Why the Child Cannot Be Happy

Joakim Andreassen Vea

Master thesis in linguistics (LING350)

Department of Linguistic, Literary and Aesthetic Studies
University of Bergen

2017

To all the squirrels who shared my life.

Contents

List of Figures

List of Tables

Sammendrag

Intetkjønnsformer av adjektiver i norsk og svensk dannes ved en regelmessig og produktiv regel der man legger et intetkjønnsmorfem /-t/ til stammen av adjektivet. Det finnes unntak i at enkelte adjektiv ikke følger denne regelen og videre synes å ikke ha noen brukbar intetkjønnsform i det hele tatt. Denne masteroppgaven undersøker disse manglende intetkjønnsformene ved bruk av "lexical decision"-oppgave (Meyer and Schvaneveldt, 1971) der manglende intetkjønnsformer ble brukt som primingstimulus for sine stammeformer. Maskert priming (Forster, 1998) ble brukt som metode. Eksperimentet måler i hvor stor grad deltakere kobler sammen intetkjønnsformer av adjektiv med stammeformer som ble brukt som målstimulus i oppgaven.

Adjektivene som ikke har intetkjønnsformer har likheter seg imellom i fonologisk form og semantisk innhold, og utgjør en forholdsvis liten men tilsynelatende produktiv gruppe. Adjektiver med blokkerte intetkjønnsformer i eksperimentet ble kontrollert opp mot adjektiver med regelmessige intetkjønnsformer og pseudoord med liknende fonologisk mønster.

30 morsmålstalende av norsk ble valgt ut til å delta i eksperimentet. Resultatene fra eksperimentet viste statistisk signifikante forskjeller mellom stimulusgruppene. Regelmessige intetkjønnsformer og konstruerte intetkjønnsformer for pseudoord forårsaket raskere reaksjonstider hos deltakere, mens de blokkerte intetkjønnsformene hadde ingen effekt på reaksjonstid. Dette peker på at fasilitering på grunn av formlikhet blir motvirket av en hemmende effekt som reaksjon på ulovlige intetkjønnsformer. En tendens til å oftere feilklassifisere adjektiver med manglende intetkjønnsformer som ikke-ord, også i de lovlige stammeformene, var også statistisk signifikant. Dette peker videre på mer kompleks prosessering av disse adjektivene. Analysen gjorde bruk av blandede regresjonsmodeller (cf. Baayen et al., 2008) med faste effekter for stimulusgruppe og priming, og tilfeldige effekter for individuelt stimulus og deltaker.

En gruppe med andrespråkstalere (L2) deltok også i eksperimentet. Denne gruppen bestod av 11 deltakere. Der var en del informasjonstap hos L2-gruppen på grunn av vanskelighetsgraden i oppgaven. Resultatene peker i samme retning som for morsmålstalere (L1), som viser at også talere med mindre eksponering for språket lærer legger merke til de manglende intetkjønnsformene.

De manglende intetkjønnsformene i norsk og svensk har ingen reparasjon. Man kan bytte ut med andre adjektiv, eller bytte substantivet med et som ikke krever samsvar med intetkjønn. De manglende formene har videre implikasjoner for grammatisk kjønn i norsk og svensk. Likheten i det semantiske innholdet antyder at intetkjønn i disse språkene har en korrelasjon med semantiske egenskaper som ikke-individ og ikke-animat (Johansson, 2003). Disse egenskapene forutsetter et visst nivå av bevissthet. Dette kan sammenlignes med semantiske effekter i samsvarsbøying der intetkjønnsformer av adjektiv brukes for å modifisere substantiv med spesifikke semantiske egenskaper men er av grammatisk hankjønn eller hunkjønn (cf. Enger, 2013), og det er mulig at disse speiler en større tendens i kontinentalskandinaviske språk mot semantisering av intetkjønn (Josefsson, 2014).

Abstract

Neuter forms of adjectives in Norwegian and Swedish are formed through the regular and productive process of suffixing the neuter morpheme /-t/ to the adjective stem. A number of adjectives however do not follow this rule and appear to have no usable neuter form at all. The master thesis investigates these missing neuter forms through the use of a lexical decision task (Meyer and Schvaneveldt, 1971) where expected but missing neuter forms of adjectives where used as priming stimulus for their stem forms. Masked priming (Forster, 1998) was the method used. The experiment measures how well participants can connect neuter forms to their stem forms which were used as target stimuli for the lexical decision.

The adjectives that do not form neuters share commonalities in phonological form and in semantic content, and constitute a fairly small but seemingly productive class. Adjectives with blocked neuter gender forms in the experiment were controlled against adjectives that regularly form neuters and pseudowords with corresponding neuters of similar phonological form.

30 native speakers of Norwegian were selected to participate in the experiment. Results from the experiment showed statistically significant differences between the stimulus classes with regular neuters and constructed neuters for pseudowords causing facilitation in reaction times when used to prime their corresponding common gender forms, while the blocked neuter forms caused no such facilitation. This suggests any facilitation due to form-similarity is canceled out by an inhibitory process in response to ill-formed neuters. A tendency to more frequently misclassify adjectives with missing neuter forms as nonwords, even in their well-formed stem forms, was also statistically significant and points towards more complex processing of these adjectives. The analysis made use of linear mixed effects models (cf. Baayen et al., 2008), where stimulus class and the priming condition were treated as fixed effects and participants and stimulus items were treated as random effects.

An second-language group (L2) was also participated in the experiment, consisting of a total of 11 participants. There was some loss of information for the L2 group due to the difficulty of the task. Results point in the same direction as for the native group (L1), which provides support for a view that the neuter gaps are recognized also by speakers with limited levels of exposure to the language.

The ill-formed neuters in Norwegian and Swedish have no repair. Different adjectives must be used instead, or the noun can be replaced with a noun that does not require neuter agreement. The missing neuter forms have further implications for grammatical gender in Norwegian and Swedish. The commonality in their shared semantics suggest that neuter gender in these languages is correlated with certain semantic properties such as non-individuality and non-animacy (Johansson, 2003). These are properties that presume some level of sentience. This can be compared to semantic agreement effects where neuter gender forms of adjectives are used to modify nouns with specific semantic properties but masculine or feminine gender (cf. Enger, 2013), and may reflect a deeper trend towards semanticization of the neuter gender in Mainland Scandinavian (Josefsson, 2014).

1 Introduction

This thesis investigates a case of paradigmatic gaps in Norwegian, although the phenomenon exists in Swedish as well (Petterson, 1990). These gaps have earlier been investigated and described by (Johansson, 2003). It aims to achieve both a description of these gaps, and an explanation for why they occur. Paradigmatic gaps are known by other names in the literature, such as *accidental gaps* or *holes in the pattern* (Crystal, 2003), however, in order to be consistent with Johansson (2003) they will be referred to as paradigmatic gaps in this thesis. Specifically, the gaps investigated in this thesis are a case of missing neuter forms in the morphological paradigm of Norwegian adjectives. This could also be called a morphological gap, to distinguish it from other types of gaps that occur in languages such as *lexical gaps* or *phonological gaps* (Crystal, 2003). An example is provided below.

(1) En sta unge.
A.C stubborn.C child[MASC].
'A stubborn child'.

(2) Et *statt barn.
A.N stubborn.N child[NEUT].
'A stubborn child.'

The gaps being confined to neuter forms means the problem is also one of gender agreement. The removal of the form occurs when the adjective is required to agree in gender with the noun being modified. The suggestion is that there is a semantic dimension to neuter gender (Corbett, 1991) in Norwegian and Swedish (Johansson, 2003) and that this property may prevent the missing forms from being well-formed. The inflectional morpheme itself is also strongly associated with neuter gender and occurs in no other inflections.

The thesis describes a psycholinguistic experiment that was used to investigate how speakers process neuter forms of adjectives. It continues a line of inquiry started by Johansson (2003), but there are a number of key differences in design and methods. An elaboration on the possible mechanisms underlying the paradigmatic gaps in Norwegian and Swedish is also provided.

There are a number of key differences from previous work on the topic. Paradigmatic gaps in neuter adjective formation occur in both Norwegian and Swedish and according to similar rules, and there is a fairly extensive literature on the phenomenon as it occurs in Swedish (Johansson, 2003; Petterson, 1990; Raffelsiefen, 2004; Rice, 2007; Fanselow and Féry, 2002). The phenomenon does not appear to have been studied as extensively in Norwegian. While the differences between the languages are small, there are in particular two differences that are relevant to these paradigmatic gaps. There are small differences in phonology between the languages which need to be taken into account when explaining the gaps. Further, while there is considerable overlap, the languages do not fully agree on which adjectives do not form neuters.

Another point of difference with Johansson's work is the use of masked priming (Forster and Davis, 1984b) for the lexical decision task. The masked priming paradigm typically provides different effects of priming such as a weaker effect of repetition (Forster and Davis, 1984b), clearer effects from form-priming (Forster, 1998) and weaker semantic effects (Perea and Gotor, 1997).

Finally, the experiment recruited both L1 and L2 speakers of Norwegian as participants. There may be differences in how the gaps are processed, and they may not be fully formed in L2 speakers who have not had sufficient exposure time to the language.

Purely phonological explanations of why the paradigmatic gaps occur in Norwegian and Swedish encounter problems due to an inability to properly apply the proposed rules to both languages. The proposed interaction between morpho-phonology by Johansson (2003) and Petterson (1990) is more succesful and may further be related to more general properties of the neuter gender in Mainland Scandinavian languages. The thesis will examine the proposed semantic rules and modify them if appropriate in response to information from Norwegian.

Results from the experiment show that there is a difference in experimental participants' reaction to regular neuter forms and missing neuter forms in Norwegian. The results are also further contrasted with the expected neuter forms of *pseudowords* within the same experiment which also found a clear difference between these item classes. Overall the results indicate that the gapped neuter forms in Norwegian cause an inhibitory mechanism to activate in participants. An examination and discussion of neuter gender in Norwegian in light of the results from the experiment is also provided in order to show how these paradigmatic gaps may have arisen in the language. The presence of these gaps along with a number of other factors indicate that grammatical gender is not arbitrary but in fact governed by subtle cues.

2 Background

Gaps in morphology occur in a number of languages. In English, one such example is deverbal nouns derived with the affixes *-al* or *(t)ion* (Kerstens et al., 2001). English has both *recital* and *recitation*, however, only *arrival* and *derivation*. The forms *arrivation* and *derival*, while seemingly permitted by the paradigm, are missing from the language, whatever the cause. This is an example of a morphological paradigm, however it occurs in the derivational morphology of English. What we are looking at in Norwegian is an example of a gap in the inflectional morphology of the language. Such gaps can occur in other languages as well, and are well documented for instance in Greek and Spanish (Sims, 2009). For example genitive plurals in Modern Greek can fail to form under certain conditions

Within the Scandinavian languages, the gaps in the adjective paradigm appear to be most salient in Swedish and Norwegian, but are supposedly also present to some extent in Danish (Johansson, 2003). In all these languages, neuter forms of adjectives are formed by adding the suffix *-t* to the stem of the adjective. This can be illustrated with the following example in

Norwegian:

(3) En kald dag.
A.C cold.C day[MASC].
'A cold day.'

(4) Et kaldt hus.
A.N cold.N house[NEUT].
'A cold house.'

Exceptions to this rule do occur, as some adjectives use the same form in the neuter as in the common or masculine and feminine genders, but the neuter form rule is generally quite predictable and productive. A smaller group of adjectives appear to have no usable form in the neuter however and these are the ones that are interesting to the research described in this thesis. The adjective *glad* is part of this group, as shown in the following is an example:

(5) En glad unge.
A.C happy.C child[MASC].
'A happy child.'

(6) Et *gladt barn.
A.N happy.N child[NEUT].
'A happy child.'

(7) Et *glad barn.
A.N happy.C child[NEUT].
'A happy child.'

The occurrence of adjectives with missing neuter forms in the Scandinavian languages has still not been fully explained. It is not immediately apparent why example 6 is so dubious. Which adjectives cause this conflict is not immediately apparent as well. Native speakers will use other adjectives when necessary. However, missing neuter forms can cause problems for language learners because of the opaqueness of this system. Learners of Norwegian and Swedish do acquire the missing neuter forms, however, as can be observed from how children successfully acquire the "missingness" of these forms. The paradigmatic gap is sustained for new speakers as they learn the language. Since the neuter forms in question are completely missing, it indicates that speakers are able to infer the existence of the gap through information that is explicitly missing.

While this inference may not be formed through the application of some "rule", there are nevertheless conditions that need to be satisfied for the gaps to occur. These conditions will be described and examined in the thesis.

The missing forms are not due to a restriction on position of the adjective either. The gaps occur in both attributive positions, as in the previous examples, and in predicative positions as in the following examples. The examples are in Norwegian, but would also work for Swedish.

(8) Ungen er glad.
Child[MASC]-n.M.SG.DEF COP happy.C.
'The child is happy.'

(9) Barnet er *gladt.
Child[NEUT]-et.N.SG.DEF COP happy.N
'The child is happy'

The neuter forms should be possible according to the rules of adjective inflection in both Norwegian and Swedish, and speakers will typically realize that these forms should be possible. The adjective cannot remain uninflected in neuter position either, so the formation is simply blocked. On the one hand, the neuter form that occurs in 6 and 9 is the expected neuter form, and speakers will generally agree that this is what the neuter form should look like. It is difficult to justify the grammaticality of these bad neuter forms even though they are theoretically possible, and this can be a confusing realization for speakers. Certainly, a non-agreeing form would appear to be just as bad or even worse in these sentences as shown in example 7 which violates the requirement for gender agreement and is if anything even less acceptable.

Note that the example nouns used, *barn* and *unge*, are roughly synonymous. There are small differences though such as *unge* more generally referring to offspring. It is not an uncommon situation when paradigmatic gaps force speakers to use different words that perfect synonyms are not available. This is further elaborated on in later examples.

A number of adjectives in Norwegian, and to a lesser extent in Swedish, can occur in when neuter gender is required without the inflectional ending *-t*. Those adjectives are further restricted by phonological rules however, and the adjectives with missing neuter forms fall outside of that restriction. For the adjectives with bad neuters, the form is removed altogether. The adjective cannot occur in the neuter form at all. In keeping with Johansson's terminology, we may refer to the adjectives that do not form neuters as *problematic* to distinguish them from adjectives that regularly form neuters. The label *problematic* is chosen in reference to the dubious status of their neuter gender forms.

The adjectives with missing forms have overlapping features in Norwegian and Swedish. Their phonological form is similar in both languages. Moreover, the languages typically agree on which exact adjectives have bad neuter forms. The neuter-forming rule is otherwise the same in both languages, but there are small differences in the phonological consequences of this rule that will be elaborated upon in the chapter on adjectives. The differences between the language, while small, should be factored into any attempt at explaining the paradigmatic gaps.

The experiment presented in this thesis makes use of the lexical decision task and priming methodology first introduced by Meyer and Schvaneveldt (1971) in order to investigate whether *priming* a target word with a related word facilitates retrieval of the target word in the experiment. The experiment in this thesis follows in the same tradition within experimental psycholinguistics. The experiment made use of masked priming (Forster and Davis, 1984b)

which was developed to reduce the effect that lexical frequency has on repetition priming. The masked priming paradigm has been shown to have a number of other advantages as well, such as a clearer effect of form-priming (Forster, 1998) and a smaller effect of semantic priming (Perea and Gotor, 1997).

The experiment described herein uses neuter gender forms, or expected neuter gender forms, of adjectives as priming stimulus for the same adjective as target stimulus (cf. Johansson and Torkildsen, 2005). From this we might learn more about how speakers process paradigmatic gaps in real time, ie. whether it leads to facilitation or inhibition in the lexical decision or if there is nothing at all. The experiment further differs from previous experiments (Johansson, 2003) in that both L1 speakers and L2 speakers of Norwegian were recruited to participate in the experiments. We were also interested in how successful L2 speakers are at figuring out these gaps and whether they are internalized in the same manner as for L1.

While the possibility that the neuter gender gaps in Norwegian and Swedish are an accident of phonology has been suggested (Fanselow and Féry, 2002; Rice, 2007), there might also be a semantic component interacting with the morpho-phonological configuration of these forms as suggested by Johansson (2003). The thesis also examines hypotheses and possible explanations for why the paradigmatic gap in adjectives in Norwegian and Swedish occur and continue to occur, with reference to suggested phonological and semantic explanations that have previously been used. The missing neuter forms are obviously linked to gender, which also warrants a discussion on the properties of grammatical gender within the language.

One suggestion is that neuter gender in Norwegian and Swedish implies certain semantic characteristics, two important ones being non-animacy and non-individuality. This line of reasoning was introduced by Petterson (1990) who suggested that the adjectives with missing neuter forms typically describe features that are not directly observable. The thesis will further explore any potential connections this has to other structures within the Norwegian language, and whether there is a tendency for grammatical gender to reflect some property in itself rather than being completely arbitrary. A number of examples of this will be shown in the thesis, as well as an elaboration of previous semantic explanations for the occurence of paradigmatic gaps in Norwegian and Swedish. This small set of adjectives may reflect an overall more general trend in the North Germanic languages in the properties of grammatical gender.

2.1 Paradigmatic gaps and accidental gaps

The difference between paradigmatic gaps and accidental gaps should be highlighted. Simply put, an accidental gap is a case where a hypothetical word or inflectional form could exist in a language but for any number of reasons, it does not. A paradigmatic gap on the other hand is systemic. The inflectional form does not exist in the language through the application of one or more rules.

It is not immediately obvious that the paradigmatic gaps in Norwegian and Swedish in fact are systematic. Several attempts have been made at locating and describing the set of rules that generates the gaps in the paradigm. In the following sections, the proposed rules will be shown and evaluated.

2.2 Definition of terms

For the purpose of this paper, *paradigmatic gap* is restricted to the gaps in morphological inflectional paradigms that occur in Norwegian and Swedish adjectives. Gaps in derivational paradigms are not touched upon in this research into Norwegian and Swedish adjectives.

Adjectives that exhibit no good neuter form in the singular indefinite will also occasionally be referred to as *problematic* adjectives occasionally, and particularly in the data set. This is in contrast to *regular* adjectives which form expected neuter forms.

In the data, pseudowords were referred to as *nonwords* since they are used for the not-a-word decision. While pseudowords and nonwords can mean different things, and pseudowords specifically are a special case of nonwords, the term *nonword* will refer to pseudowords in this thesis unless otherwise noted.

Norwegian, in contrast to Swedish, can have a three-gender system that distinguishes masculine, feminine and neuter genders. In Standard Swedish, this has for the most part been reduced to an distinction between common and neuter genders. A simplification of the grammatical gender system into two categories, common and neuter, can occur in Norwegian Bokmål as well however. Even in forms of Norwegian that distinguish three grammatical genders, there are few adjectives that distinguish masculine and feminine in written Norwegian[1]. Since the difference between masculine and feminine is not typically reflected in adjective morphology in Norwegian, it is simpler to refer to a common gender form of the adjective for the masculine\feminine as distinct from the neuter form.

2.3 Examples

The following list illustrates phonological variation within the group of problematic adjectives: *glad*, *lat*, *kry*, *sky*, *kåt*, *ru*, *slu*, *vred*, *redd*, *staut*, *distré*, *sta tru*, *solid*, *gravid*.

As we can see, problematic adjectives all have the quality that the syllable coda in the stem form ends on either a vowel or the consonants *t* or *d*. Expected neuter forms for these adjectives would be: *gladt*, *latt*, *krytt*, *skytt*, *kått*, *rutt*, *slutt*,*vredt*, *redt*, *stautt*, *distrétt*, *statt*, *trutt*, *solidt*, *gravidt*.

Neuter forms of these adjectives typically do not occur either in written media or in spoken language. The cause of problematicness does not appear to be strictly phonological. The group of problematic adjectives are an exception, as phonologically comparable adjectives such as

[1]In modern written Norwegian forms, only *liten* has distinct forms in the masculine and feminine genders. Older nynorsk and dialects may have more.

våt, *ny*, *blå*, *het* regularly form neuters *vått*, *nytt*, *blått*, *hett*. Johansson (2003) and Petterson (1990) suggest that the rule is more complex than a phonological explanation, and that in order to fully understand these paradigmatic gaps we must consider the semantics of the adjectives. The adjectives with missing neuter forms all tend to describe features which are not directly observable through sensory experience, often internal mental states or properties such as happiness, laziness, pride and so forth. This will be further discussed in following sections.

2.4 Morpho-phonological account

While there is a striking phonological similarity within the class of adjectives with missing neuter forms, the phonological similarity cannot on its own explain the missing forms. There is a much larger group of adjectives within both Norwegian and Swedish that exhibit similar phonology but that readily form neuters. The adjectives with missing neuter forms are a subset of this class. This is illustrated in the following examples:

(10) En våt hund.
A.C wet.C dog[MASC]
'A wet dog.'

(11) Et vått dyr.
A.N wet.N animal[NEUT].
'A cold animal.'

(12) En kåt hund.
A.C horny.C dog[MASC].
'A horny dog.'

(13) Et *kått dyr.
A.N horny.N animal[NEUT]
'A horny animal.'

The examples illustrate both common gender and neuter (or potential neuter) gender forms of the lexical items *våt* and *kåt*. The phonological form is similar, and if the case was that missing neuter forms were caused exclusively by phonological rules, as suggested by Fanselow and Féry (2002), we should expect it to apply to both of these adjectives.

Indeed, the mechanisms suggested by Raffelsiefen (2004) fails upon more in-depth scrutiny. One suggestion is that for the adjectives *lat* and *flat*, the gemination of the word final *-t* in the neuter form and the change in vowel quality from [ɑ:] to [a] is to complicated and therefore the neuter form is selected against. At first glance, this seems to work, but when we look beyond the boundaries of Standard Swedish, the rule falls apart.

When we introduce data from Norwegian, we do however observe that the proposed rule does not hold: *flat* readily forms neuters in Norwegian while *lat* does not. Further, the particular vowel represented by *a* is a special case in Norwegian as its quality typically remains

the same whether it occurs as a short vowel or as a long vowel (Kristoffersen, 2000). This is a point of difference between Norwegian and Swedish, at least as far as the standard languages are concerned. This rule is not absolute however. A number of Norwegian dialects have the same quality split between long and short ”A” as Swedish (Sandvik, 1979), and the extent of the quality split between long and short ”A” can also vary in Swedish dialects (Riad, 2014). The change (or lack thereof) of vowel quality due to long or short allophones of the vowel do not appear to have any consequence for the removal of neuter forms however. This is not a surprising revelation since the allophonic difference in vowel quality is secondary to the difference in vowel quantity in both languages.

Considering how similar the two languages are and how the paradigmatic gaps in adjectives occur in both, we have to ask if it is reasonable to assume that there are different mechanisms causing it in Norwegian and Swedish. Also worth noting here is that the languages demonstrate overlap in the adjectives that lack neuter forms to a large extent. Idiosyncratic behaviour within this group typically occurs where the adjective only exists in one of the languages, although *flat* is a notable case where an adjective in both languages is only problematic in this way in one of them.

The adjective *glad* displays similar behaviour but in the opposite direction, where the neuter form is accepted in Swedish but not in Norwegian. The acceptability of the neuter form of *glad* in Swedish, *glatt*, is suggested in Fanselow and Féry (2002) to be due to it being the participle of the verb *glädja*. As Johansson (2003) has pointed out, however, when the neuter form occurs it typically describes a type of ”superficial” property. It is not *true* happiness, but *observed* happiness. This is also a property of a number of other adjectives, that their neuter forms lend themselves to descriptions of properties that are immediately observable, while the common gender form can have a wider interpretation.

The identification of the shift in vowel quality in Swedish as being a potential trigger for the missing neuter form is dubious for other reasons as well. In Standard Swedish (as well as in Norwegian), the length and quality are mutually dependent. Typically, length is considered to be the primary distinctive feature while quality is secondary to this. We find that every vowel in Swedish is distinguished for length(Andersson, 2002) *and* quality, where these two features covary, so the rule proposed for the ill-formedness of the neuter *flatt* requires that the vowel represented by the grapheme ”A” in this respect is somehow different from every other vowel in Swedish. The dependency between quality and length is observed in vowels Norwegian as well, although ”A” stands out in that in Norwegian it typically exhibits little to no change in quality when its quantity changes.

The following examples in Swedish illustrate that the neuter form *hett* is allowed while *flatt* is not. The change in vowel quality should is potentially of importance to the phonological account.

(14) En flat tallrik.
A.C flat.C plate[COM].

'A flat plate.'

(15) Ett *flatt hus.
A.N flat.N house[NEUT].
'A flat house.

(16) En het potatis.
A.C hot.C topic[COM].
'A hot potato.'

(17) Ett hett ämne.
A.N hot.N topic[NEUT].
A hot topic.

The adjectives share a number of phonological similarities. Most important for evaluating the proposed phonological mechanism for the blocking of neuter forms however is that for both adjectives the neuter form exhibits a short vowel with the expected quality change. In *het* -> *hett* the change in quality is from /e:/ to /ɛ/. If lengthening of the final consonant and a change in vowel quality really is what blocks *flatt* from appearing, we ought to expect the same behaviour for *hett*. The latter form is readily used in both spoken and written language however.

A similar change is observed in the examples in the beginning of this subsection (14, 15, 16 and 17). Here the change in vowel quality is (or would be) from /o:/ to /ɔ/ when the neuter is formed, concomitant with the change from a short consonant /t/ to a long consonant /t:/. The words are almost identical in their phonology, yet the neuter form **kått* is blocked, while the neuter form *vått* is not. This might even be an argument strongly in favour of the semantic hypothesis, as there seems to be little else than semantics that can explain the fact that one adjective has a well-formed neuter and the other does not. The alternative is that the gaps truly are random, or *accidental*.

The proposed vowel change rule as one part of a phonological explanation, and the fact that we find instances where it does not apply to Norwegian weakens its potential as a complete explanation in its own right to account for the missing neuter forms.

A number of adjectives that do not have usable neuter forms however do appear to fall neatly within a morpho-phonological rule. These are adjectives with a stressed second syllable on *-id*. This is however complicated by the occurence of some of these adjectives with unmarked forms where a neuter form is expected.

Adjectives with stressed *-id* in the coda, such as *solid* (solid), *gravid* (pregnant) are another matter. The adjectives provide us with a good example to compare contrasts since they have different meanings but similar phonology. They do not appear to occur very commonly at all. For instance, *gravid* tends not to occur at all with a neuter form, whereas *solid* can occur in a position that would require a neuter form, but with no overt neuter-marking. This might be evidence further for a semantic rule, but there is clearly a phonological component at play

here as well. The expected neuter forms of the adjectives *gravidt* and *solidt*[2] are both blocked and the morphological blocking may still be partially active for *solid* even though it can occur with the stem form in the neuter.

A possible alternative explanation for the lack of a neuter form for *gravid* is that the word may be strongly associated with feminine biological sex, and by extension, feminine grammatical gender. The word does however fit into the semantic rule by its properties not being immediately verifiable through sensory information. The adjective *rigid* (rigid) is similar however and also demonstrates a missing neuter form.

Adjectives with the coda *-dd* have also been proposed as being governed by a morphophonological rule. Notably, the adjective *redd* does not occur in the neuter gender in both Norwegian and Swedish. This generalization is however unsteady, as the adjective can be contrasted with neuter participles that are well-formed in the neuter in spite of similar phonology, and the fact that it obeys the semantic rule discussed in the next section.

2.5 Semantic account

Johansson's suggestion that the missing neuter adjectives are due in part to semantics is worth investigating further. We may do so by first looking at the core of the grammatical gender system in our languages, the nouns. The notion that grammatical gender contains semantic cues is not immediately obvious for either Norwegian or Swedish. What these semantic cues might be is not immediately obvious either. It should be noted that Norwegian has three grammatical genders; masculine, feminine and neuter while Swedish only has common and neuter[3].

Petterson (1990) first suggested that the paradigmatic gaps are caused by a distinction between qualities that are immediately observable and qualities that are not immediately observable. He reports a similar feature in Russian where the adjective *krasiva/krasivaja* (beautiful) receives a different inflectional morpheme depending on whether it refers to a type of beauty that is readily observable or a type of "internal" beauty that would not be immediately apparent.

Each of the gender categories contain an assortment of nouns with gender assignment that would seem arbitrary at first. There is no obvious reason why a noun such as *bil* (car) should be masculine, or why *bok* should be feminine. There are no strict semantic regularities along the lines of biological sex that prevents nouns from entering these categories, although nouns referring to biologically masculine entities and nouns referring to biologically feminine entities typically occur as grammatically masculine or feminine. This is simplified in Swedish where the historical masculine and feminine categories have collapsed into one.

Semantic cues in gender assignment for the Scandinavian languages have been discussed in the literature, and it has been shown that there are in fact rules governing gender assignment

[2]Note that a neuter form in Danish is registered in for instance Den Danske Ordbog (DDO) amongst others.

[3]Some variants of Norwegian have a common-neuter system as well, and examples of this are also found in written Bokmål.

for nouns (Trosterud, 2001), in spite of the seemingly arbitrary categorization of nouns that refer to entities without biological sex.

This is further supported by, and aligns with the observation that for many European languages, the gender system could be thought of as consisting of a semantic ”core” (Corbett, 2013). Outside of this ”core”, assignment is usually arbitrary, but those nouns referring to entities with biological sex will tend to follow a trend in which grammatical gender category they fall into. These two factors taken together, that gender assignment is rule-based and that there is a semantic core to grammatical gender in Norwegian and Swedish is likely to have consequences for modifiers that need to agree in gender with the head noun they are modifying.

The adjectives with no good neuter forms can be categorized according to criteria that go beyond their phonological form. As mentioned, *petterson* suggests that they all typically describe properties that are not immediately observable. Johansson (2003) further suggests that they describe states that are internal to the noun they modify. Many of the adjectives do indeed describe mental states. The adjectives have both similar phonology and similar semantics. The introduction of a semantic rule allows us to avoid some of the problems associated with a purely phonological account by proposing that there is another mechanism to how the paradigmatic gaps occur. The only required phonological commonality is that the *-t*-ending is assimilated into the stem.

A further example where the semantic hypothesis applies is in adjectives that can apply to both properties that are immediately observable and those that are not. One example is the adjective *god* which can mean either *morally good* or *tasty*. In the neuter form however, the second reading is strongly preferred and the first reading might even be completely excluded. This might provide further support for the semantic hypothesis as we would expect such readings to not be compatible with the typical features of neuter gender as well.

How might neuter gender fit into a gender classification system that at its semantic core tends towards biological sex? We might at first assume that it should contain only nouns that refer to sexless entities, but this is not the case. We find counterexamples in highly frequent words such as *barn* ”child”, *dyr* ”animal” and *menneske* ”human”. These nouns do refer to entities that by default have biological sex, but the nouns do not specify the biological sex of their referents. These three nouns, *barn*, *dyr* and *menneske* are both frequent and salient members of the neuter gender in Norwegian, and they enable speakers to make reference to animate entities that typically do have biological sex without specifying it. This makes them good candidates for prototypical neuter nouns in that they cannot be identified with masculine or feminine characteristics without further specification. Their high frequency would also allow them to ignore the general rule for neuter nouns referring to inanimates. This is comparable to how highly frequent lexical items can ignore regular inflectional rules and maintain irregular paradigms.

It turns out that there are in fact few neuter gender nouns that refer to animates in the

Scandinavian language. This lends some support for the semantic hypothesis. If neuter gender nouns that refer to animates are uncommon, then adjectives that require the entity it is modifying to have internal mental states would also be uncommon in the neuter form. An issue is that, as mentioned, while few neuter nouns refer to animate entities, the ones that do tend to be rather frequent.

The neuter gender underrepresents the living category, but it also underrepresents the individual category. The general trend in Norwegian and Swedish seems to be to avoid animate nouns of neuter gender. Gender assignment for new words in the languages typically defaults to masculine/common as well. Such avoidance is further reinforced by neuter gender nouns referring to people frequently having derogatory content, ie. *hespetre*, *mannfolk*, *fruentimmer*. These nouns also contain a level of depersonalizing content and imply that the referent is exactly like something else, or a set of objects with similar properties.

To summarize, the suggestion is that neuter gender in Norwegian and Swedish implies both non-animacy and non-individuality, and these two features are distinct from each other. Non-individuality is also a feature that occurs with neuter gender inflection of adjectives in both languages in a number of other cases, which will be discussed in a later section.

2.5.1 Semantic issues

The semantic account provides us with fewer immediate problems than a purely morpho-phonological account, it does however have a few issues of its own that need to be accounted for. In both Norwegian and Swedish we can find adjectives that do not neatly fit into the rule. They may be thought of as more marginal members of the category, and this is also where the two languages show different tendencies in which adjectives have gaps.

One example of this is the adjective *flat* (flat) which in its most basic sense describes the property of flatness. This adjective does not form a neuter in Swedish, but it does in Norwegian. The issue then is whether flatness is a property that is immediately observable or not. The property cannot be verified without careful measurement in any case, but the appearance of flatness can be observed without too much effort on part of the experiencer. This indicates that some members of the *problematic* category of adjectives are more marginal than others. The fact that Norwegian and Swedish do not agree on the status of *flat* further evidences that the category has fuzzy boundaries.

While the adjective *flat* typically describes the property of flatness, it can also be used to describe other properties euphemistically. This usage is more common in Swedish where descriptions such as *"en flat attityd"*, literally "a flat attitude", occurs. The original meaning of this adjective has to some extent also been replaced by the newer adjective *platt* in Swedish. This also explains the differences between Norwegian and Swedish on this account, with *flat* having drifted (or is currently drifting) towards a different semantic domain.

This seeming inconsistency in the semantic account in fact constitutes evidence in favour of it. If the adjective experiences semantic drift and as a consequence of this the neuter form

falls into disuse, the semantic account of paradigmatic gaps is very well founded indeed. The only difference between the word *flat* in Norwegian and Swedish other than the fact that the neuter does not occur in Swedish is the tendency for it to be used in a literal sense in Norwegian and a figurative sense in Swedish.

As mentioned previously, adjectives with stress on the second syllable *-id* do not regularly form neuters independent of their semantic content, in Norwegian, although some of them can occur in their uninflected stem forms.

Adjectives with *-dd* in the coda have also been suggested (Rice, 2007) as candidates for a complete failure of phonological processes. We run into a problem upon further inspection however. For one, there are not many members of the category that obey the criterion. The frequent adjective *redd* (afraid; scared) falls within the combined semantic and phonological rule, and there are few pure adjectives of similar phonological form. There are a number of participles that have similar phonology however, but it might not be possible to generalize the rule to participles in any case, as we shall see.

Any phonological restriction appears to be restricted to adjectives however, and the participle *brydd* (concerned) may occur as the neuter *brytt*. The neuter form of participles would not only be generated through gender-agreement when used as a modifier, but also in the verbal conjugation of the verb. Participles in Norwegian and Swedish frequently take the neuter form when they are used in the supine. Since the neuter form in this case is supported through the application of another rule it may override the tendency for adjectives to be removed through the semantic removal rule.

The participle may still be problematic in the sense that it will typically not occur in attributive or predicative positions that require a neuter form, but a form that is identical to the neuter form can still be generated through the application of a different rule. A similar condition exists for adverbs, and this will be further discussed in a later section. Neuter morphology occurs in a number of instances in both Norwegian and Swedish where this morphology is not generated through agreement with a head noun. This will also be addressed in more detail in later sections.

2.6 Interaction of morpho-phonology and semantics

The morpho-phonological accounts and semantic accounts have one thing in common. Neither account suffices to explain the gaps in their own right. This has been discussed in some detail in the section on morpho-phonological accounts. A purely morpho-phonological rule would overgeneralize and cause gaps where they do not occur.

The same is also true for a semantic account on its own. In contrast to the morpho-phonological rule, any attempt to apply an exclusively semantic criterion fails once we realize the following. The adjective *glad* (happy) has a number of synonyms and these synonyms, such as *munter* and *lystig* readily form neuters.

The adjectives that do not form neuters are in fact restricted by their phonology. But this restriction is not on its own enough to cause the deficiency in the paradigm. Therefore, the semantic criterion is on top of the morpho-phonological criterion. This reduces the complexity of the phonological rules we have to postulate in order for the gaps to arise.

Both the semantic condition and the phonological condition have to be present for the neuter form to be missing in the paradigm. Adjectives for which only one of these two conditions apply will be able to form neuters without any problem. The examples of adjectives used throughout this thesis highlight this fact. Adjectives that obey the morpho-phonological criterion but not the semantic criterion are of special interest since they can be used as a control condition in the experiment. This is detailed in the section on experimental methods.

2.6.1 Productivity

The combined morpho-phonological and semantic rule is very regular, and so it could potentially generalize to new vocabulary items that enter the language. The experiment described in this thesis relies on expected neuter forms of pseudowords as one control condition in the analysis. These pseudowords are constructed according to the same phonological template as the adjectives that do not form neuters. The assumption is that since they do not contain any semantic information, the neuter-removal rule should not apply to such adjectives.

While it is beyond the scope of the experiment in this thesis, it may be possible to construct pseudowords that follow the morpho-phonological rule and give them semantic context. The expected result would be that dependent on the semantic context the new lexeme is presented in, the neuter form should fail for those adjectives that describe non-observable properties.

It is worth noting that the *problematic* adjectives also form a diverse group as to their origins. A number of them such as *glad* and *redd* descend from Old Norse, whilst other members of the category such as *gravid* and *ru* are relatively younger items, having been borrowed from Latin and Low Saxon respectively. This raises two possibilities. Either the rule has historically been generalized to new items in the language, or alternately, the neuter-removal rule is a fairly recent development itself. We do find numerous examples of neuter forms in Old Norse for adjectives that do not form neuters in modern Norwegian and Swedish, which would suggest that the rule has developed after that period.

2.7 Summary

The adjectives that fall into the problematic category all have similar form, ending either on a vowel or a dental consonant /d/ or /t/ in the stem form. In order for adjectives to lack a proper neuter form in Norwegian or Swedish, this seems to be a necessary condition. Because of this phonological similarity, it is tempting to look for some phonological rule that makes the neuter forms illegal. This has been suggested by (Fanselow and Féry, 2002; Rice, 2007; Raffelsiefen, 2004) amongst others. As discussed in the next chapter, however, phonology cannot in itself

explain the missing neuter forms. Any of the rules suggested so far will either overgeneralize or fail to account for all the missing forms. The constraint does not appear to be purely phonological as many adjectives have similar phonological form but readily form neuters. It is a necessary condition, but not a sufficient condition.

There is a semantic component to missing neuter forms in addition to a phonological rule (Johansson, 2003). The adjectives that have problematic neuter forms show a clear commonality in their semantic content. They typically describe observable or non-observable states and in many cases more specifically describe internal mental states. Internal mental states also have the property of being not immediately observable. Some adjectives however, for instance *flat* and *ru* appear to be more marginal members of the category. *flat* has no usable neuter form in Swedish, but regularly forms neuters in Norwegian. Likewise, *ru* has no usable neuter form in Norwegian, but the word is absent from Swedish altogether. The variation between the two languages, while small, suggests that the paradigmatic gap rule is productive. Further, it constitutes strong favour of a semantic account if the same adjective is used in different meaning contexts in the two languages and neuter-formation is different as a consequence.

Also according to Johansson (2003), there is a tendency for the neuter gender in both Norwegian and Swedish to typically imply non-animacy or non-individuality for the nouns that are members of this category. This is arrived at through the observation that there are very few nouns within the neuter gender that refer to animates or individual entities and also through the behaviour of certain adjectives when they occur in the neuter.

The typically internal or mental states described by the adjectives with missing neuter forms are not generally compatible inanimate or non-individual entities. It would then be the case that the semantics of these adjectives is not compatible with the semantics of the Scandinavian neuter gender, assuming that there is a semantic content to grammatical gender in the first place (Corbett, 1991). In our case, if neuter gender suggest non-animacy or non-individuality to its referent, neuter gender would act as a semantic block for the adjectives that do not form neuters. The properties described by them are incompatible with the properties assumed by a neuter gender noun.

3 Adjectives, adverbs and participles

This section will deal with adjective morphology in Norwegian and Swedish. Adverbs and participles will also be discussed on account of the similarities in their inflectional morphology to adjectives. Adverbs in particular also demonstrate another type of relationship than just morphology which will further be discussed in this section.

Example adjectives that do not form neuters have already been provided. A full table of adjective inflection in Norwegian is provided below to give a complete picture.

	Masculine	Feminine	Neuter
Sg. Indef.	våt	våt	vått
Sg. Def.	våte	våte	våte
Pl. Indef.	våte	våte	våte
Pl. Def.	våte	våte	våte

Table 1: Inflectional paradigm for the adjective *våt*

The table illustrates that there is very little inflectional morphology in adjective declension. Only the singular indefinite has separate forms between the genders, and then only between masculine and feminine. The plural and definite forms are all identical. The paradigm is only slightly more complicated in Swedish and in a number of Norwegian dialects, but the forms affected are not relevant to this study.

The following table illustrates where the gap occurs.

	Masculine	Feminine	Neuter
Sg. Indef.	kåt	kåt	∅
Sg. Def.	kåte	kåte	kåte
Pl. Indef.	kåte	kåte	kåte
Pl. Def.	kåte	kåte	kåte

Table 2: Inflectional paradigm for the adjective *kåt*

What makes this clear is that neuter forms are not blocked in the entire paradigm. It is specifically only the one form that has specific neuter-marking morphology that is blocked. The neuter forms are only restricted in the indefinite singular. As shown in the examples below, the adjectives are well-formed when they occur in the definite form.

(18) Et *redt barn.
A.N frightened.N child[NEUT]
'A frightened child.'

(19) Det redde barnet.
The.N frightened.DEF.SG child[NEUT]-et.SG.DEF
'The frightened child.'

Plural forms, both indefinite and definite, are also well-formed. One can easily use the adjective *redd* in its plural form *redde* to describe a group of frightened children in Norwegian, as an example, both in the definite form and the indefinite form. The same is true of Swedish. The main difference between the two languages is that Swedish non-singular-indefinite forms typically are formed with the vowel *-a* where Norwegian has *-e*.

It might at first seem inconsistent that an indefinite child cannot be frightened but that a definite child can. This observed inconsistency is likely due to the unmarked morphology of anything but the singular indefinite forms of adjectives in Norwegian and Swedish. There is no form in any other part of the paradigm that specifically invokes neuter morphology.

One of the features mentioned that characterizes the neuter gender in Swedish and Norwegian is that it implies non-individuality, and so the fact that these adjectives do have usable forms in the neuter plural is expected. The singular definite of the neuter is well-formed. This should however be seen in light of the morphology of adjectives in both Norwegian and Swedish having no unique gender-marked forms in anything but the singular indefinite.

A case could be made that gender is simply ignored in anything but the singular indefinite. At least in the morphological paradigms of both standard Norwegian and Swedish, the forms are all similar throughout the paradigm. This could perhaps override any semantic tendencies or rules implied by grammatical gender.

Related to this is that there are also a number of adjectives which have identical forms for all genders also in the singular indefinite, for instance *trett* (”tired”) or *lett* (”easy, light”). Due to their semantics we might expect them to not produce any neuter forms as well, based on the previous criteria. However, there is no specific inflectional form for the neuter gender in this case either, which would likely override the neuter removal rule.

3.1 Morphology

The way in which neuter forms of adjectives are formed in Norwegian is through a simple and predictable inflectional rule. In the singular indefinite most adjectives have the morpheme *-t* added to the stem, marking the neuter form. The rule is slightly complicated by adjectives with only a stressed vowel in the syllable coda. The written standards Bokmål and Nynorsk have some minor variations in how this rule is applied. In Bokmål, if the stressed vowel in syllable coda is a simple vowel, the morpheme becomes *-tt*. This usually causes a change in vowel quality and/or quantity. If the stressed vowel is a diphthong however, the morpheme remains as *-t* and causes no change in vowel quantity. In Nynorsk, the rule is simpler and the morpheme is *-tt* for all adjectives with stressed vowels in this position, including diphthongs. The thesis focuses on the morphology Bokmål which is the more commonly used standard. The results are likely to be valid for Nynorsk as well where there is overlap. The paradigmatic gaps occur in both standards of written Norwegian.

A number of phonological processes occur in neuter-formation. A long vowel is usually shortened and in the case where a pre-existing final */t/* is present in the coda, the neuter-marking morpheme is assimilated and the consonant is geminated. Refer to section 2.4 for a more detailed discussion of these features. A point of difference between Norwegian and Swedish is that diphthongs are more common in Norwegian.

We may consider the Norwegian adjective *staut* (”stout”) as a potential candidate item for the study. It fulfills the basic phonological criteria, and a corpus search returns very few instances of it being used in the neuter gender. The adjective has a number of meanings, some of which do fall within the semantic criterion. The expected neuter form would be *stautt*, but the presence of a dipthong may go some way to explain the absence of a specific neuter form

as well as shown by the previous criterion. Although they are usually treated as long vowels, it is possible for dipthongs to be shortened in Norwegian, and neuter formation should be no exception to that.

3.2 Unmarked neuter forms

A number of exceptions to the neuter-forming rule occurs other than the paradigmatic gaps. For two classes of adjectives following regular derivational patterns, this non-marked neuter form is regular. These are adjectives formed with the suffix *-ig* as in the adjectives *blodig* and *kunnig* and adjectives formed with the suffix *-lig/-leg* such as *vanskelig* and *daglig*. The inflectional paradigm is otherwise intact for these adjectives. These adjectives however do occur as modifiers for neuter gender nouns, but without any specific neuter-marking morphology. This sets them apart from the adjectives that have paradigmatic gaps where the adjectives that have neuter gaps simply cannot be used at all when the neuter form is expected. While there are some rules such as the ones shown here for unmarked neuter adjectives in Norwegian, there are a number of other exceptions. Adjectives such as *bra* (”good”) and *(lilla* (”purple”) have unmarked neuter forms as well, but these adjectives are irregular in that there is no inflectional paradigm for them. The stem form is always used.

3.3 Adjectives or adverbs

Adverbs in Norwegian and Swedish can be formed in a similar way to neuter forms of adjectives, by taking the stem of an adjective and adding the morpheme *-t*. This is also a highly productive process and introduces a slight complication. That is, a number of adverbs are derived from adjectives, and these adverbs are phonologically identical to the neuter forms of these adjectives.

Some lexical items occur frequently as adverbs in their expected neuter-gender forms, but very rarely or not at all as adjectives. Two examples of this are the adjectives *blid* and *flau*, their neuter or adverb forms being *blidt* and *flaut*. The adjective *blid* occurs in Johansson (2003) as one of the adjectives that do not have neuter forms. It does however appear to be a legal form as an adverb[4], although fewer examples are found where it occurs as an adjective.

(20) Han smilte blidt.
He smile.PAST happily.
He smiled happily.

(21) Et ?blidt smil.
A.N happy.N smile.
A happy smile.

[4]At least in Norwegian.

The form with the morpheme *-t* is not completely excluded from the grammar in that it can exist as an adverb. This is comparable to the situation for participles described earlier. The application of a different morphological rule produces the same form as the neuter-generating rule. The adjectival use may not be completely excluded on this basis that if the form already exists elsewhere in the grammar, it may feel less awkward. Such adverbs are only partially blocked.

3.4 Solutions to gaps

The paradigmatic gaps in Norwegian and Swedish mean that language users must find alternative ways to express information. A number of creative behaviours occur in an effort to avoid the gaps.

The adjective may be substituted by another adjective that is synonymous. The solution is imperfect since two lexical items are rarely fully synonymous. In the following examples, one such possibility is highlighted:

(22) Et *gladt barn.
A.N happy.N child[NEUT].
'A happy child.'

(23) Et muntert barn.
A.N happy.N child[NEUT].
'A happy child.'

The well-formed example replaces the original adjective *glad* with the adjective *munter* which falls outside of the morpho-phonological criterion even though its semantics are approximate. There is a certain similarity to the process of suppletion (Corbett, 2009) whereby a defective paradigm can be filled in with forms from a different paradigm, as in English *good - better* in comparison. Such paradigms are highly irregular. But there is a difference in that the "suppletive" forms in Norwegian and Swedish are full adjectives in their own right. The paradigmatic gaps in Norwegian and Swedish are do however share the commonality of being a very salient regularity in an otherwise regular and productive paradigm. The comparison is noteworthy however as it has been suggested that suppletive forms are provided from the paradigm of a different lexical item (Corbett, 2009), and the situation in Norwegian and Swedish may well provide us with an example of what such a historical process may look like in the early stages.

Substituting adjectives with other adjectives is a solution that may occur with some frequency, but is difficult to track since the adjective we would like to study is replaced by a different adjective. This can also cause slight differences in meaning, as mentioned previously. Speakers might therefore attempt other solutions if precision is desired, and this has also been demonstrated in examples earlier in this thesis.

Synonymy occurs not only with adjectives but also with nouns. Speakers can in some cases replace the neuter noun with a non-neuter noun to avoid any problems due to adjective agreement. This is observed in the examples below:

(24) En glad unge.
A.C happy.C child.
'A happy child.'

(25) Et *gladt barn.
A.N happy.N child.
'A happy child.'

The nouns *unge* and *barn* both mean "child". Since the noun *unge* is common gender and *barn* is neuter, using *unge* solves the problem of missing neuter forms by changing the gender agreement of the adjective into a form that is allowed. This also can occur where synonymous nouns are available.

There is a difference in style in that *unge* is a more colloquial way to refer to children. Additionally, *unge* can also be used in reference to the offspring of animals whereas *barn* is restricted to human children. Particularly in Swedish, this leads to *unge* being a somewhat less polite term. This may apply to some registers of Norwegian as well. We therefore see that the synonym-replacement strategy does not provide a perfect solution in this case either. It should be noted that both words for child, *barn* and *unge*, are frequent in both written and spoken language both in Norwegian and Swedish.

Johansson (2003) mentions that when trying to elicit missing neuter forms from children, they would demonstrate other ways to avoid the gapped neuter form. One example of this is shown in the following:

(26) Lejonet han var rädd.
Lion.DEF he was afraid.C.
The lion he was afraid.

The following example shows what would be expected if the child had not picked up on the paradigmatic gap.

(27) Lejonet var *räddt.
Lion.DEF was afraid.N
The lion was afraid.

It has long been observed that children are prone to overgeneralizing grammatical rules such that irregular patterns are replaced with more regular patterns (cf. Marcus, 1996, inter al.). While developmental data on the acquisition of paradigmatic gaps in Norwegian and Swedish is lacking, the fact that gaps have been observed in children suggests that there is some saliency to the gapping rule that speakers are able to pick up on.

3.5 Lexical properties

Frequency does not appear to be a good predictor of whether or not an adjective will lack a neuter form. The frequency of adjectives within the group spans a large band, ranging from high frequency to low frequency. The same is also true of adjectives of comparable phonology with well-formed neuters. The low frequency of some members in the apparently irregular neuter gap category points towards an analogical mechanism. This is discussed in further detail in section 8 of the thesis.

A final possibility that we must consider is that the paradigmatic gaps are learned simply by rote. This would require speakers to memorize the exception for each individual adjective where it occurs. Such an explanation already faces problems in that a number of adjectives with missing forms have very low frequency of occurrence. Memorization may be one possible mechanism by which irregularly inflected forms of words are stored in memory as separate lexical entries (Pinker and Prince, 1988). As an example the English verb *go* has an irregular past tense *went*, and these are stored as separate lexical items in the mental lexicon. If we attempt to apply this to the adjectives with missing neuter forms in Mainland Scandinavian, we come to the conclusion that they are irregular in that they are missing a neuter form. It is not however entirely clear if there exists a representation at all in memory for the missing neuter form. If no such representation exists, then we would expect the neuter to form regularly through application of the productive rule for neuter-formation.

Johansson (2003) mentions that the neuter form fails to form within a context that is unlikely to occur and that it is therefore unlikely that the missing neuter forms are memorized individually. That is, the missingness of the neuter form is not likely to be marked for each lexical entry but arrived at through a process of inference.

3.6 Problematic adjectives in North Germanic?

The feature of some adjectives lacking a well-formed neuter in this chapter has focused on Norwegian and Swedish. A question remains though whether the missing neuter forms exist in Danish as well, as neuter forms of adjectives in Danish are formed in the same manner as in Norwegian and Swedish, by adding a final *-t* to the stem of the adjective. More distantly, the languages Faroese[5] and Icelandic would be interesting candidates for comparison as well.

4 Experimental methods

The missing neuter form experiment consists of a lexical decision task (Meyer and Schvaneveldt, 1971). This type of experiment has been used for some time within linguistics and

[5]An informal interview with a Faroese speaker in fact did find that the speaker reacted negatively to neuter forms *glatt* and *rætt* in the nominative for the items *glaður* and *ræddur*. This speaker did have years of exposure to Norwegian as well, however.

related disciplines. The basic task consists of the participant classifying words or nonwords. The participant's reaction time and error rate are recorded and can tell us how these words or nonwords are processed.

The dependent variables in these experiments are primarily the reaction times, and secondarily the error rate. The reaction time data allows detection of any differences in processing in accordance to how our stimulus is grouped. The basic division relevant for this study is between ”regular” adjectives and adjectives with missing neuter forms. Further variables are considered however, and outlined in the methods section of the thesis.

While any differences in reaction time to stimuli does measure processing, the measurement is indirect as it does not show us the process as it happens in the participants mind. It does however provide us with information that may give insights into how speakers process language, and both lexical decision tasks and priming effects are well-established measures (Meyer, 2014). Moreover, data gathered from such experiments may also tell us how language behaves which is another concern to the study of linguistics.

The experiment presented in this thesis have a number if key differences to the one in Johansson (2003). A simplified method where neuter forms or expected neuter forms of adjectives and pseudowords was used as priming stimuli was devised. There was opportunity to gather data from several groups, and slightly different lists of adjectives were used for different groups as well. For instance, certain adjectives were discovered to be more difficult for participants to classify or exhibited otherwise odd behaviour in earlier stages of the experiment. Such stimuli were phased outand replaced with ”better” adjectives.

The design where speakers react only to the adjectives themselves isolated from any context also makes for a relatively simple task where participants are only required to classify the target stimulus as a word or not a word without having to judge its grammaticality within a given context. Agreement should therefore not play a role in their reactions.

Several groups of Norwegian natives were tested, however, L2 speakers were brought into the experiment as well as a separate group. We wanted to test how efficiently L2 speakers learn these paradigmatic gaps. Paradigmatic gaps are a problem for a those that learn the languages as the rule will overgeneralize. In line with this, learners may also not be overly sensitive to systematically missing forms and may therefore incorrectly apply the neuter-forming rule to forms that should not occur. Due to the simplicity of the Scandinavian neuter-forming rule, we are interested in whether L2 speakers will overgeneralize the rule. This was tested by subjecting L2 speakers of Norwegian to the same experiment as native speakers and comparing the results, where if the L1 and L2 groups show different responses to neuter forms, we may be able to conclude that there is in fact a difference in how the rule is applied. The selection process aimed at recruiting learners. Even so, L2 speakers of varying levels of ability ended up participating.

This chapter will deal with experimental methods and theory.

4.1 Problem and hypotheses

Experimental results (Johansson, 2003) showed that subjects reacted differently to regular neuters and missing neuter forms in a lexical decision task. The work in this thesis expands upon this work by introducing masked priming and a comparison between L1 and L2 subjects. A further comparison with the expected neuter forms of constructed *pseudowords* is also made. Both the neuter forms of *problematic* adjectives and the neuter forms of *pseudoword* adjectives are missing in that they do not occur, but they may be missing in different ways.

The actual gaps in the paradigm cannot be easily observed through corpus studies, owing to their general state of non-existence in written language. Another problem is that our corpus may not be big enough to capture the full breadth of the language being studied and may therefore be insufficient to conclude any negative effect or negative existence of these forms. The lack of neuter forms of such adjectives could be explained through accident, or the low frequency of nouns of neuter gender referring to entities to which they would apply.

Further, an introspective method has a number of limitations. For one, grammaticality or acceptability judgements might not actually reflect actual language usage when they are carefully deliberated upon. Second, the data gathered from such methods are not usually able to be generalized across subjects since introspection typically relies on the researcher evaluating the grammaticality of a feature by themselves. The method does however have its place in that it can lead to testable hypotheses.

To test these hypotheses, we can turn to methods from experimental psychology. We can test whether the missing neuter forms are reacted to differently than regular neuter forms, and if this is the case, the experiment will provide us with a set of data gathered from a group of speakers. While the participants are consciously aware of the task they are performing, and of the stimulus they are classifying, the short amount of time permitted for each classification suggests the classification is the result of an automated lexical process.

Participants are expected to have difficulty in classifying expected but missing neuter forms in a lexical decision task, and this is the primary motivating reason for using common gender forms as targets in the experiment. We want to avoid relying on speakers' judgements, and instead rely on their reactions. This also allows for the removal of subjective evaluations on the part of the investigator, although the experiment necessarily relies on the assumption that there may be different behaviours expected for the different groups of adjectives.

The experiments were designed to test a number of different hypotheses. The priming effect of expected neuter forms of the adjectives that do not regularly form neuters may differ from the priming effect of neuter forms of adjectives that regularly do form neuters. Additionally, there may be a difference between L1 and L2 speakers of Norwegian in this effect. Further, there may also be a difference between how the expected neuter forms of these *problematic* adjectives and the expected neuter forms of *pseudoword* adjectives prime their target stimulus, and this is also tested in the experiment.

The reason for testing second language speakers is the expectation that the existence of paradigmatic gaps might be difficult to detect. The neuter-forming rule is quite simple, and we might expect L2 speakers to overgeneralize this rule in the same way that they may overgeneralize regular inflectional paradigms to irregular lexical items. Likewise, the rule governing the occurence of the neuter gaps in adjective declension may not be immediately obvious have not detected the rule. and so it might be slower to learn. One expectation that arises from this is that the L2 group may not have fully formed the paradigmatic gaps in their mental grammars. It may therefore be easier for L2 speakers to make the connection between neuter forms of *problematic* adjectives and their common gender forms, since the neuter-removal rule might not be fully active for this group. This is another hypothesis to be tested in the experiment.

In all, there may be effects present for item class (*normal*, *problematic* or *nonword*), there may be effects for priming (*primed* or *unprimed*), and there may also be an effect for the language background of participants (*L1* or *L2*). All of these variables could potentially have interactions between them as well, and between the levels within each variable. This is further elaborated on in the analysis.

Finally, there are a number of null hypotheses that must be accounted for. A simple summary would be that the null hypothesis assumes that there would be no difference between any experimental condition or across experimental conditions.

H0: The priming effect is the same for all classes of stimulus items.

H0: There is no difference in reaction time across stimulus item classes.

H0: There is no difference between L1 and L2 groups.

As mentioned, there may be an effect for L1 or L2 groups. What this means is that we would expect L2 speakers to have slower performance overall on the task. We would also assume a higher level of misclassification of target items for this group, which in turn would render results less reliable as it becomes more difficult to separate low task compliance from difficulty in performing the task.

4.2 Introduction

The use of experimental methods allows us to observe language processing as the speaker is exposed to language. We might observe some pattern in a corpus, such as the fact that neuter forms of certain adjectives are missing. But it might just be the case that our corpus is not big enough to capture potential usage of these forms.

Since the *problematic* neuter forms of adjectives discussed in the previous chapter do not occur in regular usage of the language, we would expect them to behave like a *pseudoword* or *logatome*. In order to investigate whether this proposition holds, we need to test the behaviour of these neuter forms on language users as they are processed in real time. We would expect

language users to process missing neuter forms differently from regular neuter forms, and a simple way to test this is with priming. Moreover, we want to know if they elicit behaviour in participant that is different from that of potential neuter forms of ”regular” *pseudowords*. That is, are they truly gaps in the lexicon, or is there a process removing them from the lexicon?

The experiment is designed to provide answers to these questions. If problematic adjectives are reacted to differently than both regular adjectives and pseudowords it would suggest that they are in fact different. The experimental condition is divided into three classes. These are referred to as *normal*, *problematic* and *nonword* in the dataset, although these labels may be misleading.

The *normal* category are those adjectives which form neuter forms as expected and thus make up the control condition. The *problematic* adjectives are those that do not form neuters. Finally, the *nonword* category consists exclusively of pseudowords. All three categories consist exclusively of members that follow the same phonological pattern, as described in the section on adjectives. They are either single-syllable adjectives with either no coda or either of the consonant /t/ or /d/ in the coda, or bisyllabic with a stressed second-syllable.

4.3 Lexical decision task

The experiment is based around the lexical decision task (Meyer and Schvaneveldt, 1971; Schvaneveldt and Meyer, 1973; Meyer et al., 1975). In this task, participants are asked to classify a target stimulus as either a word or a nonword. This provides two measurable responses. The most important of these two is the time it takes participants to classify an item. Secondarily, the way items are classified provides us with some information as well. We would for instance expect items with lower lexical frequency to be classified slower and be more prone to misclassification than lexical items of higher frequency (Grainger, 1990; Whaley, 1978).

The lexical decision task measures the speed at which participants retrieve the lexical item that they are asked to classify (Meyer and Schvaneveldt, 1971). Typically the participant will receive a very short amount of time to classify the target word, which suggests an automated process in making the decision is at work.

The lexical decision task allows us to apply manipulations to the experimental condition. In this experiment, priming was used to manipulate participants' reactions. Specifically, neuter forms of adjectives or expected neuter forms were used to prime the same adjective in its common gender form.

In the presently described experiment participants were given a total of 1000 milliseconds (1 second) to make their decision on whether a given stimulus was a word or not a word. If they failed to make a decision within that time, the stimulus was discarded and the response recorded as ”NR” (no response), before the stimulus was presented.

4.3.1 Priming

Priming was used in the lexical decision task in this experiment. Specifically, neuter forms of adjectives or expected neuter forms of adjectives were used to prime the common gender form of the same adjective. Similarly, constructed neuter forms were used for the pseudoword adjectives. As mentioned previously, there is reason to suspect that grammaticality judgments on expected neuter forms of *problematic* adjectives may not be reliable. Moreover, asking participants to judge the word-status of a dubious or missing form is not too different from asking a group of people to use introspective methods. Another issue is that there is also reason to suspect that participants would be confused by such forms and thus take much longer to classify them. This would lead to unacceptably long reaction times, and the would reduce the automatic processing component of the decision. From the point of view of the experimental design as well, it is not entirely intuitive whether expected neuter forms should be classified as words or nonwords.

Priming and lexical decision research has been conducted ever since the noteworthy experiment by Meyer and Schvaneveldt (1971). Put simply, priming is an implicit memory effect where when a person is exposed to some stimulus, it affects how they process later stimulus. This effect can be observed through various measures, although in this thesis the focus is on reaction time which is also the measure in the experiment detailed. Priming will typically facilitate (reduce) reaction times to a target stimulus and in some cases inhibit (increase) reaction times to a target stimulus.

Different types of priming have been document. *Perceptual* priming is when the stimuli are physically similar to each other, and this has an effect on processing(Tulving and Schacter, 1990). For the purposes of this experiment, the most relevant effect is that of graphemically similar words priming each other. The word *nature* may for instance prime the word *mature* on account of their perceptual similarity. These two words however are only a visual match as the intonation pattern and phonemes in the words are different. There is therefore less phonological matching between these two words. Similar phonology between two items can also elicit priming reactions. This is also called *form* priming.

Another type of priming is conceptual priming. In this type of priming, meaning processing is activated. A typical example might be how the word *nurse* is primed by the semantically related *doctor*, speeding up processing of the target word. The two words are related in two ways. First, there is a semantic connection between nurse and doctor, that is, their properties overlap to some extent. They both work in hospitals and they are both healthcare professionals. Similarly, dogs and wolves have a degree of overlap in their features as well, and so we would expect priming between those two concepts as well.

There is another connection too which is termed *associative* or *collocational*. The words nurse and doctor will frequently occur within some proximity of each other in texts, and in addition to this, doctors and nurses frequently occur within some proximity of each other in

real life. Once again, in a similar way, we can consider the words *mouse* and *cheese*. While mice and cheese have few obviously overlapping characteristics, they might still prime each other due to these concepts being commonly associated with each other. We can see from these examples that words and concepts can prime each other in more ways than one, and often simultaneously. It becomes important to control for these factors.

Priming across modalities can also occur. A simple example would be that the word *red* can prime the colour red or an object that is red, or vice versa. This experiment does not make use of this however, and all priming occurs within the same modality, written words. More complex combinations can even make use of different sensory input such that auditory primes can facilitate or inhibit recognition of visual primes. While the experiment in this thesis does not make use of auditory (spoken word) stimulus, it may be of some relevance for further study of the paradigmatic gaps in Norwegian and Swedish.

A simple illustration follows below of a what one sequence within a lexical decision task might look like, incorporating a prime stimulus and a target stimulus. First, the screen displays a fixation, then the prime stimulus is displayed for a very short amount of time, typically no more than 100 milliseconds, and then the target stimulus is displayed at the end. There may also be blank screens in between each part of this sequence. It is the target stimulus that the participant is asked to classify as a word or nonword. The figure below shows how a basic sequence of presentation might look like as the different frames are presented to the participant.

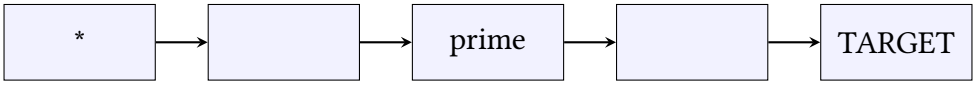


Figure 1: Sequence in a lexical decision task with priming.

For this experiment, since we are using neuter forms of adjectives to prime common gender or stem forms of adjectives, we would expect to see semantic facilitation since the primes and targets are the exact same adjective, albeit with different morphology, meaning the exact same thing. A word or concept primes itself, in the same way that it primes semantically related words or concepts. In addition to this, we would likely also see some effect of perceptual priming since the words are graphemically similar. That is, *kåt* and its neuter form *kått* are visibly similar. They are also auditively similar, differing mainly in the pronunciation of the vowel, but this might be less relevant due to the experiment only using visual stimulus.

In order to deal with the problematic status of the missing neuter forms, the neuter forms were used as prime stimulus in the experiment. Common or stem forms of the adjectives and the neuter forms of the same adjective were used for the target stimulus. This eliminates any uncertainty on the part of participants on whether or not expected neuter forms are words,

since they are not required to evaluate such forms. The same procedure applied to the *nonword* adjectives in the experiment, and expected neuter forms of these were used as priming stimulus as well.

4.3.2 Pre-activation and lexical access

The implicit memory effect observed with priming may be explained by pre-activation. We can understand this process through models of lexical access and spreading activation. When participants perceive the stimulus in the lexical decision task, this activates a cognitive process that looks for the relevant word in the mental lexicon.

The act of recognizing targets presented in the experiment as words or nonwords requires participants to search through part of their memory, the mental lexicon. This process can be described as lexical access (Levelt, 2001; Levelt et al., 1999). A memory search is performed for the relevant lexical item, and then the form is encoded phonologically.

This process of lexical encoding is an important consideration in the design of the experiment described in this thesis. Priming stimuli are presented in all lowercase, while target stimuli are presented in all uppercase. This was done so that there would be no confusion for participants on which stimulus to classify. A consequence however is that priming stimuli will have notable[6] visual differences from target stimuli. In order for form-priming to occur in our experiment then, the underlying phonology of the priming stimuli would have to be activated instead of a simple visual orthographic match. This is particularly relevant for pseudowords where wep would not expect any semantic or associative activation.

Spreading activation models (Collins and Loftus, 1975; Anderson, 1983) can account for the role of the priming stimulus in our experiment. The observation that priming stimuli facilitates the retrieval of related items suggests a partial activation of those items due to the presence of the priming stimulus.

As mentioned previously, priming typically facilitates processing of stimuli that are related to the prime in some way. When the mental representation of the prime stimulus is accessed, the activation of that representation creeps to other parts of memory. The spreading activation model accounts for this by representations that are related to the prime being partially activated during the information retrieval. Using the examples from the previous section, the item *doctor* activates both its own representation and it partially activates the representation of the item *nurse*.

In the experiment in this thesis, neuter gender forms of adjectives are used to prime their corresponding common gender forms. The priming stimulus thus activates its own lexical representation. Pseudowords on the other hand have no lexical representation, but there is still a visio-phonological match with the target stimulus. For these reasons we expect to observe facilitation for participant decisions under the priming condition.

[6] To the extent that participants notice the priming stimulus.

It is unlikely that separate entries for the common gender form and the neuter gender form exist in the mental lexicon since the neuter form is derived through the application of a regular and productive rule (Pinker and Prince, 1988). The priming stimulus therefore is likely to activate its own representation as part of the lexical access process. Since the semantic representation of the adjective is the same in both the neuter gender and common gender, the semantics are subject to repetition for each adjective since we expect gender to be primarily a syntactic phenomenon.

4.3.3 Masked priming

A basic paradigm has already been described in the previous chapter. Various techniques can be used to change how the prime is presented however. The experiments described within these thesis all made use of a masked priming paradigm where additional visual stimulus is presented to distract the participant's conscious attention from the prime.

The masked priming paradigm has a number of advantages (Forster et al., 1987; Forster and Chambers, 1973). One of these is that it obscures the prime stimulus to such an extent that it makes it invisible to most participants. In priming experiments without visual masking, such as the one described in the previous chapter, it is in many cases possible for the participant to fully observe the prime stimulus. With masked priming this is avoided for most participants, and the prime stimulus never enters into consciousness. Any observed effect of priming cannot therefore be ascribed to the participant consciously being aware of the prime stimulus. This can, as expected, also have the effect of reducing the size of the priming effect.

Masked priming also typically modifies the effect that priming has on processing of the target stimulus. Perceptual or form priming generally becomes more salient in lexical decision tasks when the masked priming paradigm is applied. The opposite might also be the case, as masked priming compared to unmasked priming typically does not modulate the N400 effect (Brown and Hagoort, 1993), which is an effect that has been associated with semantic and associative processes (Kutas and Federmeier, 2011). It does not appear to completely eliminate semantic or associative priming however (Greenwald et al., 1996; Forster, 1998; Perea and Gotor, 1997; Carr, 1990), although semantic effects are observed to be weaker within the masked priming paradigm.

The prime stimulus is also presented for a short amount of time (Forster et al., 1987). Typically no more than 60 milliseconds. For this experiment, the time the priming stimulus was presented was set to 50 milliseconds. The stimulus onset asynchrony (SOA) can be kept very short as well, particularly if the target stimulus is presented directly after the priming stimulus instead of a dedicated backward mask. The small amount of time the stimulus is presented for introduces increased variance due to a number of factors. Participants may potentially miss the priming stimulus. Effects will typically also be weaker due to the shorter amount of time that the priming stimulus is presented. In order to alleviate this, a larger group of participants is usually required to establish proper confidence intervals.

Another consideration is the typically decreased magnitude of semantic effects. This means that any facilitation due to the priming effect is likely to be smaller for a masked priming experiment. This consideration further means that more data points are required to detect the effect. Another option is to additionally increase the amount of items that the participants are required to classify, although such a procedure is difficult for the experiment described in this thesis as there by design a limited amount of lexical items to choose from.

The priming stimulus being presented for a very short time is also an attempt to prevent participants from fully processing the stimulus. The addition of masking is supposed to interrupt processing by providing a different visual stimulus and thus replacing any ”echo” representation of the priming stimulus in short term memory or sensory memory.

Below is a model of what a masked priming experiment might look like to a participant. First a fixation point is presented. This is followed by a mask which can be any string of symbols or graphical figure. A series of # marks are one option, which is also what has been used in this experiment. This is called the forward mask. The prime stimulus is presented next, usually for a very short duration. This can then be followed by a backward mask, or the target stimulus can act as a backward mask on its own. Using a backwards mask that is separate from the target word also increases the stimulus onset asynchrony.

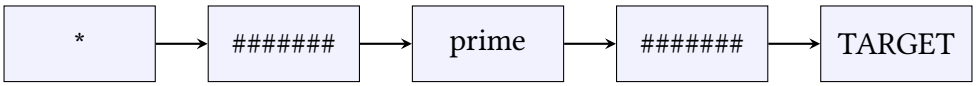


Figure 2: Sequence in a lexical decision task with a masked priming paradigm.

The end effect is that the prime stimulus ends up being ”sandwiched” between two masks. As mentioned, the target stimulus can act as a backward mask on its own, though using a separate backward mask is also common. This added visual stimulus distracts the participant from the presence of the prime stimulus as any short term visual memory effects of it are replaced by the masks. The latter method is the one used for these experiments. The figure can be contrasted with the previous figure 1 on page 27.

4.3.4 Stimulus onset asynchrony (SOA)

The stimulus onset asynchrony was kept the same for the entire experiment. With the priming stimulus being presented for 50 milliseconds and the mask being presented for 100 milliseconds before the onset of the target stimulus, the total SOA between priming stimulus and target stimulus is therefore 150 milliseconds. This allows some time for automatic lexical processes to occur before participants classify the stimulus, although there is also visual interference from the masking stimulus.

SOA may be a variable itself as well. Different SOAs for the same stimulus can elicit different behaviours from participants (cf Baayen et al., 2008, inter al.). This is controlled for in the experiment by keeping the SOA constant, but may be a topic for further research.

4.3.5 Baseline

A baseline condition was also used in order to observe any effects that are not due to priming. For this experiment, all items were presented twice, once in the primed condition and once in the unprimed condition. This applied to all participants. The primed and unprimed portions of the experiment were presented separately, and the sequence of presentation was switched around between participants. Some participants performed the *unprimed* portion of the experiment first whilst other participants performed the *primed* portion of the experiment first.

In addition to establishing a baseline, the different orders of presentation of the two tasks among participants should cancel out training effects under either experimental condition.

For consistency between the experimental conditions, and to avoid any potential confusion for participants, the forward and backward masks were presented under the unprimed condition as well. The only difference being that there is no priming stimulus sandwiched between. The onset of target stimuli between each decision is therefore the same in both parts of the experiment.

This means that each target stimulus was presented twice to each participant, which means that repetition effects are likely. In order to balance this out between the two conditions, primed and baseline, the aforementioned variation in sequence of presentation was introduced. The next effect should be that any effect due to repetition is canceled out for both experimental conditions. Training effects, where participants become better at performing the task as they become more accustomed to it, can also be controlled for in the same manner.

4.3.6 Priming effect

The priming effect is an effect where the presentation of a stimulus influences the perception of a following stimulus. If the priming stimulus causes a change in how participants react to the target stimulus compared to the baseline condition without priming, we observe a priming effect. Since we are measuring reaction time in the decision task, the change in reaction to the target stimulus is measured as either facilitation, where the participant's reaction time is decreased, or as inhibition, where the participant's reaction time is increased.

Put simply, the priming effect is the difference for any given stimulus under the primed condition compared to the unprimed condition. Priming effects for connections between the prime and target stimuli that are semantic, associative or perceptual in nature are well-documented (Meyer, 2014). The expectation is therefore that we will detect a priming effect in analysis of the results in the experiment.

The priming effect may be stronger for some target stimuli than others. As has been mentioned, various types of priming can occur. Semantic priming and form priming acting together on one stimulus would be expected to produce a stronger effect than for form priming or semantic priming alone. For this reason, pseudowords which have no semantic or associative

content are expected to have a weaker priming effect than lexical items, since they cannot be primed by semantic or associative content. They can still be primed through repetition or form priming however.

Therefore we would expect lexical items to elicit a stronger priming effect than pseudowords in participants. The priming effect can be subject to variation in size due to lexical frequency as well (Ostergaard, 1998), where less frequent items can exhibit a stronger priming effect than more frequent items (Scarborough et al., 1979). Such effects are typically not very large, however. In the analysis, we are able to account for this to some extent by introducing a random effects term for items, which estimates variance due to specific images.

4.3.7 Repetition effect

Since each target stimulus was presented twice during the course of the experiment, repetition effects are expected. When participants encounter the same stimulus more than once within the same task, an effect akin to priming occurs (Forbach et al., 1974) where the response to the stimulus is facilitated. This effect suggests a mechanism whereby exposure to a stimulus word causes a temporary alteration in memory, and the lexical retrieval becomes an easier operation.

Such an effect would potentially introduce undesirable levels of noise in our data. In an attempt to eliminate repetition effects in the experiment, the presentation order of unprimed and primed conditions was balanced so that one group of participants would perform the unprimed section first and another group would perform the primed section first. Repetition effects will still be present for each individual participant. The effect should cancel out over items however, since each priming condition is subjected to repetition in equal amounts.

This is not a perfect solution, since we cannot guarantee equal numbers of correct classifications for each stimulus under each condition. Subjects are unlikely to perform identically in their ratio of correctly classified and misclassified stimuli, so a small amount of imbalance may remain. The method was judged preferable to having different participants for the primed and the unprimed section of the experiment, however.

Another repetition effect that will occur in this experiment is the repeated semantic content of the priming stimulus and target stimulus. Since neuter forms of adjectives (or constructed neuter forms) are used as the priming stimulus for each item, any semantic activation caused by the priming stimulus would presumably be the same as for the target stimulus. Any differences observed in the priming effect would therefore indicate the strength of the connection between neuter and common gender forms.

4.4 Stimulus selection

The vocabularies of Norwegian and Swedish overlap to a large extent, and the group of adjectives that do not form neuters are no exception to this rule. As such, a number of adjectives

from Johansson (2003) could be applied to this experiment as well. Some items are of a much lower frequency in Norwegian however, and these were eventually removed and replaced with other adjectives. This was also shown in early runs of the experiment where a number of items displayed higher than expected misclassification. This problem was magnified for the L2 group where participants had variable levels of competency in Norwegian, although participants with complete fluency were excluded. While the early runs of the experiment did yield usable data, this data had a high amount of errors and gaps, and the statistical analysis would have suffered as a result. The stimulus item list was therefore refined and tailored for Norwegian conditions.

The group of adjectives with missing neuter forms appears to be a small group. There is therefore from the start a limited pool of test items to select from. There are two important factors here, discussed in the previous chapter of the thesis. Their phonological properties are restricted and their semantic content is likewise restricted. The language only has so many adjectives that belong to this category. Since the experiment contrasts these to regular adjectives, their similarity in phonology also restricts the number of test items available for the control group of regular adjectives, although the semantic constraint is lifted here. Further factors reduce the amount of viable candidates for the experiment as well.

One problem for the selection of stimulus is that a number of adjectives have corresponding adverbs, as mentioned in section 3.3. Norwegian and Swedish allow for formation of adverbs from adjective stems by the same morphology as the neuter gender forms of these adjectives. It is therefore also important to avoid adjectives that display low frequency as neuter adjectives, but high frequency as adverbs. Such items would introduce an unnecessary complication in the experiment, whether they are classed in the same group as adjectives with missing neuter forms or as regular adjectives.

Since the expected neuter form and the adverbial form of these stems is identical, and we would expect a stem to prime the same stem, that is to prime itself. We might therefore risk detecting an adverb priming an adjective, and this is something we want to avoid as far as possible since it might obscure any gaps in the adjective paradigm. Our methods do not allow us to observe directly what exactly is going on inside the mind of the participant.

Another problem is that adjectives can have several different readings. This is also the case for a number of the adjectives which are relevant for this experiment, as touched upon in the previous section on adjectives. The adjective *god*/good is one such example. It is somewhat vague and can have a number of different meanings. It obeys the phonological rules, being a single-syllable adjective that assimilates the neuter forming *-t* into its stem to form the neuter. A number of its meanings also obey the semantic rule we have postulated, but not all, and the adjective can give different readings. This is illustrated in the following examples:

(28) Et godt syltetøy.
A.N good.N berry/fruit preserve[NEUT].
'A tasty berry/fruit preserve.'

(29) Et godt barn.
A.N good.N child[NEUT].
'A tasty child.'

(30) ?Et godt barn.
A.N good.N child[NEUT].
'A child of upstanding moral character.'

The reading of the adjectives *god/godt* in its neuter forms implies "tasty", to the exclusion of properties that are less directly observable. Example 28 is unproblematic and the reading is straightforward. The second example, example 29, but the meaning of the sentence is awkward, describing a tasty child. The sentence in example 30 is also well-formed, but the reading is not particularly common and might not even be possible (Johansson, 2003).

The adjective *god* is therefore not a very good candidate as it is unclear what its status would be within the experiment. In order to not complicate the experimental parameters needlessly, we want to eliminate potentially dubious adjectives. It is therefore removed from consideration for the stimulus list for the experiment, as are any potential adjectives like it. Even though it does appear to have a well-formed neuter, the behaviour of this neuter form might not be completely straightforward. This further narrows down the amount of total adjectives that are available for the experiment.

4.4.1 Nonwords and pseudowords

In order to present an actual task for participants there also has to be stimulus to classify as not-a-word. For this experiment, the not-a-word target stimuli consisted of pseudowords. Pseudowords are commonly described as constructed words that obey the phonotactic rules of the language but contain no meaning (Simos et al., 2002; Stark and McLelland, 2000; Specht et al., 2003). This is a special case of *nonword* stimulus, as nonwords more generally can also have implausible phonology.

A few examples of pseudowords used in this experiment are *fled*, *glu* and *vrut*. These are all permitted within the phonotactics of Norwegian. While they lack semantic content, they are different from strings of non-permitted phonotactics such as *rlyeh* or *fthagn*. The latter two examples can be superficially judged as not words based on their phonology. The use of phonologically plausible pseudowords therefore removes one obvious cue to their status as nonwords, which is likely to make the task more challenging for participants.

Constructed neuter forms of pseduwords of the examples mentioned would be *fledt*, *glutt* and *vrutt*, in accordance with the rule for forming neuter gender forms of adjectives in Norwegian. As with the lexical items, these constructed neuter forms were used as priming stimulus for the corresponding target stimulus in its hypothetical common gender form.

Pseudowords have some similarities with the missing neuter forms we are examining in this experiment. For one, they are a type of gap (Trask, 1996) since they do not occur even

though they would be phonologically permitted. For this experiment, they also provide a contrast to the rule-based paradigmatic gaps that have been described for adjectives in Norwegian.

These pseudowords were constructed with the same restrictive phonological pattern used for the lexical items in the experiment. As a consequence, all stimuli in the experiment are of similar phonology and similar length, which should also limit potential effects of word length in performing the task. Pseudowords were primed by constructed neuter forms, applying the same rule for generating neuter forms as is common for Norwegian adjectives.

In the data, pseudowords were referred to as *nonwords* since they are used for the not-a-word decision. Unless otherwise noted, the term *nonword* refers to pseudowords in this thesis.

An equal amount of lexical items and pseudowords were used in the experiment so that participants would not be biased towards either category. The presentation of the target stimuli was also randomized for each participant.

4.5 Setup and programming

The experiment was conducted inside of a soundproof chamber. The only source of noise being a ventilation duct. This source was controlled for during the experiment. The experiment was programmed in SuperLab which ran on OS X 10.9 on a Mac Mini. The hardware used included an ultra HD monitor on which stimulus was presented, with a frame rate of ~4ms.

The priming stimulus in the experiment was presented for only 50ms, which is on the shorter end for priming experiments, although it is fairly common with masked priming (Forster et al., 1987). It is therefore to our advantage to reduce the potential effect that latency in the monitor can have on stimulus presentation. Although controlling it to the precise millisecond should be unnecessary, we do want to avoid displaying the stimulus for longer than 60 milliseconds (Forster, 1998).

The experiment was programmed in SuperLab. Instructions on how to perform the task were presented on screen before the start of the task. Participants were also informed that all participation was voluntary, and that they could quit the experiment at any time for any reason. Participants were also debriefed after the experiment and were also given the opportunity to ask detailed questions about the experiment after having participated.

Participants were asked to classify stimulus as words or non-words using a Cedrus RB-540. The response pad's right button was marked with green and the left button was marked with red. The green button was used to classify as *word* and the red button was used to classify as *nonword*. Participants were instructed to make the decision as quickly as possible. Instructions on how to perform the task were also presented to participants orally.

The total amount of time spent on the experiment by each participant was subject to some variation, although it typically took one participant less than 10 minutes to complete the task.

5 Results

The following section deals with the experimental data collected and the statistical analysis methods employed.

Reaction time data was gathered from 30 participants who were L1 speakers, recruited to participate in the experiment. We were also interested in doing a comparison with L2 speakers, however, only 11 participants were recruited before the conclusion of the experiment. The variance for L2 speakers is expected to be higher, as we assume they will have more errors and perform slower on the task on average. This would mean that preferably, we would have to recruit more L2 speakers in order to compensate for the higher variance. The data gathered from L2 participants did provide some usable results however, and will be presented in this thesis.

5.1 Outlier analysis

In order to ensure that the final analysis is accurate, we combed our dataset for outliers in our random variables, *items* and *subjects*. Any extreme outliers in the data runs the potential of skewing our model significantly as it is quite a sensitive method.

Exploratory data analysis methods can be used to visualize and understand the characteristics of a data set. Such methods were employed for the outlier analysis, using plots and figure to visualize the distribution of responses by item and subject. Reaction time data was also used secondarily to observe if any items were on average processed quicker or slower than others, or more importantly if some subjects were notably slower or faster than others. We set a cut-off for correct answers at 350 milliseconds, as any response faster than that might indicate a random key press before the item is fully processed.

5.1.1 Items

Various factors can contribute to how lexical items or even non-lexical items are processed in a lexical decision task. For this reason, an analysis of each individual item used in the lexical decision task is necessary to control for outliers. Any item being overly prone to misclassification by participants would make results from that item unreliable, since we cannot be sure if such items have been reacted to correctly. It is also expected that *pseudowords* may be inherently more difficult to classify than other items and this has been taken into account in the analysis as well.

The classification data was broken down into three different sets. One for each group of adjectives. This allows us to control for the aforementioned expected higher difficulty in classifying *pseudowords*. It also allows us to see if there are any differences between our adjectives that regularly form neuters and those that do not form neuters, as either stimulus class could elicit different behaviour compared to the others. They are therefore analyzed separately.

The data is visualized in association plots in the following section. Any effects observed in the final analysis should not be driven by only a few stimulus items, so it is important that items to not deviate too strongly in their rate of classification from the rest of their respective stimulus classes.

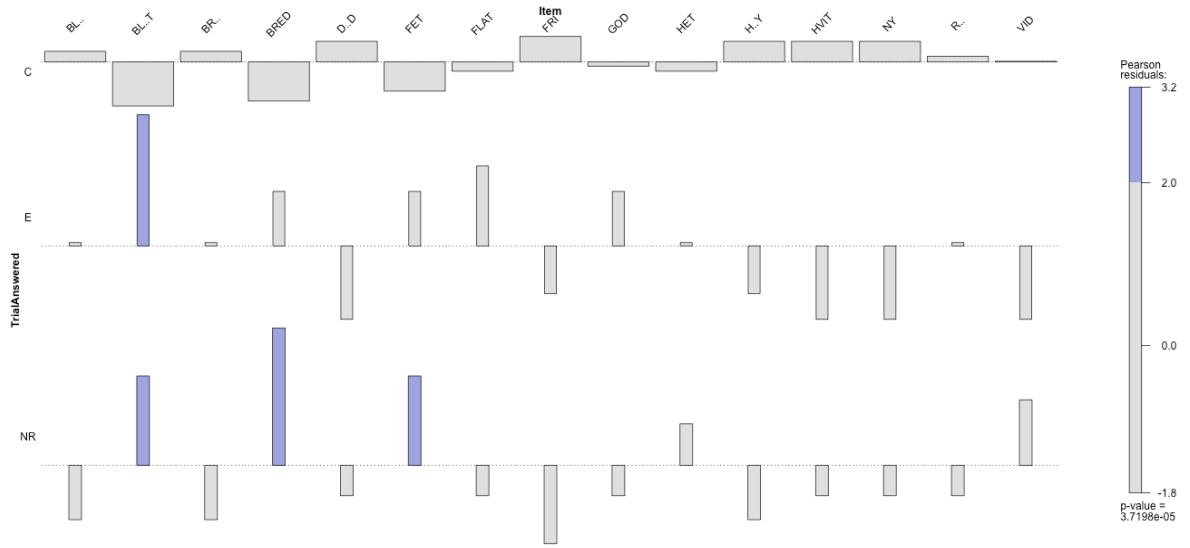


Figure 3: Association plot for responses by item, for *normal* adjectives.

We see no major outliers in the *normal* category. For correct answers, no item stands out, although there are very small effects for some items in the *no response* and *error* categories. This indicates that participants had no major difficulties processing any of the items in this category.

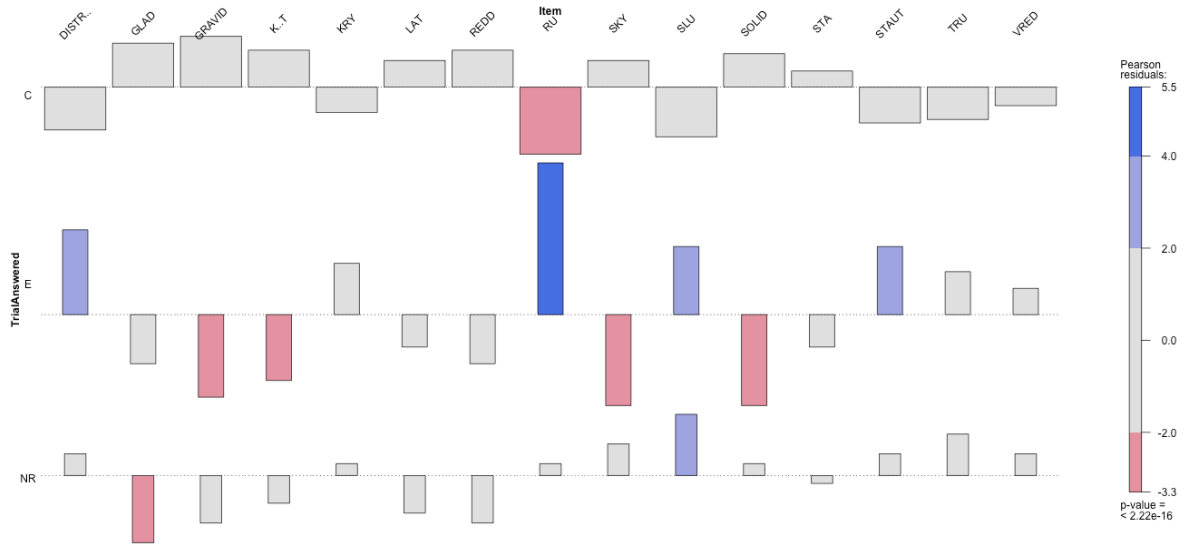


Figure 4: Association plot for response by item, for *problematic* adjectives.

In the *problematic* category we observe slightly more errant behaviour although the only important item that stands out is the adjective *ru* (rough; coarse) with a disproportionately large amount of incorrect classifications. A very large amount of participants classified it as not-a-word. For this reason, the item is removed from the final analysis as it could potentially skew the analysis.

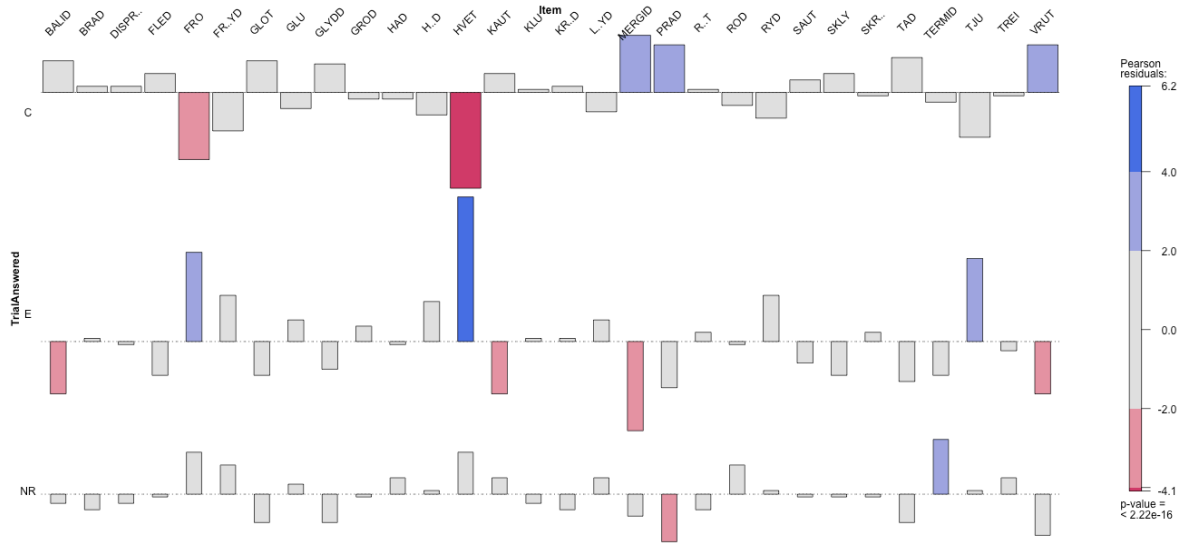


Figure 5: Association plot for responses by item, for *nonword* adjectives.

In the *nonword* category we might inherently expect more difficulty for participants in classifying them correctly, however, this should apply across the board. As mentioned, this motivates splitting our categories into different data sets for the outlier analysis.

Notably, the item *hvet*[7] was frequently misclassified, not entirely unlike *ru* and so this item was also removed from the final analysis.

In order to further balance our data for stimulus class membership, we removed one more item from the *nonword* category. One item from the *normal* category in spite of that category displaying a fairly high degree of uniformity in how often its members were classified correctly. This leaves us with a total of 56 stimulus items in the data, with 14 in the *normal* and *problematic* categories and 28 in the *nonword* category.

A complete visualization of the classifications for every item in the experiment is provided below. Note that nonwords indeed do display a tendency to be misclassified more frequently than the lexical items we selected, showing that there are aggregate differences between the groups of stimulus. A more in-depth analysis of this effect is provided in the discussion section.

[7]It turns out that this nonword item is not an entirely uncommon mispelling of *vet* (know) which might be another reason to remove it from the final analysis.

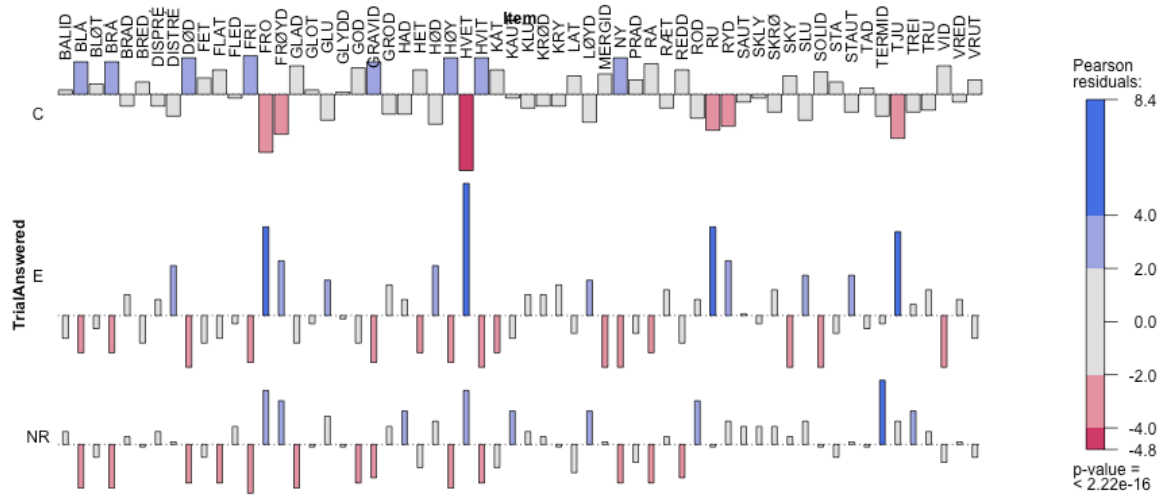


Figure 6: Association plot for responses for all items.

5.1.2 Participants

While there are numerous factors that can contribute to participants performing different from each other, we do not have the means to control for all of them. Generally, we expect our participants to perform at a fairly similar level, but any major outliers could skew the analysis in much the same way that outliers in the items could skew the analysis. The data was checked for outliers amongst subjects as well. A high incidence of incorrect classifications by subjects could also potentially skew the analysis. Further, we need to control for the effects of any potential misunderstanding of how the task was to be performed.

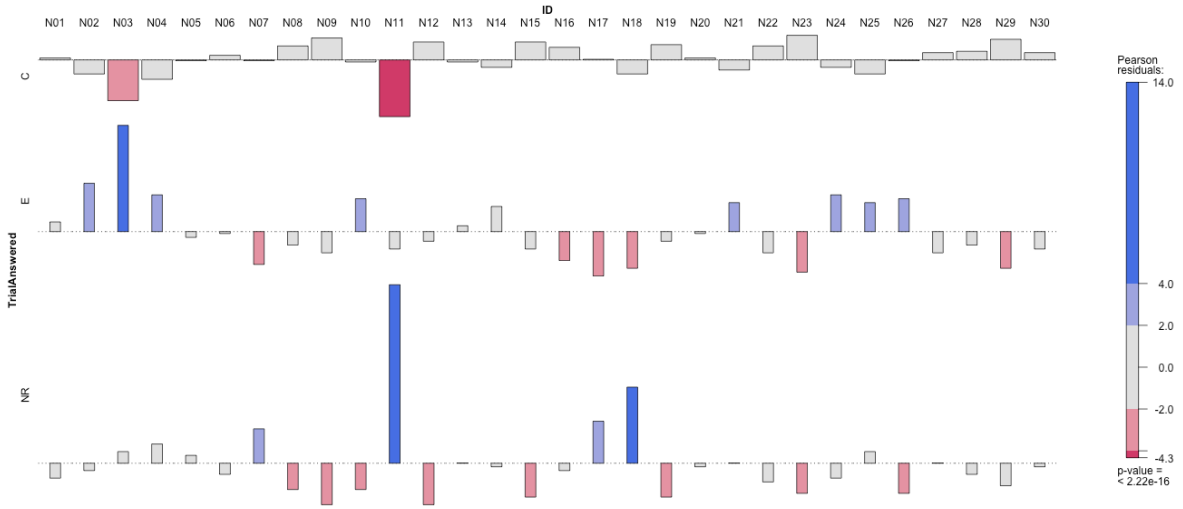


Figure 7: Association plot for responses by subject.

We see that subject N11 stands out pretty clearly from the less, with a very high amount of

no responses in addition to a lower than average score for correct answers. Subject N03 also exhibits a lower than average number of correct answers, but not quite to the same level as N11. Uniquely, subject N11 was the only participant to have fewer correct than non-correct responses with an absolute number of 54 correct responses, 7 errors and 59 no response. Subject N03, while performing slightly better, still only had 66 correct responses, which is just slightly above 50%. All other participants, for comparison, were above 70% correct responses. In order to further investigate this anomaly, we can also look at reaction time data for each individual participant.

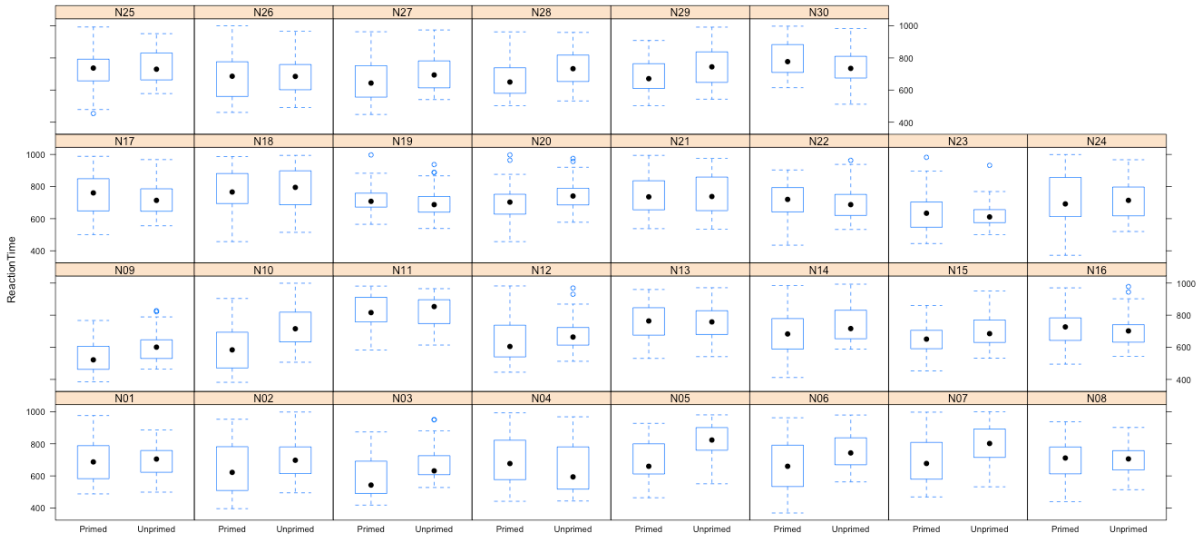


Figure 8: Boxplot of reaction times per subject, divided into primed and unprimed segments.

The plot makes it clear that subject N11 performed slower than the average trend in both segments of the task. It is unclear why this may be the case. Subject N03 displays no anomalous behaviour in their reaction time data on the other hand, but the higher incidence of incorrect classifications makes it uncertain if the instructions were understood correctly. The subject may have been confused or uncertain. Or in the case of too slow response, may have taken too much time to think about how to classify the word. In either case, this reduces the reliability of any data gathered from those participants.

We cannot expect our participants to perform perfectly or even close to perfectly on the task, due to the nature of its design. Participants only receiving one second to make their decision is bound to introduce some error. If participants have too few correct answers however, it becomes difficult to distinguish their results from that of a hypothetical participant that just presses keys randomly in response to each stimulus. Therefore, subjects N11 and N03 are excluded due to low task compliance, as their results are not reliable. This leaves us with a grand total of 28 native speakers of Norwegian for the final analysis.

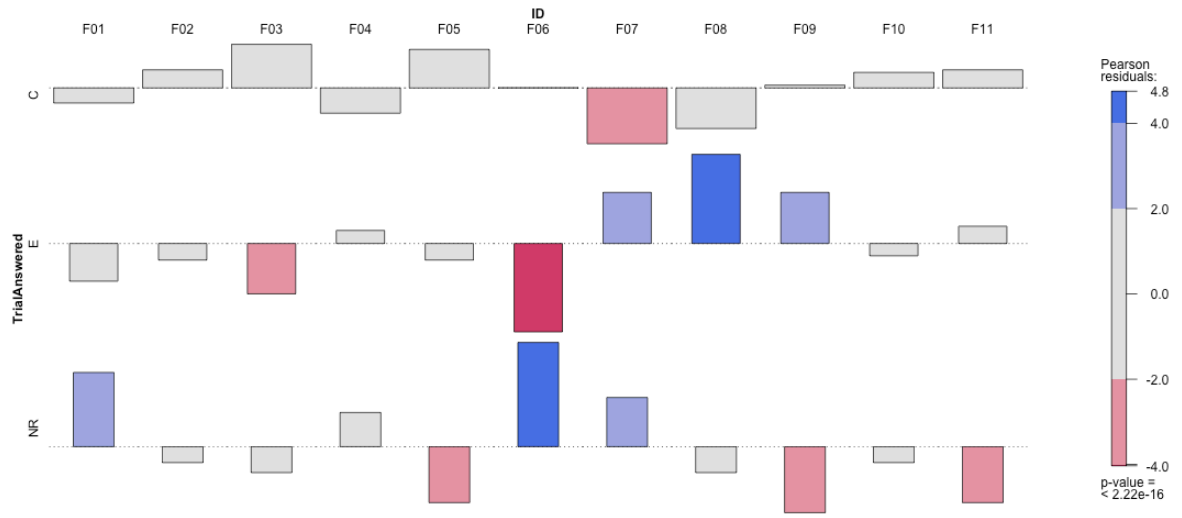


Figure 9: Association plot of responses for L2 participants.

5.1.3 L2 group

An analysis of correct answers by individual participant was performed for the L2 group as well.

We find one participant that stands out with a notably lower-than-average amount of correct responses. While there are larger differences in the no response and error categories, this would indicate that our participants for the most part achieved a roughly equal level of success on the task. Investigating further however, we see that our L2 speakers indeed did make more mistakes than the L1 group. Two L2 (F07, F08) participant had correct answers on less than half of the stimuli and two more (F01, F04) were only slightly above half for correct classifications. This casts serious doubt on the validity on any data gathered from these participants, and so discarding them would be preferred. That would leave us with only 7 participants in the L2 group however, which makes any analysis of the results from that experiment dubious. Participant F06 also stands out in the diagram, but this participant has an average number of correct answers. The participant stands out due to a much higher than average number of time-outs in proportion to incorrect classifications.

This further serves to highlight the issue with higher variance in the L2 group. Preferably, the group would contain more participants than the L1 group to counteract the lower levels of task compliance.

5.2 Comparison of correct answers between language groups

The proportion of correct answers by subject language background and the group that real adjectives belonged to (i.e. *normal* or *problematic* was tested to see if they were distributed evenly.

A chi-square test with Yates correction for continuity found a very slight imbalance in the distribution of correct answeres between L1 and L2 speakers over regular and problematic adjectives (χ^2 = 6.25, df = 1, p-value = 0.01, ϕ = 0.06). Figure 10 below visualizes the distribution of correct answers between language groups.

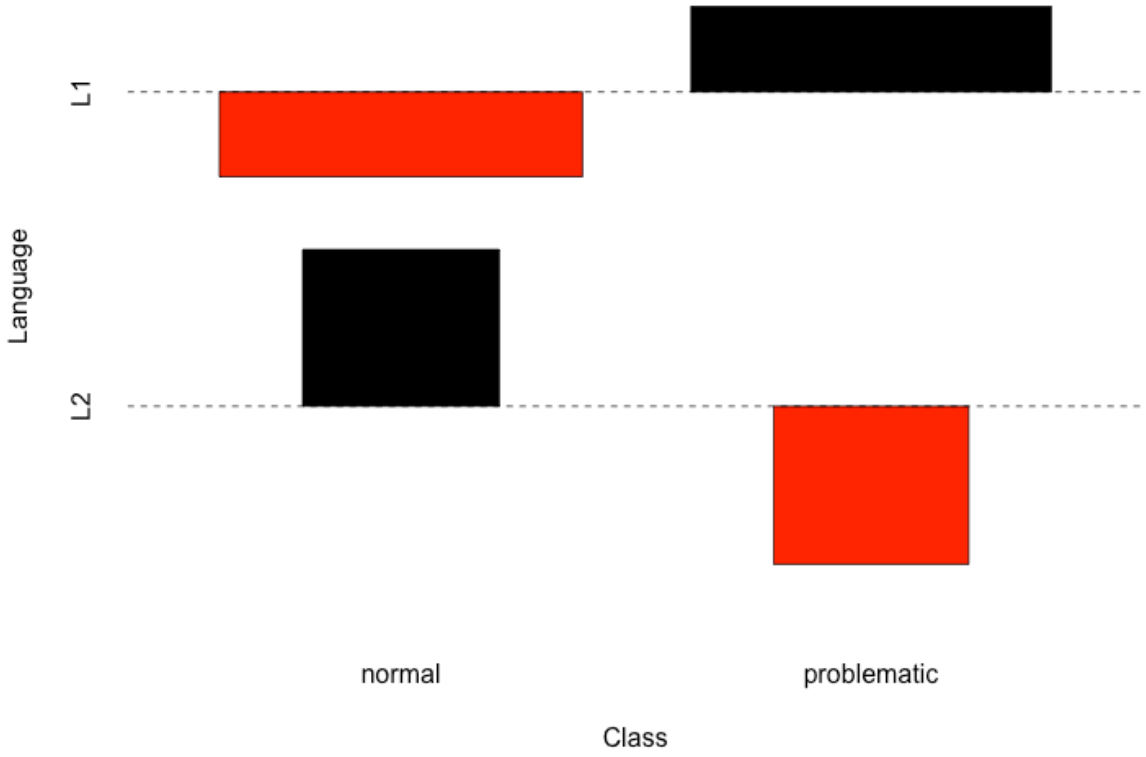


Figure 10: Association plot for correct answers of adjectives by language group.

We notice that L2 speakers are prone to more commonly misclassifying *problematic* adjectives than L1 speakers, while L1 speakers appear to be more prone than expected to misclassifying *normal* adjectives than *problematic* adjectives. While significant (χ^2 = 6.3, df = 1, p < 0.05), the association is very weak at Φ=0.06. This makes it difficult to interpret the results. It may be the case that L2 speakers have more difficulty processing the *problematic* target stimuli which could be due to a variety of factors. Frequency is one possible explanation as the stimuli used in the experiment are not perfectly balanced for token frequency. As mentioned previously, it would be very difficult to achieve this due to the limitations imposed by the restrictive phonology of the adjective group we are investigating. The data from the experiment cannot in itself make this conclusion however.

Greater difficulty in processing the *problematic* adjectives is elaborated on in the regression analysis. We observe that the *problematic* class of adjectives are processed more slowly than

the *normal* class of adjectives by quite a strong margin. Another factor that may be influencing this processing is more complex semantics than the *normal* adjectives.

A full list of absolute frequencies of correct answers per language group is provided in the tables below, with further analysis. The data was split into one set for *unprimed* and one set for *primed* items in order to provide a more detailed analysis, so we can investigate whether priming has an effect on participants' ability to correctly classify lexical items.

	L1	L2
normal	363	117
problematic	383	91

Table 3: Table of association for *unprimed* items.

The first table shows data from the unprimed portion of the experiment. Note that we do observe a higher proportion of correct answers for problematic adjectives for L1 speakers, while the opposite is true for L2. The data from the unprimed portion set of targets was found to be not significant (χ^2 = 3.5, df = 1, p > 0.06) for any imbalance in distribution between L1 and L2 speakers.

	L1	L2
normal	353	123
problematic	368	99

Table 4: Table of association for *primed* items.

The same trend is observed in the data set for primed items. L1 speakers are overrepresented for correct answers in the *problematic* category and L2 speakers are underrepresented for correct answers in the same category. Once again the results are not significant (χ^2 = 2.3, df = 1, p > 0.13).

Finally, a table of correct answers for both data sets added together is provided.

	L1	L2
normal	716	240
problematic	751	190

Table 5: Table of association for both primed and unprimed items.

Once again, the aggregate data was found to be significant (χ^2 = 6.3, df = 1, p < 0.05) though the weak association makes it unclear what is happening. Further analysis of the data did not shed any further light as to what might be causing the imbalance.

5.3 Regression analysis

The reaction time data for L1 speakers was analyzed with a mixed-effects model with participants and items as crossed random effects (Baayen et al., 2008) with priming and item group

(*normal*, *problematic* or *nonword*) as fixed effects. Correct classifications of the *nonword*[8] group was kept for part of the analysis since we are interested in how pseudowords behave when controlled against our lexical items.

We are interested in modelling the behaviour of just lexical items by themselves, and in modeling the behaviour of pseudowords compared to lexical items. For this reason, the analysis will be done in several steps. The L2 group as well gets its own model, and we will also use a total model for all data to control the effects of language background of the participants.

The regression analysis was implemented in R through the lmerTest 2.0-33 package (Kuznetsova et al., 2016). This package implements Satterthwaite approximations for degrees of freedom (Satterthwaite, 1946) which allows us to examine specific effects in the model. Coefficients for fixed variables in the model do not provide sufficient detail to examine the differences between each level of the variables however. Therefore, when significance was detected in the model (using ANOVA), further post-hoc testing was performed. The post-hoc analysis was performed using Tukey's range test. This allows us to do pairwise comparisons of means for each level of each fixed variable in the model.

The choice of method for the post-hoc analysis is motivated by the data from the experiment containing gaps so that each mean will not necessarily have the same amount of data points for each item or subject. As per the outlier analysis described in the previous section however, the most extreme examples have been removed and the arguments contain fairly similar levels of information.

5.3.1 Defining models

Individual differences are a source of variation for linguistic features (Clark, 1973) and so we need to account for these differences in our modeling in order to minimize the chance of false positives in our analysis. Therefore we use mixed-effect models that incorporate both fixed and random effects. These need to be handled with some care as the random effects must be specified precisely. Random effects allows our models to account for variation within variables that are repeated, both for individual participants and for individual items which in our case is lexical items. For modeling random effects, we define both random intercepts and random slopes (Baayen et al., 2008).

Note that while random intercepts are called 'random', they are actually a random effect for the intercept, which allows us to control for variance in a more specific way than without the random effect term. Random slopes are similar, but control the variance in a closer manner by assuming that each level of the slope is a repeated measure for the level of that factor.

Random intercepts allow the intercept to vary across the random effects specified. In our case, that is participants and target stimulus (words and nonwords). Instead of estimating an intercept from the sample drawn, our model attempts to estimate an intercept based on

[8]Note that the nonword group consists entirely of pseudowords.

the population that the sample is drawn from, although this does require the model to make assumptions (Barr et al., 2013). It should be noted that random intercepts on their own run the risk of producing anti-conservative estimates (Schielzeth and Forstmeier, 2009) since they cannot account for independence between each individual item or subject in the study.

Our ability to define models depends on what our data will allow, but another concern is finding a model that best fits the data. While it has been argued that models should be maximally fitted if at all possible (Barr et al., 2013) this must also be balanced against the risk of overspecified models producing type II errors (Matuschek et al., 2017).

Only correct answers are used for this analysis as incorrect responses are unreliable.

The mixed-effects models that were fitted to the data will be presented in each section of the analysis using syntax from R.

5.3.2 Lexical items for L1

Lexical items consist of the stimuli in the *normal* and *problematic* category. For this analysis the nonword stimuli were excluded. The model specified for this data treats the priming condition and the class membership of items as fixed effects, with crossed random effects for items and subjects and an interaction term for items and subjects. Random slopes for priming condition and item class were used for subjects (ID). The model is specified with the following syntax:

ReactionTime ~ Priming * Class + (1|Item) + (Class * Condition|Subject) + (1|Item:ID)

Other models were considered. While it was possible to fit a slope for class membership by item, this had very little effect on the final analysis. A model comparison of the two was performed with ANOVA in R. The model with only the intercept for items is labeled Model 1 whilst the model with the random slope term is labeled model 2.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr (>Chisq)
Model 1	17.00	17502.26	17591.80	-8734.13	17468.26			
Model 2	19.00	17504.78	17604.85	-8733.39	17466.78	1.48	2.00	0.48

The difference between models is not significant. Moreover, the simpler model actually appears to have a better fit than the more complex model according to both AIC and BIC parameters. These numbers can give us some indication as to which model has the better fit, the underlying assumptions in the model have to be considered as well. The inclusion of a random slope for class membership does not seem to be strongly motivated from the analysis. There may be further theoretical considerations here. Words are presumably nested within word categories (Barr et al., 2013) and we would expect a random slope term to reflect that. We may therefore, in fact, prefer the simpler and more parsimonious model.

Inclusion of a random slope for priming condition was not successful. Such models either failed to converge or in the best case scenarios were unidentifiable. This is presumably due to

a limitation in the data gathered in the study. The problems with fitting a random slope effect for priming in the model is likely due to loss of data when removing outliers as this reduces the amount of pairs in the data for the priming condition. The lack of a random slope for priming is considered in the discussion. The results are likely to still be tolerable, since we expect most of the variance to stem from participants and their reactions to items and the priming condition.

Running an ANOVA on the model yields the following results:

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
Class	60951	60951	1.00	27.76	6.45	< 0.05
Condition	81224	81224	1.00	26.90	8.60	< 0.01
Class:Condition	68358	68358	1.00	37.12	7.23	< 0.05

All levels are significant. Class membership at $p < 0.05$, priming condition at $p < 0.01$ and the interaction between these two at $p < 0.05$. The reaction times estimated by the model are 624 milliseconds for *primed* items in the *normal* category ($p < 0.001$), 673 milliseconds for *primed*items in the *problematic* category ($p < 0.01$), 671 milliseconds for *unprimed* items in the *normal* category ($p < 0.001$) and 689 milliseconds for *unprimed* items in the *problematic* category ($p < 0.05$). The p-values tell us that there is a statistically significant effect compared to the baseline.

Model coefficients are presented in greater detail in the table below.

	Estimate	Std. Error	df	t value	Pr (>|t|)
(Intercept)	623.50	15.23	41.53	40.95	< 2e-16
ClassProblematic	48.87	14.78	36.65	3.31	< 0.01
ConditionUnprimed	47.30	12.34	27.62	3.83	< 0.001
ClassProblematic:ConditionUnprimed	-31.25	11.62	37.12	-2.69	< 0.05

After having performed an ANOVA, the further post-hoc testing was performed on the model. In this process, we compare each combination of categorical fixed variables in the model to see which means differ significantly from each other. This allows us further insight into what factors cause the effects we observe in the ANOVA. A table of comparisons is presented below.

	Estimate	Std. Error	z value	Pr (>|z|)
Unprimed.Normal - Primed.Normal	47.30	12.34	3.83	< 0.001
Primed.Problematic - Primed.Normal	48.87	14.78	3.31	< 0.01
Unprimed.Problematic - Primed.Normal	64.92	18.05	3.60	< 0.01
Primed.Problematic - Unprimed.Normal	1.57	15.83	0.10	0.99
Unprimed.Problematic - Unprimed.Normal	17.62	13.85	1.27	0.55
Unprimed.Problematic - Primed.Problematic	16.05	12.20	1.32	0.52

We notice that the comparison between unprimed *problematic* and primed *problematic* finds no significant difference between the means. That is, we cannot the null-hypothesis that there is no facilitation or inhibition due to priming with neuter gender forms for the *problematic* group. This contrasts with the *normal* group where we do observe facilitation due to

priming. We also note that the difference in means between unprimed *normal* and unprimed *problematic* is not significant, meaning that the difference between the two classes is primarily driven by the interaction between priming and class membership. That is, the facilitation observed due to priming for *normal* adjectives sets them apart from the *problematic* adjectives which experience little or no facilitation.

5.3.3 L1 with pseudowords

A separate analysis for data including pseudowords was performed. In this way we can control lexical items against just each other in addition to controlling against the effect of non-lexical items, ie. pseudowords. The procedure used for modelling this data is the same as previously, and the data from the lexical-only model is nested within the data used in this part of the analysis. One difference is that this time, the random slope for class by item caused the model to fail to converge, and so it is left out of the analysis. The final formula for the model is the same as previously:

ReactionTime ~ Priming * Class + (1|Item) + (Class * Condition|Subject) + (1|Item:ID)

The major difference is that *class* is now a factor with three levels whereas it previously had two. The analysis is therefore slightly more complex although for the most part it returns very similar results. The interaction effect for class and condition is split into two, however, and this has some consequences for our model. The coefficients for fixed effects in the model is presented in the table below.

	Estimate	Std. Error	df	t value	Pr (>|t|)
(Intercept)	745.29	14.27	34.99	52.25	< 2e-16
ClassNormal	-121.69	15.72	48.36	-7.74	5.19e-10
ClassProblematic	-73.22	15.13	50.11	-4.84	1.29e-05
ConditionUnprimed	34.54	11.32	27.49	3.05	0.005
ClassNormal:ConditionUnprimed	12.70	12.76	27.65	0.99	0.33
ClassProblematic:ConditionUnprimed	-18.30	11.53	33.31	-1.59	0.12

In this case the model has estimated the intercept from the average time for pseudowords. The effects for class membership and priming are still significant in this model, however, we also see that there are two interaction effects and that they are smaller and not significant. This model estimates the intercept from primed pseudowords whereas the previous model that only uses lexical items estimates it from primed *normal* adjectives. This causes the effects to be treated differently in the models. For a visualization of this, refer to figure 11. The lack of significance in the table of coefficients is less interesting as simply redefining which factor the model estimates the intercept from will give different results.

A full account of the effects estimated in the model are as follows: 623 milliseconds for primed *normal* adjectives ($p < 0.001$), 672 milliseconds for primed *problematic* adjectives ($p <$

0.001), 745 milliseconds for primed *pseudowords* ($p < 0.001$). For unprimed stimuli, the model estimates 671 milliseconds for *normal* adjectives ($p > 0.05$), 689 milliseconds for *problematic* adjectives ($p > 0.05$) and 780 milliseconds for *pseudowords* ($p < 0.01$). Note that the estimates of effects for unprimed *normal* and unprimed *problematic* may not be reliable in this model but the estimates are identical to the model for lexical items in which those effects are significant. On the other hand, there is a strong and consistent effect for item class membership.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
Condition	102652.77	102652.77	1.00	26.97	11.16	< 0.01
Class	640774.52	320387.26	2.00	56.96	34.83	< 0.001
Condition:Class	64132.02	32066.01	2.00	33.46	3.49	< 0.05

All three variables are shown to be significant in the ANOVA. This gives us an indication that something is happening in each category. In order to investigate further, a post-hoc analysis to see which means significantly differ from each other. The procedure is quite the same as for the lexical items only, but with one more level for item class. A table is provided below.

	Estimate	Std. Error	z value	Pr (>|z|)
Normal.Primed - Nonword.Primed	-121.69	15.72	-7.74	< 0.001
Problematic.Primed - Nonword.Primed	-73.22	15.13	-4.84	< 0.001
Nonword.Unprimed - Nonword.Primed	34.54	11.32	3.05	0.025
Normal.Unprimed - Nonword.Primed	-74.46	16.33	-4.56	< 0.001
Problematic.Unprimed - Nonword.Primed	-56.99	16.76	-3.40	0.008
Problematic.Primed - Normal.Primed	48.47	13.88	3.49	0.006
Nonword.Unprimed - Normal.Primed	156.23	18.09	8.64	< 0.001
Normal.Unprimed - Normal.Primed	47.24	12.33	3.83	0.002
Problematic.Unprimed - Normal.Primed	64.71	17.45	3.71	0.003
Nonword.Unprimed - Problematic.Primed	107.76	16.94	6.36	< 0.001
Normal.Unprimed - Problematic.Primed	-1.23	14.88	-0.08	1
Problematic.Unprimed - Problematic.Primed	16.24	12.39	1.31	0.74
Normal.Unprimed - Nonword.Unprimed	-109.00	15.14	-7.20	< 0.001
Problematic.Unprimed - Nonword.Unprimed	-91.53	13.78	-6.64	< 0.001
Problematic.Unprimed - Normal.Unprimed	17.47	13.12	1.33	0.73

Note that we again observe that there is no significant difference between the means for primed and unprimed *problematic* adjectives. Once again we cannot reject the notion that priming problematic adjectives with potential neuter forms causes any facilitation or inhibition in the processing of such forms. We do however found a significant effect for facilitation from priming for the *normal* and pseudoword items with neuter forms. This indicates that there are differences in processing between classes.

Further, there is a significant difference in means between *normal* adjectives and pseudowords in all configurations. This effect is very likely driven by the class parameter alone however as pseudowords are processed much more slowly than both categories of lexical items.

There also appears to be no significant difference in means between unprimed *normal* and unprimed *problematic* adjectives in this model as well. Showing that the difference in class for the lexical items is mainly due to the different effect of priming that they experience.

The line plot below is provided to visualize the differences between the varying classes and priming. While the trend points in the same direction for all classes, the priming effect for *problematic* adjectives is dubious and not significant as previously mentioned. The pure difference in means between primed and unprimed items in this category is not more than 18 milliseconds which in any case would be a very weak effect.

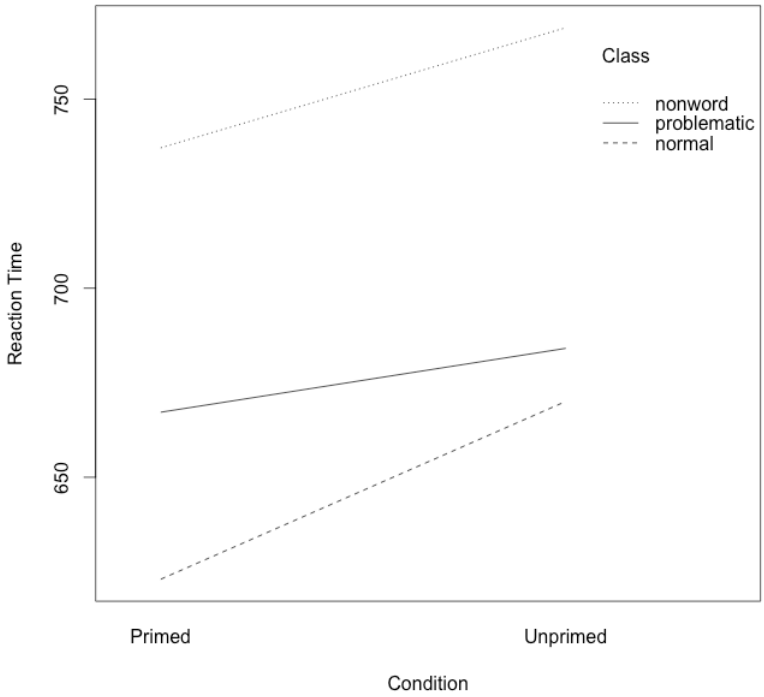


Figure 11: Line plot for mean reaction times by item class and priming condition.

5.3.4 Association of correct responses for L1

Further analysis was performed on whether class membership or priming influences the amount of correct and incorrect responses from participants. We test whether there are any differences between groups of stimuli. The total amount of responses (n = 3136 after removing outliers) is presented in the table below. Note that for the regression analysis performed previously, only correct responses were used.

	Nonword	Normal	Problematic
C	1209	747	685
E	174	15	52
NR	185	22	47

At first glance we notice that nonwords are more frequently classified incorrectly or receive no response as compared to lexical items. There is also a smaller difference between *normal* and *problematic* items in the proportion of correct to incorrect answers, with *normal* adjectives having a higher total amount of correct answers as well. The association is significant (χ^2 = 139, df = 4, p < 0.001, ϕ = 0.21) but is mainly driven by the imbalance in the nonword category. A test of association where the lexical items are treated as one vector yields a similar result (χ^2 = 119, df = 2, p < 0.001, ϕ = 0.20). This would indicate that pseudowords are more difficult to process than lexical items, which is consistent with the overall slower reaction time to pseudowords in the regression analysis.

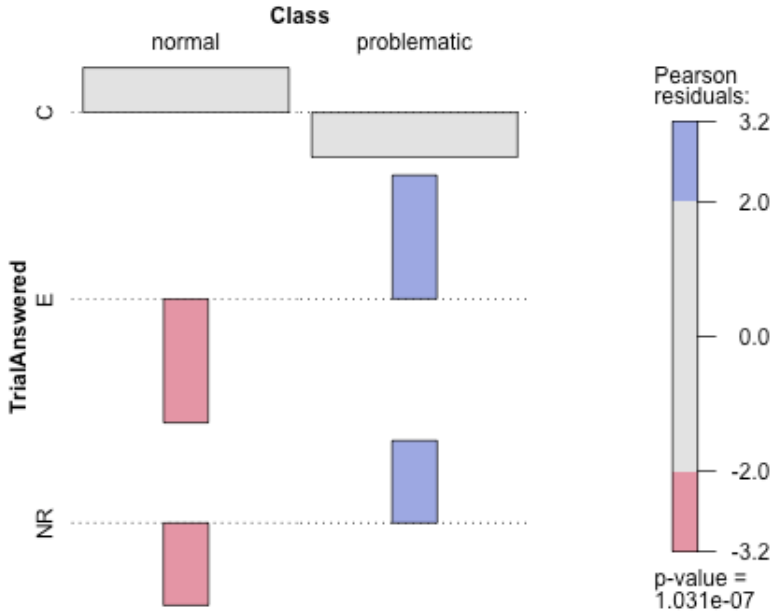


Figure 12: Association plot for classification of lexical items for L1.

Further testing is performed for the lexical items. The data shows that *normal* adjectives have a lower proportion of errors and no responses then expected, when compared to *problematic* adjectives, and this is shown to be significant (χ^2 = 32, df = 2, p < 0.001) although the effect (ϕ = 0.14) is weaker than for the comparison with nonwords. This is also an indication that participants encountered slightly more difficulty with the *problematic* class than the *normal* class of stimuli. This is also in accordance with the slightly slower reaction times to *problematic* adjectives in the regression analysis, though it should once again be noted that the difference in means between unprimed *problematic* and unprimed *normal* was shown to be not significant in the post-hoc analysis. We do observe that *problematic* adjectives are different in the baseline, by the association of correct to incorrect answers when compared to the *normal* category, but we do not observe any such effect in the mean reaction times for

these two categories. This is in contrast to pseudowords which more clearly display effects that indicate that they are more difficult to process.

The *problematic* group of adjectives did also display slightly more misclassified items under the primed condition than under the unprimed condition. This effect however was not significant (χ^2 = 4, df = 2 p = 0.13).

5.3.5 Analysis for L2

While the data gathered for L2 is severely limited as shown in the outlier analysis, it might nonetheless be of some use to perform exploratory analysis of the usable data we do have. There might be trends or tendencies that we can discover. The main question of interest is whether L2 speakers will process items in the task in the same way as L1 speakers, or if there is a difference between the two. Language background is itself a factor as shown in the outlier analysis with L2 speakers in general exhibiting lower task compliance than the L1 group.

We proceed by fitting a model for the lexical items in the L2 group first. The model is fitted according to the same criteria as the one used for L1 speakers for consistency. Running an ANOVA on the model yields the following result:

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
Class	39309.82	39309.82	1.00	26.20	4.44	0.04
Condition	50395.31	50395.31	1.00	9.01	5.69	0.04
Class:Condition	7075.80	7075.80	1.00	13.28	0.80	0.39

We observe significant effects for priming and class membership ($p < 0.05$) but the interaction term between the two is not significant. A further pairwise post-hoc analysis is unable to detect any significant difference in means apart from one, as shown in the table below.

	Estimate	Std. Error	3z value	Pr (>|z|)
Problematic.Primed - Normal.Primed	49.96	22.20	2.25	0.10
Normal.Unprimed - Normal.Primed	54.68	22.74	2.41	0.07
Problematic.Unprimed - Normal.Primed	85.88	27.75	3.10	0.01
Normal.Unprimed - Problematic.Primed	4.73	26.32	0.18	1.00
Problematic.Unprimed - Problematic.Primed	35.93	20.60	1.74	0.28
Problematic.Unprimed - Normal.Unprimed	31.20	21.65	1.44	0.45

The lack of significant effects is likely to be at least partially due to a too small sample size of participants. The only significant difference in means detected in the post-hoc analysis is between unprimed *problematic* adjectives and primed *normal* adjectives. The estimated effects are in fact larger than for the L1 group, but we do not have enough data to tell whether it falls within the confidence interval of the model.

A further analysis including the *nonword* condition yields no results beyond this. The L2 group is, just as the L1 group, slower in classifying the *nonword* category than lexical items.

This contrast does not indicate much more than that they take longer to recognize pseudowords than lexical items however.

5.3.6 Association of responses for L2

As with the L1 group, an analysis of the proportion of classifications was done for the L2 group as well. As with the L1 group, a table of absolute numbers of responses is provided. It is apparent that the L2 group exhibits a larger proportion of incorrect classifications or no-response than the L1 group from a cursory glance.

	nonword	normal	problematic
C	337	288	188
E	184	21	91
NR	139	21	51

The tendency to misclassify pseudowords is magnified for the L2 group. As was mentioned in the outlier analysis, a lot of data was lost from the L2 group. A test of association shows a significant (χ^2 = 129, df = 4, p < 0.001, ϕ = 0.31) imbalance in the distribution of answers. In fact, for the L2 group we observe that pseudowords have 337 correct classifications and 323 incorrect classifications. This may suggest that pseudowords may be too challenging for L2 participants. This is further elaborated on in the discussion section.

The L2 group also has a tendency to misclassify the *problematic* adjectives compared to the L1 group. This effect was also larger for for L2 than for L1 (χ^2 = 69, df = 2, p < 0.001, ϕ = 0.34) as shown in the association plot below.

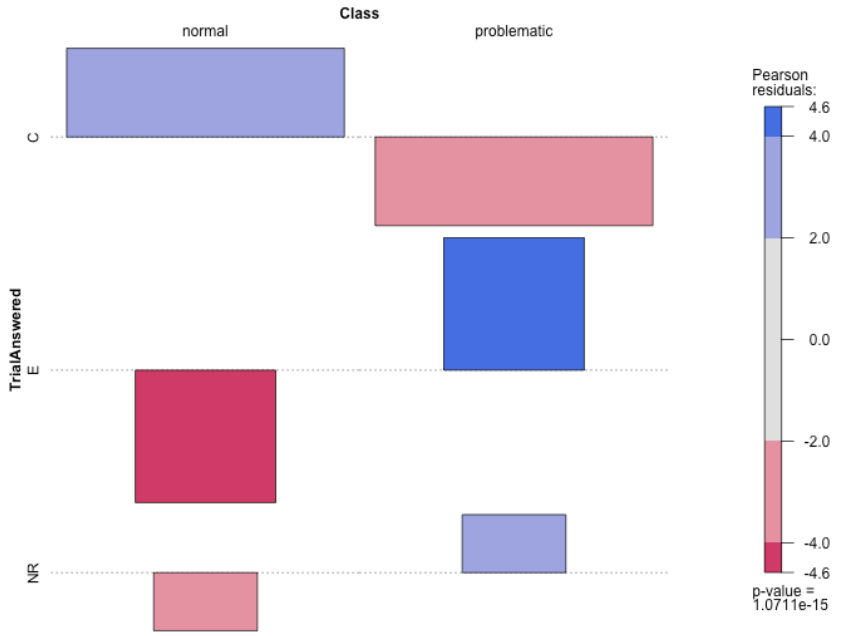


Figure 13: Association of responses for lexical items, for L2

The stronger effect for L2 may suggest that frequency plays a role. If L2 speakers are not as familiar with *problematic* words as they are with *normal* words, they would be more prone to misclassifying them.

5.3.7 L1 and L2 compared

We are also interested in comparing L1 and L2 groups against each other. For this analysis, we add the factor *language* as another variable in the model to account for the language background of our participants. This further means that we have another effect with which to specify a random slope for the by-item and by-subject random effects. Doing so however might not be an advantage, as shown in the analysis below.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr (>Chisq)
Model 1	36.00	40830.40	41050.69	-20379.20	40758.40			
Model 2	38.00	40832.63	41065.15	-20378.31	40756.63	1.77	2.00	0.41
Model 3	43.00	40842.33	41105.45	-20378.16	40756.33	0.30	5.00	1.00

Models 2 and 3 contain slopes for the effect of language for the by-subject and by-item random effect respectively. Both in one model was not viable with our data. Model 1 for comparison has no random slope for language, and we see that the differences between all three models are very small. The AIC and BIC is also lower for the less specific model, indicating that there is less information loss in the model when a random slope term for language is defined. This might not be entirely unexpected since language competence is a factor that could already be nested within subjects. Once again, the more parsimonious model is preferred. The formula is specified with the following R code:

```
ReactionTime ~ Class * Condition * Language + (1|Item) + (Class * Condition|ID) + (1|Item:ID)
```

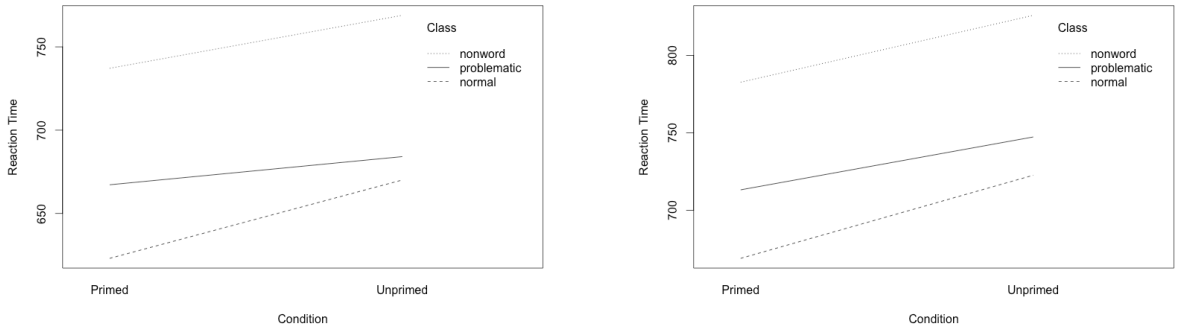
Once again, defining slopes for the by-item random effect causes convergence errors for the priming condition, while a random slope for class membership is excluded on the same basis as in previous models. The analysis contains both pseudowords and lexical items. Running an ANOVA on the model gives us the following results, shown in the table.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
Class	700096.58	350048.29	2.00	72.95	38.32	< 0.001
Condition	138382.29	138382.29	1.00	37.32	15.15	< 0.001
Language	109282.10	109282.10	1.00	37.21	11.96	< 0.01
Class:Condition	40064.94	20032.47	2.00	56.70	2.19	0.12
Class:Language	7360.37	3680.18	2.00	49.56	0.40	0.67
Condition:Language	3770.53	3770.53	1.00	37.31	0.41	0.52
Class:Condition:Language	1498.93	749.46	2.00	56.65	0.08	0.92

We only find significant effects for the individual fixed effects while none of the interaction terms are significant in this model. We see in this model that there is a significant effect for

the language background of participants. In order to examine this further, we do a post-hoc analysis to see which terms are significant and what the effect sizes are.

While L2 is on average about 50 milliseconds slower than L1, the only notable difference when removing the effect of language background is that the *problematic* line is slightly steeper for the L2 group, although as shown in the post-hoc analysis, the difference in means was found to be not significant (p = 0.28). A number of factors may contribute to this. There may simply be no or very little difference between L1 and L2 groups in how they react to the priming stimulus, or the failure to obtain the desired amount of L2 participants may be to blame. Further investigation may be warranted. The figures below are used to visualize the differences between L1 and L2 groups.



(a) L1

(b) L2

Figure 14: Line plots for reaction times for L1 and L2 groups.

5.4 Model tests

During the course of analyzing the data, the models used were further tested for goodness-of-fit by testing if the residuals were normally distributed. These tests were perfomed for each model, but for ease of reference the data is provided in this section for all models. This was performed using the Kolmogorov-Smirnov test on model residuals. The means of residuals for each model were also calculated, although we expect them to be close to 0 in all cases.

Model	D	p	Mean
L1lex	0.067	< 0.001	~0
L1pseu	0.047	< 0.001	~0
L2lex	0.037	0.62	~0
L2pseu	0.021	0.9	~0
L1L2lex	0.055	< 0.001	~0
L1L2pseu	0.037	< 0.001	~0

Table 6: Table of residuals for all models.

The code *lex* refers to models that are fitted to lexical items only while the code *pseu* refers to those models that include pseudowords. We observe that there are differences between the

models in this case too.

All models have a very low D, which tells us that they are all very good fits for the data. What little discrepancy occurs is towards the tail-end of the distribution for the L1 data, whereas for L2 this is distributed across the entirety of the data.

6 Discussion

This section is dedicated to discussing the implications of the results from the experiment, given the theoretical basis provided in the previous sections.

6.1 Different effects for priming

Perhaps the most significant result is the observation that for the L1 group, we find significant effects for priming for *normal* adjectives and for *nonword* adjectives, but not for the *problematic* group. This suggests that the status of expected neuter forms of these items is somehow different from both of these two groups in a fundamental way. We expect there to be some effect of priming due to similarity of phonological form between priming stimulus and target stimulus. One of the advantages of masked priming is that it makes clear the effects of visual priming as compared to unmasked priming (Forster and Davis, 1984a; Forster, 1998).

The priming effect we observe for *nonwords* is likely due to visual matching, since they contain no semantic or associative content. Morphologically related forms and in particular forms that are related through inflection (ie. walks - walked) have been shown to facilitate lexical decision when used as priming stimulus (Feldman, 1994; Fowler et al., 1985). Pseudowords do not typically display this property however[9], and so we can exclude that as an explanation for the facilitation we observe in our *nonword* group. This leaves us with form priming as the likely cause of facilitation. We can use this as a baseline to compare against the *normal* and *problematic* lexical items.

It is unlikely that visual matching occurs unevenly between the groups due to difference in forms being identical for all items. This means that we must explain the stronger facilitation for *normal* adjectives and the lack of facilitation observed for *problematic* adjectives. The observed differences in effects is likely due to other factors playing a role. The suggestion is that presentation of the priming stimulus can cause activation of several different processes that act together to facilitate or inhibit retrieval of the target stimulus.

As has been mentioned in the chapter on methods, there is some evidence that semantic activation does not occur for masked primes (Bentin et al., 1995; Deacon et al., 2000) or that it is typically quite small (Forster, 1998). Note that semantic effects from priming in lexical decision tasks are magnified when increasing the total amount of stimuli when using unmasked priming but that this does not occur with masked priming (Perea and Gotor, 1997). With neuter

[9]There would be no obvious paradigm to place the prime stimulus into.

gender forms being used to prime common gender forms in this experiment, it is likely that the lemma representation of lexical items are partially activated by the priming stimulus. The word activates itself rather than a related semantic connection.

6.1.1 Differences between lexical items

The analysis for lexical reveals two important findings. There is a difference in how the two classes of lexical items, *normal* and *problematic* are processed, and that this difference is driven primarily by participants' reactions to their neuter forms. The *normal* group shows a fairly strong (47ms) and significant ($p < 0.001$) facilitation effect due to priming by their corresponding neuter forms. For the *problematic* group of adjectives that avoid neuter-formation, there was no significant effect for priming. There was a small difference in means between the priming condition for the *problematic* group (16ms) but this difference was not significant either.

There is little difference between the two groups in the unprimed condition as well. There is a small (17ms) but insignificant ($p = 0.52$) difference between means for the two groups. While the reaction time data cannot tell us much about any difference between the groups in this condition, participants did have a propensity for misclassifying the *problematic* group as compared to the *normal* group as shown in the analysis of association of correct answers by group. While weak (ϕ=0.14), the effect was significant ($p < 0.001$) and hints at some underlying tendency for the *problematic* group to be slightly more difficult than the *normal* group. This might also be the cause of the difference in means of reaction time between the groups. The aggregate frequency of lexical items in the *normal* and *problematic* groups is unlikely to be completely identical, and so some discrepancy due to different frequency effects is expected. The slightly slower reactions to the *problematic* group in the baseline condition is consistent with that group having slightly lower aggregate frequency in language use (Grainger, 1990).

The difference between the two groups is very salient in the primed condition however. There is facilitation for *normal* adjectives but not for *problematic* adjectives from the neuter gender form. As mentioned, we would expect some level of visual matching to occur for both groups. The neuter forms and common gender forms are graphemically similar, with the neuter form only adding a *-t* or in a few cases *-tt* to the end of the word. We must therefore conclude that there is some other process affecting how these two categories are processed in response to priming in addition to any visual matching that occurs.

For the *problematic* adjectives, there may be an effect that inhibits the retrieval of the common gender form when exposed to a neuter gender form. This inhibition would be canceled out by any visual form matching, which may be why we observe no effect of priming for this group. Since several types of priming can interact and produce stronger effects (Forster, 1998), the opposite may also be true for an interaction. Indeed, this is what we observe in the model where the interaction between item class and unprimed condition yields a significant result in the opposite direction of what we would expect for *problematic* adjectives.

Negative priming, where a reaction to target stimulus is inhibited by reacting to a priming

stimulus has been described in the literature variously as a distraction effect (Tipper, 2001) or as an episodic memory effect (Mayr and Buchner, 2007; Neill et al., 1992). Neither of these models would seem to account for the observations in our data however. Distractors were not used in the experiment, and episodic memory seems like an unlikely cause of a general negative reaction to inflected forms.

6.1.2 Pseudowords

A facilitation effect for priming was also found to be significant for the pseudowords used as *nonword* stimuli in the decision task. Pseudowords should contain no semantic or associative content, since they do not occur in language use. We therefore can eliminate any potential facilitation effect due to these two factors, and we are left with a facilitation effect due to either form-priming or morphological priming. The form-priming effect cannot be due to simple visual matching however, since as mentioned before, lowercase letters were used for priming stimuli and uppercase letters were used for target stimuli. This suggests that the facilitation we observe happens at the phonological level.

It may be the case that priming a nonword with a nonword facilitates a decision for nonword. Participants would have to recognize that the priming stimulus is a nonword however, and in order to do so they would have to decode its phonology, which is similar to the target stimulus.

As mentioned previously, pseudowords do not exist in the lexicon and as such we would not expect them to fall into a morphological paradigm. This might however be a somewhat naive assumption as participants are exposed to inflectionally related forms throughout the experiment for the lexical items. As such we cannot completely discount a morphological effect in addition to any visual matching for the pseudowords as participants may adopt a strategy where they generalize the morphological paradigm to nonwords. Put simply, if the designer of the experiment can easily generate expected neuter forms for nonwords it seems likely that participants who know Norwegian would be able to as well. Such an effect may play a part in what causes the different behaviour observed for *problematic* adjectives in the experiment.

It is unlikely that the difference in effects is due to a different frequency for the category members. The strength of the facilitation effect is expected to decrease for less frequent items, but the frequency of pseudowords can be assumed to be equal or close to 0, with lexical items in effect being infinitely more frequent.

The fact that the expected neuter forms of pseudowords are reacted to differently as priming stimulus than expected neuter forms of *problematic* adjectives suggests that their status in the mental grammars of participants is different. A pseudoword simply does not exist and the lexical lookup yields an empty result. Comparing this against the lack of facilitation observed for *problematic* adjectives suggests that the expected neuter forms of these adjectives having some sort of nebulous existence even though there is a gap in the paradigm, and that they are reacted to negatively when the participant associates it with the common gender form.

The way the task was set up, any effects from grammatical gender agreement or disagreement would not be expected. The neuter and common gender forms were isolated from any syntactic context and so participants reacted only to the forms themselves. This is also why common gender forms of the adjectives were chosen as target stimuli, since the potential neuter forms are so problematic. This shows that the inflected form itself can cause odd behaviour, and not just its context within a phrase or sentence. While more frequent misclassification of *problematic* adjectives may also point towards a frequency effect, the classification trend in the other direction for *normal* adjectives is puzzling.

6.1.3 Correct and incorrect answers

The experiment did find a small imbalance in the amount of correct answers by adjective class membership. Although this would be expected for *pseudowords* and is also reflected in how they are processed much more slowly than lexical items in both conditions, the imbalance between *normal* and *problematic* adjectives may warrant further examination. The *problematic* adjectives were misclassified more frequently than *normal* adjectives under both primed and unprimed conditions. The effect is weak, and can probably to some extent be explained through an imbalance in frequency, but another possibility is that they are in fact somewhat more complex to process. The association effect is complicated by the failure to detect any differences in reaction time due to class membership alone between the lexical item groups.

There was a very small effect for different classifications between the groups of participants as well, with the L2 group showing a propensity for more frequently erroneously classifying *problematic* adjectives but also for more frequently correctly classifying *normal* adjectives than L1 participants. We would expect L2 participants to have a higher proportion of incorrect classifications in general due to a lower level of familiarity with the language. This is also reflected in the L2 group being on average slower than the L1 group overall in the experiment. The overrepresentation of incorrect classifications of *problematic* could therefore be explained as a frequency effect. Such an effect would however not be congruent with the observation that the L2 participants have a tendency to correctly classify *normal* adjectives than L1 participants. While the effect is very small, it replicates the observation in a similar experiment (Vea and Johansson, 2016).

The *problematic* may somehow be more difficult to process than the *normal* adjectives regardless of lexical frequency, and this may be what we observe with the imbalance of correct answers in the L1 and L2 group. The effect may be due to more complex semantics, but it is beyond the scope of the methodology used in this experiment to further investigate such a claim. The observed imbalance is therefore left unexplained.

Data from L2 however show a very strong tendency for *problematic* to be misclassified. This would indicate that participants were less familiar with this group of adjectives, which would in turn suggests that the *problematic* group is in fact less frequent than the *normal* group. At least this would be a valid inference about the items selected for this experiment.

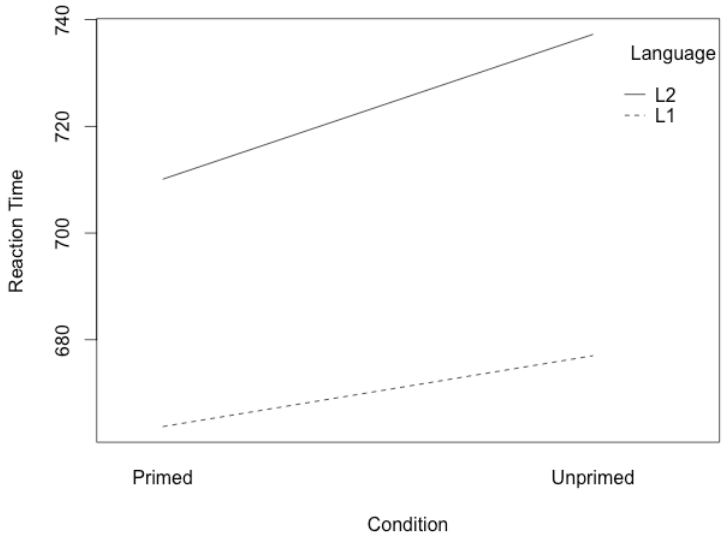


Figure 15: Line plot for *problematic* adjectives only, by language group.

It is not entirely clear however how such an effect can be eliminated. Due to the restrictive phonological conditions placed on stimulus items, there is a fairly limited set of items that would be appropriate for the control condition in both Norwegian and Swedish. Items have to be both adjectives and obey the phonological rules described in the section on adjectives.

6.1.4 L2

Results for L2 were similar to what we observe for L1, although the interaction term between class membership and priming condition was not significant for L2, although the post-hoc test did find a significant difference of means between unprimed *problematic* and primed *normal* adjectives. This indicates the same general trend as for the L1 group. The *problematic* adjectives receive weaker facilitation due to priming than do the *normal* adjectives also for L2 speakers, although the effect for this group is weaker and not significant in the ANOVA.

Overall, the priming effect was observed to be slightly stronger for the L2 group than the L1 group within all three classes of stimulus. The trend was the same as for the L1 group though, with *normal* adjectives receiving the strongest facilitation effect due to priming, *pseudowords* receiving slightly less and *problematic* adjectives receiving the least or no facilitation due to priming.

This makes for a notable difference with the L1 group, which had a significant effect ($p < 0.05$) of inhibition in the model for priming by *problematic* adjectives canceling out the facilitation observed for *normal* adjectives. This would indicate that the neuter removal rule is not as strong for our L2 group since they react less negatively to expected neuter forms for *problematic* adjectives. Consistent with this, we also see the difference in the line plot below.

While the trend lines point in the same direction, the L2 line is steeper than the L1 line,

indicating that there may be something different happening between the groups. This would seem to be confirmed by the model failing to detect a smaller reaction to priming compared to the L1 group. The L2 data has a much smaller number of data points than the L1 data however, in part due to the smaller number of participants in the L2 group. But the L2 participants on average produced less reliable data as well, which is not entirely unexpected.

A further consequence of this is that we can state with some certainty that the missing neuter forms in fact are removed due to a rule. They are not *(*accidental) but systematic in a subtle way. Such a rule would have to be internalized by language users and a language learner may take some time to detect the presence of these gaps.

6.2 Limitations

While the experiment did show some results, it does have a number of limitations, and these should as a matter of good practice be pointed out.

The models used to analyze the data were not ”maximally” defined in the sense that they included all possible random slope terms imaginable. As was mentioned in the analysis, a slope for the by-item random effect does not seem justified even though the model did converge. The inclusion of such terms either explained very little or no variance, or in the worst cases made the model a worse fit for the observed data.

Perhaps a more notable limitation was the failure to fit a model that included a random slope for the priming condition for the by-item effect. As mentioned previously in the section on results, the failure to fit such a term is likely caused by a lack of pairs within the priming condition after removing outliers and incorrect data. While most of the random variance is expected to be generated by participants, it is possible that stimulus items individually can cause different reactions to the priming condition as well. It is beyond the scope of this experiment to investigate whether this is the case. A larger sample of participants or a larger sample of items per participant would potentially alleviate this problem.

The results still hold at the level of specification used in the model. Further post-hoc testing also confirmed statistical significance for a number of variables. Importantly, lack of priming for *problematic* adjectives when compared to *normal* adjectives was present in this analysis as well.

Another limitation is the low amount of reliable data for L2 participants. There were fewer L2 participants to begin with, but a compounding factor is that the L2 group was also more prone to misclassification than the L1 group. The lack of significant effects for L2 in the model may be accounted for by the lower quality and quantity of data obtained from this group. In other words, there is a risk of type II error in the data.

The trend observed is that priming is weaker for the *problematic* group of adjectives also for L2, but the effect is not as pronounced as for L1. But with the data we have, it is impossible to conclude with certainty.

The low rate of succesful performance in the lexical decision task for the L2 group is worth consideration as well. The task used in this experiment may in fact be too difficult for L2 participants of variable proficiency to perform at an acceptable level. For further studies on the topic, it may be advantageous to devise a lexical decision task, or use some other experimental paradigm, that is in some way easier to perform than the present task in order to reduce the rate of misclassifications. A possibility would be to use syllable structures that are illegal in Norwegian, as this would provide participants with another cue to the nonword status of the stimulus. Since we were interested in how participants responded to pseudowords of similiar phonological form to the lexical items however, using illegal syllables was excluded in the present experiment.

6.3 The status of neuter gaps

The experiment revealed that participants process the expected neuter forms of *problematic* adjectives differently from both adjectives that regularly form neuters and pseudowords primed by their expected neuter forms. Specifically, the analysis shows that there is some level of inhibition for the *problematic* adjectives when they are primed by corresponding neuters, as compared to regular adjectives and pseudowords. This can be arrived to through the observation that there is a lack of effect for priming, even though we would expect some level of priming due to similariy of form alone.

This inhibition is unlikely to be due to gender agreement, since the experiment only tests the neuter and common gender forms isolated from any wider context. There is nothing within the experiment that would require gender agreement for either the priming stimuli or the target stimuli. For this reason, we can exclude effects at the phrase-level.

This presents a very strong case for the missing neuter forms not being a case of difficulty in assigning gender agreement, but the neuter forms in themselves are reacted to negatively. This negative reaction is not present for pseudowords, which do display facilitation in processing due to priming. This suggests that neuter forms of *problematic* adjectives and neuter forms of *pseudowords* are two different kinds of missing adjectives.

There was a small difference in means for primed and unprimed *problematic* items, but this difference was not significant ($p > 0.7$) in the L1 group. It should be noted that this does not in any way confirm the null hypothesis, since we can only ever reject the null hypothesis or fail to reject the null hypothesis. The only inference we can make is that under the exact same conditions as adjectives that regularly form neuters and pseudowords with corresponding neuters, the *problematic* adjectives behave differently in that priming does not reliable facilitate reaction times in participant responses. This is not a trivial observation however, and it does strongly suggest that not only are the missing neuter forms missing, but that there is a process blocking their formation.

The methods used in this experiment cannot provide any more detail than the observation

that they are processed differently. We can make a number of inferences as to why that may be the case. The fact that a highly productive and regular rule in Norwegian and Swedish is violated in this way is something to be taken into consideration, which it has. It is unlikely that a similar lexical decision task where reaction times are measured will produce very different results from those presented here, however. There is room for improvement and further research in the L2 domain, but this would likely require the lexical decision task to be reworked. The L2 group demonstrated less reliability in the performing the task. This is particularly notable in the classification of pseudowords where the L2 group where almost split down the middle with 337 correct classifications compared to 323 errors or no responses. The pseudoword stimuli may be too challenging for the L2 group, but the design of the experiment relied on participants processing pseudowords.

6.4 Gender congruency or inflectional form

A further topic we must consider is whether the paradigmatic gaps arise due to an issue with gender congruency or due to the morphological form of the adjective. The experimental data sheds some light on this.

There is no head noun that licenses gender agreement in the experiment. In light of the results, we can make the inference that the ill-formedness of the missing neuter forms persist even when we isolate the neuter form from any syntactic context. While this would suggest that congruency at the phrase-level is not the cause of the paradigmatic gaps, it is impossible to separate the morphology from neuter gender. The *-t* inflection occurs exactly only in the neuter, and so it is strongly associated with the neuter gender, and would only occur where neuter gender is required.

It is therefore not entirely clear how to separate the inflectional form of the adjective from gender congruency properties.

7 Further considerations

The experiment provides evidence that speakers of Norwegian in fact react negatively to the expected but missing neuter forms of adjectives, and that some notion of a gap exists in the mental grammars of speakers. A mechanism by which speakers acquire and sustain this gap needs to be investigated. Further, the missing neuter forms likely reflect a wider tendency within the Mainland Scandinavian languages, and this will also be discussed in the following sections.

7.1 Frequency and collocation

Further attention should be given to the possible effect of frequency on the occurrence of the paradigmatic gaps in Norwegian and Swedish. As was mentioned in the discussion section,

frequency is likely to be a factor in explaining some of the difference in both reaction times and classification of stimuli. The *normal* group being on average faster than the *problematic* group. The distinction between *type* and *token* becomes important. As mentioned in the section on adjectives, we find items of high frequency within the *problematic* group.

The adjective *glad* being one such example. A search in Leksikografisk Bokmålskorpus (Knudsen and Fjeld, 2013) returns 12457 hits for the common gender form *glad*, but 0 hits for the expected neuter *gladt*. We also find items of a lower frequency within this category. The adjective *vred* is one such example, with 1405 hits in the same corpus for the common gender form *vred*, and likewise 0 hits for the neuter gender form *vredt*. Similar searches between the categories of adjectives that have well-formed neuters and those that lack well-formed neuters show that there is considerable variation in frequency within each category. The adjective *god* along with its well-formed neuter *godt*[10] returns a little over 100000 hits in the same corpus, while *bløt* with its well-formed neuter *bløtt* only has 650 matches[11]. Definite and plural forms of adjectives have different morphology, and are always allowed in the neuter, as per the description in the section on adjectives, and are therefore excluded.

We do however find the expected neuter forms of *problematic* adjectives have a token frequency that is at or close to zero. One mechanism then for how the neuter gap is acquired by new learners is that the token frequency of the neuter form of the adjective is much lower than what speakers would expect from how frequently the adjective occurs in general, i.e. its type frequency. That is, the token frequency of the neuter indefinite singular is strongly under-represented compared to the token frequency of all other forms in the morphological paradigm. Speakers have no feedback from the occurrence of an irregular inflected, and so the missing neuter rule must be learned through observing that there is no form where one would be expected.

Since a semantic motivation for the missing gaps has been proposed and is supported by the data, collocation would further provide a mechanism by which speakers learn the gap. If the adjectives imply animacy and individuality, and the neuter gender at the same time under-represent such items in its noun membership, then the adjectives with missing neuter forms would be less likely to be used with neuter agreement from the outset. The mechanism reinforces the lower-than-expected token frequency of neuter forms for these adjectives.

From the experimental data it appears that the paradigmatic gap also exists for the L2 group, although it is weaker. This indicates that L2 speakers become aware of the neuter gap with relatively lower levels of exposure to the language. The very low or zero occurrence of neuter forms of adjectives that are otherwise frequent would give speakers an indication that the neuter should not occur.

Speakers are further able to generalize the neuter gap to a class of adjectives that share similarities. The neuter gap is apparently productive. This can be observed through new lexical

[10] 37934 matches for the common gender *god* and 62475 matches for the neuter gender *godt*, respectively.
[11] 359 for the common gender *bløt* and 291 for the neuter gender *bløtt*, respectively

items being able to enter the category and also in how the same word (the example *flat*) are able to attain different status in whether it has a neuter gap or not in Norwegian where *flatt* is well-formed, but not in Swedish where it is ill-formed.

The neuter gap appears to have some level of productivity too, however. Even adjectives of very low frequency such as *sky* or *slu* also have no well-formed neuter[12], which points towards the neuter gap being able to be generalized across a class of adjectives. The phonological and semantic features that are common across this class of adjectives have been discussed previously in the thesis. The phonological criterion proposed in the is immediately apparent when studying these adjectives, but as has been shown, a phonological account would overgeneralize. The combination of phonology and semantics is required for the gap to be formed. This fact constitutes strong evidence that irregular inflections cannot be accounted for only by phonological mechanisms, as was noted for English strong verbs in Pinker and Prince (1988).

We can observe that the proposed phonological and semantic must both be present for a neuter form of the adjective to fail to form. As an example the adjective *flink* (skillful) is assumed to occur infrequently in the indefinite neuter due to the semantic property of skillfulness requiring a certain level of sentience, which implies animacy. A search in Leksikografisk Bokmålskorpus returns 3238 matches for the common gender form *flink*, but only 17 matches for the neuter gender form *flinkt*. The neuter gender form is not blocked in the same way as our *problematic* adjectives however, since *flink* does not fulfill the phonological criterion.

The implication for adjectives in Norwegian and Swedish is that the neuter gap should be generalizable to a new adjective entering the language. If we introduce a pseudoword with appropriate phonology and within a context where its semantics would be within the same category as the *problematic* adjectives. We can use the example pseudoword *krød* from the experiment.

(31) En gutt som tenker mye på enhjørninger kan sies å være en krød gutt. Et barn som tenker mye på enhjørninger kan sies å være et *krødt barn.

'A boy who thinks a lot of unicorns can be said to be a *krød* boy. A child who thinks a lot of unicorns is a *krødt* child.'

(32) En gutt som har brunt hår kan sies å være en krød gutt. Et barn som har brunt hår er et krødt barn.

'A boy who has brown hair can be said to be a *krød* boy. A child who has brown hair is a *krødt* child.'

The prediction is that *krød* in example 31 would not be able to form the neuter *krødt*, but in example 32 due to its semantics describing a property that is immediately verifiable the neuter form would be more acceptable. While it is beyond the scope of this project to test the processing of such sentences using pseudowords, but the prediction may prove useful for

[12]At least not in corpus search.

further research into language processing. Specifically in processing irregular morphological patterns.

This suggestion is that the lack of a neuter form in adjectives is not completely contained within the adjective itself, but also emanates from a wider property of neuter gender in nouns in the language.

7.2 Gender semantics

A semantic account, such as the one proposed by Johansson (2003) and Petterson (1990), has a number of advantages to a purely phonological or accidental account of the paradigmatic gaps described in this thesis, as mentioned previously in the thesis.

The observation that *problematic* adjectives are more frequently misclassified in the experiment is surprising given that their common gender forms are well-formed and regularly occur in usage. This suggests that in the experiment, they are negatively impacted by the presence of their neuter gender forms. This is further supported by the lack of facilitation observed in reaction time due to priming, which may explained in two ways. One is that the *problematic* neuter form does not occur, and so it primes a decision for nonword for the common gender target stimulus. Another is that the connection between neuter gender form and common gender form elicits a negative reaction due to ill-formedness. In either case, the negative reaction is caused by a feature of grammatical gender.

The adjectives with paradigmatic gaps in the neuter typically imply *animacy* and *individuality* to the head nouns that they modify. Moreover, many of them imply sentience and the capacity for mental states. There are a number of other cases in both Norwegian and Swedish where the use of neuter gender relates to semantic features, and these will be discussed in the following section. The suggestion is that there is a semantic core to the grammatical gender classes in the Mainland Scandinavian languages (Corbett, 1991, 2013) and further that these semantic features can occasionally surface in syntactic features in the languages.

This relates to the larger problem of gender semantics. If the neuter forms of these adjectives are incompatible with the semantic content of these adjectives, the suggestion is that neuter gender infers semantic properties as well. In the semantic account discussed previously, non-individuality is shown for a number of neuter nouns that refer to animate entities. The examples *mannfolk* (man) and *kvinnfolk* (woman), respectively refer to male or female humans, which are typically individuals and typically animate. But the use of these terms specifically is frequently disparaging or at least emotive. The compound with the element *folk* may suggest that these individuals are in fact part of a larger interchangeable mass, not unlike how an individual blueberry may be interchangeable with any other blueberry. This quality of being part of a *mass* of something occurs with neuter gender in other cases as well.

A less emotionally loaded or derogatory example of the non-individual property of neuter gender can be shown with the example word *politi* (police), which can be both masculine/common

gender and neuter gender in Norwegian. When it occurs as a masculine noun, it unambiguously refers to an individual police officer. As a neuter noun however, it will more generally refer to the institution of the police force and the people that make up this institution. That is, the neuter gender lexical item has the ability to refer to the police as a mass entity.

It is not within the scope of this project to attempt to give a complete account of semantic features of the grammatical gender system as a whole in Norwegian, although a short overview will be given. The neuter gender in particular is what we are interested in.

7.3 Theoretical considerations

An account where grammatical gender correlates with semantic features is not immediately obvious due to the status of grammatical gender in both Norwegian and Swedish. The masculine and feminine[13] contain nouns that cannot be classified semantically with reference to masculine or feminine (biological) features. It has however been demonstrated that gender assignment in Norwegian (and in Swedish) is governed by a number of rules (Trosterud, 2001). Upon closer examination of grammatical gender classes in many European languages, we notice that there is a semantic core to each gender class (Corbett, 2013).

It is not necessary to posit a strict classification criterion where grammatical gender must necessarily correspond to biological sex. Such an account of gender in the Scandinavian languages would be false from the outset since we can find numerous example nouns within the masculine gender referring to entities that cannot have biological sex, and we also find numerous examples within the feminine gender of nouns referring to entities that cannot have biological sex. It is not entirely clear how a neuter gender should fit into such a paradigm, but prototypically the most common nouns referring to animates within the category to so without invoking biological sex[14]. As has been previously noted, there is a tendency for the neuter gender to under-represent animate and individual categories in the Mainland Scandinavian languages. This is elaborated upon in the next section.

It is enough for our purposes however to note that the neuter gender under-represents the animate and individual categories. This tendency can be observed in all of the Mainland Scandinavian languages, and is further reinforced by assignment of gender to new nouns entering the languages defaulting to masculine or common gender (Kilarski, 2003). There is in fact a tendency for mass nouns specifically to more commonly be assigned neuter gender as they enter the languages, as compared to other types of nouns (ibid), though masculine or common gender still dominates. This provides a mechanism by which the amount of nouns within the neuter gender refers to animate or individual entities remains relatively small and constant.

The neuter gender containing nouns that refer to animate or individual entities though, language users may not associate those features with the neuter gender. Since gender is innate to nouns, and since the gender system of the languages in question is mostly a formal gender

[13]Common gender for Swedish and some varieties of Norwegian.

[14]The examples *barn* "child", *dyr* "animal" for instance

system, it is likely more fruitful to observe modifiers such as adjectives, which is what has been done in the experiment.

7.3.1 Semantic agreement effects

The observation in Johansson (2003) is that neuter gender in Norwegian and Swedish is associated with *non-animacy* and *non-individuality*. We find examples of this in the lexicon as previously described, but there are also examples where this feature enters into the syntax other than the paradigmatic gaps investigated in this experiment.

A number of examples of apparent gender discongruency occur in the Mainland Scandinavian languages. In these cases, adjectives assume neuter gender morphology in spite of modifying head nouns that would also be able to license masculine/feminine/common gender. Examples are given below:

(33) Pannekaker er gode.
Pancakes[MASC].PL COP tasty.M.PL
'Pancakes are tasty.'

(34) Pannekaker er godt.
Pancakes[MASC].PL COP tasty.N.SG
'Pancakes are tasty.'

Example 33 shows a neuter form being used to modify a masculine noun. This is a common type of construction in Norwegian, Swedish and Danish and has been referred to as *pancake* sentences in the literature (Enger, 2013), as per the example. This type of semantic agreement (Enger, 2004) is in the languages under consideration restricted to predicative positions, and cannot occur in attributive positions. Note that it is specifically the singular indefinite neuter form that is used in these cases. This is the only neuter-form in the paradigm that is explicitly marked for neuter gender as previously discussed.

This provides us with another example of how grammatical gender can be subject to interactions from semantic effects in the Scandinavian languages. Neuter agreement with non-neuter nouns typically occur when the noun is read as referring to a mass entity or to an event. The mass reading is particularly relevant to the previous observation that the neuter gender implies non-individuality (along with non-animacy) in Scandinavian.

The neuter-agreement is further strengthened when the noun is innately a mass noun. A pancake at least has the possibility to be an individual pancake, but that would not apply to everything. As in the following example:

(35) Vodka er sunt.
Vodka[MASC].SG COP healthy.N.SG
'Vodka is healthy.'

In this case, the mass-entity quality of vodka is so unambiguous that the same sentence would be grammatically incorrect if the adjective was to have masculine agreement with the masculine noun.

Neuter-agreement as semantic agreement is not restricted to mass reading. Example 34 may not only refer to pancakes as a mass entity where each individual pancake is indistinguishable, but it may also refer to an event that is defined by the presence of pancakes. Enger (2004) in fact ascribes the neuter-agreement specifically to the fact that the controller (the pancakes) are low on the individuation scale. This concords with the notion in (Johansson, 2003) that the neuter gender under-represents the individual category.

Head nouns that can accept this type of semantic neuter-agreement do not seem to be able to refer to entities that are capable of internal states. The animate category is therefore precluded from the outset.

There thus seems to be a connection between these two agreement-phenomena in Norwegian and Swedish. The semantic properties motivating both pancake sentences and missing neuter forms are identical. The inhibitory priming effect observed in our experiment highlights this as well, where the inhibition is likely to stem from a processing conflict between gender-agreement features (the neuter-marking */-t/* and semantic features. Both pancake sentences and missing neuter forms reflect an underlying property of the neuter gender in Norwegian and Swedish.

7.3.2 Diachronic tendencies

A relevant question to ask is why similarly missing neuter forms do not occur in closely related languages. German can be used as an example. It has preserved three grammatical genders just as with Norwegian, and its morphology is fairly similar[15]. One notable difference however is that German uses neuter gender as a type of default assignment (cf. Johansson, 2003), and we find that many animals are neuter in German. We find examples such as 'das Zebra', 'das Pferd', 'das Krokodil' and many more. Neuter gender does not appear to under-represent the animate category to the same extent as do the Mainland Scandinavian languages.

There appears to be a wider tendency towards semanticization of the neuter gender in the Mainland Scandinavian languages (Josefsson, 2014), and that this development is particularly advanced in variants of Danish where for instance in West Jutlandic the neuter gender has been completely semanticized (ibid). The semantic properties specifically related to non-individuality or mass readings, and these are properties we observe in the aforementioned pancake sentences. Moreover, the implication of individuality is proposed as the mechanism by which the group of adjectives in our experiment fail to form neuters, so there appears to be an interesting parallel.

We can compare with more closely related languages than German as well. The missing

[15]Neuter gender is not marked with a final -t however.

neuter forms described for Norwegian and Swedish are well-formed in the closely related Icelandic language. As in the following example where the neuter form of the adjective *glaður* (cognate with Norwegian and Swedish *glad*) is used:

(36) Fólkið er glatt.
People[NEUT].SG.NOM.DEF COP happy.NEUT.SG.NOM.INDEF
'The people are happy.'

The neuter-forming morphological rule is similar to what we have observed in Norwegian and Swedish in this particular case[16]. For comparison, the equivalent sentence in Norwegian with the ill-formed adjective is provided below.

(37) Folket er *glatt.
People[NEUT].DEF.SG COP happy.NEUT.SG.INDEF
'The people are happy.'

The suggestion is that missing neuter forms are an innovation in the Mainland Scandinavian languages, and this would concord with the observation in (Josefsson, 2014) that pancake agreement is an innovation, suggesting a further semantic parallel between the systems.

It might therefore be fruitful to conduct a study on whether the missing neuter forms occur in Danish as well. (Johansson, 2003) mentions that this does appear to be the case, but there does not appear to be any experimental testing of the neuter gaps in Danish in the literature.

The suggestion is that the process where the neuter gender has come to be used as semantic agreement in certain contexts reflects a deeper tendency where semantic qualities implied by the neuter gender surfaces in syntactic phenomenon such as agreement. While the missing neuter forms appear to be problematic when they occur isolated from any grammatical context as shown in the experiment, the neuter-morpheme is still an agreement feature.

It must also be noted that these neuter gender effects only seem to appear where neuter gender is explicitly marked in agreement. That is, only forms marked by */-t/* are affected, since neuter gender-marking for the most part is identical with other genders in plurals or in the definite singular.

8 Conclusion

This thesis has given an in-depth account of adjectives with missing neuter forms in Norwegian. While the experiment specifically investigated Norwegian language, the problem is present in Swedish as well, and possibly in Danish. The data presented gives strong evidence for language users being able to notice that certain neuter forms are missing from regular language use, and they are further able to make inferences about adjectives of similar characteristics based on this information.

[16]Only in the nominative and accusative cases. Icelandic retains four distinct cases, and the inflectional system is considerable more complicated than we find in the Mainland Scandinavian languages.

There is some similarity to irregular inflections in that a productive and regular rule for forming neuters in Norwegian and Swedish is blocked, but instead of there being negative input from a competing irregular form, there is instead a gap in the paradigm. The implication is that negative reinforcement from a competing irregular form in the paradigm is not strictly necessary for speakers to acquire irregular inflections, but that the expected form simply being absent may in some cases be enough.

The experimental methods used have found a processing conflict caused by missing neuter forms. The lack of facilitation in classifying their corresponding common gender forms when the neuters are used as priming stimulus would suggest that speakers do not connect the forms. No facilitation due to similarity of phonological form is observed either. This is unexpected and suggests an inhibitory process is competing with any facilitation due to form-similarity.

There may therefore be several competing processes occurring at the same time as the adjectives are processed. The present experiment was concerned with how the neuter gaps are processed compared to regular neuters and pseudoword baselines, but manipulating stimulus onset asynchronies may be one method by which we could observe different processes. Shorter SOAs may produce different results for priming than longer SOAs, if facilitation and inhibition processes happen at different time intervals. More technologically advanced equipment such as EEG could also be considered for further research into this area, if we are interested in more detailed observations of the processing.

The experiment also found that speakers exhibited facilitation for psuedowords when those pseudowords were primed by their expected neuter forms. This points towards the neuter forms not being naively missing, but that speakers have an active process that causes the paradigmatic gap. If a lack collocation was the only factor, as observed in corpus searches, then the expectation would be that neuter forms should be possible due to the productivity of the neuter-forming rule, as long as their semantics are applicable. The aforementioned experimental findings strongly suggest that this is not the case however.

The inclusion of an L2 group also points towards the paradigmatic gap being noticed fairly early by speakers. Loss of data was larger in the L2 group however, and further research may be warranted. The L2 group could be better balanced for proficiency as well, although it became clear in the analysis that the task may be too difficult for some L2 participants and particularly language learners. An easier task would have the potential to reduce some of these problems.

We can therefore conclude that the gap in the neuter form actually has some existence in the mental grammars of speakers of Norwegian and Swedish. This can be thought of as a null form. There is no repair to the gap, although imperfect solutions where either the noun or the adjective is replaced with a synonym exist.

The combined phonological and semantic mechanism proposed in this thesis is able to fully account for the paradigmatic gap. Purely phonological explanations are unable to account for the full range of adjectives that fall within the *problematic* group and properly separating them from adjectives that have well-formed neuters. The semantic criterion cannot account

for the ill-formedness of neuters on its own either, as observed with adjectives such as *flink* which have well-formed neuters in spite of a very low frequency of occurrence in the neuter gender. The semantic mechanism stems from a conflict between the presumed semantics of the adjective and the implied semantics of neuter gender, which prevents the neuter form of adjectives to appear in contexts where a neuter form is expected.

The paradigmatic gaps being restricted to neuter gender is further of relevance when investigating the properties of grammatical gender. The ill-formed neuters being caused by semantics provides evidence that formal gender systems with seemingly arbitrary assignment are correlated with semantic properties.

This thesis has given an account of the paradigmatic neuter gaps in Norwegian. The neuter gaps have been verified through experimental methods which indicate that not only are they missing but that they are also reacted to negatively by speakers. An explanation for the occurrence of the gaps that fits the data has also been given, and the explanation should further be applicable to Swedish. The status of neuter gaps in Danish still remains untested, but the semantic properties of neuter gender are applicable in Danish as well, so there is reason to suspect that there may be a similar blocking process in Danish. This would be a topic for further research. The fact that the paradigmatic gaps are restricted to neuter gender, and that the gaps are motivated by semantic features is further of interest when considering the properties of grammatical gender. Particularly, it provides further support for the notion of (Josefsson, 2014) that neuter gender in the Mainland Scandinavian languages has a tendency towards semanticization.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior 22*, 261–295.

Andersson, E. (2002). Swedish. In E. König and J. van der Auwera (Eds.), *The Germanic Languages*, Routledge language family descriptions, pp. 271–312. Routledge.

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59*, 390–412.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language 68*(3), 255–278.

Bentin, S., M. Kutas, and S. A. Hillyard (1995). Semantic processing and memory for attended and unattended words in dichotic listening: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance 21*, 54–67.

Brown, C. M. and P. Hagoort (1993). The processing nature of the n400: evidence from masked priming. *Journal of Cognitive Neuroscience 5*, 33–44.

Carr, T. H. (1990). Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition 16*(2), 341–350.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics on psychological research. *Journal of Verbal Learning and Verbal Behavior 335-359*.

Collins, A. M. and E. F. Loftus (1975). A spreading-activation theory of semantic processing. *Psychological Review 82*(6), 407–428.

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

Corbett, G. G. (2009). Suppletion: Typology, markedness, complexity. *Trends in Linguistics: Studies and Monographs*, 25–40.

Corbett, G. G. (2013). *Sex-based and Non-sex-based Gender Systems*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics*. Malden: Wiley-Blackwell.

Deacon, D., S. Hewitt, Y. Chien-Ming, and M. Nagata (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the n400 does not reflect a post-lexical process. *Cognitive Brain Research 9*, 137–146.

Enger, H. O. (2004). Scandinavian pancake sentences as semantic agreement. *Nordic Journal of Linguistics 27*(1), 5–34.

Enger, H. O. (2013). Scandinavian pancake sentences revisited. *Nordic Journal of Linguistics 36*(3), 275–301.

Fanselow, G. and C. Féry (2002). Ineffability in grammar. In *Resolving conflicts in grammars: optimality theory in syntax, morphology, and phonology*, pp. 265–397. Hamburg: Buske.

Feldman, L. B. (1994). Beyond orthography and phonology: Differences between inflections and derivations. *Journal of Memory and Language 33*(4), 442–470.

Forbach, G. B., R. F. Stanners, and L. Hochhaus (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition 2*(2), 337–339.

Forster, K. I. (1998). Pros and cons of masked priming. *Journal of Psycholinguistic Research 27*(2), 203–233.

Forster, K. I. and S. M. Chambers (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior 12*(6), 627–635.

Forster, K. I. and C. Davis (1984a). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language 30*, 1–25.

Forster, K. I. and C. Davis (1984b). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory & Cognition 10*(4), 680–698.

Forster, K. I., C. Davis, C. Schoknech, and R. Carter (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology 39*(2), 211–251.

Fowler, C. A., S. E. Napps, and L. Feldman (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition 13*(3), 241–255.

Grainger, J. (1990). Word frequency and neightborhood frequency effects in lexical decision and naming. *Journal of Memory and Language 29*(2), 228–244.

Greenwald, A. G., S. C. Draine, and R. L. Abrams (1996). Three cognitive markers of unconscious semantic activation. *Science 273*, 1699–1702.

Johansson, C. (2003). How is grammatical gender processed? In *Computational Linguistics in the Netherlands 2002*, pp. 65–76. Rodopi.

Johansson, C. and J. C. Torkildsen (2005). Reaktionstider för genuskongruens: hur processeras genus? In S. Lie, G. Nedrelid, and H. Omdal (Eds.), *MONS 10: utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003*.

Josefsson, G. (2014). Pancake sentences and the semanticization of formal gender in mainland scandinavian. *Language Sciences 43*, 62–76.

Kerstens, J., E. Ruys, and J. Zwarts (2001). *Lexicon of Linguistics*. Utrecht: Utrecht institute of Linguistics OTS.

Kilarski, M. (2003). Gender assignment in danish, swedish and norwegian: a comparison of the status of assignment criteria. *Nordlyd 31*(2).

Knudsen, R. L. and R. V. Fjeld (2013). A balanced; annotated national corpus for norwegian bokmål. In *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA*. NEALT Proceedings Series 19.

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kutas, M. and K. D. Federmeier (2011). Thirty years and counting: Finding emaning in the n400 component of the event related brain potential. *Annual Review of Psychology 62*, 621–647.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2016). Tests in linear mixed effects models. *CRAN, R-Project*.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America 98*(23), 13464–71.

Levelt, W. J. M., A. Roelofs, and A. S. Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences 22*, 1–75.

Marcus, G. F. (1996). Why do children say "breaked"? *Current Directions in Psychological Science 5*(3), 81–85.

Matuschek, H., R. Kleigl, S. Vasishth, H. Baayen, and D. Bates (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language 94*, 304–315.

Mayr, S. and A. Buchner (2007). Negative priming as a memory phenomenon: A review of 20 years of negative priming research. *Journal of Psychology 215*, 35–51.

Meyer, D. E. (2014). Semantic priming well established. *Science 345*, 523.

Meyer, D. E. and R. W. Schvaneveldt (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology 90*(2), 227–234.

Meyer, D. E., R. W. Schvaneveldt, and M. G. Ruddy (1975). *Attention and performance V*, pp. 98–118. London: Academic Press.

Neill, W. T., L. A. Valdes, K. M. Terry, and D. S. Gorfein (1992). Persistence of negative priming: Ii. evidence for episodic trace retrieval. *Journal of Experimental Psychology: Learning, Memory & Cognition 18*(5), 993–1000.

Ostergaard, A. L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory & Cognition 26*(1), 40–46.

Perea, M. and A. Gotor (1997). Associative and semantic priming effects occur at very shot stimulus-onset asynchronies in lexical decision and naming. *Cognition 62*, 223–240.

Petterson, T. (1990). Varför barnet inte kan vara latt[why the child cannot be lazy]. *Svenskans Beskrivning*, 293 – 302.

Pinker, S. and A. Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition 28*(1-2), 73–193.

Raffelsiefen, R. (2004). Absolute ill-formedness and other morphophonological effects. *Phonology 21*(1), 91–142.

Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.

Rice, C. (2007). Gaps and repairs at the phonology-morphology interface. *Journal of Linguistics 43*(1), 197–221.

Sandvik, O. H. (1979). *Talemål i Rogaland: i går, i dag og i morgon: om uttale, bøying og ordtilfang i Rogalands-målet*. Stavanger: Rogalandsforskning.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin 2*, 110–114.

Scarborough, D. L., L. Gerard, and C. Cortese (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition 7*, 3–12.

Schielzeth, H. and W. Forstmeier (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology 20*(2), 416–420.

Schvaneveldt, R. W. and D. E. Meyer (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV*, pp. 395–409. New York: Academic Press.

Simos, P. G., J. I. Breier, J. M. Fletcher, B. R. Foorman, E. M. Castillo, and A. C. Panaicolaou (2002). Brain mechanisms for reading words and pseudowords: an integrated approach. *Cerebral Cortex 12*(3), 297–305.

Sims, A. D. (2009). Why defective paradigms are, and aren't, the result of competing morphological patterns. *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*.

Specht, K., C. Holtel, R. Zahn, H. Herzog, B. J. Krause, F. M. Mottaghy, I. Radermacher, D. Schmidt, L. Tellmann, S. Weis, K. Willmes, and W. Huber (2003). Lexical decision of nonwords and pseudowords in humans: a positron emission tomography study. *Neuroscience Letters 345*(3), 177–181.

Stark, C. E. L. and J. L. McLelland (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory & Cognition 26*(4), 945–972.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? a review and integration of conflicting views. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology 54A*, 321–343.

Trask, R. L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.

Trosterud, T. (2001). Genustildelning i norsk er regelstyrt [gender assignment in norwegian is regular]. *Norsk Lingvistisk Tidskrift 19*, 29–58.

Tulving, E. and D. L. Schacter (1990). Priming and human memory systems. *Science 247*, 301–306.

Vea, J. and C. Johansson (2016). Beginner learners react to paradigmatic gaps: A masked priming experiment. *Poster presented at Psycholinguistics in Flanders (PIF).*

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior 17*, 143–154.