be interpreted with caution, however, as the second experiment of the same article used a slightly less natural artificial language and did not find learners without awareness performed better than chance level.

While these studies were discussed in Leow and Bowles (2005) were not necessarily based on incidental learning experiments, a very recent study by Andringa (2020) does tie together theories about awareness and incidental learning in SLA. In this paper, Andringa aims to investigate whether awareness emerges in participants who are not instructed about the syntactic rule they are learning. Moreover, he maps out the learning process using eye-tracking to see whether implicit awareness is a prerequisite for implicit learning to occur.

Participants did a visual world eye tracking experiment while learning a novel miniature language based on Esperanto, in which determiners were marked for distance and animacy. All participants were coded for their awareness of the pattern based on their answers to the debriefing questions (Andringa, 2020). This method of awareness measurement is not ideal because it relies on participants' memories of when they were and were not aware of the rule during the experiment, but other methods would have disrupted learning.

Results show that most participants never became aware of the rules in the determiner system. They did not show signs of predictive eye movements in experimental trials which would have indicated awareness and learning. However, the eye movements of the small number of participants who did indicate awareness of the rules did show an ability to predict the target noun before onset. Furthermore, there was a significant difference between their behaviour prior to and after becoming aware of the rule. Therefore, the correct use of the determiners was fully dependent on awareness of the rules involving these determiners (Andringa, 2020). This supports Schmidt's noticing hypothesis (Schmidt, 1990), as participants who remained unaware of the rules did not seem to implicitly still be able to use them and awareness was therefore required for learning.

1.1.2 Learning styles

In the studies mentioned up to this point, researchers studied specific processes and sequences found among participants overall and drew conclusions that they hoped to be able to generalise to all language learners. In linguistics in general, the study of universals is a very common goal. However, individual differences between learners in SLA are equally or more important in predicting a learner's success in second language learning (Dörnyei & Skehan, 2003).

An early review of research on individual differences in SLA was written by Skehan (1991). He specifically focused on four areas of impact regarding differences between individual learners: language aptitude, motivation, learner strategies and learner styles. Skehan speculates that learner strategies and learner styles may be able to mediate the influence of aptitude and motivation on learner success.

Arguably the most influential factor to predict success in SLA is language aptitude, which is a talent for learning language independent of intelligence and previous learning experience. (Skehan, 1991). Skehan (1989) proposed two profiles of language aptitude: analytic and memory-oriented aptitude. For the former, language learning is a case of pattern recognition and analysing and acquiring rules. The latter group views language as a collection of chunks that have communicative functions by themselves. These learners produce language by storing the chunks and swapping lexical information in and out. The aptitude profiles determined learners' preferences for instructional methods as well. Skehan's choice to gather the scale of analytic vs. memory-oriented under language aptitude profiles is noteworthy, as other researchers who use a similar model classify it as a model on learner styles. This highlights the lack of a single focused definition of the term "learning style" in the literature.

Skehan (1991) defines learning styles as "a general predisposition [...] towards processing information in a particular way" (p. 288). He describes how early research mostly separated field dependent learners, who deal with information as whole structures and like collaborative learning, from field independent learners, who analyse information into component parts and work independently (Dörnyei & Skehan, 2003). Skehan sees links between the learner styles of field dependence and the aptitude type: there seems to be a clear similarity between the field independent identity and the analytic aptitude type, but field independence and a memory-oriented aptitude do clearly not correspond. However, he differentiates between the two models, stating that field dependence is a binary measure, while the aptitude types analytic and memory-oriented can exist in a single individual, being a two-dimensional scale.

In another work, Dörnyei and Skehan (2003) again provide an overview of studies on individual differences in second language learning, on the same four topics as Skehan (1991). They are the first to differentiate between cognitive and learner styles, describing cognitive style as a "predisposition to processing information in a characteristic manner" and the latter as a "typical preference for approaching learning in general" (Dörnyei & Skehan, p. 602).

Based on more recent research, Dörnyei and Skehan (2003) critically evaluate the classic field dependence model. Critics call the model outdated, showing that there are very low correlations between field dependence and learning and that the model always favours field independent learners (Griffiths & Sheen, 1992). Proponents of the model extend the classic model to more clearly reflect current knowledge on learner styles (Chapelle & Green, 1992).

Skehan (1998) himself has criticised the field dependence/independence model for its binary nature. Rather, he prefers differentiating between an analytic orientation and a memory orientation. Instead of a binary scale, this model consists of two dimensions: a learner can be

strong or weak in both the analytic orientation and the memory orientation. This approach is striking, as Skehan has previously described this analytic vs. memory orientation contrast as a difference in aptitude profiles rather than a cognitive style (Skehan, 1991). This illustrates once again that the lack of a precise definition of learner styles and cognitive styles is one of the shortcomings of the field, because it leads to such a wide range of learning patterns being described as the same thing.

An investigator whose work is related to learning styles, but who does not explicitly name any of the models is Ullman, who described his declarative/procedural model in a 2016 paper. This is a neurological model of language, which has parallels to Skehan's (1998) analytic vs. memory orientation learning style split.

In his theory, Ullman (2016) theorises that the declarative and procedural memory systems must play a large role in language acquisition, representation and production. The declarative memory system is critical for learning explicit new information while the procedural memory system plays an important role in implicit learning (Ullman, 2016). According to the redundancy hypothesis (Ullman, 2004), both systems can acquire the same knowledge and skills, making both at least partly redundant. Various factors determine which system is used for a certain task. For example, new information may be acquired first through the declarative memory system and become ingrained by being taken over by the procedural memory system. Ullman (2016) defines a number of predictions for language based on established knowledge of memory systems that should, according to his theory, also apply to language, after which he presents evidence from empirical language studies.

Ullman (2016) predicts linguistic knowledge, like semantic memory and grammatical rules, is (at first) rapidly learnable using declarative memory systems. This knowledge may be explicit or implicit. Further, procedural memory systems underlie learning of sequences and probabilistic rules. This is the system that acquires syntax implicitly, or it should play a role in the production of syntax once the process has been automated. For his evidence, he cites behavioural studies that found correlations between vocabulary size and declarative memory learning abilities, while syntax acquisition is correlated to performance on procedural memory tests (Kidd, 2012).

The procedural/declarative model states syntax can be acquired by both the procedural and the declarative memory systems (Ullman, 2016). Individual preference for or predisposition to one of the two systems could be an influential factor in which system is used to acquire certain grammatical forms. Ullman's theory could then be used as neurological basis for the learner types, but this remains speculative.

1.2 The present study

With this study I aim to explore the incidental acquisition of German morphosyntax. Specifically, this study examines the main research question: "How successful is incidental grammar acquisition in adult Dutch native learners of German?" I will study this question with regards to acquisition of the German nominative, accusative and dative cases in interaction with gender as marked in the definite article and vowel change rules in verb inflection. In descending order of importance, the following sub-questions will be addressed:

1. How much grammar do learners incidentally acquire in an implicit learning context? Is there a difference between lexical learning (vowel change) and rule learning (case)? I expect some incidental learning will take place, expressed in above-chance proportions of correct answers on a GJT. This hypothesis is based on other incidental

learning studies in natural languages and artificial grammars finding above-chance accuracy on similar tests (e.g. Denhovska & Serratrice, 2017). Moreover, I hypothesise that rule learning will be more effective than lexical learning, which is more arbitrary and has no influence on the perceived semantics of a sentence for this group of learners.

2. If there is evidence for learning: is learning more strongly based on analogy to previously seen material (data-driven) or is it rule-driven? Does this differ between individual learners, i.e. can we find evidence of learning styles (analytical vs. memory-oriented) based on these data?

Evidence from the literature on this topic is inconclusive. I will be testing this hypothesis by testing for an old-new effect in the GJT response accuracy. If participants perform better in a testing phase on material they have previously seen in a training phase, learning seems to be data-based. On the other hand, an absence of this effect (given a general above-chance performance) would indicate participants have learnt rules rather than instances of specific sentences, because they are able to generalise the rules to unseen material. A presence of an old-new effect in combination with above-chance performance on the GJT in previously unseen sentences would indicate that participants *have* abstracted the rules but still use data-based learning in their acquisition.

However, there might also be a large variety between participant's individual old-new effects, dividing the participant set into two groups. Because I will be unable to use procedural and declarative memory tests, it is impossible to determine without question whether the participants can be divided into analytical and memory-oriented learners based on their learning style. These memory tests would then be able to integrate Ullman's (2016) declarative/procedural model with Skehan's (1991) memory-oriented vs analytical learners model, with memory-oriented learners relying more heavily on the declarative memory system while analytic learners rely on the procedural memory system. Results would show a correlation between participants' preferred memory system and the old-new effect tested in this experiment.

As it is, data visualisation and individual differences between participants in this experiment may suggest the presence of the learner style division among participants but the engagement with this question can only be an tentative inquisition and any outcomes must be considered with caution.

3. How do the specific grammatical rules in focus compare to one another with respect to difficulty of acquisition for adult Dutch learners of German?

Dutch does not have an active case system outside of its personal pronouns. Because of the stark difference between nominative pronouns and accusative or dative pronouns in Dutch, I hypothesise that participants will experience more difficulty with the perception of the distinction between accusative and dative German cases than the distinction between nominative and accusative or dative case.

In addition to these three main sub-questions, the following minor questions will be investigated:

4. What is the relationship between the development of different language modalities? Is there a correlation between proficiency in the receptive and the productive modalities?

It is plausible to expect that higher receptive proficiency promotes higher productive proficiency. If that is the case, I would expect a correlation between the percentage of incorrect answers on a receptive task and a productive task. However, as these tasks are also very different in nature and training is only provided as a receptive task, this correlation could very well be absent. This would be in accordance with Denhovska and Serratrice's (2017) findings, in which participants did show learning on a receptive task, but not on a productive task.

5. What is the role of awareness in the incidental acquisition of grammar?

Based on Schmidt's (1990) noticing hypothesis, I expect participants who later indicate awareness of the purpose of the study or who demonstrate conscious awareness of the relevant syntactic rules in the debriefing questions (without having demonstrated this awareness in the screening) will perform better on the GJT. However, as this measure will be based on debriefing answers, it will not be possible to deduce at which point in the experiment awareness emerged in the participant and the number of participants who reach awareness cannot be predicted. Therefore, any analyses of this topic will only be explorative.

To answer these questions, I will be conducting an online behavioural experiment with a training phase in which participants listen to grammatical German sentences while performing a simple task in which they need to select the video that corresponds to the semantics of the sentence out of two alternatives. After completing the training phase, participants continue to a production task in which they produce German sentences based on clips from the same database (Muylle et al., 2020) and a grammaticality judgement task.

Chapter 2: Methods

2.1 Participants

46 participants (nine male, 37 female, age range: $M=23.4$ years old, $SD=4.9$ years) participated in the experiment. They were recruited using the DCC's SONA participant database and paid €12.50 or 1.25 participant credits for their time spent filling out the online survey. All participants were native speakers of Dutch. Five screening questions (see procedure section for more detail) were used to confirm participants were not (too) familiar with the German case system, as a result of which four participants were removed from the dataset. All participants described their experience with the German language as little to none and their proficiency as very low to low. As German is a required foreign language class in most Dutch secondary schools for at least two years, participants reported having taken zero to three years ($M=2.4$) of German classes.

2.2 Materials

Visual stimuli

The visual stimuli used for this experiment were created by Muylle, Wegner, Bernolet and Hartsuiker (2020) for use in sentence elicitation. These videos have been experimentally proven to elicit intransitive, transitive and ditransitive sentences in English. They depict an agent, which is always a character (intransitive sentences), or an agent and a patient (both characters, in transitive sentences), or an agent, an object, and a beneficiary (ditransitive sentences). Figure 1 shows a table with a number of examples of the German translation of the sentences in this database with their English original. 18 out of the 21 German character and object nouns were cognates with Dutch. Of the set of 423 three-second long video clips, 79 were used in the implicit training phase and 18 were used in the production test, nine of which participants had already seen during the training phase and nine of which were new. A still of one of these clips is depicted in figure 2.

German sentence	English translation
Die Ärztin schläft.	The doctor sleeps.
Der Pirat winkt.	The pirate waves.
Der Matrose küsst den Lehrer.	The sailor kisses the teacher.
Die Hexe schlägt die Schwimmerin.	The witch punches the swimmer.
Der Leibwächter liefert der Kellnerin das Buch.	The bodyguard delivers the book to the waitress.
Der Boxer verkauft dem Lehrer die Tasse.	The Boxer sells the cup to the teacher.

Figure 1: examples of six sentences depicted by videos in the Muylle et al. (2020) corpus.

For use in the word learning phase, in which participants learnt vocabulary and familiarised themselves with the visual material, I used pictures of the different human figures and objects that were provided by Muylle et al. (2020) with the video database (for the full database, see <https://osf.io/4awyu/>). An overview of these characters can be found in appendix 1.



Figure 2: still of a video depicting a transitive sentence (the knight shoots the pirate) from the Muylle et al. (2020) corpus.

Sentence selection

The sentences corresponding to the videos from the Muylle et al. (2020) database were translated into German for use in this experiment. All of the characters were divided into three sets, to be divided over the implicit training phase and the two test phases. All characters could occur in the nominative case in the training phase and both testing phases. Set A contained five characters (3 of which masculine) that could occur in the accusative and dative cases in the implicit training phase, as well as being added to both testing phases to act as old (previously seen in the training phase) material. Set B contained five characters (three of which masculine) that could occur in the accusative and dative cases in the GJT. Set C contained four characters (two masculine) that could occur in the accusative and dative cases in the production task. Three of the characters (a policeman, a pirate and a sailor) were not assigned to any of the sets because their case inflections in the dative and accusative cases were not only reflected on the article but also on the nouns, which would have complicated the rule system for participants. The specific distribution of each character into sets and the number of occurrences per character per experiment phase can be found in appendix 2.

79 sentences were selected for the implicit training phase. Of these, eight were intransitive, 31 were transitive and 40 were ditransitive. The sentences were balanced as well as possible with respect to the number of occurrences of the three main case-gender interaction rules I was testing. In this system, the article inflects for gender. Assuming participants would easily and successfully acquire the nominative articles, which are *der* (masculine) and *die* (feminine), the most important rules are the following three:

1. The masculine article in the accusative case is *den*.
2. The feminine article in the dative case is *der*.
3. The masculine article in the dative case is *dem*.

Sentences for which these rules are used occurred 21 times (for the third rule) and 19 times (for the first and second rules) in the implicit training phase. Another 20 sentences that did not fit into any of these categories were used as fillers. These sentences were either intransitive, or had a feminine NP as the object. For this experiment, I was not specifically interested in the feminine accusative case, because its article, *die*, is the same as the feminine nominative article. Notably, the first rule is technically also presented to participants in 20 ditransitive trials in addition to the 19 transitive trials it was used for, because two of the objects in the sentences are masculine nouns in German. Moreover, eight out of 79 sentences are presented in a slightly marked, but still grammatical word order, which does not correspond to the standard Dutch word order (for example: *Den Hut liefert der Polizist der Kellnerin*). This was done to prevent participants relying purely on word order in their possible extraction of the rules. Three out of the total of twelve verbs in the dataset are stem-changing verbs in German, which is a lexical rule rather than a grammatical one. These have a different vowel when inflected for the third person singular than the vowel in the infinitive. These verbs (*schlafen*, *schlagen* and *geben*) occurred a total of 17 times in the implicit training phase.

For both comprehension and production testing phases, the occurrence of each of the rules was approximately balanced as well. Of the 47 correct sentences in the GJT, thirteen were classed as rule 1, thirteen as rule 2, 14 sentences were classed as rule 3 and eight belonged in none of the categories. 23 of these 47 correct sentence stimuli were new, in the sense that participants had not previously encountered them in the training phase. That also means that the direct and indirect objects in these sentences are from set B, hence I can test whether participants can generalise the rules to new nouns. In the production task, each of the rules was represented by four sentences, supplemented by six sentences that did not belong in any category. Of this total of 18 sentences, nine were new material. The direct and indirect objects in these sentences were characters from set C. Appendix 3 contains the specific numbers for the occurrences of each grammatical rule per experiment phase.

The incorrect sentences for the GJT were not reliant on video material from Muylle et al. (2020). They were created to be approximately balanced for the number of violations per rule (rule 1 occurred twelve times, rule 2, thirteen times and rule 3 occurred 14 times) using the characters from sets A and B and the verbs from the dataset. Thirteen sentences contained a stem-changing verb. Furthermore, the way in which they are ungrammatical is based on likely mistakes Dutch learners of German are likely to make, for example incorrect use of the vowel change in the verb, using the nominative article in the dative or accusative case and using the feminine article *die* for masculine words. All sentence stimuli used in all experiment phases are recorded in appendix 4.

Stimuli recording and experiment distribution

The German words and the sentences for the word learning phase, implicit training phase and the GJT were recorded by a native speaker of German using Voice Record Pro for iPhone.

Because of preventative measures to limit the spread of COVID-19, the experiment was distributed using a Qualtrics survey instead of an in-person experiment. Audio recordings, images and video files were imbedded in the survey. The production task, which would have been recorded in case of an in-person experiment, was changed to a typed task.

2.3 Procedure

Participants could access the Qualtrics survey via the SONA system after signing up. In the study description, I described the aim as “researching how well Dutch people learn German vocabulary”, to ensure participants paid attention to the meaning of the words rather than the syntax of the sentences and to ensure learning is incidental. After providing their informed consent and some demographic information, participants proceeded to the experiment itself.

First, in the screening phase, participants translated five sentences from Dutch into German. The target sentences contained a variety of gender and case combinations, to be able to control for participants’ level of German. Participants were asked to translate the sentences to assess the size of their German vocabulary and to guess the words if they did not know them. Four participants who showed (some) knowledge of at least one of the rules in question were excluded from the rest of the study.

Next, in the word learning phase, participants were shown images of the characters and objects, with the German and Dutch word below. Both the German word and the translation were presented with the article. The verbs were not depicted, only presented in German with a Dutch translation. Each screen contained a single word and participants would click through to the next screen in their own time. Participants studied a total of 33 vocabulary items this way. They studied all words from the old stimuli (presented in sentence training and in the testing phase) and the new stimuli (only presented in a sentence in the testing phase).

Then, in the implicit training phase, participants completed 79 trials of a two-alternative forced choice task. They listened to a German sentence and had to choose which of the two clips they saw depicted that sentence. The clips differed from one another in characters involved and verbs to ensure participants did not need to rely on the NP inflections to answer the question. They increased in both vocabulary and grammatical difficulty: the first block contained only intransitive sentences, the second contained transitive, and the third ditransitive sentences. Participants thought the purpose of this phase was to test how well they had learnt the vocabulary, but it was a way to very implicitly present grammatical features and the German case system.

The final phase was the testing phase, which was twofold. First, in the production task, participants saw a video and were asked to write what happens in one sentence. These videos from the Muylle et al. (2020) corpus elicited sentences similar to the ones participants had previously listened to in the implicit training phase. At this point, they were still unaware of the fact that the study (also) focussed on the syntax of the sentences. The second part of the testing phase was a GJT in which participants listened to a sentence and were asked whether it was grammatically correct. Appendix 5 contains screen captures of what participants saw during the implicit training phase, production task and GJT.

2.4 Data-analysis

Data were extracted from Qualtrics and edited in Excel to combine the scores per participant per category of trial. I coded participants for evidence of knowledge of each grammatical rule judging from their answers to the screening questions. Based on participants' answers to the debriefing questions, they were coded for two types of awareness: awareness of the study's aim after finishing it and awareness of the correct grammatical rules during the experiment. For the first type of awareness, participants who thought the study focused on the German case-gender system (or something similar) were classed as aware. The second type of awareness was to be based on awareness of the correct grammatical rules. However, participants reported such highly varied levels of understanding of the rules in question that this measure was not used in any further analyses.

For the production task, I coded each of the trials for each participant for mistakes. These mistakes were categorised according to each of the six gender and case rules (two genders combined with three cases) that could be violated. In addition, two mistake categories were added for "vowel change" and "other". The only sentences that were included in the vowel change category were the sentences containing stem-changing verbs. For each trial, I noted whether participants had made errors in the mistake categories that were relevant for that trial. GJT items were coded per participant for total percentage incorrect and percentage of hits and correct rejections. The grammatical items were further divided into percentage of old (previously heard) items incorrect and new items incorrect. The ungrammatical items were grouped by grammatical rule violated to form five mistake type categories, three for each of the central case-gender rules with two added for the masculine nominative and vowel changes. Participants were coded for the proportion of items answered incorrectly per mistake type.

To examine whether participants had been successful at the incidental acquisition of the grammatical features at focus, I performed a one-sample t-test comparing the participants' results in the GJT to the 50% chance level overall. A binomial distribution calculator was used to investigate how well individual participants performed compared to chance level.

Using a paired samples t-test, I compared participants' performance on the GJT items that were new to the GJT items that were old stimuli (previously seen in the implicit training phase). This way, I could demonstrate whether or not there was an old-new effect among learners, and whether or not participants were able to generalise their learnt knowledge to new lexicon. A similar old-new effect test inspected the proportion of mistakes made on old and new items in the production test. If learning occurred and there were large differences between participants for this old-new effect, this could be a tentative indication of the presence of learning styles as well. However, in the end learning styles could only be tentatively analysed using descriptive measures and data visualisation.

Furthermore, a paired t-test determined whether there is a difference between learning case inflections (rule learning) and learning vowel changes (lexical learning), comparing the

proportion of incorrect GJT answers on sentences with the wrong case inflection to the proportion of incorrect answers on sentences with the wrong presence or absence of a vowel change. To inquire into the more specific difference in the level of difficulty between the different grammatical rules, a repeated measures ANOVA established whether the proportion of incorrect answers differs between the different rules. A similar repeated measures ANOVA was performed on the production task to compare the relative difficulty of sentences containing each rule per experiment task.

After these main research questions have been tested, two smaller tests explored awareness and the relationship between scores on the GJT and the production task. Awareness was tested using a paired samples t-test on the different awareness groups. I calculated Pearson correlations to inspect the relationship between the GJT and the production task.

Chapter 3: Results

The main part of this analysis is based on participants' performances on (a portion of) the grammaticality judgement task. An overview of the proportion of incorrect answers to correct answers on the complete GJT (94 items) per participant can be found in figure 3. Clearly, most learners scored near the 50%-chance mark.

Some of the analyses also focus on performances on the production task. An overview of the proportion of incorrect answers on the production task (18 items) per participant is given in figure 4. To obtain this total, the mistakes made in all of the applicable mistake categories per sentence were added up for each participant and divided by the total number of possible mistakes over all sentences.

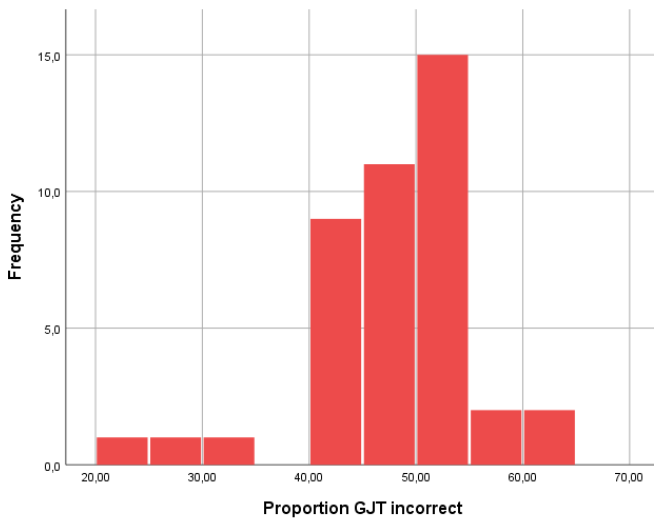


Figure 3: histogram of the proportions of incorrect GJT answers per participant

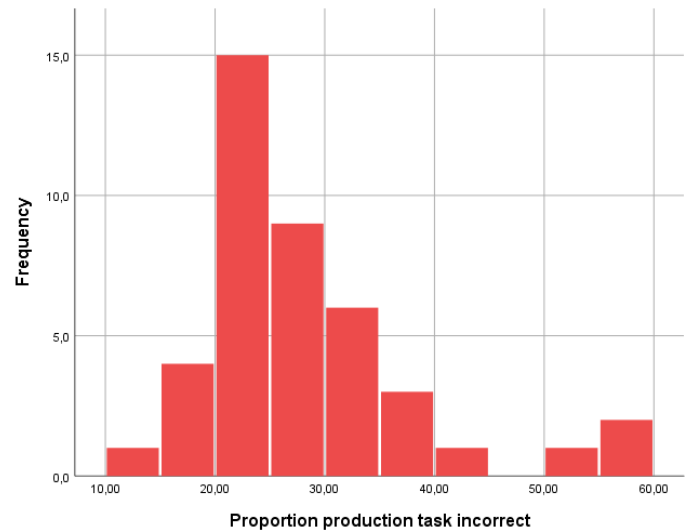


Figure 4: histogram of the proportions of incorrect production task answers per participant

3.1 Learning effect GJT

The presence of a learning effect was tested using a one-sample t-test, comparing the average percentage of incorrect answers on the GJT (total score) to chance level. Because of the number of participants ($N=42$) this test was robust to the assumptions of normality and homogeneity of variance. The mean score ($M=48.00$, $SD=7.89$) was slightly higher than chance level. However, this difference, -2.00 with 95% CI $[-4.46, 0.46]$ was not statistically significant $t(41)=-1.64$, $p=.108$, indicating that participants did not acquire the rules through incidental learning.

I used a binomial calculator to find the number of correct answers participants needed to achieve in order to exceed chance level. This showed that 56 correct answers on the GJT (out of a total of 94) surpassed statistical significance. Only three participants attained this number of correct answers and are therefore the only ones who seem to have learnt the rules incidentally, as they did not show indications of knowing them during the screening.

Interestingly, there is a clear contrast between participants' scores on ungrammatical items compared to grammatical items, as shown in figure 5. Participants have a strong bias towards "yes" answers to the question "is this sentence grammatically correct?". A paired t-test showed that on average, participants scored higher on the percentage of hits ($M=69.30$, $SD=16.11$) than the percentage of correct rejections ($M=34.70$, $SD=14.65$). This difference, 34.60, was significant $t(41)=8.48$, $p<.001$.



Figure 5: mean proportion of correct answers for hits and correct rejections on the GJT. Error bars show 95% confidence interval.

3.2 Comparison of old and new incorrect items

Another paired samples t-test was used to determine whether there was a difference between performances on old stimuli in comparison to new stimuli in the grammatical items of the GJT and in the production task. Based on the percentage of incorrect answers in the GJT, participants made fewer mistakes in the new stimuli ($M=29.61$, $SD=17.42$) than the old stimuli ($M=32.64$, $SD=15.96$). This difference of 3.03, 95% CI [-0.65, 6.72] was not significant in a two-tailed t-test $t(41)=1.66$, $p=.104$.

Surprisingly, participants performed better on new stimuli in the production task ($M=26.01$, $SD=10.76$) compared to old stimuli ($M=30.71$, $SD=10.29$). This time, scores were measured as the percentage of mistakes made out of a total of possible mistakes per sentence. This difference, 4.70, was significant $t(41)=4.22$, $p<.001$. Contrary to my hypothesis that participants would retain some memory of previously heard sentences, participants performed significantly better on new stimuli rather than old stimuli in the production task, while the GJT showed no difference between the two categories.

The old-new effect was also relevant to find any possible indications of the presence of learning styles. While learning styles cannot be reliably tested in a group of participants that did not show learning, there was a remarkably great variability in the size of the old-new difference in the GJT. Figure 6 shows that some participants performed over 20% better on the new items than the old items, while others did 20% better on the old items.

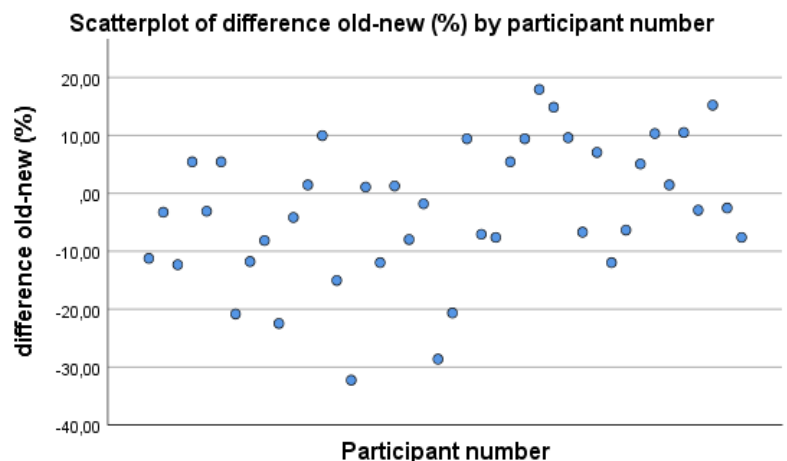


Figure 6: scatterplot of the difference between GJT score on old and new items per participant.

3.3 Analysis of mistake types

A more detailed inquisition into the types of mistakes made on the GJT lead to a comparison of rule-learning to lexical learning. The ungrammatical GJT sentences were categorised into five groups based on the rule they violated. Mistake types 1 up to 4 are rules of case and gender while mistake type 5 was a wrong absence (usually) or presence (rarely) of a vowel change on the 3rd person singular form of the verb. A paired t-test was used to compare the average proportion of errors made on trials for mistake type 1-4 ($M=59.83$, $SD=17.26$) compared to trials for mistake type 5 ($M=79.10$, $SD=14.97$). Results show participants perform better on grammatical rule trials compared to lexical learning trials. This difference, -19.27 , CI $[-26.04, -12.50]$, was significant $t(41)=-5.75$, $p<.001$.

As this is an average of all four case and gender rules violated in the ungrammatical trials, I performed a more complete comparison of all grammatical rule mistakes (mistake types 1-4) separately using a repeated measures ANOVA. Mistake type 1 to 3 were mistakes in masculine articles, respectively in the nominative, dative and accusative cases, while mistake type 4 was a mistake in feminine dative articles. Figure 7 shows a table of mean proportions of violation of each of the four grammatical rules in the ungrammatical component of the GJT.

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5)=24.10$, $p<.001$. Therefore, degrees of freedom were corrected using a Huynh-Feldt correction ($\epsilon=.75$). Unsurprisingly, results show a significant effect of the type of mistake on the proportion of incorrect answers to the GJT, $F(2.26, 92.78)=48.89$, $p<.001$. Further pairwise comparisons based on estimated marginal means indicate GJT scores do not differ between mistake types 3 and 4, $p=.436$. However, participants performed significantly better on trials with mistakes in masculine nominative articles (mistake type 1) in comparison to all other mistake types, all $p<.001$. Trials with mistakes in masculine dative articles (mistake type 2) were significantly more difficult than trials with mistake type 1 ($p<.001$) but significantly easier than trials with mistake type 3 ($p=.001$) or mistake type 4 ($p<.001$).

A similar comparison of all grammatical rule mistakes was performed on the production task scores using another repeated measures ANOVA. Mistakes types 1 to 4 were the same as on the GJT. Mistake type 6 was a mistake in feminine nominative and mistake type 7 in feminine accusative. Figure 8 shows a table of mean proportions of errors made in each of the mistake categories in the production task. To do this, the number of sentences in which participants made that particular mistake were divided by the total number of sentences in which they could have made that mistake.

Mauchly's test indicated that the assumption of sphericity had been

GJT: percent incorrect answers per mistake type			
Category		M	SD
Grammatical	% mistake 1	31.54	24.72
	% mistake 2	59.30	19.38
	% mistake 3	72.91	23.89
	% mistake 4	75.54	26.06
	average 1-4	59.83	17.26
Lexical	% mistake 5	79.10	14.97

Figure 7: table of mean percentages of incorrect answers and standard deviations per mistake type on the ungrammatical portion of the GJT. Mistake types: 1= m. nominative, 2=m. dative, 3= m. accusative, 4= f. dative, 5= vowel change.

Production task: percent incorrect answers per mistake type			
Category		M	SD
Grammatical	% mistake 1	23.44	30.00
	% mistake 2	88.60	24.76
	% mistake 3	93.11	18.04
	% mistake 4	91.44	19.09
	% mistake 6	9.73	24.77
	% mistake 7	16.51	26.63
Lexical	% mistake 5	65.76	26.92
Other	% mistake 8	3.23	4.63

Figure 8: table of mean percentages of incorrect answers and standard deviations per mistake type on the production task. Mistake types: 1= m. nominative, 2=m. dative, 3= m. accusative, 4= f. dative, 5= vowel change, 6= f. nominative, 7= f. accusative, 8=other

violated, $\chi^2(14)=77.08$, $p<.001$, so degrees of freedom were corrected using a Greenhouse-Geisser correction ($\epsilon=.57$).

Results show a significant effect of the type of mistake on the proportion of incorrect answers on the production task, $F(2.88, 103.82)=111.06$, $p<.001$. Further pairwise comparisons based on estimated marginal means indicate that the masculine nominative case (rule 1) was significantly easier than the masculine dative and accusative and feminine dative (rules 2, 3 and 4, all $p<.001$), and relatively more difficult than the feminine nominative (rule 6, $p=.023$). It was, however, not significantly more difficult than the feminine accusative (rule 7, $p=.250$). There was no significant difference between masculine dative (rule 2) and accusative (rule 3) and feminine dative (rule 4) in the proportion of mistakes made (all $p>.05$), but they are all significantly more difficult than the masculine nominative, feminine nominative and feminine accusative (rule 1, 6 and 7, all $p<.001$). There was no significant difference between the proportion of mistakes made on feminine nominative (rule 6) and feminine accusative (rule 7), $p=.009$.

3.4 Comparison of language modalities

Furthermore, a Pearson correlation was calculated to detect a correlation between participants' performance on the GJT and the production task. Participants' mean score on the GJT was 48.00% incorrect, $sd=7.89$, while the mean proportion of errors on the production task was 28.36%, $sd=9.89$. A positive correlation would be expected if participants who perform better on the one task also perform better on the other. However, Pearson's correlation showed there was no significant relationship between scores on both tasks, $r=.071$, $p=.653$. Figure 9 shows production task scores in relation to GJT scores. In general, most scores are clustered around the mean with a few outliers. Interestingly, the three participants who performed above chance level on the GJT (the three lowest scores), also did quite well on the production task.

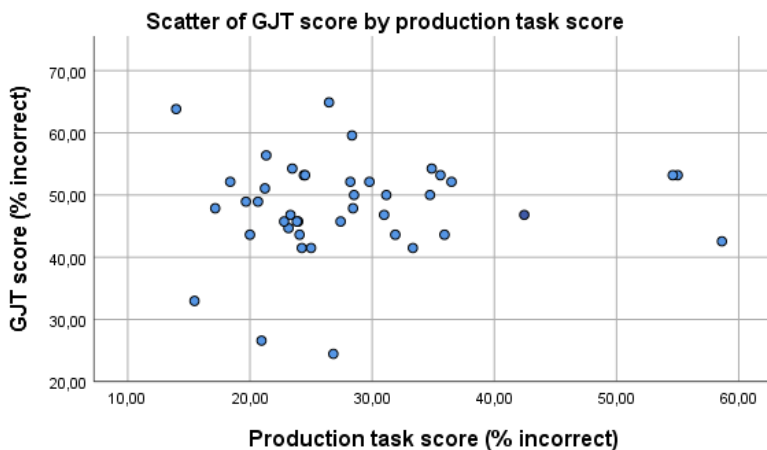


Figure 9: scatterplot of production task score in relation to GJT score.

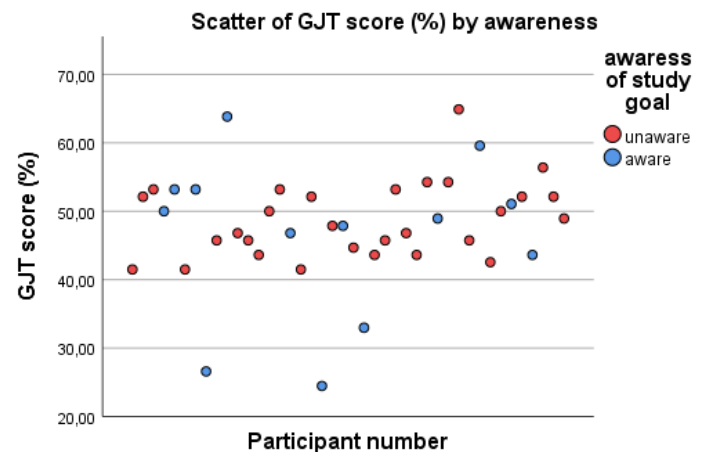


Figure 10: scatterplot of GJT score (% incorrect) per participant grouped by awareness.

3.5 Effects of awareness

Finally, I performed an independent samples t-test to test for an effect of awareness on the total proportion of incorrect answers on the GJT. Awareness was based on participants' answers to the question "what do you think is the goal of this study?", with 13 participants being classed as aware and 29 as not aware. The distribution of participant scores on the GJT when separated for awareness is shown in figure 10, which shows there does not seem to be a clear relationship between awareness and GJT score. On average, participants who were aware that the study's goal was acquisition of the case system performed slightly better ($M=46.32$, $SD=11.81$) than participants who said they did not know or who gave the wrong answer ($M=48.75$, $SD=5.43$). This difference, -2.44 , 95% CI $[-4.90, 9.77]$ was not significant at $t(14.32)=-0.71$, $p=.489$. Therefore, awareness as indicated in the debriefing questions did not have a significant effect on performance on the GJT. However, the only three people who scored above chance level did all report being aware of the study goal, as can be seen in the scatterplot.

Chapter 4: Discussion

The study described in this thesis aimed to investigate incidental grammar learning in Dutch learners of German. Specific attention was paid to three main research questions.

Firstly, I studied how much grammar adult learners incidentally acquire when implicitly presented with grammatically correct German sentences and whether there is a difference between lexical learning (vowel change) and rule learning (case). Results based on the grammaticality judgement task show that on average, participants did not show an incidental learning effect. Furthermore, only three participants performed significantly above chance level. An interesting fact was that performances were considerably better on grammatical items of the GJT in comparison to ungrammatical items indicating a strong “yes”-bias. In addition, participants achieved higher scores on ungrammatical items that violated case-gender interaction rules (grammatical rule learning) in comparison to ungrammatical items that violated vowel changes in German (lexical learning).

The second research question asked whether learning was more strongly based on analogy to previously seen material or whether it was rule-based. This was operationalised and tested for by comparing previously heard grammatical GJT items to new grammatical GJT items and by comparing scores on old and new production task items. There was no significant effect on the GJT, with participants even performing slightly better on the new material. Surprisingly, scores on the production task were significantly higher for new items in comparison to old items. The data collected were insufficient to accurately explore learning styles (analytic vs. memory-oriented) among participants because of the absence of a learning effect and because I was unable to perform declarative and procedural memory tests. However, there was a remarkably sizable variation in difference between scores on old and new stimuli on the GJT for different participants, with some participants scoring considerably better on old stimuli and some scoring considerably better on new stimuli.

The third main research question explored the differences in difficulty of acquisition for Dutch learners of German for specific grammatical rules. Accuracy scores show similar patterns of relative difficulty on the ungrammatical portion of the GJT and the production task: participants found the masculine and feminine dative and masculine accusative case rules significantly more difficult to acquire than the masculine nominative case. Furthermore, participants found feminine nominative case more easier to learn than masculine nominative, but equally easy as feminine accusative.

The final two research questions were more explorative in nature. The first inquired into the relationship between productive and receptive proficiency. Results showed that there was no correlation between scores on the GJT and the sentence production task. The second question explored the role of awareness in incidental acquisition of grammar. This was investigated using the overall GJT scores for participants and an awareness measure based on a correct answer to a debriefing question about the study’s purpose. These results showed no effect of awareness of the experiment goal on the GJT score.

4.1 Incidental learning

The absence of incidental learning effects means that participants were unable to incidentally acquire German case rules through implicit auditive presentation of correct sentences. This contradicts my earlier hypothesis that learning would occur, which was based on a number of previous studies on incidental grammar learning that did find above-chance accuracy rates or better performances than control groups (e.g. Denhovska & Serratrice, 2017; Franck et al., 2016; Williams, 2010). While this outcome was surprising, there are a number of possible reasons for its divergence from most previous studies:

One possible explanation is the number of stimuli in the implicit training phase. Participants were presented with 79 grammatical German sentences while they were asked to watch two videos and select the one that corresponded to the spoken sentence they heard. To avoid a ceiling effect, which I expected to occur if participants were presented with too many sentences, the number of stimuli in the implicit training phase was limited to 79. This number may have been too low for participants to abstract the rules, resulting in the absence of learning effects. For example, Williams (2010) exposed his participants to either 194 or 388 sentences of a language that was a mixture of English and Japanese, with participants benefitting from a 388-sentence exposure phase in comparison to 194 sentences for some sentence structures. Yet, a study investigating incidental acquisition of a similar grammatical element to the one currently being studied (namely gender agreement, but in adjectives) found positive results when exposing participants to only 36 stimuli in the implicit training phase (Denhovska & Serratrice, 2017). One should note that the learners in both of these experiments were complete beginners while the participants in the current study had taken some German classes in the past. The number of stimuli required will be largely dependent on the grammatical element to acquire, which was considerably more complex in Williams' study in comparison to Denhovska and Serratrice.

Based on DeKeyser (2005), acquisition of the German case system for Dutch speakers might be particularly difficult because of 'complexity of form': a gendered case system does not exist in Dutch outside of pronouns, which is why the concept can be both new and abstract for learners, which hinders acquisition. Thus participants might have needed more sentences in the implicit training phase to first understand the concept of grammatical case and to then extract the rules. It is also entirely possible that a gender and case agreement system like the one in the current study is too complex for Dutch learners to acquire through implicit exposure regardless of the number of stimuli in the training phase. These gender and case categories might be more suitable for explicit learning, which would be in agreement with DeKeyser's statement that meaningful grammatical elements were most successfully acquired through explicit learning, especially in cases of such limited exposure, and with Robinson's (1996) finding that instructed learners outperform those in incidental learning conditions on both complex and simple structures.

A second possible explanation is the method of presentation of the stimuli in the implicit training phase. Sentences were presented in auditory form for two reasons: firstly because hearing sentences is the most natural format of processing language in everyday contexts. Secondly, I expected to find a ceiling effect if sentences were presented in written form, in which it would have been easier to distinguish the different articles from one another. Previous studies, like Denhovska and Serratrice (2017) and Williams (2010) provided participants with written sentences (sometimes in addition to spoken sentences). In this case, participants may specifically have had difficulty distinguishing the masculine accusative article *den* from the masculine dative article *dem*. Orthography would also have drawn more attention to the variability in article form, which may have alerted participants to the relevant rules earlier.

Moreover, the task in the implicit training phase was quite simple to complete successfully. It was designed to not be overly complicated so participants would not have to consciously pay attention to the grammar and sentence structures to determine, for example, which character was the agent and which was the patient. I aimed to not distract them from the focus on semantics and in natural contexts it would be rare to need case or gender to disambiguate meaning. Thus, characters in the distractor items were always entirely distinct from characters in the target items and they were performing a different action. To be able to select the target item, it would have been sufficient to focus solely on the meaning without

considering grammar. This shallow processing of the grammar in the sentences may have led to them not subconsciously storing the syntactic structures.

The specific form taken by the absence of the learning effect can be found in the clear bias found among participants for answering “yes” to the question “is this sentence grammatical?” in GJT trials. The percentage of correct answers on grammatical trials was significantly higher (and far exceeds chance level) than the percentage of correct answers on ungrammatical trials. Rather than being too critical, participants are overly tolerant of multiple forms. This yes-bias may occur because participants’ individual grammars have stored multiple articles as optional for a single gender and case combination. This optionality is a phenomenon that has been observed before in non-native grammars, when L2 speakers perceive two variants of a given construction to both be grammatical (and therefore both variants are optional), while in native grammars only one of the variants is correct (Sorace, 2000). If this is the case for these learners of German, optionality may explain why they have a strong bias for judging multiple variants of a single gender and case combination as grammatically correct.

In summary, to make incidental learning more successful, researchers may choose to expose participants to a larger number of grammatical sentences in written form. If such a study has positive results, it could ascertain that the German case system can be acquired by native speakers of Dutch through implicit exposure only. However, if participants have still not successfully acquired the rules after exposure to a high number of stimuli, it would be short-sighted to conclude it is still due to a lack of exposure. An additional factor to enhance grammar learning would be to make the distraction task during the implicit training phase more syntax-dependent. For example, the addition of a small number of trials in which the videos contain the same characters in different grammatical positions would require more of participants’ attention and more conscious processing of the sentences. If proof for incidental learning is found, subsequent attention may be paid to questions whose answer is contingent upon the presence of learning.

4.2 Data-based or rule-based learning?

With regards to the question whether learning is based on previously heard material (data-based) or rule-based, a conclusive answer cannot be given. Because participants did not learn, the underlying type of learning to occur can also not be determined without question. However, tests for an old-new effect in both the GJT and the production task did return interesting results. As previous research was inconclusive on the topic (McAndrews & Moscovitch, 1985 found learning was neither fully example-based nor fully rule-based and Williams, 2010 also argues for data-based learning), there was no clear direction for this hypothesis. The presence of an old-new effect would indicate learning is data-based, while an absence points towards rule-based learning. Specifically, these new sentences contained characters in object or beneficiary positions that participants had never heard in that grammatical role before, while old sentences were literal repeats of the implicit training phase. The old-new effect on the GJT was not significant. This could have been because the new sentences (which contained new nouns) were not considerably different from old sentences, in which different nouns with the same gender had appeared in the applicable grammatical role. It may have been too easy for participants to see the analogies between the old and new sentences.

The production task did show an new-old effect, which was opposite to what was expected. This time performances on the new material were significantly better than the old material. There is no precedent in the literature for significantly better accuracy scores on new trials in comparison to previously heard trials. Attention factors may explain the remarkable

effect found in the production task which was opposite to the expected old-new effect. For example, newness could lead participants to pay closer attention to sentences containing new nouns in particular grammatical roles. This is something future studies could explore further.

Additionally, the absence of a learning effect and the fact that only three participants exceeded chance level in their GJT scores makes it impossible to draw any conclusions on learner styles based on the old-new effects. However, similar to McAndrews and Moscovitch (1985), my results show large differences between participants in their scores on old items compared to new items; some participants score over 20% higher on old items while other score over 20% higher on new items. This discrepancy may be an indication of the presence of learning styles in the way they are described by Skehan (2003). The people who score substantially higher on old items may be memory-oriented learners, while others could be analytical learners and focus more on abstraction of the rules. Therefore, the participants in this last group would be able to perform equally well or better on new items. Learning styles could not be investigated fully in this study because I was unable to perform procedural and declarative memory tests on participants (Ullman, 2016). A preference for either of these memory systems could be a neurological proof for the learning styles. For future research with an objective to detect learning styles among participants it would be advisable to determine which type of memory is most developed, which should then correlate to whether a participant shows an old-new effect. This then indicates which learning style a participant has.

4.3 Grammatical rule learning or lexical learning?

Investigation of participants' scores on the different types of items in the ungrammatical portion of the GJT shows that vowel rules in this experiment were more difficult to acquire than grammatical rules. This was expected, as lexical learning is fully arbitrary, while rule learning is based on an abstract system. However, rules of vowel changes in German are only one form of lexical learning, as I did not measure how well participants retained the vocabulary they studied during the first phase of the experiment. Moreover, it is a type of lexical learning that does not exist in Dutch, which, according to DeKeyser (2005) will make it particularly difficult to acquire.

However, as this measurement is based on an average of four measurements compared to a single measurement, a more detailed investigation of the difference between the difficulty levels of the grammatical rules is required, using both the production task and the GJT. While these are very different tasks, the relative performance on each mistake type is very similar in both tasks, highlighting that the same grammatical rules are difficult regardless of the task. I hypothesised that participants would experience more difficulty distinguishing between accusative and dative case in German in comparison to nominative vs. accusative and dative case stimuli, because the distinction between Dutch nominative and accusative or dative personal pronouns is very distinct, so Dutch learners are used to differentiating between these cases.

Out of masculine nominative case, masculine accusative case and masculine and feminine dative case in the GJT, participants found the nominative case the easiest to learn, as expected. Similarly, masculine and feminine nominative were both significantly easier than masculine and feminine dative and masculine accusative in the production task. Participants were most exposed to this case, as it was not balanced like the other three rules, which were specifically in focus in this study. Furthermore, participants were also exposed to the nominative articles in print during the word training phase and as the nominative case is the citation form in vocabulary lists, most had frequently heard these articles before in the one to three years of German lessons they took in secondary school.

Out of the other three rules on the GJT, namely that the masculine dative article is *dem*, the masculine accusative article is *den* and the feminine dative article is *der*, the first was significantly easier to learn than the latter two. This may be the case because the types of ungrammatical forms in the place of the correct articles for the masculine accusative and feminine dative case articles were mostly nominative articles with the correct gender agreement. Dutch learners of German are more likely to accept feminine nominative articles in the place of feminine dative articles and masculine nominative articles where masculine accusative articles were supposed to be, because that article would have been correct in Dutch, which lacks this case distinction in articles. Some of the ungrammatical forms in GJT trials which should have had a masculine dative article were similar, but others were replacements with masculine accusative or feminine nominative articles. The fact that they were more likely to mark these types of mistakes as grammatically incorrect may be related to Sorace's (2000) previously discussed theory of erroneous optionality in non-native grammars; it is an indication that the options for one type of construction stored in these learners' individual grammars may always be: the grammatical article for the syntactic function, as well as the nominative article.

This distinction between masculine accusative and feminine dative being more difficult than masculine dative was not significant in the production task results, probably due to the fact that participants were not limited to the ungrammatical forms offered by the GJT. They could make whichever mistakes were correct according to their internal grammars, often opting to use the nominative case article of the correct gender rather than the grammatical gender-case combination. Moreover, the reason the feminine accusative was relatively easier than the masculine accusative can also be explained using Sorace's (2000) optionality theory: because the feminine accusative article *die* takes the same form as the nominative article, learners whose internal grammar contains two options (the nominative and the grammatical form) will always be correct in case the NP is feminine.

For future research into grammatical rule learning versus lexical learning, it would be interesting to increase the number of types of lexical learning in the study, to be able to have the same number of grammatical rules as lexical "rules" to be acquired, particularly to investigate the variance between the different types of lexical learning. Furthermore, a stronger balance between the types of mistakes made for each grammatical rule violated when creating the stimuli would result in a more honest depiction of difficulty levels of each rule. While slightly unrealistic, asking participants for the reason they accepted or rejected a GJT item would also provide a lot of insight into the individual grammars on which they base their decisions, which would show which rules participants have and have not successfully acquired and the origin of their mistakes.

4.4 Explorative questions

The first of the explorative research questions was the correlation between performance on the productive and the receptive tasks. The fact that this correlation was not significant at all could be because of the deeply differing nature of the receptive and productive tasks. Participants performed a receptive task in the training phase, rather than a productive one. Then, when unexpectedly tasked to produce correct full sentences, participants who had taken German lessons for longer or who had been more recently exposed to German may have performed slightly better than others, even after more extreme cases of proficiency had been removed from the data. At this point in the study, the apparent study goal was still to study how well people could acquire vocabulary, with the actual goal of studying syntax only becoming clear after the introduction of the GJT. This may have been another cause for the dissociation between both tasks. Participants were also commonly unable to produce a full

sentence in response to a production task prompt, let alone use the articles that agreed with the correct case. Therefore, a lot of responses in the production task were unusable for this particular scoring system and objective of study. This could be remedied in future studies by preparing participants better for a productive task through more required engagement with the material and more exposure.

The question of awareness cannot be fully answered using these data, mainly because of the absence of a learning effect in all but three participants. Results show no effect of awareness, which contradicts Schmidt's (1990) noticing hypothesis. This study would not be the only study to provide this contradiction, but the measure of awareness used may not have been entirely reliable, as it is based on debriefing answers and on participants who did not successfully acquire the rules. Most effects of awareness have been found when participants report awareness while performing the training task (e.g. Leow, 2001). Studies that were unable to find awareness effects often asked participants to report awareness in a debriefing questionnaire (as mentioned in Leow & Bowles, 2005) like the one used in this study. Moreover, the measure of awareness was based on participants' answers to the question "what do you think is the goal of this experiment?", with participants who answered correctly being categorised as aware. Despite having this type of awareness, however, some participants were unable to list the grammatical rules they had used during the study or reported not having consciously used any rules at all. As found in Leow and Bowles, there are many types and levels of awareness and my usage of a lower level may be the reason why there was no significant effect of it. Something else to be examined in future studies is the possible interaction between awareness and the correlation between multiple task types; the correlation between receptive and productive task scores is likely to be higher in participants who indicate awareness of the purpose of the study.

Chapter 5: Conclusion

Incidental learning of second language grammar was investigated using an experiment in which Dutch adult learners of German were implicitly presented with correct German sentences with target structures involving the German case system and vowel changes in the third person singular verb. Learning was tested using a receptive GJT task and a productive task in which participants described the actions occurring in a short video, all based on a dataset of short virtual reality clips by Muylle et al. (2020). Results show participants were unable to incidentally acquire the particular gender-case rules in focus, which illustrates that these structures are more difficult for Dutch native speakers to internalise through implicit exposure to a limited set of stimuli than previously expected. Because of the absence of similar structures in the Dutch grammar, this indicates the difficulty of acquiring grammatical elements that are novel and abstract for learners if their native language does not distinguish between certain categories. These structures seem to be more suited for explicit instruction, in which learners are consciously made aware of the new rule before they attempt to internalise it in their individual grammars. Incidental acquisition may be possible as well but would most likely need more exposure in multiple modalities. Which structures in particular are experienced as more difficult and which mistakes learners make, seems to be mostly based on interference of the native language, and does not differ between language modalities. A structure that is similar in the native language is easily learnt, while learners are likely to incorporate native structures into their individual grammars for grammatical rules they have not acquired yet.

While evidence for either data-based learning or rule-based learning is still largely inconclusive, large differences between learners have been observed with respect to old-new effects. Future research into individual differences and preferences for either data-based or rule-based learning would be a highly valuable addition to this discussion.

This study highlights once again the difficulty of acquiring a grammatical structure in a second language that does not exist in the learner's native language. Case systems may therefore not be learnable through mere exposure in natural contexts, particularly if correct understanding of case rules is hardly ever needed to understand meaning in conversation.

References

- Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, 31, 175–198.
- Chapelle, C., & Green, P. (1992). Field independence/dependence in second language acquisition research. *Language Learning*, 42, 47-83.
- DeKeyser, R. M. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning*, 55(s1), 1-25.
- Denhovska, N., & Serratrice, L. (2017). Incidental learning of gender agreement in L2. *Journal of psycholinguistic research*, 46(5), 1187-1211.
- Dörnyei, Z., & Skehan, P. (2003). Individual Differences in Second Language Learning. In C.J. Doughty & M.H. Long (eds.), *The Handbook of Second Language Acquisition*, 589-630. Malden, MA: Blackwell.
- Franck, J., Rotondi, I., & Frauenfelder, U. H. (2016). Learning structure-dependent agreement in a hierarchical artificial grammar. *Journal of Memory and Language*, 87, 84-104.
- Griffiths, R., & Sheen, R. (1992). Disembedded figures in the landscape: a reappraisal of L2 research on field dependence/independence. *Applied Linguistics*, 13, 133-48.
- Hulstijn, J. H. (2003). Incidental and Intentional Learning. In C.J. Doughty & M.H. Long (eds.), *The Handbook of Second Language Acquisition*, 349-381. Malden, MA: Blackwell.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48(1), 171-184.
- Leow, R. P. (2001). Attention, awareness, and foreign language behavior. *Language Learning*, 51(Suppl. 1), 113–155.
- Leow, R. P., & Bowles, M. A. (2005). Attention and awareness in SLA. In C. Sanz (ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice*, 179-203. Georgetown, DC: Georgetown University Press.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender marked articles by native and nonnative Spanish speakers. *Journal of Memory and Language*, 63, 447–464.
- McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory & Cognition*, 13(5), 469-475.

- Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure - An ERP study in adult learners. *Bmc Neuroscience*, *10*(89).
- Muylle, M., Wegner, T. G., Bernolet, S., & Hartsuiker, R. J. (2020). English norming data for 423 short animated action movie clips. *Acta Psychologica*, *202*, 102957.
- Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in second language acquisition*, 27-67.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.
- Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11-26.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.
- Skehan, P. (1991). Individual differences in second language learning. *Studies in second language acquisition*, *13*(2), 275-298.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.
- Sorace, A. (2000). Syntactic optionality in non-native grammars. *Second Language Research*, *16*(2), 93–102.
- Taraban, R. (2004). Drawing learners' attention to syntactic context aids genderlike category induction. *Journal of Memory and Language*, *51*, 202–216.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231-270.
- Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953-968). Academic Press.
- Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.
- Williams, J. N. (2004). Implicit learning of form-meaning connections. In J. N. Williams, B. VanPatten, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Lawrence Erlbaum.
- Williams, J. N. (2010). Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning*, *60*, 221-244
- Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, *118*, 522–553.

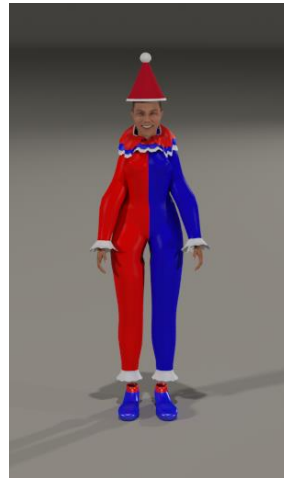
Appendix 1: Images of characters and objects



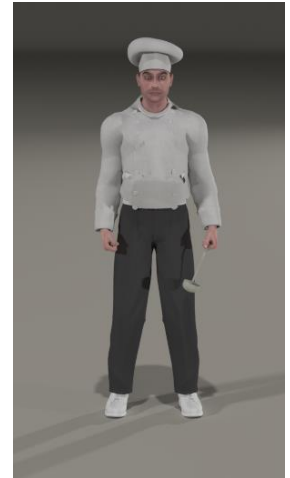
Der Leibwächter
The bodyguard



Der Boxer
The boxer



Der Clown
The clown



Der Koch
The cook



Der Cowboy
The Cowboy



Die Tänzerin
The dancer



Die Ärztin
The doctor



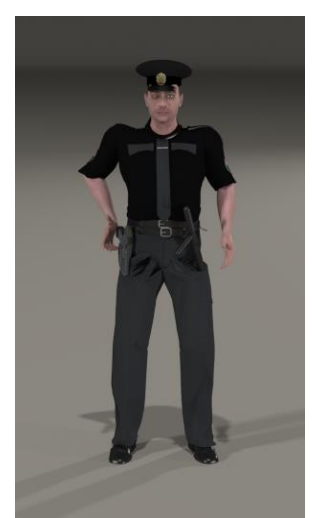
Der Ritter
The knight



Die Nonne
The nun



Der Pirat
The pirate



Der Polizist
The policeman



Der Matrose
The sailor



Der Lehrer
The teacher



Der Mönch
The monk



Die Schwimmerin
The swimmer



Die Hexe
The witch



Die Kellnerin
The waitress



Der Ball
The ball



Das Buch
The book



Die Tasse
The cup



Der Hut
The hat

Appendix 2a: Number of occurrences per character per experiment phase

Set	Character	nom exposure phase	nom GJT		acc exposure phase	acc GJT	acc Production	dat exposure phase	dat GJT	dat production
			Old	New						
A	Der Clown	4	1	2	6	2	1	7	2	
A	Der Koch	4	1	2	7	2	1	7	2	1
A	Der Lehrer	3	2	2	6	2		7	3	1
A	Die Kellnerin	4	2	2	6	3	1	11	3	1
A	Die Schwimmerin	4	2	2	6	1	1	8	3	1
B	Der Boxer	4	1	1		2			2	
B	Der Cowboy	4	2	2		2			3	
B	Der Leibwächter	4	1	1		2			2	
B	Die Hexe	3	2	2		1			3	
B	Die Tänzerin	3	1	1		2			3	
C	Der Mönch	4	1	1			1			1
C	Der Ritter	4	1	1			1			1
C	Die Nonne	4	2	1			1			1
C	Die Ärztin	4	2	2			1			1

Appendix 2b: Specification of character occurrences in the GJT

Set	Character	nom GJT correct		Nom GJT Incorrect	acc GJT correct	acc GJT incorrect	dat GJT correct	dat GJT incorrect
		new	old					
A	Der Clown	1	2	3	2	1	2	2
A	Der Koch	1	1	3	2	2	2	2
A	Der Lehrer	2	1	3	2	2	2	2
A	Die Kellnerin	2	1	3	2	2	3	3
A	Die Schwimmerin	2		3	1	2	3	3
B	Der Boxer	1	1	3	2	2	2	2
B	Der Cowboy	2	1	3	2	1	2	2
B	Der Leibwächter	1	2	3	2	2	2	2
B	Die Hexe	2	2	3	1	2	3	3
B	Die Tänzerin	1	1	3	2	2	3	3
C	Der Mönch	1	1	3				
C	Der Ritter	1	1	3				
C	Die Nonne	1	1	3				
C	Die Ärztin	2	2	3				

Appendix 3: Occurrences per rule per experiment phase

Rule	Exposure phase 79	GJT			Production task	
		Correct new 23	Correct old 24	Incorrect 47	New 9	Old 9
m.acc = den	19	6	7	12	2	2
m.dat= dem	21	7	7	14	2	2
v.dat = der	19	6	6	13	2	2
other	20	4	4	8	3	3

Appendix 4: All stimuli sentences

Used in experiment phase	Sentence
Exposure phase	Die Nonne springt
Exposure phase + GJT old	Der Cowboy springt
Exposure phase	Die Tänzerin rennt
Exposure phase	Der Koch rennt
Exposure phase	Die Ärztin schläft
Exposure phase	Der Ritter schläft
Exposure phase	Die Tänzerin winkt
Exposure phase	Der Pirat winkt
Exposure phase	Der Polizist küsst den Lehrer
Exposure phase	Der Leibwächter küsst die Schwimmerin
Exposure phase + GJT old	Der Matrose küsst die Kellnerin
Exposure phase	Die Kellnerin küsst den Lehrer
Exposure phase	Den Clown küsst der Cowboy
Exposure phase	Die Tänzerin küsst den Koch
Exposure phase	Der Pirat küsst den Koch
Exposure phase	Der Matrose küsst den Lehrer
Exposure phase + GJT old	Der Leibwächter schlägt den Koch
Exposure phase + GJT old	Die Ärztin schlägt die Kellnerin
Exposure phase	Der Matrose schlägt den Clown
Exposure phase + GJT old	Die Hexe schlägt die Schwimmerin
Exposure phase	Die Hexe schlägt die Kellnerin
Exposure phase	Der Pirat schlägt die Schwimmerin
Exposure phase + GJT old	Der Clown erschießt den Lehrer
Exposure phase	Der Matrose erschießt den Lehrer
Exposure phase	Die Ärztin erschießt den Koch
Exposure phase	Die Schwimmerin erschießt der Polizist
Exposure phase	Der Pirat erschießt die Kellnerin
Exposure phase	Der Lehrer erschießt den Clown
Exposure phase	Der Lehrer erschießt die Kellnerin
Exposure phase + GJT old	Den Koch erschießt die Kellnerin
Exposure phase + GJT old	Die Nonne erschießt den Koch
Exposure phase	Der Clown kitzelt den Koch
Exposure phase	Die Nonne kitzelt den Clown
Exposure phase + GJT old	Der Koch kitzelt den Clown
Exposure phase	Den Lehrer kitzelt der Mönch
Exposure phase	Der Polizist kitzelt die Schwimmerin
Exposure phase	Der Mönch kitzelt die Schwimmerin
Exposure phase + GJT old	Der Pirat kitzelt den Clown
Exposure phase	Die Kellnerin kitzelt der Polizist
Exposure phase + GJT old	Der Boxer liefert dem Lehrer das Buch
Exposure phase	Der Pirat liefert der Schwimmerin den Ball
Exposure phase	Den Hut liefert der Polizist der Kellnerin
Exposure phase + GJT old	Der Leibwächter liefert der Kellnerin das Buch
Exposure phase	Die Schwimmerin liefert der Kellnerin den Hut
Exposure phase	Die Tänzerin liefert der Schwimmerin die Tasse
Exposure phase	Der Mönch gibt den Ball dem Clown
Exposure phase + GJT old	Der Mönch gibt der Kellnerin den Hut
Exposure phase	Die Nonne gibt dem Clown den Ball
Exposure phase + GJT old	Der Matrose gibt der Schwimmerin die Tasse
Exposure phase	Die Kellnerin gibt der Schwimmerin den Hut
Exposure phase	Die Schwimmerin gibt den Hut der Kellnerin
Exposure phase	Der Clown gibt dem Koch den Ball
Exposure phase + GJT old	Der Clown gibt der Schwimmerin die Tasse

Exposure phase	Der Polizist gibt der Schwimmerin den Hut
Exposure phase	Die Schwimmerin verkauft der Kellnerin den Ball
Exposure phase	Der Koch verkauft der Kellnerin den Hut
Exposure phase	Der Matrose verkauft der Schwimmerin den Hut
Exposure phase	Der Pirat verkauft dem Lehrer das Buch
Exposure phase	Der Boxer verkauft dem Lehrer die Tasse
Exposure phase + GJT old	Die Ärztin verkauft dem Koch die Tasse
Exposure phase + GJT old	Den Ball verkauft der Polizist dem Lehrer
Exposure phase + GJT old	Die Hexe verkauft dem Clown die Tasse
Exposure phase	Der Polizist verkauft den Hut der Kellnerin
Exposure phase	Den Ball verkauft der Matrose dem Koch
Exposure phase + GJT old	Der Ritter verkauft dem Clown die Tasse
Exposure phase	Die Schwimmerin verkauft den Ball dem Lehrer
Exposure phase	Die Schwimmerin zeigt dem Koch das Buch
Exposure phase	Der Ritter zeigt den Ball dem Clown
Exposure phase + GJT old	Die Kellnerin zeigt der Schwimmerin das Buch
Exposure phase	Der Leibwächter zeigt dem Koch das Buch
Exposure phase	Der Koch zeigt dem Clown die Tasse
Exposure phase	Der Boxer zeigt dem Koch die Tasse
Exposure phase	Der Mönch zeigt dem Clown das Buch
Exposure phase + GJT old	Der Lehrer zeigt dem Koch die Tasse
Exposure phase	Der Matrose zeigt der Kellnerin den Ball
Exposure phase	Der Boxer zeigt dem Lehrer das Buch
Exposure phase + GJT old	Die Tänzerin zeigt dem Lehrer die Tasse
Exposure phase	Der Ritter zeigt der Kellnerin den Ball
Exposure phase + GJT old	Der Matrose zeigt der Kellnerin die Tasse
GJT new	Der Leibwächter winkt
GJT new	Die Tänzerin küsst den Boxer
GJT new	Der Boxer küsst den Cowboy
GJT new	Der Cowboy küsst die Tänzerin
GJT new	Die Kellnerin schlägt die Hexe
GJT new	Der Koch erschießt den Leibwächter
GJT new	Die Ärztin erschießt den Boxer
GJT new	Die Kellnerin erschießt die Tänzerin
GJT new	Der Matrose kitzelt den Leibwächter
GJT new	Die Hexe kitzelt den Cowboy
GJT new	Der Lehrer liefert dem Boxer das Buch
GJT new	Die Schwimmerin liefert der Tänzerin die Tasse
GJT new	Die Nonne liefert die Tasse dem Cowboy
GJT new	Der Mönch liefert der Hexe die Tasse
GJT new	Die Schwimmerin gibt der Hexe den Ball
GJT new	Die Hexe gibt dem Cowboy das Buch
GJT new	Der Clown gibt dem Cowboy die Tasse
GJT new	Der Lehrer verkauft der Tänzerin das Buch
GJT new	Die Nonne verkauft dem Leibwächter das Buch
GJT new	Der Ritter verkauft dem Boxer die Tasse
GJT new	Der Cowboy zeigt der Tänzerin das Buch
GJT new	Der Koch zeigt dem Leibwächter das Buch
GJT new	Der Pirat zeigt der Hexe die Tasse
GJT incorrect	Der Ritter schläft
GJT incorrect	Die Schwimmerin schläft
GJT incorrect	Der Koch kitzelt der Clown
GJT incorrect	Dem Lehrer kitzelt die Kellnerin
GJT incorrect	Die Kellnerin kitzelt dem Lehrer
GJT incorrect	Der Schwimmerin kitzelt der Koch
GJT incorrect	Die Hexe schlägt die Schwimmerin
GJT incorrect	Die Boxer schlägt die Schwimmerin
GJT incorrect	Die Hexe schlägt dem Boxer

GJT incorrect	Der Mönch schlägt die Hexe
GJT incorrect	Der Cowboy schlägt der Lehrer
GJT incorrect	Der Nonne erschießt die Tänzerin
GJT incorrect	Der Mönch erschießt der Leibwächter
GJT incorrect	Der Clown erschießt der Boxer
GJT incorrect	Der Koch erschießt der Cowboy
GJT incorrect	Die Cowboy küsst die Hexe
GJT incorrect	Der Leibwächter küsst der Koch
GJT incorrect	Der Nonne küsst die Tänzerin
GJT incorrect	Der Ärztin küsst der Leibwächter
GJT incorrect	Der Boxer zeigt die Schwimmerin die Tasse
GJT incorrect	Die Tänzerin zeigt die Clown die Tasse
GJT incorrect	Die Ärztin zeigt den Cowboy das Buch
GJT incorrect	Der Matrose zeigt die Hexe das Buch
GJT incorrect	Der Cowboy zeigt der Leibwächter die Tasse
GJT incorrect	Der Leibwächter zeigt die Koch die Tasse
GJT incorrect	Die Nonne liefert der Boxer das Buch
GJT incorrect	Der Clown liefert die Kellnerin die Tasse
GJT incorrect	Die Kellnerin liefert den Lehrer das Buch
GJT incorrect	Der Boxer liefert die Tänzerin das Buch
GJT incorrect	Die Hexe liefert den Clown die Tasse
GJT incorrect	Die Kellnerin liefert die Cowboy das Buch
GJT incorrect	Der Matrose verkauft den Lehrer die Tasse
GJT incorrect	Der Clown verkauft die Kellnerin das Buch
GJT incorrect	Der Pirat verkauft die Schwimmerin das Buch
GJT incorrect	Die Schwimmerin verkauft die Koch die Tasse
GJT incorrect	Der Koch verkauft der Boxer das Buch
GJT incorrect	Der Leibwächter verkauft die Tänzerin das Buch
GJT incorrect	Die Tänzerin gibt die Hexe die Tasse
GJT incorrect	Der Mönch gibt dem Leibwächter die Tasse
GJT incorrect	Der Ritter gibt der Kellnerin den Hut
GJT incorrect	Die Ärztin gibt die Schwimmerin das Buch
GJT incorrect	Der Lehrer gibt die Hexe den Ball
GJT incorrect	Der Cowboy gibt die Tänzerin die Tasse
GJT incorrect	Die Cowboy kitzelt die Kellnerin
GJT incorrect	Der Ritter verkäuft der Nonne die Tasse
GJT incorrect	Die Kellnerin verkäuft dem Boxer das Buch
GJT incorrect	Die Hexe verkäuft dem Cowboy das Buch
Production task new	Die Schwimmerin springt
Production task new	Der Leibwächter küsst den Mönch
Production task new	Der Boxer erschießt die Nonne
Production task new	Der Clown kitzelt den Ritter
Production task new	Der Mönch kitzelt die Ärztin
Production task new	Der Cowboy liefert der Nonne die Tasse
Production task new	Die Nonne gibt dem Mönch den Ball
Production task new	Der Clown verkauft dem Ritter die Tasse
Production task new	Der Ritter zeigt der Ärztin das Buch
Production task old	Die Ärztin schläft
Production task old	Der Leibwächter schlägt den Koch
Production task old	Die Hexe schlägt die Schwimmerin
Production task old	Der Pirat erschießt die Kellnerin
Production task old	Die Nonne kitzelt den Clown
Production task old	Der Boxer liefert dem Lehrer das Buch
Production task old	Der Matrose gibt der Schwimmerin die Tasse
Production task old	Der Leibwächter zeigt dem Koch das Buch
Production task old	Der Matrose zeigt der Kellnerin die Tasse

Appendix 5: Examples of task layouts

Luister naar de audio en bekijk beide video's. Welke video geeft de inhoud van de zin het best weer?



video 1



video 2



video 1

video 2

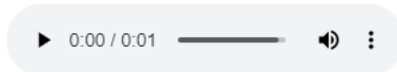
Figure 11: screenshot of an implicit training phase question in the Qualtrics survey.

Bekijk de video. Typ in één zin in het Duits wat er in de video gebeurt.



Figure 12: screenshot of a production task question in the Qualtrics survey.

Is de volgende zin grammaticaal correct?



ja

nee

Figure 13: screenshot of a GJT question in the Qualtrics survey.