

**Learning words with or without orthography:  
The influence of incongruent orthographic information and rule-  
based inconsistencies on second language word learning.**

Sanne A.W.M. de Kroon

S4587677

28th of February 2021

Master's thesis

MA Linguistics: Nederlands als Tweede Taal: Docent en Expert

Primary supervisor: J.P.M. Fikkert

Secondary supervisor: M.E. Broersma

## Abstract

In this study, we investigated whether there is an effect of offering orthographic forms during second language auditory word learning. Previous research showed that there can be a positive effect of orthography on word learning when grapheme-to-phoneme mappings are the same in the new language as in the native language (*congruent* mappings). However, when mappings are not the same (*incongruent* mappings), offering orthographic forms may hinder word learning. In that case, offering auditory forms only leads to better performance compared to offering both auditory and orthographic forms.

In addition, the present study further considers the role of *inconsistent* mappings, i.e. mappings where the grapheme-to-phoneme mapping is not fixed, but varies within a language. Only mappings that vary based on phonological rules (*rule-based inconsistencies*) were considered here. As inconsistent mappings lead to incongruencies, these type of mappings are expected to hinder word learning as well.

Two groups of Dutch participants that had no experience with the Danish language, were taught Danish words on an online platform. They were trained using auditory word-picture pairs either with or without orthographic forms. The group that received orthographic forms, encountered items in two conditions: a 'baseline' condition in which mappings for consonants were both congruent with Dutch and consistent within Danish and a 'rule-based' condition including rule-based inconsistencies in the Danish orthography based on the rule of consonant gradation. The inconsistent mappings varied in such a way that they were sometimes congruent with Dutch and sometimes incongruent. To test whether participants had learned the words, they were asked to judge matching and mismatching auditory word-picture pairs.

The present study aimed to answer several questions. The first was whether participants receiving auditory forms only would generally outperform those that received orthographic forms additionally to the auditory forms (RQ1), which would replicate prior findings. Moreover, it was expected that the effect of orthography would be different for the consistent baseline compared to the inconsistent rule-based condition. In other words, an interaction between group (with vs. without orthography) and item type (baseline vs. rule-based) was expected. Orthography should hinder more in the rule-based condition because this condition included more incongruencies, as a result of rule-based inconsistencies (RQ2). The factor sleep was investigated as well. This was the third research question, concerning whether participants would perform better after a night of sleep, as sleep was predicted to consolidate the learning effect (RQ3).

None of the hypotheses was met. There was no main effect of group, nor an interaction between group and item type. There was an effect of sleep, but this was not in the expected direction. The most striking effect was the large ceiling effect in the data. Possible reasons for this effect are discussed.

## Table of contents

|  |    |
|--|----|
| 1. Introduction .....  | 4  |
| 2. Background .....  | 6  |
| 2.1 Interactions between phonology and orthography .....             | 6  |
| 2.2 Orthographic depth .....   | 7  |
| 2.3 Orthographic effects in word learning in a second language ..... | 8  |
| 2.4 Grapheme-to-phoneme correspondence (GPC) rules .....             | 12 |
| 2.4.1 GPC rules and second language word learning .....              | 12 |
| 2.4.2 Danish GPC rules .....   | 13 |
| 2.5 The role of sleep in word learning .....                         | 14 |
| 3. Research questions and hypotheses .....                           | 15 |
| 3.1 Research questions .....   | 15 |
| 3.2 Hypotheses .....   | 15 |
| 4. Method .....  | 15 |
| 4.1 Participants .....   | 16 |
| 4.2 Material .....   | 16 |
| 4.2.1 Words .....  | 17 |
| 4.2.2 Audio files .....  | 18 |
| 4.2.3 Pictures .....   | 19 |
| 4.3 Procedure .....  | 19 |
| 4.3.1 Selection questions .....                                      | 20 |
| 4.3.2 Instructions .....   | 20 |
| 4.3.3 Word learning phase .....                                      | 21 |
| 4.3.4 Testing phase .....  | 21 |
| 4.3.5 Language background questionnaire .....                        | 21 |
| 4.3.6 Manipulation check .....                                       | 21 |
| 4.4 Design and analyses .....  | 22 |
| 5. Results .....   | 22 |
| 6. Discussion .....  | 23 |
| 7. Conclusion .....  | 33 |
| 8. References .....  | 35 |
| 9. Appendices .....  | 37 |
| A: Full material .....   | 37 |
| B: Full set of pictures .....  | 39 |
| C: Naming phase (pre-test) .....                                     | 43 |
| D: Instructions .....  | 44 |
| E: Mismatching word-picture pairs .....                              | 58 |
| F: All estimated percentages correct .....                           | 59 |

## 1. Introduction

All second language learners have at one point sighed in defeat after learning the pronunciation of a particular word. For one thing, the pronunciation could be unexpected from the perspective of their first language. A Dutch speaker beginning to learn Danish may be surprised to find that the grapheme 'g' is pronounced [g] instead of [ɣ] or [x]<sup>1</sup>. This difference between grapheme-to-phoneme mappings in two languages is called *incongruency*. After learning Danish for a while, this Dutch speaker will moreover find that 'g' can also be pronounced as [j] or [w] depending on the position in the word and the phoneme that precedes it. Thus, mappings may also vary within a language, which is called *inconsistency*. In this particular example, the inconsistent mappings are rule-based, because 'g' always maps onto [g] in initial position and in final position it always maps onto [j] following front vowels or /l/ and onto [w] following back vowels or /r/. The rule in question is called consonant gradation (Basbøll, 2005). To an initial learner, however, it might seem as if there is no system to these mappings. Learning a language with many inconsistencies, rule-based or not, may feel daunting. You might wonder if it would be easier if you did not have to deal with orthography at all. Whereas if you are learning a language with consistent mappings, preferably one that has mappings that are congruent with your first language, then you might find orthography helpful.

There has been a lot of research into the role of orthography in auditory word learning in a second language. The aim of these studies has been to find out to what extent orthography can aid the learning of phonological word forms, while considering congruency as a factor. Results show that orthographic forms can help with the learning of phonological contrasts that are orthographically represented (Escudero, Hayes-Harb & Mitterer, 2008). The congruency of the orthographic information was found to play a role though. While congruent orthographic information can have a beneficial effect on word learning (Escudero, Simon & Mulak, 2014; Escudero, 2015) or have no effect (Hayes-Harb, Nicol & Barker, 2010), incongruent orthographic forms might hinder word learning compared to a situation in which only auditory forms are presented (Escudero et al., 2014; Hayes-Harb et al., 2010).

None of these studies manipulated consistency as a factor. Research into first language reading acquisition found that children in orthographically deep languages, which are languages with many inconsistent mappings, were slower to acquire reading skills. They were thus affected by the inconsistencies in the mappings in their language (Seymour, Aro & Erskine, 2003). As inconsistent mappings affect first language users, it is likely that second language learners are affected by inconsistencies within the language they are learning as well. If not, then they should at least be affected by the incongruencies that might result from certain inconsistencies.

The present study will start to investigate the effect of inconsistency by including *rule-based inconsistencies* in mappings in the material. These are inconsistencies that arise as the result of a phonological rule. Research into second language productive word learning has shown that learners are affected by rule-based inconsistencies in mappings (Hayes-Harb, Brown & Smith, 2018). This might be the case for receptive language learning as well. In the current study, Dutch learners that are inexperienced with the Danish language, will learn Danish words that undergo consonant gradation. In short, this phonological rule causes the pronunciation of some consonants to change in final position compared to the initial position. The Dutch language also has a phonological rule that affects certain consonants in final position, namely final devoicing. This rule causes voiced obstruents to become voiceless

---

<sup>1</sup> When speakers from the South are concerned, the expected mappings would be onto [ç] or [j].

in word final or syllable-final position (Grijzenhout, 2000). As Dutch language users are used to taking word position into account, they should be able to learn the rule of consonant gradation. Initially, the rule-based inconsistencies that arise from consonant gradation might hinder their word learning though.

To research the influence of consonant gradation, the material will consist of two conditions. The first condition is the 'baseline' in which the mappings for all consonants are both consistent and congruent. In the second condition, the rule-based condition, mappings for consonants are subject to consonant gradation. Mappings for vowels include (rule-based) inconsistencies in both conditions. To allow for the testing of orthographic effects, this study will use two groups of participants. One group will receive auditory forms only (AF only), while the second group will be trained on auditory and orthographic forms (AF & OF).

The present study aimed to answer the following questions. Firstly, will participants in the AF only group perform better than those in the AF & OF group? Secondly, will participants perform worse on the rule-based words compared to the baseline words? Based on previous research, the expectation is that participants in the AF only group would indeed perform better. Although this may be the case for all items (as mappings are never completely congruent with Dutch orthography), we expect that the inconsistent mappings in the rule-based items particularly hinder learners in the AF & OF group. As a result, we expect the AF & OF group to perform better on baseline items compared to rule-based items. The rule-based inconsistencies are expected to hinder word learning.

The third question regards the role of sleep in word learning. Previous research revealed a role for sleep in lexicalisation: sleep was found to be necessary for words to become lexical competitors (Dumay & Gaskell, 2007). Moreover, words are differently represented in the brain after a night of sleep (Davis, Di Betta, Macdonald & Gaskell, 2009). This leads to better performance when it comes to word recognition after a night of sleep, compared to word recognition within one session. In the present study, a word-picture matching task was used to test word learning. The study included multiple testing days, in order to be able to consider sleep as a factor. The expectation was that participants would perform better on the second session compared to the first one, because of the night of sleep in between those sessions.

The present study is theoretically relevant because it takes a first step towards investigating more complex effects of orthography on word learning in a second language, by taking rule-based inconsistencies into account. Moreover, real language material was used instead of pseudowords, so it will provide an opportunity to replicate orthographic effects found in previous studies with an approach that is closer to natural language learning. Additionally, the factor sleep will be considered, which will help to answer the methodological question of whether multiple testing days should be included in a word learning experiment. The field of research considering orthographic effects in language learning, should eventually lead to an advice for the language teaching practice.

In the following sections, the present study will be described in detail. First a theoretical background will be provided in section 2, beginning with interactions between phonology and orthography and leading all the way up to effects of orthographic congruency and consistency in second language word learning. In section 3, the research questions and hypothesis will be listed. Section 4 will serve to describe the methodology used in the present study, including a description of the participants, material, procedure and design. Section 5 will contain the results and in section 6, these results will be discussed. In section

7, the conclusion of the present study will be presented, along with recommendations for future research.

## **2. Background**

This section will provide a theoretical background for the present study, that investigates orthographic effects in word learning in a second language. Section 2.1 will describe interactions between phonology and orthography and how language users are affected by inconsistencies in how phonemes map onto graphemes and vice versa. Section 2.2 will address inconsistencies at a language level and two ways to determine the orthographic depth of a language will be described. In section 2.3, previous research will be discussed that focussed on orthographic effects on second language word learning. Specifically, effects of the incongruency of grapheme-to-phoneme-mappings between the first and second language are central to this section. Section 2.4 will go into grapheme-to-phoneme correspondence (GPC) rules. Section 2.4.1 will introduce the topic of GPC rules in second language word learning and in section 2.4.2 the Danish GPC rules that are relevant for the present study will be discussed. Finally, in section 2.5 the role of sleep in word learning will be considered.

### **2.1 Interactions between phonology and orthography**

Phonology and orthography are not two completely separate sources of information for word recognition, but they are highly interactive. Frost and Ziegler (2007) wrote a chapter about the interdependence of visual and auditory word recognition. They show that phonology is activated during word reading. They cite, among others, a study done by Ziegler and Jacobs (1995). In this study, participants were asked to identify a target letter in a masked letter string that was only briefly presented. When the target letter was 'i', the string 'brane' led to more false alarms than the string 'brate'. The researchers conclude that this is the case because the phonological form of 'brane' activates the lexical entry 'brain' and its orthographical form, which does include the letter 'i'. There is moreover evidence that orthography is activated during spoken word recognition. For example, Seidenberg and Tanenhaus (1979) found that literate adults found it harder to determine whether two spoken words rhymed when the rimes were spelled differently compared to when they were spelled the same, even though orthographic forms were not provided to them.

Furthermore, these interactions between phonology and orthography lead to consistency effects in word recognition. For example, Glushko (1979) showed that it takes longer to name inconsistent words than it does to name consistent words with comparable word frequency. Often, only the consistency of grapheme-to-phoneme mappings is considered. This is called feedforward consistency. The current state of research assumes a bidirectional flow of information, which leads to the assumption that feedback consistency should matter too (Frost & Ziegler, 2007). This refers to consistency in phoneme-to-grapheme mappings. Stone, Vanhoy and Van Orden (1997) found both effects of feedforward and feedback inconsistency in a lexical decision task. Both words with inconsistent grapheme-to-phoneme mappings ('pint' [paɪnt] versus 'mint' [mɪnt]) and words with inconsistent phoneme-to-grapheme mappings ('heap' [hi:p] versus 'deep' [di:p]) lead to slower 'yes' responses than fully consistent words ('duck' [dʌk] versus 'luck' [lʌk]). Language users thus seem to be influenced by how consistently graphemes map onto phonemes and vice versa.

## 2.2 Orthographic depth

Languages vary based on how consistently the graphemes correspond to the phonemes in that language. This is called orthographic transparency or orthographic depth and it applies to both directions: from grapheme to phoneme and from phoneme to grapheme (Borleffs, Maassen, Lyytinen, & Zwarts, 2017). Languages can be placed on a continuum ranging from consistent or shallow languages like Finnish or German, to deeper languages such as Danish and English (Seymour et al., 2003). Languages are not always as consistent in the feedforward as the feedback direction. When comparing French and English for example, one finds that these languages are of similar consistency in the feedback direction: 79.1% of French words are inconsistent (Ziegler, Jacobs & Stone, 1996) compared to 72.3% of English words (Ziegler, Stone & Jacobs, 1997). The feedforward direction paints a different picture though, with 12.4% of French words being inconsistent compared to 30.7% of English words.

Seymour et al. (2003) found that children learning to read in orthographically deep languages were slower in acquiring reading skills compared to children that learned to read in shallow languages. Especially English children were weighted down by the inconsistency of their native language, but also children with a native language in which most inconsistencies are in the feedback direction, such as French, performed worse than the children with consistent native languages. Most children showed a near ceiling effect of performance on both word reading and non-word reading, while English (34%), Danish (71%), Portuguese (73%) and French (79%) children had lower or, in the case of English, much lower accuracy scores.

In order to establish orthographic depth, different approaches can be taken. Two will be explained here: the *consistency approach* and the *regularity approach*. The consistency approach is perhaps the most straightforward approach, if you look at consistency as a dichotomous notion. A grapheme or phoneme can be considered consistent when there is only one mapping possible and inconsistent when multiple mappings are possible (Borleffs et al., 2017). Larger units, like words, could be considered as well. For example, Ziegler et al. (1996) and Ziegler et al. (1997) calculated the percentage of consistent words for French and English (monosyllabic) words respectively, with the help of a database.

A graded analysis of consistency is possible as well. In this case the relative frequency of each possible mapping is taken into consideration. Consistency can then be calculated as the proportion of words in which a grapheme is pronounced a certain way, divided by the total number of words that include that grapheme regardless of how it is spelled. This leads to a feedforward consistency ratio ranging from 0 (fully inconsistent) to 1 (fully consistent) (Lété, Peereman & Fayol, 2008). The feedback consistency can be calculated in a comparable way, when the perspective of the phoneme is considered.

The regularity approach is unique in the sense that its main focus is on grapheme-to-phoneme mapping rules and the violation of these. In order to assess the orthographic depth of a language, pronunciation rules or grapheme-to-phoneme correspondence rules (GPC rules) are formulated. With these rules, the regularity of a language can be calculated as the proportion of words for which the pronunciation according to the GPC rules matches the lexical pronunciation (Ziegler, Perry & Coltheart, 2000). An orthographically deep language is then a language that includes many GPC rule violations.

There are three types of GPC rules: single-letter, multi-letter and context-sensitive rules. The position, being the beginning, middle or end of the string, is specified for each rule. Next to rules for specific graphemes, general rules or 'super rules' can be formed. An example of a rule like this, is final devoicing. In some languages, voiced obstruents become

voiceless when they occur in syllable-final position (Grijzenhout, 2000). This is the case for Dutch but also for German, which is the language for which Ziegler et al. (2000) formed GPC rules. So instead of forming separate context-sensitive rules for each obstruent, one super rule could describe final devoicing. This way, fewer rules need to be formed and you do not end up overestimating the complexity of a language by overlooking systematicities.

Keep in mind that the regularity approach is used for researching visual word recognition. It only looks at the feedforward direction, from grapheme-to-phoneme. What is not taken into account here, is that some GPC rules might result in difficulties in the feedback direction. If we again take a look at the super rule final devoicing, then this can be illustrated. In Dutch, the phonological form [rat] could be spelled both like 'rat' (meaning 'rat') or 'rad' (meaning 'wheel') depending on which word is intended. So as a result of final devoicing, the phoneme [t] could map onto both 't' and 'd'. The question is whether it is justified that GPC rules do not contribute to the orthographic depth of a language in the regularity approach, when it is clear that these rules can lead to feedback inconsistencies. In the consistency approach, GPC rules do add to the orthographic depth of a language to some extent because feedback inconsistencies are included here.

In the present study, the consistency approach and the regularity approach will be combined. Even though the present study does focus on grapheme-to-phoneme mappings only, the consistency approach will be used as a starting point because its bidirectional approach is preferred to the feedforward focus of the regularity approach. The notion of GPC rules will be incorporated by using the term 'rule-based inconsistencies' to set grapheme-to-phoneme mappings that vary due to GPC rules apart from *random inconsistencies*. Random inconsistencies are inconsistencies in grapheme-to-phoneme mappings that cannot be explained by rules. These can occur for example in loanswords, like in the Danish word 'giro' [ɕi:ʁo] which includes a mapping for 'g' that is different from the regular mapping of 'g' onto [g] in initial position. It is likely that rule-based inconsistencies do not hinder word recognition as much as random inconsistencies, but based on the research into feedback inconsistencies it seems unjustified to assume that language users are not affected by them at all. Especially second language users, that might not yet be familiar with the GPC rules of the language that they are learning, are bound to be affected by rule-based inconsistencies. These effects might decrease or disappear as language users become more competent. The present study will begin to investigate what effect GPC rules have on beginning language learners. Do these rules hinder word learning? Can learners derive these rules without explicit instruction?

### **2.3 Orthographic effects in word learning in a second language**

The consistency of grapheme-to-phoneme mappings is important for written word recognition in a first language, as well as in a second language. Where a second language is concerned however, there is another factor that should be considered, which is congruency. This term is used to compare the grapheme-to-phoneme mappings between a first language and a second language or pseudowords used in a study. When these mappings are the same, they are considered 'congruent' and when they are different they are 'incongruent'. Table 1a shows how congruency is different from consistency. In table 1b, two different types of inconsistencies are explained: rule-based and random inconsistencies. Even though random inconsistencies do not play a role in the present study, it is important to be aware of the difference between rule-based and random inconsistencies.

**Table 1a***Overview of terminology concerning grapheme-to-phoneme mappings*

| <b>Mapping type</b> | <b>Description</b>                                 |
|---------------------|--|
| Congruent           | Mappings in the L1 match the ones in the L2        |
| Incongruent         | Mappings in the L1 do not match the ones in the L2 |
| Consistent          | Mappings are always the same within a language     |
| Inconsistent        | Mappings differ within a language                  |

**Table 1b***Overview of the types of inconsistent grapheme-to-phoneme mappings*

| <b>Type of inconsistency</b> | <b>Description</b>   |
|------------------------------|--|
| Rule-based inconsistencies   | Mappings within a language vary, but this is in accordance with GPC rules                    |
| Random inconsistencies       | Mappings within a language vary in a way that is random and cannot be explained by GPC rules |

Incongruency can come in different forms as well. Sometimes a grapheme maps onto one phoneme in one language and onto a different one (or two) in another language. Take for example ‘j’ which maps onto [j] in Dutch and [dʒ] in English. The situation becomes more complex when a grapheme has multiple mappings in one or both languages, as a result of different GPC rules. In Dutch for example, the obstruent ‘d’ can map onto both [d] and [t] as a result of final devoicing, while in English this rule does not exist and ‘d’ only maps onto [d]. Besides rule-based inconsistencies, random inconsistencies also cause congruency to be something that is not necessarily constant. Especially learners of orthographically deep languages and languages with many GPC rules may encounter many different forms of incongruency.

Something else that can be a factor in how much difficulties incongruency can cause to language learners, is perceptual ease. One language could have a different phoneme inventory than the other. For example, English has the phonemes [æ] and [ɛ], while in Dutch only [ɛ] exists. Besides having to learn that [æ] maps onto ‘a’ and [ɛ] onto ‘e’, Dutch listeners will have to learn to perceive the contrast between these vowels. This contrast is notoriously hard for them to master (Escudero et al., 2008). While the present study is not about phonological contrasts, much of the research into orthographic effects on word learning in a second language does include perceptual difficulty as a factor. Some of these studies will be discussed below.

Escudero et al. (2008) found that the orthographic word form helped their participants to acquire a new phonological contrast that is orthographically represented. They taught English pseudowords with [æ] versus [ɛ] to Dutch learners of English. The pseudowords were connected to ‘pseudomeanings’. Participants saw line drawings of nonobjects that were supposed to depict the words they learned. The researchers compared a group of participants that received auditory word forms only, with a group that received auditory and orthographic forms. The phonological contrast became visible in the orthographic forms: [æ] always mapped onto the grapheme ‘a’ and [ɛ] always mapped onto ‘e’. Participants had to click on the picture matching the word they heard. Both in the

learning phase as well as in the testing phase. During the testing phase, their eye-movements were tracked. Only the group that was exposed to the orthographic forms as well, showed the asymmetric word recognition that is common for Dutch learners of English. This means that when the target word contained [ɛ], these participants looked only at words with [ɛ]. When the target contained [æ] they looked at both words containing [æ] as well as [ɛ]. The group that was exposed to auditory forms only, always looked at words containing [ɛ] as well as [æ], regardless of the target. The researchers conclude that the group that received orthographic forms, (partially) acquired the phonological contrast while the other group did not. Orthography could thus help learners with acquiring phonological contrasts that do not appear in their native language.

Orthographic forms could also have a negative effect on word learning. Escudero et al. (2014) taught Dutch pseudowords to Spanish participants that were learners of Dutch or had no experience with the language. The participants were divided over two groups: one group received auditory forms only and the other group was trained on orthographic forms as well. During the learning phase, participants heard a word and saw a picture of a nonobject. Afterwards they saw two pictures and they were asked to click on the picture matching the word they had just learned. Participants in the group with orthographic forms, also saw the written word underneath it. During testing, participants were presented with two pictures and they had to click on the picture matching the word they heard. The other picture described another word they had learned that either formed a minimal pair or a non-minimal pair with the word they heard. Some of the minimal pairs were perceptually easy, while others were perceptually difficult. The difficult pairs were further divided into pairs with mappings that were congruent between Spanish and Dutch and mappings that were incongruent. In the testing phase, none of the groups was presented with the written word form. For the minimal pairs, Escudero et al. (2014) did find a positive effect of orthography when it came to words that contain grapheme-to-phoneme mappings that are congruent between Dutch and Spanish. However, in the case of incongruent grapheme-to-phoneme mappings, there was actually a negative effect of orthography. In this case, the group that received only auditory word forms actually outperformed the group that received orthographic forms as well. The learners of Dutch performed better at the word recognition test than the naïve listeners, but were not differently affected by orthographic effects. There was no interaction between Dutch exposure and condition.

In a follow up experiment, Escudero (2015) compared these results to results obtained from another set of participants with a different language background, namely native speakers of Australian English that were either monolingual or bilingual. The aim of the study was to find out whether orthographic effects would differ based on language background. No such effect was found though. Moreover, there was no difference between participants that received only auditory forms compared to those that received both auditory and orthographic forms, considering nonminimal pairs and perceptually easy pairs. There was only a difference with the two perceptually difficult pairs with the highest accuracy. For these pairs, there was a positive effect of orthography. Since these pairs were thus still relatively easy, Escudero (2015) argues against the use of orthographic information. She finds the benefits negligible compared to the negative effects that were found in other studies, like the one done by Hayes-Harb et al. (2010) that will be discussed next.

Hayes-Harb et al. (2010) took a slightly different approach to researching orthographic effects in word learning in a second language. They did not include perceptual difficulty as a factor. Central to their study was the following question: “When a written form

is available to learners, are mismatches between the grapheme-phoneme correspondences of the native and second languages reflected in learners' memory of the phonological forms of words?" (Hayes-Harb et al., 2010, p. 371).

The methodology of their study was not much different from the studies described above. Hayes-Harb et al. (2010) also taught their participants pseudowords, but they did not specify which 'language' the participants would be learning. The pseudowords were connected with line drawings of real objects, like an apple or a broom. What is particular for this study, is that congruency was manipulated between subjects as well as within. There were three groups of participants: the first group received auditory word forms only; the second group got auditory and orthographic forms, that were congruent with the English mappings; and the third group received auditory forms and orthographic forms that were sometimes congruent and sometimes incongruent with the English language. The incongruent group received three different types of items: 1) items that were congruent and matched items in the congruent group; 2) items that contained a 'silent' letter, for example 'nd' mapping onto [d]; and 3) items that contained a mismatch between a grapheme and a phoneme, for example 'b' mapping onto [v].

The experiment had three phases: a word learning phase, a criterion test and the testing phase. In the word learning phase, participants heard a word and saw the accompanying picture for four seconds either with or without orthographic form. After seeing all 24 word-picture pairs four times, in a random order, they did a criterion test. They were presented with a picture and an auditory word form and they had to decide if the pair matched. They encountered all pictures and all words twice: once in a matching pair and once in a pair that did not match. If participants scored below 90% correct, they went through the word learning phase again. Once they scored above 90% they continued to the testing phase. This phase was comparable to the criterion test, except this time participants in all three groups were also presented with similar-sounding auditory forms and they had to also judge these as incorrect. The newly constructed auditory forms reflected the incongruent orthographic forms that the incongruent group had been exposed to.

The results show several things. First of all, the incongruent group performed better on those items that were congruent with English spelling compared to those that were not. They scored similar on these items, to the congruent group. The newly constructed auditory forms in the testing phase, that ought to be judged as incorrect by all groups, were more often judged correct by the incongruent group than the congruent group. The researchers concluded that the incongruent orthographical forms that the incongruent group received in the learning phase, had apparently influenced their memory of the auditory forms. With respect to the targets in the set of incongruent forms, results showed that the 'silent' letters did not have a detrimental effect on participants' performance in the testing phase, whereas the mismatching letters did. Another result is that there seemed to be no benefit of the presentation of (congruent) orthographical forms. The congruent group performed similar to the auditory form only group. The incongruent group was not compared to the other groups for this analysis.

To sum up, what is relevant for the present study is that both positive and negative effects of orthographic forms have been found in previous research. A positive effect is, for example, that orthographic forms can help with the learning of phonological contrasts (Escudero et al., 2008). Specifically, congruent orthographic information was found to have a positive effect on word learning (Escudero et al., 2014; Escudero, 2015) or could have no

effect (Hayes-Harb et al, 2010), while incongruent orthographic information had a negative effect on word learning (Hayes-Harb et al., 2010; Escudero et al., 2014).

## **2.4 Grapheme-to-phoneme correspondence (GPC) rules**

### **2.4.1 GPC rules and second language word learning**

The present study will further investigate the effects of incongruency in mappings between the L1 and the L2. So far congruency has mainly been investigated in relation to phonological contrasts and the ease with which they can be perceived. In these studies, mappings were always consistent so that orthography could help learning certain contrasts. In the present study, a first step will be taken into investigating the effect of inconsistent mappings on second language learners, by looking at the effect of inconsistencies that are rule-based and can thus be captured in GPC rules. There is one super rule that will be focussed on, being consonant gradation. This rule will be explained in the next section.

To our knowledge, no studies as of yet have looked at the effect of GPC rules on receptive word learning in a second language. However, there has been research focussing on rule learning in productive word learning. Hayes-Harb et al. (2018) taught German pseudowords to native English participants. The material included pictures describing words that contained rule-based inconsistencies as a result of final devoicing. So the super rule that participants had to learn, was that final voiced obstruents become devoiced. A specific example is the grapheme 'd' mapping onto [t] in final position. Half of the participants were trained on the auditory word form only, while the other half received the orthographic form as well. When afterwards they were asked to name the pictures, the participants that had received orthographic forms as well as auditory forms were more likely to produce final voiced obstruents where voiceless ones were required.

In a second experiment, Hayes-Harb et al. (2018) tried to minimise the disruptive effect of the rule-based inconsistencies by explaining German final devoicing to their participants. This did not lead to more accurate pronunciation. To conclude: a GPC rule like final devoicing is difficult for learners to master in a short time frame. Even explicit knowledge does not automatically lead to correct application of the rule. Participants in the second experiment were still subject to orthographic interference, meaning that those who received orthographic forms were still more likely to produce final voiced obstruents where voiceless ones were required. Some questions for the present study are whether participants become aware of the rule consonant gradation without explicit instruction and whether the presence of this rule will hinder their word learning.

Most studies have used pseudowords for obvious reasons (e.g. experimental control), but ultimately we would like to consider the full complexity of learning words in a second language. Hence, real language material will be used in the present study: participants will be taught Danish words. Dutch learners of Danish will have to deal with incongruencies in grapheme-to-phoneme mappings that are possibly inconsistent as well. All inconsistencies within the material will be rule-based though. Participants will have to deal with the GPC rule consonant gradation, which is central to the study. There are more GPC rules that occur throughout the material, which apply to vowels. These were not used to create different conditions though. In the section below, all rules that influence the material will be discussed.

### 2.4.2 Danish GPC rules

Learners of Danish will face several difficulties. For one, listeners will have to deal with the radical reduction processes that Danish is known for. Processes that are relevant for the present study, are consonant gradation and schwa-assimilation (Basbøll, 2005). Along with these processes, also GPC rules concerning vowels will briefly be discussed.

The rule of consonant gradation causes some obstruents to change into non-lateral approximants when they occur in word final position or before a schwa (Bleses, Basbøll & Werner, 2011). Using the terminology from the regularity approach, this would be a super rule. This super rule overarches context-sensitive rules for individual graphemes. In this experiment, only consonant gradation concerning the graphemes ‘d’ and ‘g’ will be in focus. In initial position, the grapheme ‘d’ is realised as [d] and in final position as [ð]. For example, the word ‘dyne’ meaning ‘duvet’ is pronounced [dy:nə], while the word ‘gade’ meaning ‘street’ is pronounced [gæ:ðə]. The GPC rules for the grapheme ‘g’ would be that it is pronounced [g] in initial position, [j] in final position following front vowels or /l/ and [w] following back vowels or /r/ (Basbøll, 2005). Using the same example as before, the word ‘gade’ [gæ:ðə] meaning ‘street’, illustrates the pronunciation of ‘g’ in initial position. The word ‘løg’ [lɔjʔ] meaning ‘onion’ is an example of the pronunciation of ‘g’ following a front vowel and the word ‘tog’ [tɔʔw] meaning ‘train’ demonstrates how ‘g’ is pronounced following a back vowel. As the examples illustrate, an extra rule is needed in order to pronounce ‘g’ in final position correctly because the preceding phoneme has to be considered as well. Since the Dutch language also knows a phonological rule that affects consonants in final position, being final devoicing, the Dutch participants in this study might be able to derive the rule of consonant gradation from the material.

Schwa-assimilation refers to schwa not being realised, which can happen when it occurs before or after a sonorant. This context makes assimilation possible, because the following or preceding sonorant could take over the syllabicity of the schwa that is deleted. For example, the Danish word for bicycle, ‘cykel’ [sygəl], can be pronounced as [syg]. In this case, the sonorant [l] is syllabified. As the example illustrates, the number of syllables usually does not change due to this rule. The term schwa-deletion or schwa-drop could also be used to describe this process, because evidently the vowel segment schwa is no longer present after this GPC rule has been applied. However, because the syllable node is not deleted, the term ‘schwa assimilation’ was chosen here.

When schwa occurs before a sonorant, assimilation is obligatory except in highly distinct speech. When schwa occurs after a consonantal sonorant, assimilation is favoured and after a vocoid it is strongly favoured (Basbøll, 2005). Next to sonorants, the approximant [ð] can trigger schwa assimilation as well. Both a schwa preceding as well as one following the approximant [ð] can be assimilated. Schwa assimilation is not central to this study, but it is connected to consonant gradation<sup>2</sup>. If we again look at the word ‘gade’ meaning ‘street’ then it should be stressed that the pronunciation [gæ:ðə] belongs to very distinct speech only (Basbøll, 2005). Due to schwa assimilation, this word will usually be pronounced as [gæ:ð]. Note that the [ð] is now syllabified. In case of the [g], schwa assimilation can also be triggered as a result of consonant gradation since the [g] becomes [j] or [w] in final position. Both of these are sonorants. Take for example the word ‘læge’ [lɛ:jə] meaning ‘doctor’ which is in the material realised as [lɛ:j].

---

<sup>2</sup> The material in the present study also included items that allowed for schwa assimilation that did not originate as a result of consonant gradation. However, the speaker that recorded the material for this study somehow only assimilated schwa in words that did include consonant gradation.

In Danish, grapheme-to-phoneme mappings for vowels include many inconsistencies, both rule-based and random inconsistencies. For the present study though, it suffices to know the GPC rule that dictates that mappings vary based on whether the vowel appears in a phonologically open versus a closed syllable (Basbøll, 2005). This variation occurs throughout the material, regardless of item type. Mappings that were different from these were avoided when selecting the material. An exception is of course the grapheme 'e', which has three mappings: one for the occurrence in an open syllable, one for a closed syllable and the third mapping is onto schwa. In Dutch, 'e' can also map onto schwa so this should not be difficult for the participants in the present study. Another matter is that Danish vowels include graphemes that do not exist in Dutch, being 'æ', 'ø' and 'å'. This is something that participants encounter in items in all conditions in the present study, so it is controlled for.

## **2.5 The role of sleep in word learning**

In previous research, the matter of a potential role for sleep in word learning has been addressed but it has not been fully sorted. The question remains whether newly learned words are already represented in the same way as familiar words after only one session. In other words, have these words already been adopted by the mental lexicon? The studies described above all had one experimental session, in which the testing phase immediately followed the word learning phase. In order to validate this choice, it is important to know if lexicalisation can indeed happen that fast. Lexicalisation can be defined as "the process by which episodic memories for new words become lexical items" (Hayes-Harb et al., 2010, p. 380).

Dumay and Gaskell (2007) found evidence for the role of sleep in lexicalisation, in the sense that newly learned words only became competitors in a word recognition task after a period of sleep. Sleep was essential in the sense that the same period of time spent awake, did not lead to the words becoming lexical competitors.

The role of sleep in word learning was also researched by Davis et al. (2009). They took another perspective though, considering consolidation instead of lexicalisation. Consolidation is a neuroscientific term that also relates to how words are stored. It is a term that describes a shift in the area in the brain that is involved. Unconsolidated words are represented in the hippocampal region, while consolidated words are represented in the neocortical region.

The fMRI study done by Davis et al. (2009) showed that new words are represented differently in the brain after a night of sleep. They taught their participants pseudowords that were similar to existing words. Results showed that different brain regions were activated for consolidated words compared to novel ones. Untrained words led to hippocampal activity and no lexical competition. Trained but unconsolidated words, led to less hippocampal activity but still no competition. Consolidated words, which were trained words after a night of sleep, led to (reduced) neocortical activity. This is similar to the activity for existing words. These words did become competitors in a lexical competition task, so lexicalisation had taken place as well. Moreover, what is relevant to the present study is that in a forced-choice recognition task, participants scored highly on both unconsolidated words (learned on the same day) and consolidated words (learned on the day before). However, participants recognised consolidated items better than unconsolidated items.

To conclude, there is evidence that words are represented differently after a night of sleep. Of course, there is no question that (at least some aspects of) word learning can happen within one session: results from previous research using one session show that word

learning took place. The remaining question is to what extent it is necessary to include sleep as a factor in word learning research. The present study consists of multiple testing days, with a period of sleep in between each day. Each set of words will be tested twice: once right after the learning phase and once more at the beginning of the next testing day. This should show whether there is a difference between testing right away and testing after (at least) a night of sleep. As the present study does not look at brain activity, only the performance is concerned here.

### **3. Research questions and hypotheses**

The aim of this project is to further explore the role of orthography in auditory word learning in a second language. It is new in the sense that it considers inconsistencies in mappings within the second language next to incongruencies between mappings in a first and second language. Moreover, the present study uses real language material and includes sleep as a factor. For the purpose of this study, three research questions and matching hypotheses have been construed. These can be found below. Underneath is explained how the research questions lead to the hypotheses.

#### **3.1 Research questions**

- 1) Will participants receiving auditory forms only perform better than those receiving orthographic forms and auditory forms?
- 2) Will participants perform worse on rule-based items compared to items in the baseline category?
- 3) Will participants perform better after a night of sleep?

#### **3.2 Hypotheses**

- 1) Participants receiving auditory forms only will perform better than those receiving auditory forms and orthographic forms.
- 2) There will be an interaction between group and item type: Participants that receive both auditory and orthographical forms, will perform worse on rule-based items compared to items in the baseline category.
- 3) Participants will perform better after a night of sleep. In other words: they will perform better during the second test session compared to the first one, regardless of group.

Question (1) leads to hypothesis (1), because items in both conditions (baseline and rule-based) contain incongruent mappings, which was found to harm word learning. The baseline category might consist of only congruent grapheme-to-phoneme mappings concerning consonants, but mappings concerning vowels contain incongruencies in both conditions.

Question (2) leads to hypothesis (2), because there are more rule-based inconsistencies and as a result more incongruent mappings in the rule-based condition compared to the baseline. Participants receiving auditory forms only, are not influenced by orthographic effects so there should be no effect of item type for this group.

Question (3) leads to hypothesis (3) because a night of sleep has been found to benefit word recognition.

### **4. Method**

The methodology used in the present study will be explained in this section. It was largely based on the study done by Hayes-Harb et al. (2010), with some adaptations. The most

important changes were that the present study used real language material and consisted of multiple testing days, so that sleep could be included as a factor. Moreover, the participants in this study were bilingual or multilingual which means that they had experience with language learning, albeit not with the Danish language. Hayes-Harb et al. (2010) used three groups, one with phonology only, one with congruent orthography and one with incongruent orthography. Here, congruency was manipulated within subjects only. As a result, there are only two groups: one received phonology only and the other group was exposed to both phonology and orthography. Another difference is that the present study did not use a criterion test. This choice was made so that all participants encountered each word-picture pair the same amount of times. Finally, in the testing phase only words from the study were used. No additional material was recorded that reflected the incongruent mappings, in order to stay close to the natural situation. The methodology of the present study will be described in detail in the sections below.

#### **4.1 Participants**

The participants used in this experiment were native speakers of Dutch, that had no prior experience with learning Danish. They did have experience with learning several other languages though. All participants were sequential bilinguals, with English as a second language. Most participants indicated that they also spoke German or French or both. Other languages that were mentioned in the language background questionnaire were Spanish, Norwegian, Hungarian, Arabic, Korean and Frisian. No one reported a high command of any language besides English, German or French. Ten participants lived abroad for a period of time longer than three months. Two participants even lived abroad twice. Countries that were mentioned are: England (3), Sweden (3), Norway, Germany, France, Ireland, Canada and Suriname.

Participants were approached through social media. Four participants were excluded beforehand based on their answers to the selection questions at the beginning of the experiment. They either did not meet the age requirement (2), they were not native speakers of Dutch (1) or they had prior experience with learning Danish (1). In addition, two participants were excluded for not completing the first testing day and an additional two participants were excluded for not completing the second testing day within the maximum of three days after the first testing day.

In total, the data from 30 participants were used in this study. 18 of these were female and 12 were male. All participants were between 21 and 30 years old. The average age was 24 ( $SD=2.24$ ). Participants reported to be highly educated: their highest or current education level was either university (20) or HBO (10). They were randomly distributed over two conditions: one with only auditory forms (AF only) and one condition with both auditory and orthographic forms (AF & OF). 15 participants participated in the former condition and 15 in the latter.

#### **4.2 Material**

The material used in this study consisted of 36 Danish words and 36 pictures. The following three sections will explain how the words were chosen for the experiment, how they were recorded and how the pictures matching the words were selected or created.

#### 4.2.1 Words

The Danish words in the experiment were not cognates with Dutch. Since most native speakers of Dutch tend to have a fairly high proficiency in English, it was also made sure that these words were not cognates with English either. The term ‘cognate’ is used here not in the etymological sense, but is viewed as something concerning overlap in word form. Both the orthographic as well as the phonological form were considered here. Cognates are then translation equivalents that have similar word forms. It is hard to draw a line deciding how much overlap translation equivalents need in order to do or do not counts as cognates. In the present experiment, the Danish words were allowed a maximum of two phonemes or two graphemes in common with their Dutch or English translation equivalents.

All words were underived nouns that were either mono- or bisyllabic. To reduce the word learning load, the material was split into two word sets. Participants were taught the first set during the first day and the second set during the second one. The number of syllables was kept more or less equal across word sets and conditions.

The words used can be divided over two conditions: half of the words formed a ‘baseline’ in which all grapheme-to-phoneme mappings concerning consonants are congruent with Dutch and consistent within the Danish language. Take for example the word ‘lyn’ [lyːn]. The other half of the words were in the rule-based condition. Here, mappings for consonants showed rule-based inconsistencies as a result of the GPC rule consonant gradation. Word-initially, ‘d’ maps onto [d] like in [dy:nə] and in word final position or before schwa, it maps onto [ð] like in ‘gade’ [gæ:ð]. As a result, mappings were sometimes congruent with Dutch and sometimes incongruent. This is necessary to illustrate the (rule-based) inconsistency. The grapheme ‘g’ maps onto [g] at the beginning of a word, like in ‘gade’ [gæ:ð]. At the end of a word or before schwa, it maps onto [j] after a front vowel as is the case with ‘læge’ [læ:ʝ] and onto [w] after a back vowel as is illustrated by the example ‘tog’ [tɔːw]. Participants were exposed to the initial and to both final forms, in order to be able to learn this rule. In the case of the grapheme ‘g’ all mappings were incongruent with Dutch, although the mapping of ‘g’ onto [g] does occur in some loan words in Dutch. Some examples of the material and their Dutch and English translations<sup>3</sup> can be found in tables 2a and 2b. Each set of words contained 9 baseline items and 9 rule-based items. The complete material can be found in appendix A.

When grapheme-to-phoneme mappings between vowels are concerned, there is no difference between the conditions. Mappings varied, based on whether the vowel appeared in an open or closed syllable. This led to rule-based inconsistencies for vowels for both conditions. Mappings were moreover sometimes congruent, sometimes incongruent with the Dutch language. Different mappings that were the result of additional GPC rules, were avoided.

Because of the strict selection criteria for the words in the material, some word forms turned out to be quite similar and shared the same rime. In order to prevent rime being a confounding factor, an equal number of word pairs with the same rime occurred in both word sets for both item types.

---

<sup>3</sup> Danish IPA transcriptions stem from *Den Danske Ordbog* (n.d.). Although, in some cases the dictionary transcription did not include schwa assimilation while schwa was assimilated in the material. In this case, a transcription of the realisation of the word in the material was chosen. These differences are highlighted in appendix A. For the translation to Dutch, the *Van Dale Miniwoordenboek: Deens-Nederlands, Nederlands-Deens* (2009) was used. For translations from Dutch to English, *Van Dale Nederlands-Engels* (n.d.) was used.

**Table 2a***Examples of words in the baseline condition*

| Danish word | IPA transcript | Dutch translation | English translation |
|-------------|----------------|-------------------|---------------------|
| hale        | [hæ:lə]        | staart            | tail                |
| næb         | [nɛːb]         | snavel            | beak                |
| lyn         | [lyːn]         | bliksem           | lightning           |
| avis        | [aviːs]        | krant             | newspaper           |

**Table 2b***Examples of words in the rule-based condition*

| Danish word | IPA transcript | Dutch translation | English translation |
|-------------|----------------|-------------------|---------------------|
| dyne        | [dy:nə]        | deken             | duvet               |
| gade        | [gæ:ɔ̃]        | straat            | street              |
| læge        | [lɛ:j]         | arts              | doctor              |
| tog         | [tɔːw]         | trein             | train               |

The words in the experiment were controlled for word frequency. Words in both conditions were of comparable frequency. Moreover, words were divided over the two word sets in such a way that words in the first set were not more frequent than those in the second set. The word frequency in Danish was used, taken from *KorpusDK* (n.d.). The search was restricted to exact matches and nouns only. The frequencies are listed with the complete material in appendix A.

#### 4.2.2 Audio files

The words were recorded by a native speaker of Danish with a Yeti Blue Microphone in a quiet room. The speaker was instructed to read out the words slowly but in a natural manner. Before the recording, the speaker was not aware of the purpose of the study. He received no specific instructions regarding consonant gradation or schwa assimilation. In some recordings the speaker did apply schwa assimilation whereas in others he did not. The auditory word forms were selected from a recording in which schwa assimilation did occur, because this recording was judged to be more natural. The native speaker confirmed this. Audacity was used to cut the words into separate sound files and to perform noise reduction on the material.

Surprisingly, schwa was only assimilated in words involving consonant gradation. For example, the word ‘gade’ was realised as [gæ:ɔ̃] and ‘læge’ was pronounced as [lɛ:j]. In words without consonant gradation, schwa was not assimilated even when the context did allow for it. For example, in the word ‘due’ [du:ə] schwa was realised even though it occurred after a vocoid. This is a context in which schwa assimilation is strongly favoured. Moreover, in the word ‘øje’ [ʌjə] schwa was not assimilated, while [j] did inspire schwa assimilation when it occurred in a word like ‘læge’ [lɛ:j] as a result of consonant gradation of [g]. The goal of this study is not to create an account of this, though. As the speaker used to record the material was a native speaker, his use of schwa assimilation will not be questioned here.

### 4.2.3 Pictures

The pictures in the experiment were either taken from the MultiPic database (Duñabeitia, Crepaldi, Meyer, New, Pliatsikas, Smolka & Brysbaert, 2018) or were constructed in the same style. 30 of the pictures stem from the Multipic database, which means that they were pretested. Note that the pictures were available for Dutch but not for Danish. The pictures in the experiment were thus selected using the Dutch translation and a native speaker of Danish was consulted to confirm that the pictures accurately described the Danish translation as well<sup>4</sup>. The other 6 were drawn for the sake of the experiment. These were pretested informally and were adapted when they were not named correctly. Figure 1 shows two examples of the pictures used, one for each condition. The examples illustrate what participants in the AF & OF condition saw. Those in the AF only condition saw the string 'xxxx' instead of the written words. The full set of pictures was added to appendix B.

As an additional pretesting measure, participants were asked to name the pictures before they started learning them. This was done to check if participants indeed named the pictures correctly or whether some items needed to be removed. No items were removed in the end. The discussion of this pre-test can be found in appendix C.

**Figure 1**

*Examples of pictures used in the experiments*



### 4.3 Procedure

The experiment was conducted online, using the programme Qualtrics. Participants could access the experiment through a weblink on their laptop or phone. There were three testing days: the first two were estimated to take about 25 minutes and the third about 10 minutes. Participants ideally completed the testing days in three consecutive days but were included if there was a maximum of three days in between each testing day. Table 3 shows an overview of the tasks that occurred on each testing day. Participants in both groups completed the same tasks, but the word learning phase was slightly different for each group. All tasks will be explained in detail in the sections below. Moreover, all the instructions that participants received have been added to appendix D.

---

<sup>4</sup> For example, the Danish word 'kæde' can be translated to the Dutch word 'ketting'. The Dutch word can refer to both 'chain' as well as 'necklace', whereas the Danish word can only refer to a chain. A necklace would be 'halskæde'.

**Table 3***Overview of the tasks per testing day*

| <b>Testing day</b> | <b>Tasks</b>  |
|--------------------|---|
| Day 1              | -Selection questions<br>-Instruction<br>-Naming the first 18 words (pre-test)<br>-Learning first 18 words<br>-Testing first 18 words (test 1) |
| Day 2              | -Testing first 18 words (test 2)<br>-Naming the second 18 words (pre-test)<br>-Learning second 18 words<br>-Testing second 18 words (test 1)  |
| Day 3              | -Testing second 18 words (test 2)<br>-Language background questionnaire<br>-Manipulation check  |

**4.3.1 Selection questions**

The first testing day began with three selection questions concerning the background of the participants. They were asked whether they were between 18 and 30 years old, whether Dutch was their native language and whether they had experience learning the Danish language. If the first two questions were answered with 'yes' and the third with 'no', then participants were allowed to continue the experiment. If any of the questions were answered differently, then participants were told that they were not suitable for the experiment and they were thanked for their time.

**4.3.2 Instructions**

Participants were told that they would learn Danish words and that they would afterwards be tested on their knowledge of these words. This would be the case for the first and second testing day, while the third day only consisted of a test and some (language) background questions. Participants were informed that they would need to leave their e-mail address at the end of each testing day and that this address would be used to send them the weblink for the next testing day and that this would be used to connect their data. Each testing day moreover began with an overview of the tasks of that day. Both the learning phase and testing phase were preceded by instructions as well.

Before the learning phase, participants were told that they would learn 18 words by looking at pictures and listening to words. The words would follow each other automatically and they could take two short breaks. Participants in the AF & OF group were warned that they could use the written word for studying but that it would not be present in the testing phase.

Because of the automatic nature of the experiment, participants had to fill out a checklist before they could start the learning phase. They had to indicate that they were in a quiet room, that the volume on their device was loud enough and that it had enough charge left to last through the experiment. They were given the opportunity to test their volume with the help of a sound recording containing the word 'test', which was played at the same volume as the experimental items. Lastly, they had to indicate that they were 'ready'. Then they could start the learning phase.

### **4.3.3 Word learning phase**

In the word learning phase, participants saw a picture and heard the Danish word that described it. The audio was played automatically and the picture disappeared after 4 seconds. Afterwards, the next word-picture pair followed immediately. Participants in the AF & OF group also saw the written word above the picture. Those in the AF only group simply saw the string 'xxxx'.

Each word-picture pair appeared ten times. The pairs were randomised in such a way that all 18 pairs were shown once before they were repeated again. In this manner, a pair could occur a maximum of twice in a row. There were two small breaks in the experiment: one after 3 rounds of the 18 words, the second after another 4 rounds.

### **4.3.4 Testing phase**

In the testing phase, participants were confronted with 36 word-picture pairs. Half of these were matched correctly, while the other half consisted of a mismatch between the picture they saw and the word they heard. Pairs were displayed to them in a similar way as in the word learning phase: they again heard a word and saw a picture that disappeared after 4 seconds. In this phase, there was no difference between the groups though: neither group saw a written word in this phase. Above each picture stood the string 'xxxx'. After every pair, participants were asked whether the picture matched the auditory word and they could respond with 'ja' (yes) or 'nee' (no). The mismatching pairs can be found in appendix E.

Before the actual testing phase began, participants were confronted with two test items: one matching and one mismatching pair. This was done for every test session on each testing day. At the end of each testing phase, participants were asked to estimate the percentage of the pairs that they judged correctly. This way, information could be attained about how sure participants were about their responses, while still preserving the forced-choice character of the experiment. This variable was not used in the main analysis, but was subjected to exploratory analyses.

### **4.3.5 Language background questionnaire**

The main goal of the language background questionnaire was to confirm that participants indeed had no prior experience with learning the Danish language and that the rest of their (language) background was similar. The first questions concerned more general aspects, like age, gender and education level. Afterwards, participants had to fill in the languages they spoke and for each language they had to indicate how old they were when they started learning it and how many years they have used it since then. Moreover, they were asked to report their command of the language (on a 7-point scale ranging from 'very limited' to 'perfect') and the frequency of usage for each language (on a 7-point scale ranging from 'rarely' to 'every day'). Participants were also asked whether they had lived abroad for a period longer than three months and if so, which languages they used there. The questionnaire used in this study was adapted from the LHQ3 (Li, Zhang, Yu & Zhao, 2019). All questions can be found in appendix D.

### **4.3.6 Manipulation check**

At the very end of the last experiment, participants were asked whether they had suspicions about the goal of the experiment. Keep in mind that only participants in the AF & OF group could realise what the experiment was about. If they consciously learned GPC rule consonant gradation, then they might respond that they noticed a difference in mappings

between initial versus final consonants. If this rule was learned unconsciously, then still participants might mention that some words were harder to learn than others. Apart from the inconsistencies in the material, participants might also mention something about incongruencies between Danish and Dutch mappings. Participants in the AF only group should report no difference between words, since they were not affected by grapheme-to-phoneme mappings.

#### 4.4 Design and analysis

The aim of the study was to research the effect of orthographic forms on word learning. The dependent variable was the percentage of the word-picture pairs that were judged correctly as matching or mismatching in the testing phase. This variable was expected to be influenced by several independent variables. One of them was measured between subjects, which was the type of information that the participants received. One group received auditory forms only (AF only) while the other group was trained on both auditory and orthographical forms (AF & OF). The other two variables were measured within subjects. These were item type (baseline vs. rule-based) and test session (test 1 vs. test 2). The design is displayed in table 4.

The design of this study is thus a mixed design with one continuous dependent variable and with one categorical independent variable measured between subjects and two within subjects. Therefore, a factorial mixed ANOVA was used to analyse the data. Because of the relatively small number of participants, it would have been better to construct a linear mixed effect model. However, due to time constraints and a lack of expertise, the decision was made to execute the mixed ANOVA.

**Table 4**

*The independent variables that could influence the dependent variable ‘percentage correct’*

| <b>Independent variables</b> | <b>Levels</b>           | <b>Type of measurement</b> |
|------------------------------|-------------------------|----------------------------|
| Group                        | AF only vs. AF & OF     | Between subjects           |
| Item type                    | Baseline vs. Rule-based | Within subjects            |
| Test session                 | Test 1 vs. Test 2       | Within subjects            |

Another dependent variable was collected for an exploratory analysis, which was an estimated percentage correct. As some learners claim to need orthography to learn words, we wondered whether the learners’ own judgement on how well they performed might differ between groups. Based on the literature, no specific hypothesis could be set up regarding this estimated percentage correct. Therefore, the variable called for an exploratory analysis. This will be further addressed in the discussion.

### 5. Results

In this section, the results of the present study will be described. This section will solely report the result from the main analysis. The exploratory analyses will be discussed in the discussion section.

The percentages correct were calculated for each participant separately for item type (baseline vs. rule-based) and test session (test 1 vs. test 2). The percentages correct for

matching and mismatching pairs were taken together<sup>5</sup>. Table 5 shows the average percentages correct, split by group, item type and test session. The first thing that stands out when looking at the data in tables 5, is that the average percentages correct are quite high (95%-100%). The data show a ceiling effect of performance.

**Table 5**

*Percentage correct and standard deviations (in parentheses), split by group, item type and test session*

| Item type          | Test session |            |
|--------------------|--------------|------------|
|                    | Test 1       | Test 2     |
| <i>AF only</i>     |              |            |
| Baseline           | 96% (4.77)   | 96% (5.63) |
| Rule-based         | 97% (4.48)   | 96% (5.73) |
| <i>Mean</i>        | 96%          | 96%        |
| <i>AF &amp; OF</i> |              |            |
| Baseline           | 99% (1.76)   | 98% (2.71) |
| Rule-based         | 98% (3.30)   | 96% (5.34) |
| <i>Mean</i>        | 98%          | 97%        |

A factorial mixed ANOVA<sup>6</sup> was performed on the percentage correct scores, using SPSS. The results show only a main effect of test session ( $F(1,28)=5.58, p=.03, \text{partial } \eta^2=.17$ ), indicating that participants performed better overall on the first test ( $M=97.36, SE=.64$ ) compared to the second test ( $M=96.25, SE=.77$ ). None of the other effects were significant and there were moreover no significant interactions.

## 6. Discussion

The results found in this study do not confirm any of the hypotheses that were formed for the sake of this study. This section first discusses the results with reference to the hypotheses. Subsequently, we discuss possible explanations for the results. Then, the limitations of the present study will be outlined. Following this, the implications for future studies will be discussed. Finally, suggestions for future research will be made.

The first hypothesis was that participants in the AF only group would perform better than those in the AF & OF group. This was based on previous studies that found that incongruent orthography had a detrimental effect on word learning, compared to no orthographic forms at all (Escudero et al., 2014; Hayes-Harb et al., 2010). Yet, no main effect of group was found in the present study, even though items in both conditions included incongruent mappings for vowels and items in the rule-based condition included

<sup>5</sup> Before the percentages correct for matching and mismatching pairs were piled together, an analysis was done to assure that this decision was valid. Because the difference scores for the data were not normally distributed, a Wilcoxon signed-rank test was performed. This test showed that there was no difference between the percentage correct on matching pairs ( $Mdn=98.61$ ) compared to mismatching ones ( $Mdn=97.92$ ),  $z=.24, p=.81, r=-.04$ . This supports the decision to combine these scores.

<sup>6</sup> Even though the data from the percentages correct were not normally distributed, the factorial mixed ANOVA was still performed. The reasoning behind this decision will be explained in the discussion section, as will be the possible reasons for the ceiling effect in the data distribution.

incongruent mappings for consonants as well. One striking result is that the performance of both groups was very good, indicating a possible ceiling effect.

The second hypothesis built on the first one, adding that there would be an interaction between group and item type. Since the items in the rule-based condition included more incongruent mappings and rule-based inconsistencies in mappings, they were expected to hinder word learning to a larger extent. No studies as of yet have looked at the effect of rule-based inconsistencies in second language receptive word learning, but still a negative effect of (rule-based) inconsistencies was expected here. This was based on results from a study that found orthographic interference from rule-based inconsistencies in productive word learning (Hayes-Harb et al., 2018) and research concerning first language reading acquisition that found that children with native languages including many inconsistencies took longer to acquire reading skills (Seymour et al., 2003). However, the hypothesized interaction between group and item type was not found either.

The absence of a main effect of group or an interaction between group and item type, can have several reasons. It could be the case that the participants had no difficulties dealing with words that show rule-based inconsistencies and the incongruencies that come along with them. This would contradict previous research though. That no effects were found, does not necessarily mean that such effects do not exist. If these effects do exist but were merely not found in this study, then this could be due to the relatively small sample size or the ceiling effect. Another possible explanation for this, is that the participants in the AF & OF group did not pay close attention to the orthographic forms because they were aware that these forms were supplementary. At the beginning of the experiment they were warned that they had to listen carefully, since the written words would not be available to them in the testing phase. It is unclear if this was the case in previous studies as well.

The third hypothesis was that participants would perform better after a night of sleep. In other words, they would perform better on the second test compared to the first one. This hypothesis stemmed from previous research, which found that a night of sleep improved word recognition (Davis et al., 2009). A main effect of test session was found. This indicates that there is indeed a difference between performance on the first set compared to the second set. However, this difference was not in the expected direction: participants performed worse on the second test compared to the first one. This is contrary to what was expected based on previous research. It is hard to say what caused this effect. Perhaps it was some sort of fatigue effect. Two participants reported that they had trouble staying motivated because of the length of the experiment and the similarity between some of the word pairs. It is likely that more participants faced the same issue. Future research could continue to investigate the effect of sleep on word learning.

As an exploratory measure, participants were asked to estimate their own percentage correct scores after each test session on each testing day. The effect of group was investigated for the first and the last estimate only, as running more tests would increase the chances of a Type I error. For the sake of completeness, a table with all of the averages per group for each estimation was added to appendix F. No hypotheses could be formed based on previous research.

The estimated percentages correct that were collected after the first test session of the first 18 words, were not normally distributed. Therefore, a Mann-Whitney test was performed on these data. The estimated percentages correct of the AF only group ( $Mdn=85.00$ ) did not significantly differ from those of the AF & OF group ( $Mdn=84.00$ ) on the first time they were asked to estimate their performance:  $U=95.50$ ,  $z=-.71$ ,  $p=.49$ ,  $r=.13$ .

The last time participants were asked to estimate their performance, after the second test session of the second set of 18 words, the estimated percentages correct were lower for the AF only group ( $M=73.53$ ,  $SD=18.12$ ) compared to the participants in the AF & OF group ( $M=88.53$ ,  $SD=10.66$ ). This difference was significant:  $t(28)=-2.76$ ,  $p=.01$ ,  $r=0.46$ , 95% CI [-25.02, -5.87].

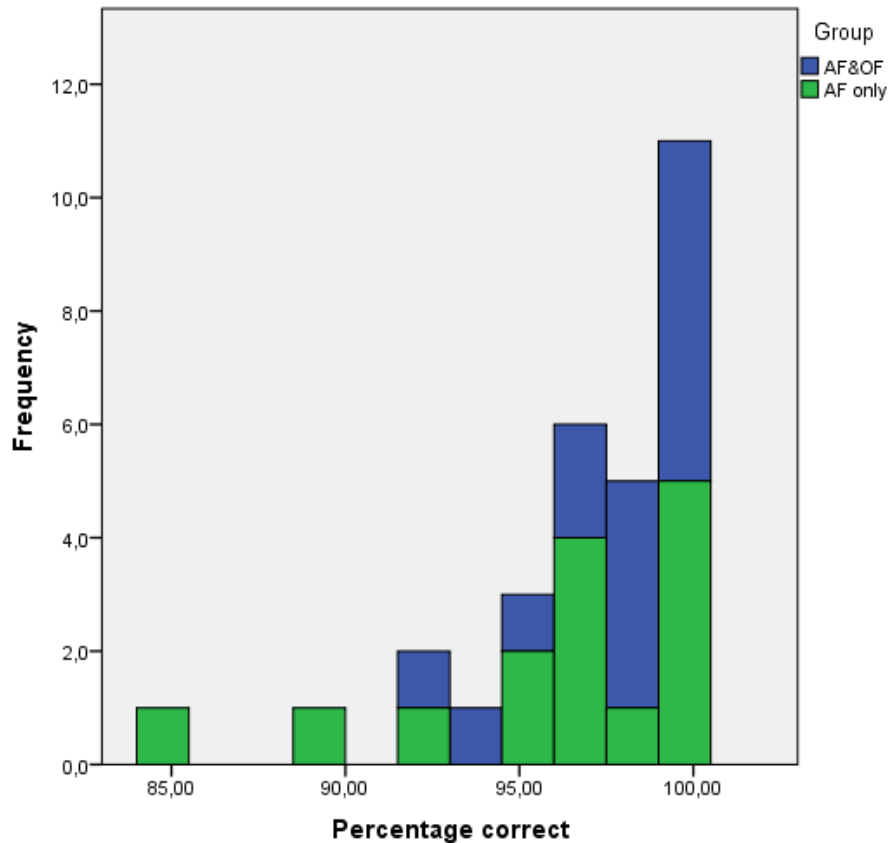
It is difficult to say what exactly caused this result. Did the lack of orthographic forms make participants in the AF only group less confident? One participant in this group did mention that she felt that it would have been easier to learn the words, had the written forms been available to her. It is unclear though, why there would only be a significant difference for the last estimate and not for the first one. Future research could further investigate the possible relationship between the group (AF only vs. AF & OF) and the confidence of the learner. Perhaps a study with a larger number of participants could shed more light on this matter.

There are several limitations to this study, which could have led to the discrepancies between the hypotheses and the results. These limitations moreover mean that the results obtained should be regarded critically. Firstly, problems with the data will be addressed and possible causes of these problems will be named. Afterwards, a more general discussion of the advantages and disadvantages of the used method will follow.

The main problem with the data, is that they are not normally distributed. The Shapiro-Wilk test of normality was highly significant for all variables ( $p<.01$ ) and the skewness and kurtosis values were for most variables not within the acceptable range. The assumption of the homogeneity of variance was moreover violated for the first test session ( $F(1,28)=10.55$ ,  $p>.00$ ) and the second test session ( $F(1,28)=5.19$ ,  $p=.03$ ) of the baseline condition. This is mainly due to a large ceiling effect, as is illustrated by figure 2. It is necessary to have normally distributed data in order to run a mixed ANOVA. Despite of this, the test was still performed. The main reason for this, was that no robust alternative for this test exists in SPSS and the combination of a lack of expertise and time constraints did not allow for running a robust mixed ANOVA in R or for creating a linear mixed effect model. The results in this study should therefore be regarded with certain care.

**Figure 2**

*Data distribution, averaged over item type and test session.*



Several possible explanations exist for the ceiling effect of performance: the type of task, the specific materials chosen and the influence of the language background of the participants. These explanations will be discussed below. The methodology in the present study was largely based on the study done by Hayes-Harb et al. (2010). Therefore a comparison between the task, material and participants in that study and the present study, should help to shed light on how the ceiling effect could have occurred.

The main difference between the current study and the Hayes-Harb et al. (2010) study, is that in the present study no criterion test was used. This way, the amount of encounters with each word-picture pair could be kept equal for all participants. This amount was set on 10 encounters, which was based on the fact that participants in the Hayes-Harb et al. (2010) study needed on average two word learning cycles before they passed the criterion test and could continue to the testing phase. This meant that they had encountered all words eight times in the word learning phase and had moreover judged them four times already in the two criterion tests: twice in matching pairs and twice in mismatching pairs. The number of encounters was thus carefully considered. How could this ceiling effect still have occurred?

Firstly, participants in the Hayes-Harb et al. (2010) study had to learn 24 words in one session, while the participants in the present study only had to learn 18 words per testing

day. Another explanation, is that the way that mismatching items were created could have had an effect. In the present study, mismatching pairs consisted of a picture matched with another word from the material. Hayes-Harb et al. (2010) only constructed a third of the mismatching items this way. In the other two thirds of the mismatching pairs, a picture was matched with an auditory word that was similar to the matching word but reflected the incongruent spellings. The goal of this was to see if participants that were trained on the incongruent forms, might have had different lexical presentations of the words. The present study only focussed on whether words would be successfully recognised, so all mismatching pairs were created with words from the material. This means that the testing phase in the present study was thus arguably easier.

Another point that ought to be made, is that the present study used real language material. Because Danish and Dutch are both Germanic languages, it was difficult to avoid cognates while controlling for phonological rules. And because of the language background of the participants, cognates with English had to be avoided as well. Therefore the criterium was that words could be selected as items if they did not exceed an overlap of two phonemes or two graphemes. Despite of this precaution, participants might still have benefited from the small portions of overlap that were present in the material. Hayes-Harb et al. (2010) used pseudowords, which allowed them to create words that differed to a larger extent from their translation equivalents. Other studies even used pictures of pseudo-objects as well as pseudowords (Escudero et al., 2008; Escudero et al., 2014; Escudero, 2015), which meant that there was no influence from the word form of the translation equivalents since there were none.

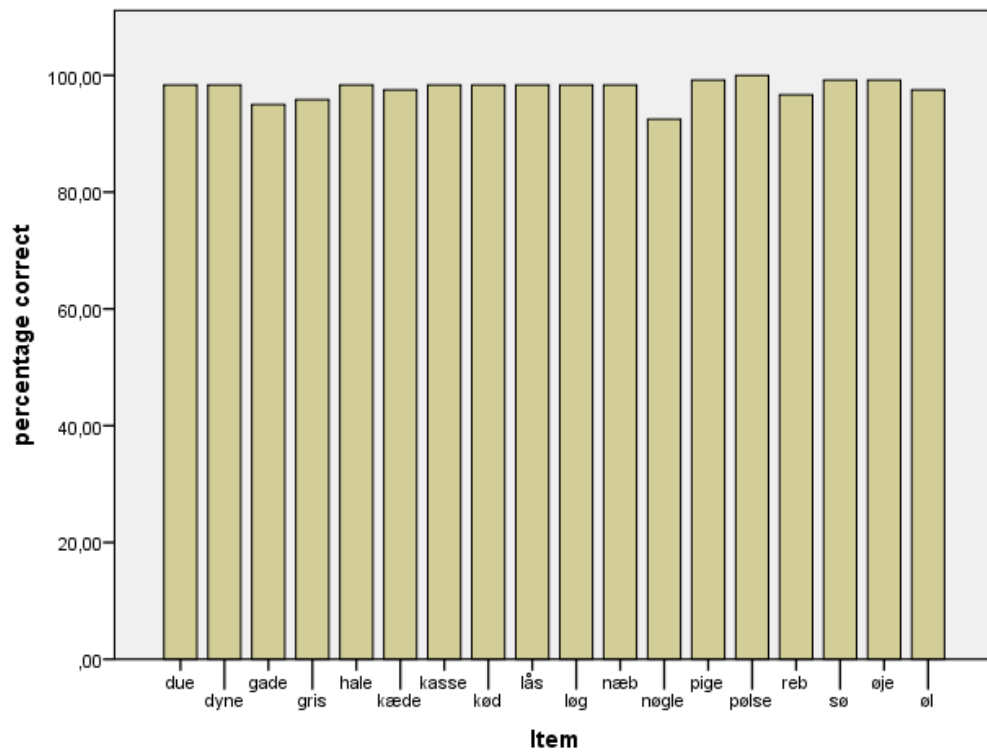
When participants were asked at the end of the experiment to mention any suspicions they had regarding the aim of the study, they did mention that they suspected that the study had something to do with similarities between Danish and Dutch or English. When inspecting the data by eye though, there does not seem to be an effect of overlap between Danish and Dutch or Danish and English, the second language of the participants. Figures 3a and 3b show the percentage correct by item, piled together for test sessions and groups. Percentages correct represent the item both in matching and mismatching pairs. All items were judged correctly by most of the participants. No specific items seem to have contributed disproportionately to the ceiling effect. To conclude, participants do not appear to have been effected by the presence or absence of overlap.

Another alternative explanation that participants proposed, was that the similarities within the material were core to the present study. The rhyming pairs within the material do not seem to have lower percentage correct scores than the other words. It is hard though to draw absolute conclusions regarding the effects of similar words within the material, because here only four rhyming pairs were concerned.

Finally, some participants also mentioned that they thought the semantic links between some of the words were part of the experiment. Again, no items seem to have had a differential effect on participants.

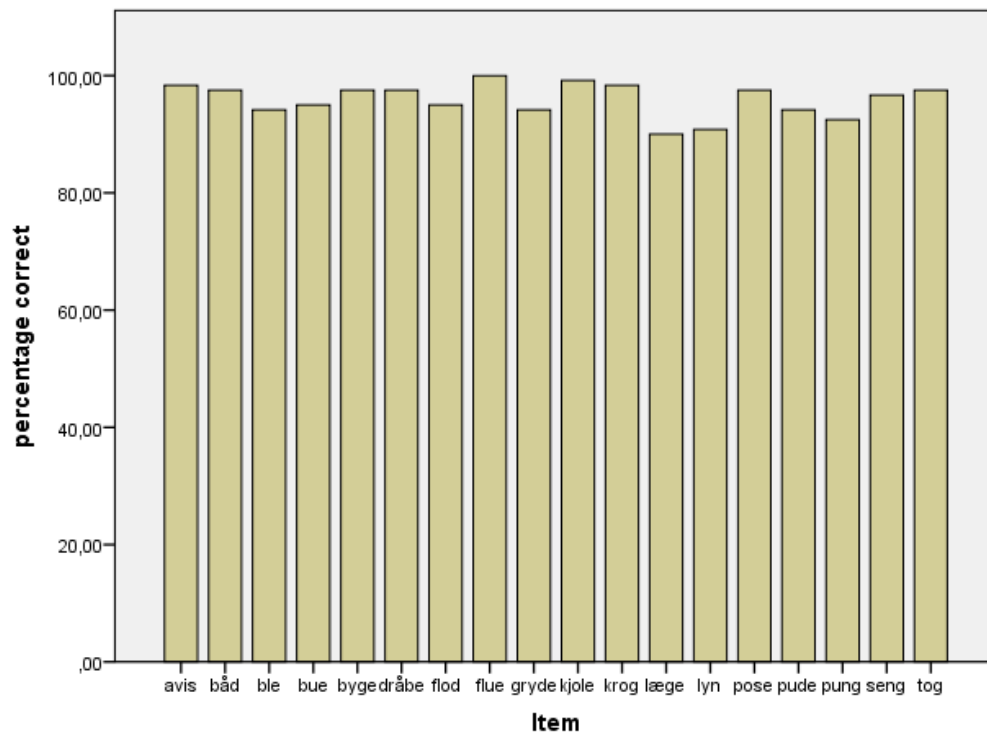
**Figure 3a**

*Percentages correct for the first set of 18 words*



**Figure 3b**

*Percentages correct for second set of 18 words*



Besides overlap, participants could have experienced another benefit of the real language material. Which is of course that they could have heard the words before. Even though participants were selected based on them having no experience learning Danish, they could still have heard Danish words before if they ever went on holiday to Denmark or if they watched a Danish crime show for example. Moreover, participants could have benefited from hearing Swedish or Norwegian before. Some of the participants reported having experience with learning Norwegian (1) or even living in Norway (1) or Sweden (3) for a period of time.

In order to check if the participants that had experience with Norwegian or Swedish influenced the results, the main analysis was run again without these participants. Excluding these participants did not change the results, besides the fact that the main effect of test session was no longer significant ( $F(1,23)=3.66, p=.07, \text{partial } \eta^2=.14$ ). None of the other effects were significant either, which is in line with the original analysis. The data distribution was still not normal. Based on these data, it cannot be concluded whether the excluded participants caused the effect in the original analysis or if the analysis without these participants no longer had enough power to detect an effect that was in reality still present. Most importantly, there is still no effect of group or an interaction between group or item type. Because the sample size was rather small to begin with, this analysis did not replace the original analysis. Still, the language background of the participants could have contributed to the ceiling effect.

The native language of the participants could have played a role as well. The Dutch language also has a GPC rule that causes graphemes to be mapped onto different phonemes when they occur in final position, namely final devoicing. It could be the case that participants had no difficulties with consonant gradation, because they are used to rule-based inconsistencies linked to the word position. If this was indeed the case, then participants did at least not seem to be consciously aware of this. At the end of the experiment, when participants were asked to write down any suspicions they had about the purpose of the study, no one mentioned anything about the rule-based inconsistencies within the material. One participant did say something about the incongruencies in the material. She mentioned that some words were pronounced the same way as she would read them in Dutch, while others were pronounced very strange or unpredictably.

Future research could further look into questions regarding how easily participants notice inconsistencies compared to incongruencies and if this relates to how much they are affected by them. Moreover, the matter of whether familiarity with certain GPC rules in their native language can help participants to spot and master GPC rules in a new language could be investigated. This could be done by comparing Dutch participants learning Danish with participants with a native language that does not include a GPC rule that is related to word position.

Another aspect of the language background of the participants that should be considered, is that the participants in this study were highly multilingual. Their experience with language learning might have given them an edge over the English monolinguals in the Hays-Harb (2010) study. Language background has been considered before in previous research. Escudero (2015) taught Dutch pseudowords to four different groups of participants. Two groups with Spanish as a native language and two groups with Australian English as a native or dominant language. Out of the Spanish groups, the participants in one group had experience with Dutch and English to various extents. Those in the second group had no experience with Dutch and only basic knowledge of English. Out of the two English

speaking groups, one group was monolingual and the participants in the other group were bilingual with either Arabic or Mandarin as a second language. She found that the Spanish groups outperformed the English groups and that the experienced group outperformed the unexperienced group. There was no effect of bilingualism or multilingualism by itself. Language background did not influence orthographic effects. The aim of this study was slightly different though than that of the current study, because it focussed on the learning of minimal pairs rather than word learning in general. Results might be different when the words that ought to be learned do not form minimal pairs. Moreover, the orthographic depth of the second language of bilingual participants could still be manipulated systematically. This is something that could be considered in future research.

The methodology used in the present study had several advantages, but also some disadvantages. These will be discussed now. Contrary to previous research, this study used was conducted online. Carrying out research online can have several advantages. For example, it makes it easier to find participants because they can do the study in their own time and do not have to come to a lab or invite someone with testing equipment into their home. Besides, it can be argued that an online environment creates a natural language learning situation. Nowadays it is quite common to learn a language online on a phone or laptop through an app. During the pandemic it was for a while moreover the only way to collect experimental data. It is important to consider the lockdown as a factor, because it is likely that the participants suddenly took part in a lot more online contact compared to their usual situation. This is due to work or school and social contact all of a sudden taking place online as well. This increase in screen time and the stress that they were likely experiencing, might have caused them to become fatigued more easily.

Another disadvantage is that online research does create a less controlled situation. It is hard to know how focussed participants are and you become dependent upon things like their Wi-Fi connection. Two participants mentioned that sound files were not always played. This happened only once during a testing phase though, so the impact on the results is limited.

A problem with the program Qualtrics specifically, is that the time stamp it collects does not specify when participants complete which part. This makes it difficult to exclude participants based on the time they spent on the study. For example, one participant had an absurdly large time stamp of over 87 hours. It is not likely that these hours were spent actively working on completing the survey. Still, it is not clear if the participant completed half of the experiment in one day and the other half two days later or if he looked at the instructions and decided that he would do the experiment some other time. If the former is true, then consolidation effects could appear within the first testing day and the data would have to be thrown away. However, if the latter is true, then there would be no issue in keeping the data.

Something else to be taken into consideration, is the length of the experiment. In order to investigate the role of sleep, the experiment needed to consist of at least two testing days. A third day was added so that more items could be included without making the individual testing days too long. Despite of this effort, the first and second day still turned out to take twenty-five minutes to complete, which is rather long for an experiment for which participants do not receive compensation. The main downside of the length of the first two testing days is that it probably made it more difficult to recruit participants. Moreover, long experiments can lead to fatigue effects. Two participants mentioned having trouble staying motivated. There does not seem to be a strong fatigue effect within these

data though. Participants made very few mistakes and there was even a ceiling effect of performance. The main effect of test session is the only thing indicating a possible fatigue effect.

The implications of the present study mostly concern methodological aspects of word learning research. Because of the small number of participants and the ceiling effect within the data, it is hard to draw conclusions about what the results of this study might mean for language learning and teaching practices.

One reason why the present study could be considered interesting, is because it is the first attempt to look at orthographic effects in word learning using real language material. The ceiling effect in this study implies that research using real words might be fundamentally different than research using pseudowords. This is possibly the case because of the potential overlap between real words and their translation equivalents or because real words can be encountered by participants before taking part in a study. Future research could further investigate how word learning with real words is possibly different from learning pseudowords.

The present study made an attempt to further investigate the role of sleep in word learning. There was a main effect of test session in the present study, but this effect was not in the expected direction: participants performed worse on the second session. Further research could continue to look into the effects of sleep. This could possibly be done in the form of an fMRI study, which would show in which brain regions words are represented after one session compared to after a night of sleep.

There are two things that stand out when looking at the language background of the Dutch participants in this study, that can be considered potentially relevant not only for research into orthographic effects in word learning, but for linguistic research in general. The first thing is that the participants are all highly bilingual and that most of them are even multilingual. Even though it is a relatively small sample, it is remarkable that a third of the participants have lived abroad. This raises two questions that could be addressed in future research. The first being to what extent the word learning of these multilingual Dutch participants is comparable to word learning done by monolingual participants. Moreover, does this mean that we no longer have a baseline of monolingual Dutch (students)?

The second thing that is worth mentioning, is that the participants in this study seem to be relatively modest when it comes to the languages they mention. Familiarity with the Dutch school system would lead to the expectation that most participants would have had both French and German in secondary school, next to English. Yet there are ten participants that do not mention any command of these languages or they refer to years of studying them in a small side note at the end. Are they being modest or is their command of French and German really not worth mentioning?

Future research could go in several directions. First of all, the factors in this study could be investigated again using more participants and fitting a more refined model to the data. A similar study could aim to avoid the ceiling effect in the present study by reducing the number of encounters participants get with the words, either by pretesting the words or by adding a criterion test to the study. Another way to prevent a ceiling effect, could be to use a more sensitive measure than the percentage correct. For example reaction times could be measured.

Another issue that future research could focus on, is how research into orthographic effects using pseudowords could be replicated with real language material. This ties in with the language background of the participants. Not only the influence of the amount of

experience with a language can be investigated, but also other aspects of language background. If participants already have experience learning a second language with a lot of (rule-based) inconsistencies, will they then be more equipped to deal with orthographic effects?

Furthermore, if another study would find an effect of rule-based inconsistencies in the material, then the role of explicit instruction could be further investigated. Hayes-Harb et al. (2018) found that even after explicit instruction, participants showed orthographic interference from rule-based inconsistencies during productive second language word learning. Perhaps explicit instruction could have a benefit though, even in receptive word learning, in the long run.

The exploratory measure, the estimated percentages correct, could also be further investigated in future research. Or other variables could be researched that represent the confidence of learners in some way. Not only the relationship between group and confidence, but also a possible relationship between group and motivation could be investigated. Are participants less motivated when they do not receive orthographic information, because they feel like they need this information in order to learn new words? Do incongruent or inconsistent orthographic forms cause participants to become demotivated?

In order to take the topic of the present study a step further, future research could also look at random inconsistencies in mappings. These are inconsistencies in grapheme-to-phoneme mappings that cannot be explained by GPC rules. This would be the moment where the orthographic depth of a language comes into play. Ziegler et al. (2000) find that for German, deviations from GPC rules usually occur in loan words, proper names and geographical terms. In Danish, loan words also form a source of irregularity. It is part of the Danish language policy to adapt the spellings of loan words once they have become part of the Danish vocabulary. An example of this is the word for mayonnaise, which is now spelled 'majonæse' (Basbøll, 2005). Despite of this policy, Danish tends to make fewer changes than Swedish and Norwegian. Look at the French loan word 'menu' for example, which is also spelled as 'menu' in Danish. Based on Danish grapheme-to-phoneme mappings, one would expect it to be pronounced as [menu] while it must be pronounced as [meny]. In Swedish and Norwegian, the spelling does represent the pronunciation. In both languages, the word is spelled as 'meny'. Loanwords can also disobey GPC rules. For example, the word 'guide' [gɑjd], which was borrowed from French through English<sup>7</sup>, violates Danish grapheme-to-phoneme mappings as well as GPC rules. Without considering GPC rules, the expected pronunciation would be [guid] and with taking these into consideration, something like [guið] would be expected. Whereas the mappings in words like 'menu' and 'guide' might be confusing to native Danish speakers, learners of Danish might be familiar with these mappings from their own language or a second or third language. Therefore, investigating these effects can be quite complex.

Another important question could be whether feedback inconsistencies, which are inconsistencies in the mappings from phonemes onto graphemes, have an effect on word learning. And if so, whether this effect is different from the effect of feedforward inconsistencies in grapheme-to-phoneme mappings. As mentioned before, languages can have seemingly inconsistent mappings that can be explained by GPC rules. However, these rules are sometimes only helpful in the direction of graphemes to phonemes. A word that was used in this study was the word 'løg' which was pronounced [lɑj<sup>7</sup>] due to consonant

---

<sup>7</sup> <https://ordnet.dk/ddo/ordbog?query=guide>

gradation. Knowing that 'g' can map onto [j] in final position might cause you to expect a [j] in final position to map onto the grapheme 'g' but this is not always the case. In the word 'tøj' [tʌj] meaning 'clothes', the [j] maps onto the grapheme 'j'. To conclude, there is a whole range of orthographic effects that can still be investigated.

## 7. Conclusion

The present study investigated the influence of orthographic forms on second language auditory word learning in comparison to word learning with auditory forms only. The aim was to find out if incongruencies in grapheme-to-phoneme mappings between the native language (Dutch) and the new language (Danish) hindered word learning. Moreover, a possible negative influence of rule-based inconsistencies within Danish was investigated. No evidence for an effect of orthographic forms was found.

Based on previous research, orthographic forms with incongruent mappings were expected to hinder auditory word learning. The present study investigated whether there was also an effect of inconsistency in grapheme-to-phoneme mappings. This has not yet been researched for second language receptive word learning. Based on research concerning first language reading acquisition and second language productive word learning, however, inconsistency was expected to hinder second language word learning as well.

The present study used two groups of participants in order to investigate the effect of orthographic forms. One group was trained on auditory forms only (AF only) and the second group on both auditory and orthographic forms (AF & OF). The effects of congruency and consistency were investigated by creating two conditions. There was one 'baseline' category in which all mappings involving consonants were both congruent between Dutch and Danish and consistent within Danish. There was also a rule-based condition in which mappings for consonants varied as a result of the GPC rule consonant gradation. These mappings were thus inconsistent within Danish and as a result of this inconsistency, these mappings became at times incongruent with the Dutch language. Mappings for vowels included incongruencies in both the baseline and the rule-based condition. In addition, the effect of sleep was investigated by means of having multiple test sessions. Previous research showed better performance on a second test session after a night of sleep, as words become represented in a different brain region.

The expectation was that participants in the AF only group would outperform those in the AF & OF group, because of the baseline level of incongruencies that occurred in both groups. In addition, an interaction between group and condition (baseline vs. rule-based) was hypothesized. Participants in the AF & OF group were expected to perform better on items in the baseline condition compared to those in the rule-based condition. Neither a main effect of group, nor an interaction between group and condition was found. The most striking result was a large ceiling effect. Regarding the effect of sleep, participants were expected to perform better on the second test session compared to the first one, because of the night of sleep in between those sessions. There was a significant effect of test session, but this effect was in the opposite direction.

The results of the present study could indicate several things. Firstly, it could be the case that learners in the AF & OF group had no issues with the incongruencies in the material nor with the (rule-based) inconsistencies in the rule-based condition. This would contradict previous research though. Another explanation of the results could be that participants in the AF & OF condition ignored the orthographic forms, which would make the groups similar. The question is to what extent orthographic information can be ignored. Previous

research found that participants used this information, even if it was detrimental to learning. In line with the large ceiling effect, it could be the case that in reality there was an effect of orthography but it was simply not detected. This could be the case because the experiment turned out to be too easy or because the sample of participants was not large enough. The effect of test session that was not in the expected direction, could indicate a fatigue effect or could be due to boredom or lack of motivation. As this was an online study, we were not able to determine this further.

Since the main result of the present study was a large ceiling effect, a lot of questions remain that could serve as inspiration for future research. The variables that appeared in this study, could be investigated again using more participants and a more refined statistical model could be applied to the data. Not only the variables from the main analysis, being group, item type and test session, but also the estimated percentages correct could be investigated again in a different setting. Future research using real language material could prevent the ceiling effect found here, for example by shortening the learning phase or by measuring a more sensitive dependent variable like reaction time instead of the percentage correct.

Another factor that could be considered in more depth is language background. A possible question could be to what extent participants can benefit from having experience with learning a second language with many (rule-based) inconsistencies. In the specific case of learning the rule of consonant gradation, future research could also investigate whether participants can benefit from already knowing a GPC rule that takes position into account like final devoicing. Another matter to consider is the potentially differential effect of random inconsistencies compared to rule-based inconsistencies. Moreover, the possible effect of feedback inconsistencies could be investigated as well. In short, all factors contributing to the complexity of second language word learning can be a topic for future research. Eventually this field of research should aim to provide insights that are relevant for language teaching.

The present study contributed to the research area concerning the influence of orthography on second language auditory word learning by looking at the effect of rule-based inconsistencies in mappings within the Danish language on Dutch participants, next to the effect of incongruencies between Danish and Dutch mappings. This could provide inspiration for future research to further investigate rule-based inconsistencies and possibly also random inconsistencies in order to come closer to considering the full complexity of learning words in a second language. This study was moreover unique in the sense that it used real language material. Even though no significant effect of orthography was found, the present study could still provide materials and methodological advice for future studies that aim to go in the same direction.

## 8. References

- Basbøll, H. (2005). *The Phonology of Danish*. Oxford: Oxford University Press.
- Bleses, D., Basbøll, H., & Werner, V. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, 26(8), 1193-1231.
- Borleffs, E., Maassen, B., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative review. *Reading and Writing*, 30, 1617-1638.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of cognitive neuroscience*, 21(4), 803–820. <https://doi.org/10.1162/jocn.2009.21059>
- Den Danske Ordbog* (n.d.) Retrieved from <https://ordnet.dk/ddo>
- Det Danske Sprog- og Litteraturselskab (n.d.). *KorpusDK sproget i brug*. Available online at <https://ordnet.dk/korpusdk>
- Dumay, N., & Gaskell, M. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18, 35-39.
- Duñabeitia, J.A., Crepaldi, D., Meyer, A.S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808-816.
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36, 345-360.
- Escudero, P., Simon, E., & Mulak, K. (2014) Learning words in a new language: orthography doesn't always help. *Bilingualism: Language and cognition*, 17(2), 384-395.
- Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied psycholinguistics*, 36, 7-22.
- Frost, R., & Ziegler, J.C. (2007). Speech and spelling interaction: The interdependence of visual and auditory word recognition. In M.G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 107–118). Oxford: OUP.
- Glushko, R.J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5(4), 104-115.
- Grijzenhout, J. (2000). Voicing and devoicing in English, German, and Dutch; evidence for domain-specific identity constraints. *Theorie des Lexikons: Arbeiten des Sonderforschungsbereichs* 282, 116.
- Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the Phonological Forms of New Words: Effects of Orthographic and Auditory Input. *Language and speech*, 53(3), 367-381.
- Hayes-Harb, R., Brown, K., & Smith, B.L. (2018) Orthographic Input and the Acquisition of German Final Devoicing by Native Speakers of English. *Language and Speech*, 61(4), 547-564.
- Lété, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on word spelling among first- to fifth-grade French children: A regression based study. *Journal of Memory and Language*, 58, 962-977.
- Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*.
- Seidenberg, M.S., & Tanenhaus, M.K. (1979). Orthographic effects on rime monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 546-554.

- Seymour, P. Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143-174.
- Stone, G.O., Vanhoy, M., & Van Orden, G.C. (1997). Perception is a two-way street: feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337-359.
- Van Dale Miniwoordenboek: Deens-Nederlands, Nederlands-Deens* (2009). Utrecht/Antwerpen: Van Dale uitgevers.
- Van Dale Nederlands-Engels* (n.d.) Retrieved from <https://run.vandale.nl/zoeken/zoeken.do>
- Ziegler, J.C., & Jacobs, A.M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language, 34*(5), 567-593.
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers, 28*, 504-515.
- Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for –OUGH and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, and Computers, 29*, 600-618.
- Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology, 12*, 413-430.

## 9. Appendices

### Appendix A: Full material

This appendix contains two tables (6a and 6b) with the full material used in the present study. The tables include the written word forms, the dictionary version of the IPA transcription and the transcription of how the word is realised in the material. Moreover, the meaning in Dutch and English can be found as well. Finally, the frequency of the Danish word is listed. This is the amount of times a word occurred in the *KorpusDK* (n.d.). In the search, only nouns that were exact matches were counted.

Within the tables, the word pairs that shared a rime can be found. For the rule-based items, this was the pair ‘kæde’ and ‘gade’ on the first testing day and ‘tog’ and ‘krog’ on the second testing day. For the baseline words, the pairs were ‘næb’ and ‘reb’ on the first testing day and ‘bue’ and ‘flue’ on the second testing day. Keep in mind that the words ‘kæde’ and ‘gade’ do not appear to rhyme when one looks at the transcription. However, as Dutch listeners have trouble differentiating between the vowels [æ] and [ɛ], the participants will probably have perceived these words as a rhyming pair.

**Table 6a**

#### *Rule-based words*

| Word  | As in dictionary | As in material | Meaning Dutch     | Meaning English    | Frequency Danish |
|-------|------------------|----------------|-------------------|--------------------|------------------|
| dyne  | [dy:nə]          | [dy:nə]        | dekbed            | duvet              | 300              |
| due   | [du:ə]           | [du:ə]         | duif              | pigeon, dove       | 154              |
| gris  | [gʁiːs]          | [gʁiːs]        | varken            | pig                | 344              |
| gade  | [gæ:ðə]          | [gæ:ð]         | straat            | street             | 1138             |
| kæde  | [kɛ:ðə]          | [kɛ:ð]         | keten,<br>ketting | chain              | 486              |
| kød   | [køð]            | [køð]          | vlees             | meat, flesh        | 2868             |
| nøgle | [nʌjlə]          | [nʌjlə]        | sleutel           | key                | 395              |
| pige  | [pi:ə]           | [pi:j]         | meisje            | girl               | 7535             |
| løg   | [lʌjʔ]           | [lʌjʔ]         | ui (ajuin)        | onion              | 1155             |
| dråbe | [dʁɑ:bə]         | [dʁɑ:bə]       | druppel           | drop               | 297              |
| gryde | [gʁy:ðə]         | [gʁy:ð]        | pan, pot          | pot                | 905              |
| pude  | [pu:ðə]          | [pu:ð]         | kussen            | cushion,<br>pillow | 184              |
| flod  | [floʔð]          | [floʔð]        | rivier            | river              | 420              |
| båd   | [bɑʔð]           | [bɑʔð]         | boot              | boat               | 1148             |
| læge  | [lɛ:jə]          | [lɛ:j]         | arts              | doctor             | 3734             |
| byge  | [by:ə]           | [by:j]         | bui               | rain shower        | 143              |
| tog   | [tɔʔw]           | [tɔʔw]         | trein             | train              | 2351             |
| krog  | [kʁɔʔw]          | [kʁɔʔw]        | haak              | hook               | 386              |

**Table 6b***Baseline words*

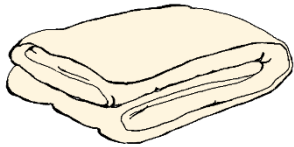
| <b>Word</b> | <b>As in dictionary</b> | <b>As in material</b> | <b>Meaning Dutch</b>  | <b>Meaning English</b> | <b>Frequency Danish</b> |
|-------------|-------------------------|-----------------------|-----------------------|------------------------|-------------------------|
| lås         | [lɔːs]                  | [lɔːs]                | slot                  | lock                   | 199                     |
| pølse       | [pølsə]                 | [pølsə]               | worst                 | sausage                | 266                     |
| reb         | [ʁɛːb]                  | [ʁɛːb]                | touw                  | rope                   | 277                     |
| næb         | [nɛːb]                  | [nɛːb]                | snavel                | beak, bill             | 354                     |
| hale        | [hæ:lə]                 | [hæ:lə]               | staart                | tail                   | 483                     |
| sø          | [søː]                   | [søː]                 | meer                  | lake                   | 947                     |
| kasse       | [kasə]                  | [kasə]                | doos                  | chest, crate           | 1070                    |
| øl          | [øɫ]                    | [øɫ]                  | bier                  | beer                   | 3240                    |
| øje         | [ʌjə]                   | [ʌjə]                 | oog                   | eye                    | 6286                    |
| ble         | [bleː]                  | [bleː]                | luier                 | diaper                 | 158                     |
| pung        | [pɔŋː]                  | [pɔŋː]                | portemonnee,<br>beurs | purse,<br>wallet       | 213                     |
| flue        | [flu:ə]                 | [flu:ə]               | vlieg                 | fly                    | 240                     |
| lyn         | [lyːn]                  | [lyːn]                | bliksem               | lightning              | 352                     |
| bue         | [bu:ə]                  | [bu:ə]                | boog                  | bow                    | 431                     |
| pose        | [po:sə]                 | [po:sə]               | zak, tas              | bag                    | 778                     |
| kjole       | [kjo:lə]                | [kjo:lə]              | jurk, japon           | dress                  | 1059                    |
| avis        | [aviːs]                 | [aviːs]               | krant                 | newspaper              | 2449                    |
| seng        | [sɛŋː]                  | [sɛŋː]                | bed                   | bed                    | 3665                    |

**Appendix B: Full set of pictures**

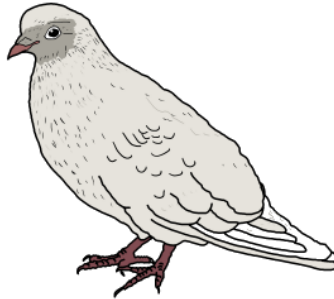
**Figure 4a**

*Pictures of rule-based words on first testing day*

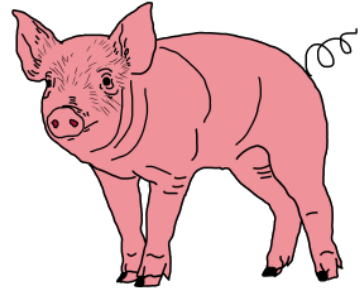
dyne



due



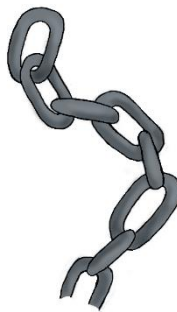
gris



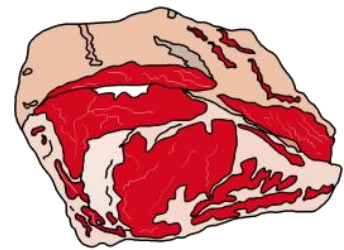
gade



kæde



kød



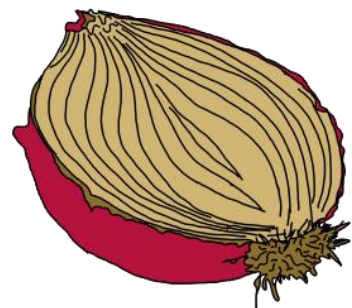
nøgle



pige



løg

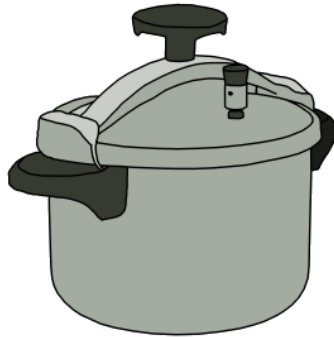


**Figure 4b**  
*Pictures of rule-based words on second testing day*

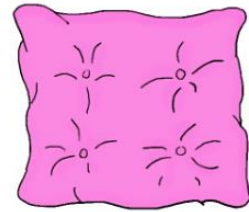
dråbe



gryde



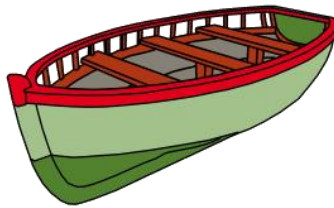
pude



flod



båd



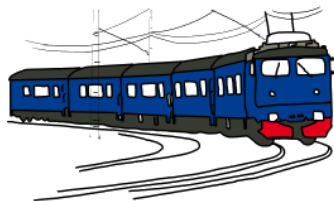
læge



byge



tog



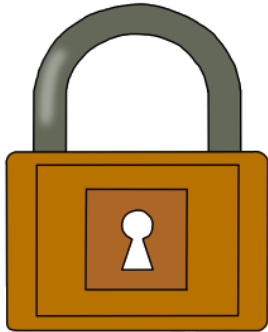
krog



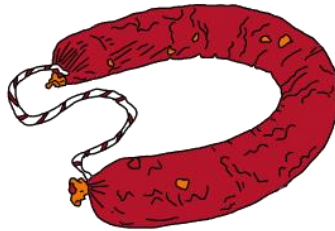
**Figure 4c**

*Pictures of baseline words on first testing day*

lås



pølse



reb



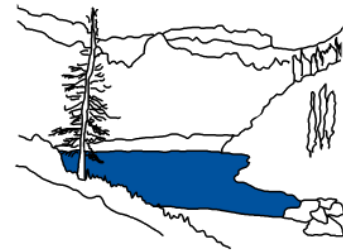
næb



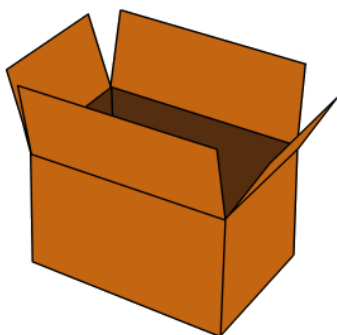
hale



sø



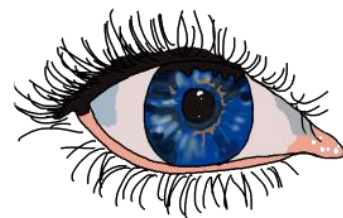
kasse



øl



øje

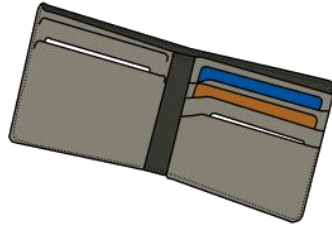


**Figure 4d**  
*Pictures of baseline words on second testing day*

ble



pung



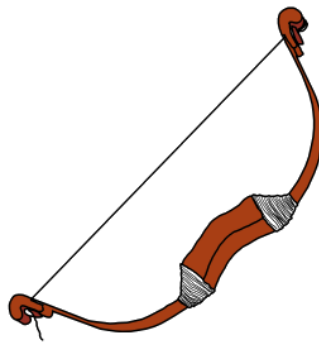
flue



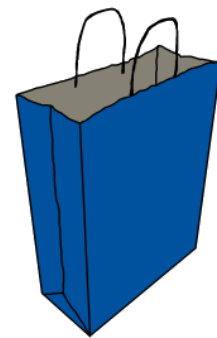
lyn



bue



pose



kjole



avis



seng



### **Appendix C: Naming phase (pre-test)**

Before participants started the learning phase, they were asked to name all of the pictures in Dutch. This was a way to check if it was clear to them which words were depicted. If a picture would be consistently named incorrectly, then it could be removed. Thirty pictures were pretested and should therefore not lead to any difficulties. This was mostly the case. Most of the participants named these correctly and some made small mistakes, like choosing the Dutch word for 'bird' when a pigeon was depicted. A picture that did turn out to be hard, was the picture describing the word 'beer'. This was named correctly by only 16 of the participants. Others described it as a tap (7), beer tap (2), drink (2), lemonade (2) or faucet (1). The pictures that were created for the sake of the experiment were also named correctly most of the time. Except for the pictures that depicted a duvet and a rain shower. The duvet was named a 'towel' (15) or a 'blanket' (15), while the rain shower was described as 'rain' (28), 'rain cloud' (1) and only once as 'rain shower' (1).

In total, participants made few mistakes in naming the words and if they did make a mistake, they usually chose a semantically related word. Items that were named incorrectly were not removed from the analysis. This had several reasons. The main reasons were the relative small amount of data and the statistical analysis. An ANOVA cannot deal with missing values all that well, so it would be necessary to remove an item for all participants even when it was named incorrectly by only a small number of participants. Moreover, even when participants learn the wrong semantic label for a picture, they are still learning to match a Danish word to a picture. There is no reason to assume orthographic effects to be different when participants choose an alternative label.

## **Appendix D: Instructions and questions**

This appendix includes a detailed listing of all the separate parts of the three testing days. Moreover, it includes all instructions and examples of the matching question and the slider that participants saw when they were asked to estimate their percentages correct. The language background questionnaire was added in full as well. All instructions and questions are presented in the same way as they were presented to the participants, which was in Dutch. In order to provide an idea of the full content of the experiment to all readers, the content is listed below in English and the headers are in English for each part.

### **1. Content of first testing day**

#### 1.1 Selection questions

#### 1.2 Skip message

#### 1.3 Information and request for permission

#### 1.4 Instructions for first testing day

#### 1.5 Instructions before naming phase (pre-test)

-Naming phase (pre-test) for first 18 words

#### 1.6 Instructions before word learning phase for first 18 words

-Word learning phase

-all 18 word-picture pairs in random order, three times

#### 1.7 First break

- all 18 word-picture pairs in random order, four times

#### 1.8 Second break

-all 18 word-picture pairs in random order, three times

#### 1.9 Instructions for first test session for first 18 words

-Testing phase

-practise items: one matching pair, one mismatching pair

#### 1.10 Matching question (after each pair)

#### 1.11 Warning that real testing phase begins

-all 18 words and pictures once in a matching pair and once in a mismatching pair in a random order with a matching question after each pair

#### 1.12 Estimation of percentage correct

#### 1.13 End of first testing day

### **1.1 Selection questions**

Welkom bij dit onderzoek!

Voordat je verdergaat, moet je eerst wat vragen beantwoorden. Zo kunnen we controleren of jij geschikt bent om mee te doen.

Ben je tussen de 18 en 30 jaar oud?

- Ja
- Nee

Ben je een moedertaalspreker van het Nederlands?

- Ja
- Nee

Heb je ervaring met het leren van de Deense taal?

- Ja
- Nee

### **1.2 Skip message**

In dit onderzoek wordt gekeken naar jongvolwassen moedertaalsprekers van het Nederlands, zonder voorkennis van het Deens. Helaas kun jij dus niet deelnemen aan dit onderzoek.

### **1.3 Information and request for permission**

#### **Onderzoek: Deense woorden leren**

Fijn dat je mee wil doen aan dit onderzoek!


#### Informatie en toestemming

Je bent uitgenodigd om mee te doen aan een onderzoeksproject in het kader van een masterscriptie op het gebied van de taalwetenschap. Dit onderzoek wordt uitgevoerd door Sanne de Kroon, masterstudent aan de Radboud Universiteit. Dit gebeurt onder supervisie van Paula Fikkert.

#### Wat wordt er van je verwacht?

Meedoen aan het onderzoek houdt in dat je drie online vragenlijsten gaat invullen, verdeeld over drie sessies. Tussen de lijsten zit minimaal 1 dag en maximaal 3. In de eerste sessie van het onderzoek zul je Deense woorden leren en je kennis van deze woorden zal getest worden. Hetzelfde geldt voor de tweede sessie. In de derde sessie leer je geen nieuwe woorden, maar wordt alleen je kennis getest. Ook zal er om aanvullende gegevens gevraagd worden over jouw (taal)achtergrond. Het invullen van de vragenlijst kost ongeveer 25 minuten. Dit geldt voor sessie 1 en 2. Sessie 3 zal maar 10 minuten duren.

#### Vrijwilligheid

Je doet vrijwillig mee aan dit onderzoek. Daarom kun je op elk moment stoppen met de vragenlijst en je toestemming intrekken. Ook na het voltooien van één of meerdere sessies, kun je je toestemming nog intrekken. Je hoeft niet aan te geven waarom je stopt. Je gegevens zullen dan worden verwijderd. Zelfs tot twee weken na deelname kun je je onderzoeksgegevens laten verwijderen. Dit kun je doen door een mail te sturen naar .

#### Vertrouwelijkheid van de gegevens

De gegevens die worden verzameld tijdens dit onderzoek, zullen door wetenschappers gebruikt worden voor artikelen en presentaties. Gegevens zullen hiervoor worden geanonimiseerd en zullen daarna op geen enkele manier herleidbaar zijn tot de persoon die de vragenlijst heeft ingevuld. De anoniem gemaakte onderzoeksgegevens zijn tenminste 10 jaar beschikbaar voor andere wetenschappers. Alle onderzoeksgegevens worden bewaard op beveiligde wijze volgens de richtlijnen van de Radboud Universiteit.

Let op: aan het einde van deze vragenlijst en aan het einde van de volgende lijsten, zal jouw e-mailadres gevraagd worden. Dit is nodig om jou de volgende lijst te kunnen toesturen,

maar ook om jouw gegevens van de verschillende lijsten te combineren. Wanneer je de laatste lijst afmaakt of wanneer je jouw toestemming intrekt, dan zal jouw mailadres direct verwijderd worden.

### Vragen of klachten?

Voor meer informatie over het onderzoek of voor specifieke vragen, kun je contact opnemen met Sanne de Kroon via het e-mailadres [REDACTED]. Ook voor klachten over dit onderzoek kun je bij haar terecht.

Door op "Ik geef toestemming" te klikken, geef je aan dat je:

- bovenstaande informatie hebt gelezen
- vrijwillig meedoet aan dit onderzoek
- minimaal 18 jaar oud bent

Als je niet wil meedoen aan dit onderzoek, druk dan op "Ik geef geen toestemming".

- Ik geef toestemming
- Ik geef geen toestemming

### ***1.4 Instructions for first testing day***

#### **Instructies sessie 1**

Vandaag ga je 18 Deense woorden leren. Je ziet steeds een afbeelding en je hoort het Deense woord dat afgebeeld is. De afbeelding blijft vier seconden in beeld. Hierna komt direct de volgende afbeelding met het bijbehorende woord. Je moet dus goed blijven opletten!

Zorg ervoor dat je in een rustige ruimte zit, waar je het komende kwartier niet gestoord wordt. Controleer of het geluid van je apparaat hard genoeg staat en gebruik indien mogelijk een koptelefoon. Je krijgt straks de gelegenheid om je geluid te testen.

Er zitten twee pauzes in het experiment. Heb je geen pauze nodig? Dan mag je ook direct verdergaan.

Nadat je de woorden geleerd hebt, volgt een korte overhoring. Hierin wordt gecontroleerd of je de woorden nu ook echt kent. Het is niet erg als je niet alle woorden kent. Het gaat erom dat je je best doet!

Succes!

### ***1.5 Instructions before naming phase (pre-test)***

Hieronder zie je afbeeldingen van de woorden die je vandaag gaat leren. Vul onder iedere afbeelding het woord in, waarvan jij denkt dat het erbij hoort. Zo kunnen we controleren of de afbeeldingen duidelijk zijn.

### ***1.6 Instructions before word learning phase for first 18 words***

Straks ga je echt beginnen met het leren van de woorden.<sup>8</sup> Let op: de woorden volgen elkaar automatisch op. Je kunt tussendoor dus niet zomaar even pauze nemen. De eerste pauze is pas na ongeveer 5 minuten.

Bekijk de checklist dus goed, voordat je verdergaat. Zo weet je zeker dat je goed voorbereid bent. Hieronder kun je testen of je geluid hard genoeg staat.

[Audio file of the word 'test', played at the same volume as the material]

#### Checklist

- Ik zit in een rustige ruimte
- Het geluid van mijn apparaat staat hard genoeg
- Mijn apparaat heeft genoeg stroom
- Ik ben er klaar voor

### **1.7 First break**

#### **Pauze**

Je bent al op een derde! Ken je de woorden al een beetje?

Nu mag je een paar minuten pauze nemen, als je wil. Zodra je op het pijltje klikt, gaat het experiment weer automatisch verder.

### **1.8 Second break**

#### **Pauze**

Je hebt nu twee derde gehad. Goed bezig!

Nu mag je opnieuw een paar minuten pauze nemen, als je wil. Zodra je op het pijltje klikt, gaat het experiment weer automatisch verder.

### **1.9 Instructions for first test session for first 18 words**

Je bent klaar met het leren van de woorden. Straks wordt er gecontroleerd of je alle woorden nu kent.

Je zult een afbeelding zien en een woord horen. Na 4 seconden zal de afbeelding weer verdwijnen. Dan moet je aangeven of de afbeelding en het woord bij elkaar hoorden. Als je denkt dat deze bij elkaar hoorden, klik dan op 'ja'. Als je denkt dat deze niet bij elkaar hoorden, klik dan op 'nee'. Iedere afbeelding zie je twee keer: één keer in combinatie met het juiste woord en één keer met het verkeerde woord.

Voordat je begint, mag je eerst twee keer oefenen. Je ontvangt geen feedback.

Succes!

---

<sup>8</sup> For participants in the AF & OF condition, this text included the warning: "Tijdens het leren zie je het geschreven woord, maar tijdens de overhoring niet. Je moet dus ook goed luisteren" (During learning you see the written word, but this will not be the case during testing. So you need to listen carefully as well.)

### **1.10 Matching question**

Paste het woord dat je net hoorde, bij de afbeelding die je zag?

- Ja
- Nee

### **1.11 Warning that real testing phase begins**

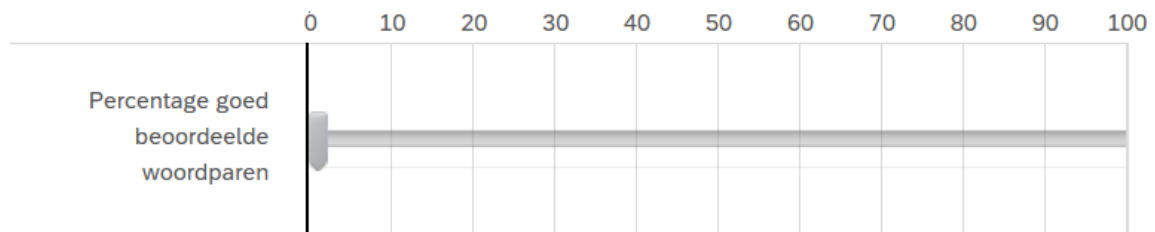
Dit waren de oefenvragen. Nu ga je echt aan de slag.

Succes!

### **1.12 Estimation of percentage correct**

Je bent klaar met de overhoring. Hoeveel procent van de woordparen denk je goed te hebben beoordeeld?

Met 'goed' beoordeelde woordparen bedoelen we zowel paren waar het woord en de afbeelding bij elkaar hoorden en je 'ja' hebt aangegeven, als de paren waar deze niet bij elkaar hoorden en je 'nee' hebt aangegeven.



### **1.13 End of first testing day**

Dit is het einde van de eerste sessie. Om deel te kunnen nemen aan de tweede sessie, moet je hieronder je mailadres invullen. Dan krijg je morgenochtend de link voor de tweede sessie toegestuurd. Er zijn in totaal drie sessies.

Als je morgen geen tijd hebt, dan is dat geen probleem. Zorg er wel voor dat je binnen drie dagen deelneemt aan de tweede sessie. Anders zijn je gegevens niet bruikbaar.

Tot snel!

---

## **2. Content of second testing day**

### 2.1 Instructions for second testing day

#### 2.2 Instructions for second test session of first 18 words

##### -Testing phase

-practise items: one matching pair, one mismatching pair

##### 2.3 Matching question (after each pair)

##### 2.4 Warning that real testing phase begins

-all 18 words and pictures once in a matching pair and once in a mismatching pair in a random order with a matching question after each pair

#### 2.5 Estimation of percentage correct

#### 2.6 Instructions before second naming phase (pre-test)

##### -Naming phase (pre-test) for second 18 words

#### 2.7 Warning for second word learning phase

##### -Word learning phase

-all 18 word-picture pairs in random order, three times

##### 2.8 First break

- all 18 word-picture pairs in random order, four times

##### 2.9 Second break

-all 18 word-picture pairs in random order, three times

#### 2.10 Instructions for first test session of the second 18 words

##### -Testing phase

-practise items: one matching pair, one mismatching pair

##### 2.11 Matching question (after each pair)

##### 2.12 Warning that real testing phase begins

-all 18 words and pictures once in a matching pair and once in a mismatching pair in a random order with a matching question after each pair

#### 2.13 Estimation of percentage correct

#### 2.14 End of second testing day

### ***2.1 Instructions for second testing day***

#### **Instructies sessie 2**

Leuk dat je er weer bent!

Deze sessie bestaat uit drie verschillende delen. Eerst word je getest op de woorden van de eerste sessie. Vervolgens zul je 18 nieuwe woorden leren. Ook op deze woorden zul je getest worden. In totaal zul je ongeveer 25 minuten bezig zijn.

Succes!

### ***2.2 Instructions for second test session of first 18 words***

Straks wordt er gecontroleerd of je alle woorden van de vorige sessie nog kent.

Je zult opnieuw een afbeelding zien en een woord horen. Je moet weer aangeven of de afbeelding en het woord bij elkaar horen. Als je denkt dat deze bij elkaar horen, klik dan op

'ja'. Als je denkt dat deze niet bij elkaar horen, klik dan op 'nee'. Je mag weer eerst twee keer oefenen.

Voordat je begint, willen we je vragen om de checklist nogmaals te doorlopen. Hieronder kun je weer testen of je geluid hard genoeg staat.

[Audio file of the word 'test', played at the same volume as the material]

#### Checklist

- Ik zit in een rustige ruimte
- Het geluid van mijn apparaat staat hard genoeg
- Mijn apparaat heeft genoeg stroom
- Ik ben er klaar voor

### **2.3 Matching question**

Paste het woord dat je net hoorde, bij de afbeelding die je zag?

- Ja
- Nee

### **2.4 Warning that real testing phase begins**

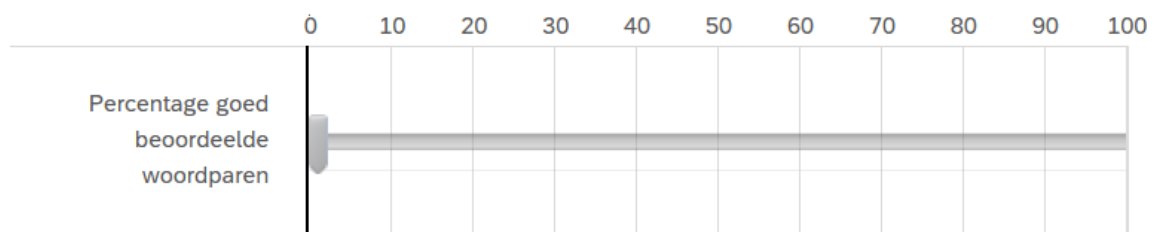
Dit waren de oefenvragen. Nu ga je echt aan de slag.

Succes!

### **2.5 Estimation of percentage correct**

Je bent klaar met de overhoring. Hoeveel procent van de woordparen denk je goed te hebben beoordeeld?

Met 'goed' beoordeelde woordparen bedoelen we zowel paren waar het woord en de afbeelding bij elkaar hoorden en je 'ja' hebt aangegeven, als de paren waar deze niet bij elkaar hoorden en je 'nee' hebt aangegeven.



### **2.6 Instructions before second naming phase (pre-test)**

Straks ga je weer 18 nieuwe woorden leren. Hieronder zie je afbeeldingen van de woorden die je vandaag gaat leren. Vul onder iedere afbeelding het woord in, waarvan jij denkt dat het erbij hoort. Zo kunnen we controleren of de afbeeldingen duidelijk zijn.

### **2.7 Warning for second word learning phase**

Let op: na dit blok begin je echt met het leren van de nieuwe woorden. De woorden volgen elkaar weer automatisch op. Dus zorg dat je klaar zit, voordat je verder klikt.

Succes!

### **2.8 First break**

#### **Pauze**

Je bent al op een derde! Ken je de woorden al een beetje?

Nu mag je een paar minuten pauze nemen, als je wil. Zodra je op het pijltje klikt, gaat het experiment weer automatisch verder.

### **2.9 Second break**

#### **Pauze**

Je hebt nu twee derde gehad. Goed bezig!

Nu mag je opnieuw een paar minuten pauze nemen, als je wil. Zodra je op het pijltje klikt, gaat het experiment weer automatisch verder.

### **2.10 Instructions for first test session of the second 18 words**

Je bent klaar met het leren van de nieuwe woorden. Nu volgt er een overhoring over deze woorden.

Je zult weer een afbeelding zien en een woord horen. Je moet aangeven of de afbeelding en het woord bij elkaar horen. Als je denkt dat deze bij elkaar horen, klik dan op 'ja'. Als je denkt dat deze niet bij elkaar horen, klik dan op 'nee'. Je begint weer met twee keer oefenen.

### **2.11 Matching question**

Paste het woord dat je net hoorde, bij de afbeelding die je zag?

- Ja
- Nee

### **2.12 Warning that real testing phase begins**

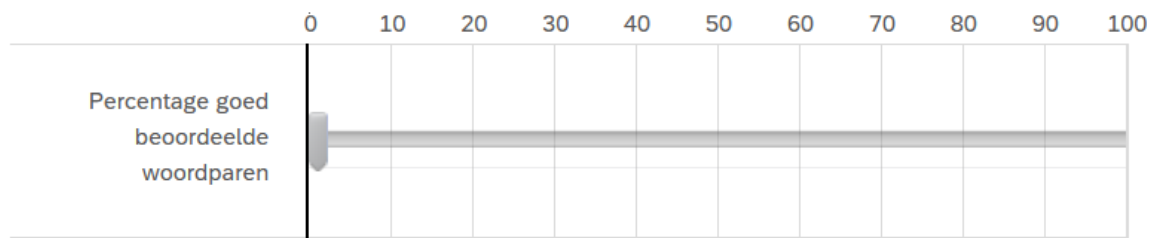
Dit waren de oefenvragen. Nu ga je echt aan de slag.

Succes!

### **2.13 Estimation of percentage correct**

Je bent klaar met de overhoring. Hoeveel procent van de woordparen denk je goed te hebben beoordeeld?

Met 'goed' beoordeelde woordparen bedoelen we zowel paren waar het woord en de afbeelding bij elkaar hoorden en je 'ja' hebt aangegeven, als de paren waar deze niet bij elkaar hoorden en je 'nee' hebt aangegeven.



### **2.14 End of second testing day**

Dit is het einde van de tweede sessie. Om deel te kunnen nemen aan de derde sessie, moet je hieronder je mailadres invullen. Dan krijg je morgenochtend de link voor de derde sessie toegestuurd. Dat zal de laatste sessie zijn.

Als je morgen geen tijd hebt, dan is dat geen probleem. Zorg er wel voor dat je binnen drie dagen deelneemt aan de derde sessie. Anders zijn je gegevens niet bruikbaar.

Let op: je mailadres wordt tevens gebruikt om je gegevens te koppelen. Vul dus hetzelfde mailadres in als bij de vorige sessie.

Tot snel!

---

### **3. Content of third testing day**

3.1 Instructions for third testing day

3.2 Instructions for second test session of second 18 words

-Testing phase

-practise items: one matching pair, one mismatching pair

3.3 Matching question (after each pair)

3.4 Warning that real testing phase begins

-all 18 words and pictures once in a matching pair and once in a mismatching pair in a random order with a matching question after each pair

3.5 Estimation of percentage correct

-Language history questionnaire

3.6 General questions

3.7 Language background questions

3.8 Manipulation check

3.9 End of third testing day

#### ***3.1 Instructions for third testing day***

##### **Instructies sessie 3**

Dit is alweer de laatste sessie! Vandaag ga je geen nieuwe woorden meer leren. Je wordt wel nog getest op de woorden van de vorige sessie. Daarna mag je wat vragen over jezelf beantwoorden en over jouw taalachtergrond. Dit alles zal in totaal ongeveer 10 minuten duren.

Veel plezier!

#### ***3.2 Instructions for second test session of second 18 words***

Straks wordt er gecontroleerd of je alle woorden van de vorige sessie nog kent.

Je zult wederom een afbeelding zien en een woord horen. Je moet aangeven of de afbeelding en het woord bij elkaar horen. Als je denkt dat deze bij elkaar horen, klik dan op 'ja'. Als je denkt dat deze niet bij elkaar horen, klik dan op 'nee'. Je mag weer eerst twee keer oefenen.

Voordat je begint, willen we je vragen om de checklist nogmaals te doorlopen. Hieronder kun je weer testen of je geluid hard genoeg staat.

[Audio file of the word 'test', played at the same volume as the material]

##### **Checklist**

- Ik zit in een rustige ruimte
- Het geluid van mijn apparaat staat hard genoeg
- Mijn apparaat heeft genoeg stroom
- Ik ben er klaar voor

### 3.3 Matching question

Paste het woord dat je net hoorde, bij de afbeelding die je zag?

- Ja
- Nee

### 3.4 Warning that real testing phase begins

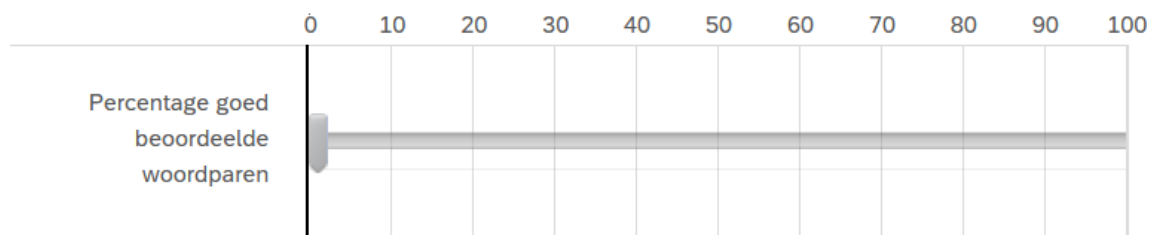
Dit waren de oefenvragen. Nu ga je echt aan de slag.

Succes!

### 3.5 Estimation of percentage correct

Je bent klaar met de overhoring. Hoeveel procent van de woordparen denk je goed te hebben beoordeeld?

Met 'goed' beoordeelde woordparen bedoelen we zowel paren waar het woord en de afbeelding bij elkaar hoorden en je 'ja' hebt aangegeven, als de paren waar deze niet bij elkaar hoorden en je 'nee' hebt aangegeven.



### 3.6 General questions

Zo meteen krijg je nog een aantal algemene vragen over jezelf.

Hoe oud ben je?

\_\_\_\_\_

Wat is je gender?

- Man
- Vrouw
- Overig
- Wil ik niet zeggen

Wat is je huidige of hoogst genoten opleiding?

- Vmbo
- Havo
- Vwo
- Mbo
- Hbo
- Universiteit

### 3.7 Language background questions

Nu komt er een aantal vragen over jouw taalachtergrond. Probeer deze zo goed mogelijk te beantwoorden. Het is niet erg als je niet alle antwoorden precies weet. Het gaat om een zo goed mogelijke inschatting. Waarschijnlijk zul je niet overal iets in hoeven te vullen, dit is geen probleem. Je mag vragen leeg laten als ze niet op jou van toepassing zijn.

De antwoorden die je geeft, komen in het volgende blok automatisch in de tabellen terecht. Let er dus op dat je alles goed invult.

Vul hieronder in welke vreemde talen je spreekt. Nederlands hoef je dus niet in te vullen. Later word je om meer details gevraagd.

- Eerste taal  
\_\_\_\_\_
- Tweede taal  
\_\_\_\_\_
- Derde taal  
\_\_\_\_\_

Spreek je nog meer talen? Vul deze dan hier in.

\_\_\_\_\_

Heb je wel eens langer dan 3 maanden in het buitenland gewoond? Zo ja, geef hieronder alle landen aan waar je gewoond hebt.

- Eerste land  
\_\_\_\_\_
- Tweede land  
\_\_\_\_\_
- Derde land  
\_\_\_\_\_

Heb je in nog meer landen gewoond? Vul deze dan hieronder in.

\_\_\_\_\_

Geef voor iedere taal hieronder aan hoe oud je was toen je deze taal begon te gebruiken op het gebied van luisteren, spreken, lezen en schrijven. Geef ook de totale gebruiksjaren aan. (Het kan zijn dat je een taal een aantal jaar niet gebruikt hebt. Zorg ervoor dat je deze jaren niet meetelt in het totaal.)

|               | Luisteren | Spreken | Lezen | Schrijven | Gebruiksjaren |
|---------------|-----------|---------|-------|-----------|---------------|
| [Eerste taal] |           |         |       |           |               |
| [Tweede taal] |           |         |       |           |               |
| [Derde taal]  |           |         |       |           |               |

Geef nu voor iedere taal hieronder aan hoe goed jouw beheersing van de taal is.

|                      | <b>Ze<br/>er<br/>beperkt</b> | <b>Beperkt</b>        | <b>Matig</b>          | <b>Middel-<br/>matig</b> | <b>Goed</b>           | <b>Ze<br/>er<br/>goed</b> | <b>Perfect</b>        |
|----------------------|------------------------------|-----------------------|-----------------------|--------------------------|-----------------------|---------------------------|-----------------------|
| <b>[Eerste taal]</b> | <input type="radio"/>        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>    | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> |
| <b>[Tweede taal]</b> | <input type="radio"/>        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>    | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> |
| <b>[Derde taal]</b>  | <input type="radio"/>        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>    | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> |

Geef vervolgens per taal aan hoe vaak je deze taal ongeveer gebruikt.

|                      | <b>Zelden</b>         | <b>Alleen<br/>op<br/>vakantie</b> | <b>Een<br/>aantal<br/>keer per<br/>jaar</b> | <b>Eens<br/>per<br/>maand</b> | <b>Eens<br/>per<br/>week</b> | <b>Meerdere<br/>keren per<br/>week</b> | <b>Iedere<br/>dag</b> |
|----------------------|-----------------------|-----------------------------------|---|-------------------------------|------------------------------|--|-----------------------|
| <b>[Eerste taal]</b> | <input type="radio"/> | <input type="radio"/>             | <input type="radio"/>                       | <input type="radio"/>         | <input type="radio"/>        | <input type="radio"/>                  | <input type="radio"/> |
| <b>[Tweede taal]</b> | <input type="radio"/> | <input type="radio"/>             | <input type="radio"/>                       | <input type="radio"/>         | <input type="radio"/>        | <input type="radio"/>                  | <input type="radio"/> |
| <b>[Derde taal]</b>  | <input type="radio"/> | <input type="radio"/>             | <input type="radio"/>                       | <input type="radio"/>         | <input type="radio"/>        | <input type="radio"/>                  | <input type="radio"/> |

Spreek je nog meer talen? Beschrijf hieronder dan kort wat over jouw beheersing en gebruik van die talen.

---

Als je ooit langer dan 3 maanden in het buitenland hebt gewoond, vul dan voor de landen hieronder in: hoeveel maanden je daar woonde en welke taal je het meest gebruikte. Indien je ook andere talen gebruikte, vul dan in welke dit waren.

|                      | <b>Lengte verblijf in<br/>maanden</b> | <b>Meest gebruikte<br/>taal</b> | <b>Andere gebruikte<br/>talen</b> |
|----------------------|---------------------------------------|---------------------------------|-----------------------------------|
| <b>[Eerste land]</b> |                                       |                                 |                                   |
| <b>[Tweede land]</b> |                                       |                                 |                                   |
| <b>[Derde land]</b>  |                                       |                                 |                                   |

Heb je in nog meer landen gewoond? Beschrijf dan hieronder kort wat over hoe lang je in deze landen gewoond hebt en welke talen je daar gebruikte.

---

Wil je verder nog iets kwijt over jouw taalachtergrond? Vul dat dan hieronder in.

---

### **3.8 Manipulation check**

Is je iets opgevallen aan de woorden die je moest leren in dit onderzoek? Met andere woorden: heb je enig idee wat het doel is van dit onderzoek? Vul hieronder je vermoedens in.

---

### **3.9 End of third testing day**

Dit is het einde van de derde en laatste sessie. Bedankt voor het meedoen!

We willen je vragen om nog één laatste keer je mailadres hieronder in te vullen. Dit is nodig om je gegevens van deze sessie te kunnen koppelen aan de eerdere sessies. Vul dus weer hetzelfde mailadres in als bij de vorige twee sessies.

Tot ziens!

---

## Appendix E: Mismatching word-picture pairs

**Table 7**

*Mismatching word-picture pairs*

| <b>Picture</b> | <b>Auditory word</b> |
|----------------|----------------------|
| dyne           | kasse                |
| due            | øl                   |
| gris           | øje                  |
| gade           | dyne                 |
| kæde           | due                  |
| kød            | gris                 |
| nøgle          | gade                 |
| pige           | kæde                 |
| løg            | kød                  |
| lås            | nøgle                |
| pølse          | pige                 |
| reb            | løg                  |
| næb            | lås                  |
| hale           | pølse                |
| sø             | reb                  |
| kasse          | næb                  |
| øl             | hale                 |
| øje            | sø                   |
| dråbe          | kjole                |
| gryde          | avis                 |
| pude           | seng                 |
| flod           | dråbe                |
| båd            | gryde                |
| læge           | pude                 |
| byge           | flod                 |
| tog            | båd                  |
| krog           | læge                 |
| ble            | byge                 |
| pung           | tog                  |
| flue           | krog                 |
| lyn            | ble                  |
| bue            | pung                 |
| pose           | flue                 |
| kjole          | lyn                  |
| avis           | bue                  |
| seng           | pose                 |

## Appendix F: All estimated percentages correct

Table 8 shows the estimated percentages correct per group for each test session. These are on average quite high (74%-89%) but they do not show the same ceiling effect as the actual percentages correct. Moreover, there is much more variation within these data, which leads to the larger standard variations here.

**Table 8**

*Estimated percentage correct and standard deviations (in parentheses), split by group, word set and test session*

| <b>Group</b> | <b>Word set 1</b> |                  |
|--------------|-------------------|------------------|
|              | <b>Session 1</b>  | <b>Session 2</b> |
| AF only      | 85% (14.47)       | 79% (16.06)      |
| AF & OF      | 84% (11.19)       | 83% (15.89)      |
| <b>Group</b> | <b>Word set 2</b> |                  |
|              | <b>Session 1</b>  | <b>Session 2</b> |
| AF only      | 75% (22.48)       | 74% (18.12)      |
| AF & OF      | 91% (12.90)       | 89% (10.66)      |