Determiner Selection in Speech Production: Evidence from Norwegian.

By

Astrid Storm Hammerstad Breien

Master thesis in Computational Linguistics and Language Technology

Section for linguistic studies
University of Bergen
April 2005

Abstract

This thesis reports on two picture-word interference experiments where I investigate the processes involved in selection of grammatical gender in Norwegian Noun Phrase production. The respondents, consisting of subjects with a two gender system (masculine and neuter) and subjects with a three gender system (masculine, feminine and neuter), produced a gender-marked demonstrative + noun NP in response to a picture shown with either a distractor word of the same grammatical gender or one with a different grammatical gender. Though there are three different genders in Norwegian, there are only two demonstratives; *den* for feminine and masculine nouns, and *det* for neuter nouns. Significant longer naming latencies were found when the picture was shown with a distractor word of different gender than if the word and picture had identical gender. This effect however, was only obtained for subjects with a three gender system. This finding suggests that the interference does not occur between the independent word forms, the demonstratives, as proposed by Schiller and Caramazza (2003), but occur when selecting the noun's grammatical gender feature as originally suggested by Schriefers (1993). This implies that selection of grammatical features, at least gender, is based on competition, and does not come as an automatic consequence of selecting the lemma. A computational model based on the findings was tested in WEAVER++ (Roelofs, 1992). The simulations showed that WEAVER++ accounts for the some of the empirical findings.

Sammendrag

I denne oppgaven rapporteres to eksperimenter der jeg ved bruk av bilde-ord-interferensmetoden, undersøkte prosessene som er involvert i framhenting av grammatisk kjønn i produksjon av norske nomenfraser. Respondentene skulle produsere en demonstrativ + nomen frase i respons til et bilde som ble vist enten sammen med et distraktor ord med samme grammatisk kjønn, eller et med et annet grammatisk kjønn. Gruppen av respondenter besto både av folk med et tokjønns-sytem og folk med et trekjønns-system. I norsk er det tre genus, men bare to demonstrativer (*den* og *det*), der demonstrativen *den* er felles for maskuline og feminine nomen. Resultatet viste at det tok lengre tid å navngi bildene hvis distraktor-ordet hadde et annet grammatisk kjønn enn bilde, enn det tok hvis bilde og distraktor-ord hadde samme kjønn. Denne effekten var ikke betydelig for respondenter med et tokjønns-system. Resultatene fra eksperimentene indikerer at interferense ikke oppstår når demonstrativene skal velges som hevdet av Schiller og Caramazza (2003), men oppstår når nomenets grammatisk genustrekk skal velges som opprinnelig hevdet av Schriefers (1993). Dette impliserer at seleksjon av grammatiske trekk er basert på konkurranse, og kommer ikke som en automatisk konsekvens av å velge lemmaet. En datamodell basert på resultatene ble testet i WEAVER++ (Roelofs, 1992). Datasimuleringene viste at WEAVER++ kan redegjøre for noen av de empiriske funnene.

Preface

The present work is a masters thesis in Computational Linguistics and Language Technology, submitted at the University of Bergen in May 2005.

The project reported here is a cooperation project initiated by Niels Schiller from the University of Maastricht and Max Planck Institute for Psycholinguistics, Herbert Schriefers at the Radboud University Nijmegen, and my supervisor Koenraad de Smedt from the University of Bergen. During a ten-day period in June 2003, I was a guest at the Max Planck Institute for Psycholinguistics which has generously lent me the necessary equipment for running the experiments and also taught me how to set up the experiments and how to use the NESU box. I performed the experiments in Bergen, and would like to thank all the the people who participated.

All in all, I did three experiments and many analyses. When Experiment 1 did not seem to provide the expected results, and some weaknesses were discovered in the material, a second experiment was run. However, Experiment 2 did not yield any useful results concerning the research question and therefore the experiment is not reported in the thesis. Experiment 3 was conducted as a last attempt to find a gender congruency effect. When that experiment gave some results, but not the ones I had expected, I did a reanalysis of Experiment 1. The new results from Experiment 1 indicated an effect of gender congruency (supporting my hypothesis). Subsequently, the research question was also investigated with the help of the Weaver++ computer model. The results of the computer simulations were then considered in relation to the experimental results.

The project has been supported by grants from the Section of linguistic studies at the University of Bergen.

I would like to thank the following people for their invaluable help and enthusiasm:
My supervisor Koenraad de Smedt at the University of Bergen for keeping me on the right track, and always asking the right questions; Niels Schiller for sharing from his enormous well of knowledge again and again; Herbert Schriefers for the same; Ardi Roelofs at the Radboud University Nijmegen for letting me use the WEAVER++ computer model; Christer Johansson at the University of Bergen for helping me with the writing and statistics; and last, but not least, Oliver Müller at the F.C. Donders Centre for Cognitive Neuroimaging and Max Planck Institute for Psycholinguistics for giving me the code of his computer model and his valuable guidance through the simulation process and after.

Many others, whom have not been named here, have helped me in everything from lending me a room to reading the text with a critical eye. Thanks to all of you also.

Determiner Selection in Speech Production: Evidence from Norwegian

CHAPER 1, INTRODUCTION

Speaking is one of the tasks that we humans do every day without giving much thought to the complex processes behind it. We do it with great ease and speed, and in normal conversation we can produce “two to three words per second, which amounts to about four syllables and ten or twelve phonemes per second” (Levelt, 1999:223).

Speaking involves many processes going from thought to sound. We start with an intention to communicate something and turn it into a “message” of what to express. Then we have to retrieve the appropriate concepts which we have words for, lexical concepts, from our mental lexicon (Levelt, 1989). The mental lexicon can be thought of as a dictionary where we have all the declarative knowledge about a word, including its sense, its syntactic properties, its morphological properties and its phonological properties. An example of its sense is knowledge such as a banana is a fruit. Syntactic properties can be knowledge that *banana* is a noun, and a morphological property can be that *banana* is the stem and in plural it gets a *-s* added to the stem. Phonological properties like the word *banana*'s syllable structure and accent structure is also knowledge we have in our mental lexicon (Levelt, 1989). Choosing the intended lexical item in the enormous mental lexicon which consists of roughly 50-100 thousand words for an adult speaker (Miller, 1991 via Levelt, 1999), is called lexical selection. The lexical concepts have syntactic properties, such as being a noun, adjective or a transitive verb, which contributes to the planning of a syntactic structure for the utterance, it being a phrase or a sentence structure. This process is called grammatical encoding (Levelt, 1999). The utterance also needs a phonological plan for syllabification and prosody based on the phonological and morphological properties of the words. This is called phonological encoding, and the phonological plan is input to the articulatory apparatus which transforms this to overt speech (Levelt, 1999).

Given this rather complex path starting from a thought, retrieving the correct words from the huge lexicon, making a phonological plan for syllabification and prosody and executing this plan with the correct articulatory gestures, we make surprisingly few mistakes retrieving and uttering the words. On average, a speaker makes about 1 or 2 errors per every 1000 words (Levelt, 1999).

Much work in psycholinguistics has been focused on phonological encoding, but during the last few years, more studies have been done on the process of grammatical encoding, and especially on the retrieval of syntactic features like grammatical gender (e.g., Schriefers, 1993). If a speaker of a language with grammatical gender, for example Norwegian, wants to produce a phrase with an agreement target (e.g., a noun and its indefinite article), the speaker would not only have to retrieve the correct lexical item from the lexicon, but also its grammatical gender information in order to produce the correct indefinite article in the noun phrase (NP).

In this thesis I investigate the mechanisms that govern selection of grammatical gender for Norwegian speakers in two experiments, inspired by studies concerning grammatical feature selection in Dutch (Schriefers, 1993; Van Berkum, 1997; La Heij et al., 1998), German (Schriefers et al., 2002; Schriefers & Teruel, 2000; Schiller & Caramazza, 2003), Italian (e.g., Miozzo & Caramazza, 1999), and French (Alario & Caramazza, 2002). First, grammatical gender will be explained in general and gender in Norwegian will be described.

1.1 Grammatical gender

Grammatical gender is a phenomenon that occurs in many languages around the world. Gender comes from the Latin word *genus* which originally meant 'kind' or 'sort'. As a starting point, I use the following definition of grammatical gender:

> “Genders are classes of nouns reflected in the behavior of associated words” (Hockett, 1958:231 via Corbett, 1991).

“In some languages, gender is central and pervasive, while in others it is totally absent” (Corbett, 1991:1). How many languages that have gender as a grammatical category is unknown, but Corbett (1991) looked at more than 200 languages with gender systems, so it is fair to say it is a well-spread phenomenon in the languages of the world. The degree it occurs in however, can vary. In some languages grammatical gender is central while in others it is hardly present at all (ibid.).

Which associated words that reflect a noun's gender is something that differs between languages. Adjectives, verbs, determiners, adverbs, numerals and even conjunctions can agree in gender (Corbett, 1991:4-5). Consider the example from Swahili which is a member of the Bantu languages found in sentence (1). The Bantu languages have a varying number of gender classes, normally between ten and twenty (Corbett, 1991). In Swahili almost all nearby linguistic constituents agree with the noun's gender, as in sentence (1) where all the words are gender marked (even the head noun itself).

(1)ki-kapu ki-kubwa ki-moja ki-lianguka
basket$_{\text{gender-7}}$ large$_{\text{gender-7}}$ one$_{\text{gender-7}}$ fell$_{\text{gender-7}}$,
'one large basket fell' (Corbett, 1991:43)

On the other end of the scale, when it comes to amount of agreement targets, we find English, where gender is almost exclusively present in pronouns (e.g. The boy....he, the girl....she)[1].

Given this diversity, one may ask what the function of grammatical gender is.

> "Gender is crucial in establishing agreement relationships, i.e., a concord between different phrasal constituents (together with number and, in some languages, case as well), or in establishing local and global coherence in discourse" (Corbett 1991; Comrie, 1999 via Cacciari & Cubelli, 2003:377-378).

Gender is an intrinsic property of a noun (Faarlund, Lie, & Vannebo, 1997). The fact that a language has the category gender means that there are categories of nouns governing other words in a special manner. For example, in Norwegian we have three different categories for nouns to fall into, called feminine, masculine and neuter. The nouns in the category feminine, are the ones that take the indefinite article *ei*, the masculine ones take the indefinite article *en*, and the neuter nouns take the indefinite article *et.*

Gender is also essentially an arbitrary property (though gender may be semantically motivated e.g. by biological sex; Corbett, 1991). What kind of gender a noun has is language dependent. For example, the word *table* is masculine in German *(**der** Tisch)* it is feminine in French *(**la** table),* and it is neuter in Norwegian (*bord**et**).* In some languages, the noun's gender is apparent from the word's semantic meaning, morphology or phonology (Corbett, 1991), or all

---

[1] This type of gender systems where only pronouns show gender agreement is referred to as "pronominal gender systems" by Corbett. Whether English has the category grammatical gender or not is debated (Corbett, 1991:5)

to some extent. In other languages again, the word itself does not have overt clues as to its gender, so that the speakers of that language must learn each noun's gender by note. For learners of foreign languages, this can be a difficult task, and to produce the correct agreements (i.e. choosing, for example, the correct article or adjective) can be challenging. For native speakers of a language with gender systems, however, gender causes few problems. They know each noun's gender and when speaking they produce the agreement targets (like articles, adjectives, etc.) correctly in real time. Van Berkum (1997) estimated that a speaker of Dutch, which has a rather poor gender agreement system (Schriefers & Jescheniak, 1999) has to retrieve gender information about once every 10 seconds.

The fact that native speakers retrieve the noun's gender so easily, raises a lot of interesting questions, among the following questions taken from Schriefers and Jescheniak (1999):

- How is linguistic information like grammatical gender stored in the mental lexicon?
- Is the mental lexicon organized so that gender is stored together with each noun and can be looked up when needed?
- Or is gender computed, based on semantic, morphological and phonological information, each time it is needed in order to produce an agreement target such as a gender marked adjective + noun NP?
- If it is stored, how is it stored and how is it retrieved?

The three models of speech production that will be described in chapter 2 all assume that gender at one point became stored in the mental lexicon and can just be looked up when needed. The motivation for this view comes from the fact that the relation between a noun and its gender seems to be arbitrary (Schriefers & Jescheniak, 1999). However, there are linguists (e.g. Corbett, 1991) and psycholinguists (e.g., Johansson, 2003; Schwichtenberg & Schiller, 2004) who believe that for some words gender can also be computed anew from “a set of rules and regularities that governs the selection of gender” (Schwichtenberg & Schiller, 2004:327) each time it is needed.

As we will see in chapter 2, the most prominent models of speech production today, make specific claims about the way lexical syntactic properties like grammatical gender are stored. The three models of word production described in chapter 2, all assume that gender is stored with the noun in the mental lexicon instead of being computed anew each time it is needed.

"Because theories of language production make clear assumptions on the storage, retrieval, and use of such lexical-syntactic properties, grammatical gender provides a promising testing ground for examining these issues" (Schriefers & Jescheniak, 1999:576).

Grammatical gender also has implications for other fields than psycholinguistic processes in speech production, for example natural language processing, where it can be a useful tool for solving local ambiguities in parsing (Corbett, 1991) as shown in sentence (2) to (5).

(2) *Det var en stor ball.*
'It was a$_{masc}$ big$_{masc}$ ball$_{masc}$' (as in the sense: *football*).

(3) *Det var et stort ball.*
'It was a$_{neu}$ big$_{neu}$ ball$_{neu}$' (as in the sense: *dance party)*

(4) *Det var en fin bord.*
'It was a$_{masc}$ nice$_{masc}$ trimming$_{masc}$.'

(5) *Det var et fint bord .*
'It was a$_{neu}$ nice$_{neu}$ table$_{neu}$.'

The ambiguities in sentence (2) and (3) can be resolved by looking at the noun's context, whereas in the English translation the two senses cannot be disambiguated just by looking at the noun's syntactic environment.

1.2 Gender in Norwegian

Norwegian has three genders: feminine, masculine and neuter. Norwegian is a language with many dialects and has two official written norms, one called Bokmål and the other Nynorsk. While Nynorsk and the traditional dialects (except the Bergen city dialect) have a thoroughly developed three gender system (Faarlund et al., 1997), Bokmål has a more complicated gender system.

'Bokmål' is the official name for 'Riksmål' since 1929. Riksmål has its roots in the Danish written language and educated urban speech, and had at first, like Danish, a two-gender system with the genders neuter and common. However, through official language reforms in 1917 and 1938 a three-gender system was introduced in Riksmål/Bokmål, resulting in much variation in actual written usage (Torp & Vikør, 1993). Hence the three gender system in Bokmål is less strict, and words with feminine gender can vary between masculine and feminine, all depending on ”dialectical, sociolectical, stylistic and language political conditions” (my translation, Faarlund et. al., 1997:151).

In Norwegian, grammatical gender shows in determiners, adjectives, possessives, definite form suffixes, and in Nynorsk, also in pronouns. I will only give examples from Bokmål, since this written norm is used in all the material in this study.

Table 1 shows four different types of singular NPs that are gender-marked. The second column shows an indefinite NP for a masculine, a feminine and a neuter noun. The indefinite article singular is *en* for masculine, *ei* for feminine and *et* for neuter nouns. In an indefinite singular NP with an adjective, the adjective is gender-marked. An example of that is also shown in the second column in Table 1. The adjectives used in the examples are the adjectives *liten* [small], which may be the only adjective that has different forms for masculine and feminine, and the regular adjective *rød* [red].

The definite article in the singular is realized as a suffix attached to the noun's stem. An example for all the three gender-marked suffixes is shown in the third column in Table 1. However, if the noun is modified by an adjective, the definite form is then realized not only through the suffix, but also through a determiner. The examples in the fourth column in Table 1 show NPs consisting of a definite adjective + noun. (Here we can see that the adjective *liten* changes to *lille* in the definite form singular).

The demonstratives have the same form in the feminine and the masculine. A demonstrative NP usually has double definiteness as in the phrase *den bilen*, which directly translated would be *that car_the.* Simple definiteness occurs with non-specific reference plus restrictive relatives ("Den student som gjør slikt, blir utvist" [the student who does these things, will be expelled]), proper names ("Det hvite hus" [The white house]), and in some other cases in formal style. An example for each gender is found in the fifth column in Table 1.

Table 1. Fractions of the gender system in Bokmål in singular.

	Indefinite singular	Definite singular	Definite singular, NP with Adj	Demonstrative singular
Masculine	**en** (liten/rød) bil [a (small/red) car]	Bil**en** [the car]	**Den** lille/rød**e** bil**en** [the small/red car]	**Den** bil**en** [that car]
Feminine	**Ei** (lita/rød) hytte [a (small/red) cabin]	Hytt**a** [the cabin]	**Den** lille/rød**e** hytt**a** [the small/red cabin]	**Den** hytt**a** [that book]
Neuter	**Et** (lite/rød**t**) tog [a (small/red) train]	Tog**et** [the train]	**Det** lille/rød**e** tog**et** [the small/red train]	**Det** tog**et** [that train]

In indefinite form plural the masculine and feminine nouns get -*(e)r* added to the stem. Most monosyllabic neuter nouns get no ending added, while polysyllabic neuter nouns ending in -*e* get -*r: værelse**r*** [rooms], *teppe**r*** [carpets]. The adjective *liten* [small] is probably the only adjective that has different form for masculine and feminine as mentioned earlier. This adjective changes to *små* in the plural. In other words, this is not a typical Norwegian adjective. The adjective *rød* [red], however, is a regular adjective and an example of the regular inflection form for adjectives in indefinite form plural, where it gets an -*e* added to the stem, is shown in the second column.

In definite form plural, all nouns get the same ending -*(e)ne,* except the neuter nouns which either can take -*a* as the plural ending or -*(e)ne.* An example for each gender is shown in the third column in Table 2.

Table 2. Fractions of the gender system in Bokmål in plural.

	Indefinite plural	Definite plural	Definite plural, NP with Adj	Demonstrative plural
Masculine	(Små/rød**e**) bil**er** [(small/red) cars]	Bil**ene** [the cars]	**De** små/rød**e** bil**ene** [the small/red cars]	**De** bil**ene** [those cars]
Feminine	(Små/rød**e**) hytte**r** [(small/red) cabins]	Hytte**ne** [the cabins]	**De** små/rød**e** hytte**ne** [the small/red cabins]	**De** hytte**ne** [those cabins]
Neuter	(Små/rød**e**) tog [(small/red) trains]	Tog**ene** [the trains]	**De** små/rød**e** tog**ene** [the small trains]	**De** tog**ene** [those trains]

If a definite NP contains an adjective, the adjective has the same form for all three genders, i.e. it gets an *-e* added to the stem as can be seen in the fourth column in Table 2. In a demonstrative NP plural, there is no difference between the masculine, feminine and neuter nouns, as shown in the fifth column.

As mentioned earlier, Bokmål has a more complex gender system than Nynorsk. In addition to the three gender system, there is also a two gender system. In this system, there is no feminine marking of the agreement targets. The feminine words behave like the masculine, so they are called common gender. Thus, we get a distinction between neuter gender on one side and a *common* gender on the other. While the three gender system is used in "radical" Bokmål, the two gender system is used in conservative Bokmål (Faarlund et. al., 1997).

Not only is there a two gender and three gender system, but also a "two-and-a-half" gender system. The three systems are contrasted in Table 3. In a two-and-a-half gender system, a speaker varies between using a specific noun in the feminine form or the masculine form depending on the noun phrase. The same speaker might say ***en*** *bok* [a book $_{\text{masc}}$] but *bok*___a___ [the book $_{\text{fem}}$]or ***min*** *bok* [my book $_{\text{masc}}$] but *bok*___a mi___ [the book of mine $_{\text{fem}}$]. In written language some constructions are more used than others. *En bok* (masc) is more common than *ei bok* (fem), *min bok* (masc) is more common than *mi bok* (fem). *Boken* [the book $_{\text{masc}}$] and *boka* [the book $_{\text{fem}}$] are about equally common and so are *boken min* [the book of mine $_{\text{masc}}$] and *boka mi* [the book of mine $_{\text{fem}}$]. (Faarlund et. al., 1997). There is some difference between written form and spoken dialect when it comes to how consistent and how often the feminine forms are used. The feminine form is more an "everyday" form and occurs more frequently in colloquial speech than in written language.

*Table 3 Differences between a three gender system, two-and-a-half gender system and a two gender system. Phrases marked with ' * ' are not likely to be used by the speaker within this gender system, or is never used.*

	3 gender system	2.5 gender system	2 gender system
Indefinite singular	*en (liten) bok	en (liten) bok	en (liten) bok
	ei (lita) bok	*ei (lita) bok	*ei (lita) bok
Definite singular	boka (mi)	boka (mi)	*boka (mi)
	*boken (min)	boken (min)	boken (min)

For the majority of Norwegian nouns, it is not apparent from its form which gender a noun has, and the assignment of gender to nouns does not seem to be rule-based (Faarlund et. al., 1997). However, this opinion has recently been challenged (Trosterud, 2001). In Nynorsk, where data about gender distribution is available, about 40% of the nouns are masculine, about 30% are feminine and about 25% are neuter (Beito, 1954 via Faarlund et. al., 1997). According to Norsk referansegrammatikk “the distribution is mostly the same for Bokmål, but that is if we count the words which can take either *-en* or *-a* in definite form in Bokmål as feminine” (my translation, Faarlund et. al., 1997:152).

In chapter 2 we will look at the processing of grammatical gender in speech for three different models of speech production. In chapter 3 I will describe the method used in my experiment, the picture-word interference paradigm. Chapter 4 describes previous studies concerning grammatical feature selection. In chapter 5, the experiments are described in detail and the results from the experiments are discussed. A computational model based on the findings from the experiments is adapted to fit WEAVER++ in chapter 6. The simulations done in WEAVER++ are also reported in chapter 6. In chapter 7, I discuss the general findings from the experiments and the computer simulations, briefly describe questions that remains unanswered, propose future work to solve them and a conclusion sums up the work in this thesis.

CHAPTER 2, MODELS OF SPEECH PRODUCTION

As mentioned in chapter 1, producing speech is a process containing many sub-processes. Part of it involves the selection of a word "(...) that is semantically and syntactically appropriate, (2) the retrieval of the word's phonological properties; (3) the rapid syllabification of the word in context; and (4) the preparation of the corresponding articulatory gestures" (Levelt, 1999:223). One of the processes involved is called lexicalization, and refers to the process of turning the semantic meaning of a content word into its phonological form (Harley, 2001). In this process three questions arise that a model of speech production has to account for. The first is: how many stages are involved in this process? The second is: what is the time course of the processes involved? The third is: do the stages involved interact with each other or are they independent? (Harley, 2001). We will return to these questions in section 2.1 to 2.3 where three different models of speech production are described. First, we will look at one of the areas where studies of grammatical gender in speech production have had an impact on theory, more specifically, the split in the lexicalization process between the word as syntactic unit and the word as a phonological unit.

Evidence from gender studies have been used as support for the assumption that the lexicalization process occurs in two steps. Within the standard theory of speech production, initially developed by Garrett (1975 cited in Dell, 1996:328) one assumes that lexicalization involves two stages going from a mental concept to its phonological word form. It is assumed that there is an intermediate state between concept and phonological form where a word's syntactic information is stored. This intermediate representation is called the lemma.

Retrieval of lemmas is an extremely rapid and accurate process. In normal conversation a speaker might produce two to three words per second (Levelt et al., 1999), and therefore lemmas are likely to be retrieved at an equal rate. Increasing this to five words per second is not particular problematic (Levelt, 1989). Though; "In connected speech lemmas can be retrieved in parallel, so the number of words per second does not reveal the speed of retrieval of a single lemma *per se* " (Roelofs, 1996:309). Given the speed in retrieving and producing words, the speaker makes noteworthy few errors. A speaker only fails to retrieve the appropriate lemma roughly once per 1000 words in normal conversation (Levelt, 1999).

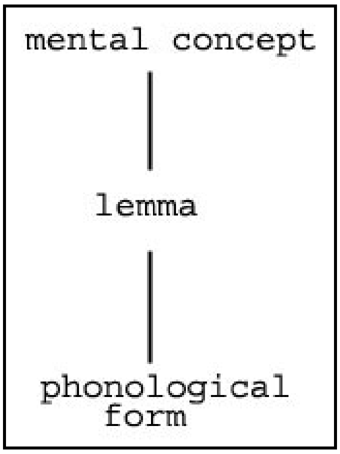


Figure 1. Two-stage model of lexicalization (inspired by Harley, 2001)

Psycholinguistic findings from studies on gender have given support to the claim that lexicalization occurs in two stages. Support for a two stage process, a lemma stage where the lexical concept's syntactic (and semantic) information is retrieved and a phonological stage where its phonological word form is retrieved, comes from the studies of the Tip-of-the Tongue (TOT) phenomenon among other. Having a word on the tip of the tongue is something we probably all have experienced. The TOT-state is a state where you cannot produce the word you want to say, even though you know that you know this word. You might even be able to hum the rhythm of it, and maybe know the first letter or its grammatical gender. One way of observing TOT-states can be by diary studies (e.g., Burke, MacKay, Worthley & Wade cited in Harley, 2001:362). However they can also be provoked in experiments. A normal way to do this is to read a definition of a rare object to the participants and ask them to name the word, as for example:

> "A navigational instrument used in measuring angular distances, especially the altitude of the sun, moon, and stars at sea." (Brown and McNeill, 1966 via Harley, 2001:362). [2]

If the respondent is put in a TOT-state, following-up questions are performed, either by a questionnaire or by interviewing the subjects, asking them if they know the initial sound of the word, how many syllables the word is made up of, its grammatical gender etc. (Schriefers & Jescheniak, 1999). One of the strongest evidence for two stages in lexicalization comes from

---

[2] The word is 'sextant'

studies of grammatical gender in TOT-states. Vigliocco, Antonini and Garrett (1997, cited in Harley, 2001) found that their Italian subjects could retrieve syntactic information about the target item, such as its grammatical gender, but without being able to retrieve its phonological form.

Studies of brain-damaged patients have also supported a two-stage process. Badecker, Miozzo and Zanuttini (1995 cited in Harley, 2001) observed an Italian person with anomia, which is an inability to name objects or to recognize written or spoken names of objects. The person, Dante, could report detailed information of grammatical gender for words even though he could not produce them. This indicates success in accessing the word's lemma (with all its syntactic information), but failure to access its phonological form.

Further evidence from gender studies in support of a two-stage model comes from electrophysiological experiments. Van Turennout, Hagoort and Brown (1998, cited in Schriefers & Jescheniak, 1999) did an experiment with Dutch subjects, where the subjects were shown a colored picture and had to name it with a gender-marked noun phrase (e.g., *red table).* Adjectives are gender-marked in Dutch. The experiment consisted of two tasks for the participants. In the first task, the participants had to give a binary push-button response on whether the noun was of common gender or neuter gender, but only if the noun started with a certain phoneme. In the second task, the respondents had to make a binary decision of the noun's first phoneme, but only if it had a certain gender. In other words, there were go and no-go trials. During the tasks, the lateralized readiness potentials (LRPs) were registered. Data from the preparation of motor movements showed that the subjects prepared for pushing the button in the first task, even in the no-go trials. In the second task, however, LRPs were only obtained for the go trials. This indicates that syntactic information, such as gender, is accessed before phonological information.

Not all current models of speech production operate with the abstract lemma level, and studies on grammatical gender have been one of the reasons why Caramazza (1997) dispenses with it. Even though most current models of speech production give different accounts for the three questions mentioned in the beginning of the chapter, they have in common that they are all network models of some kind (Levelt, 1999). According to Levelt (ibid), the majority are also “localist”, non-distributed models. This means that a node in the network represents one linguistic unit in the network, unlike a distributed model where one linguistic unit is spread over several nodes that together make up that unit. A linguistic unit can for example be a semantic feature, a morpheme, a phoneme, a mental concept or a grammatical feature. I will briefly explain three different models of speech/word production. The first is Levelt, Roelofs and Meyer's model which incorporates two independent stages in the lexicalization process, with lemma selection and retrieval of phonological form occurring in a strict serial order and without interaction between the two levels. The second model is Dell's model which also assumes the same two stages in lexicalization. However, the two stages are assumed to interact. In contrast to Levelt et al.'s model, not only the selected lemma activates its phonological form, but so do other active lemmas. The third model of speech production is Caramazza's Independent Network Model (IN model) which dispenses with the lemma as an intermediate level between a concept and its phonological form. Instead the split between phonological form and syntactic information, including gender, is contained by a construction of different networks, where syntactic information is in one network and the phonological word form information is in another independent network. While describing the production process in these three different models of speech production, I will concentrate on the production of single-word utterances.

2.1 The serial discrete two-step model

Levelt, Roelofs and Meyer's serial discrete two-step model of speech production (1999), (captured in the computer model WEAVER++ by Roelofs, 1992), is primarily based on findings from several years of reaction time experiments. Among the core findings is that there is an early stage where lemmas are selected and a later stage where the phonological forms are accessed (Levelt et al. 1991). The experiments (ibid.) showed that same category members of the target item and the target item was activated at the semantic level, the target item was phonologically activated, but the semantic alternatives were not. In other words, if the target item was CAT[3], the lexical concept of DOG, FOX etc was activated at the semantic level. CAT became phonologically activated, but not DOG or FOX. This finding suggests that only the selected lemma is subsequently passed on to the phonological word form level.

The model, schematically represented in Figure 2, operates with three main levels in the lexicalization process. The first level is the conceptual level or stratum. Each mental concept

---

[3] Lexical concepts are denoted with capital letters throughout the text, and lemmas with small letters.

represented as a holistic unit (a "whole"), unlike the two other models described below, is believed to be connected to a network of other concepts or nodes. In Figure 2, the drawing of the cat represents the target concept CAT at the conceptual layer. Circles symbolize nodes in the network, and the arrows connections. The arrow head(s) show into which direction the activation flows. Note the unidirectional arrow from lemma to word form level. The N stands for noun. The thickness of the red lines symbolizes the degree of activation. The thicker the line, the higher degree of activation.

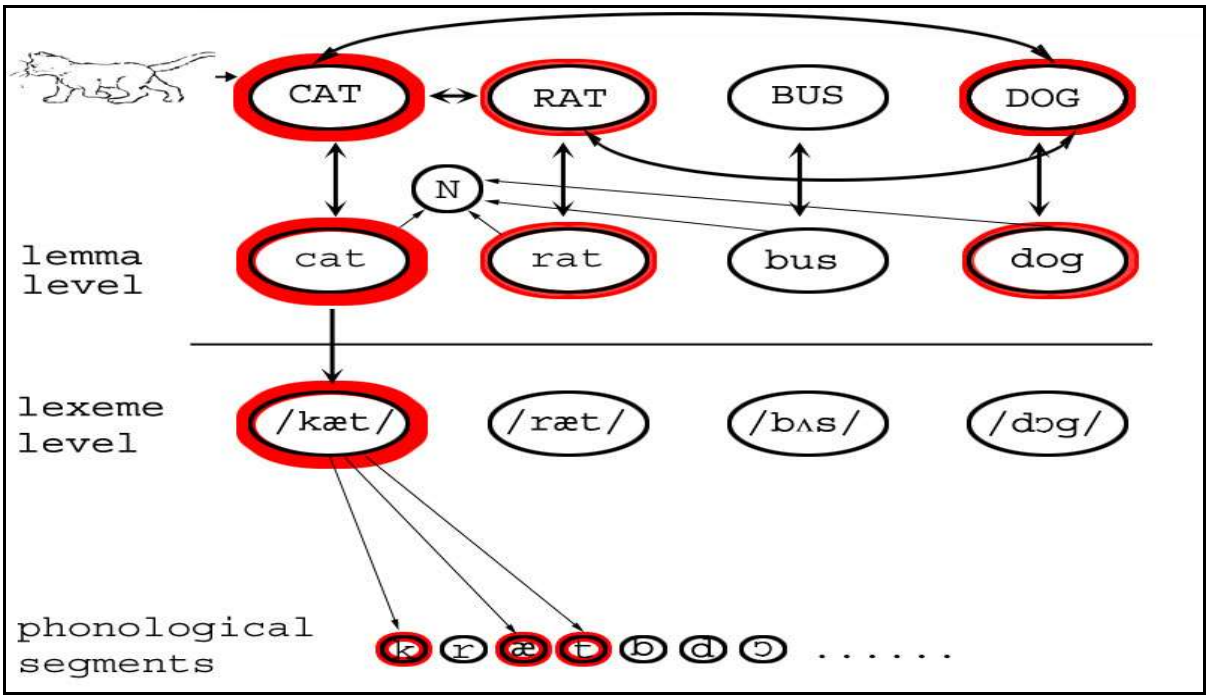


Figure 2. A simplified version of how activation flows in a serial discrete two-stage model (e.g., Levelt, Roelofs & Meyer, 1999). The phonological notation is informal.

When activating a lexical-semantic concept (such as CAT in Figure 2), activation spreads through the connections to its semantic neighbors at the conceptual level. These activated nodes are then input to the second level, the syntactic level or stratum. The nodes at the syntactic level are the lemmas. The lemmas at this level compete for selection and there are no links between the lemmas. Only the lemma with the highest degree of activation is selected and passed on to be realized phonologically at the third level, the phonological word form level (or lexeme level). This implies that a word's grammatical feature is selected prior to its phonological form.

Processing between the conceptual level, lemma level and lexeme level is thought to be spreading of activation along the links between the different stages. The target concept, and to some degree its semantic neighbors, then send activation along the links down to the lemma level where all the grammatical information about the lexical concept is retrieved.

In Levelt et al.'s model (1999), grammatical gender is stored in the mental lexicon and can just be looked up if needed. However, it is not stored once for each entry in the lexicon, but rather, it is stored so that all nouns of same gender are connected to a gender node which specifies that gender. We can think of this as an abstract gender node which we will return to in chapter 5. The gender node is in turn connected to all agreement targets, which are the concrete realization of gender (e.g., adjectives, pronouns, articles etc.). The activation flows only in one direction, so there is no activation flow from a gender node back to the lemma. "...this entails that retrieval and selection of grammatical gender of a noun and the computation of gender agreement with some agreement target (such as articles) cannot be affected by phonological form" (Schriefers & Jescheniak, 1999:579). The uni-directional activation flow from lemma to phonological form means that in serial discrete two step model, gender cannot be computed from phonological cues each time it is needed anew. None of the models described here actually assume that gender is computed, but rather they assume that it is stored with the noun.

2.2 Interactive two-step model

Dell's Interactive activation model of speech production (1986) is also a two-step model, meaning that there is a separation between the lemma level and the lexeme level. Dell's theory of speech production is also captured in a computer program. The theoretical model is primarily based on findings from speech errors unlike Levelt et al.'s model which is primarily based on findings from reaction time experiments.

Dell's model is referred to as interactive because activation flows not only top down (as the discrete two step model), but also bottom up. When activating a mental concept (represented as a set of distributed features in Dell's model of speech production in Figure 3) activation flows to other concepts that share features with the target concept. Activation then cascades from level to level for all the activated nodes. This means that not only the target lemma is passed on to activate its phonological form (as in the serial discrete model two-step model), but also all the other candidates that have received some activation at the conceptual level. In Figure 3, the drawing of the cat represents the target concept CAT at the conceptual layer. The nodes at the conceptual level represent semantic features that make up a concept. The red nodes represent some of the features activated from the concept CAT. In this simplified example, RAT and DOG share one feature with the cat and receive activation through the connections from that feature. Once a selected lemma has activated its phonological representations (e.g. /kæt/), these can then send feedback through the bidirectional links and activate lemmas to phonological similar words like rat, cap and mat.

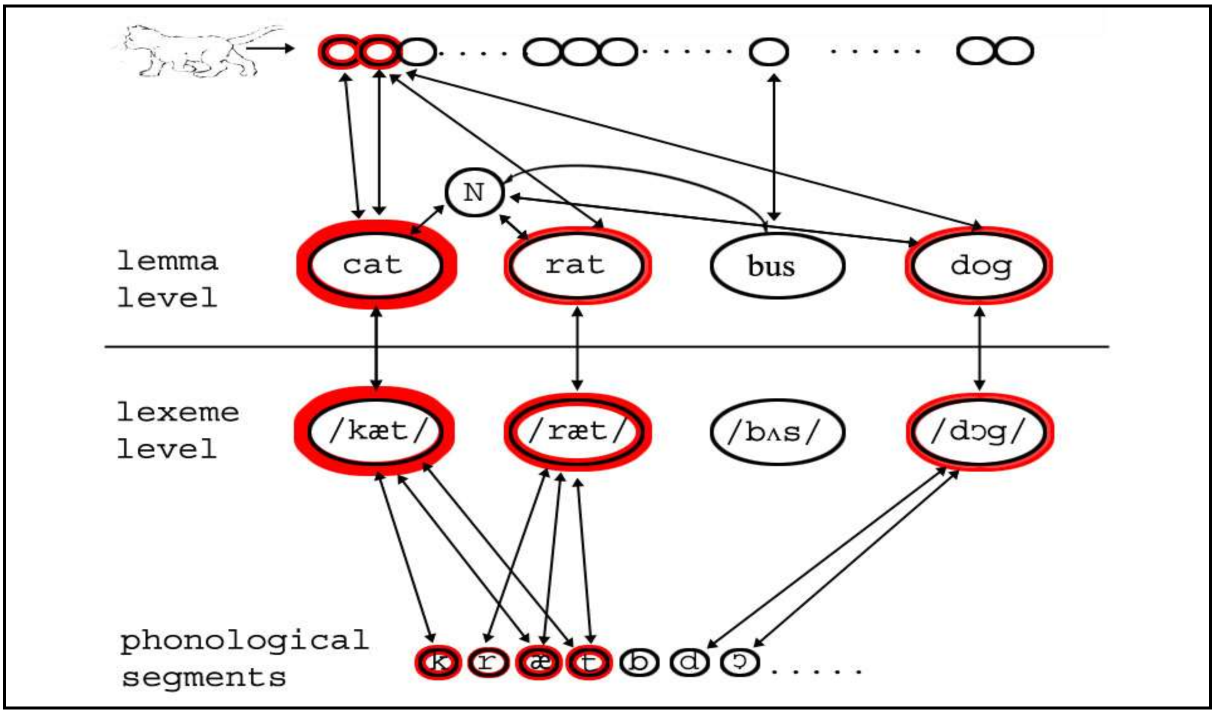


Figure 3. Activation flow according to an interactive model (e.g., Dell, 1986). Note the bidirectional arrows. The phonological representations are informal.

The motivation for incorporating the interactive aspect of the model, was to account for the greater- than-chance finding in speech error corpora of mixed errors. These are errors which has a degree of both semantic and phonological character, like producing the word *rat* instead of *cat.* The word rat is more likely to be selected than a word which is just semantically related to the target, like DOG, and a word which is only phonologically related to the target, like /mæt/ (the example is taken from Levelt, 1999).

Dell's model from 1986 was mainly focused on the phonological aspects of speech production and does not make specific claims on how grammatical gender is stored or processed. It is not clearly specified whether activation flows from lemma to gender node and back, or if it flows only from lemmas to gender node. In principle, the cascade activation would make it possible to retrieve phonological word forms without accessing gender information (Schriefers & Jescheniak, 1999).

2.3 Independent network model

Caramazza's (1997) Independent network model (IN model) was mainly developed to account for naming errors in brain damaged subjects. In the IN model, lexical knowledge is organized in sets of separate independent networks, shown in Figure 4. In the lexical-semantic network word meaning is represented as sets of features. Another network is the syntactic network which represents a word's syntactic features such as grammatical category, gender, tense, etc. The nodes in the lexical-syntactic network are organized into subnetworks. One subnetwork contains the category nodes, such as Noun, Verb, Adjective, etc. Another subnetwork represents gender nodes, such as feminine, masculine and neuter, and so on. Nodes within a subnetwork have inhibitory connections between them, because they are in competition with each other. The phonological lexemes with connections to segmental phonological information are in a different network, the lexeme network.

The connections between the lexical-semantic network and the lexeme network are strong. The connections between the lexical-semantic network and the syntactic network are weak (the weak connections are indicated as dotted lines in Figure 4). The activation flow is feed-forward. However, Caramazza (1997:204) does not make a definite claim on the dynamics of activation (discrete vs. continuous; strictly forward vs. forward and backwards propagation) and selection of representations at different levels of processing, nor on the time course aspect for the different levels of representation.

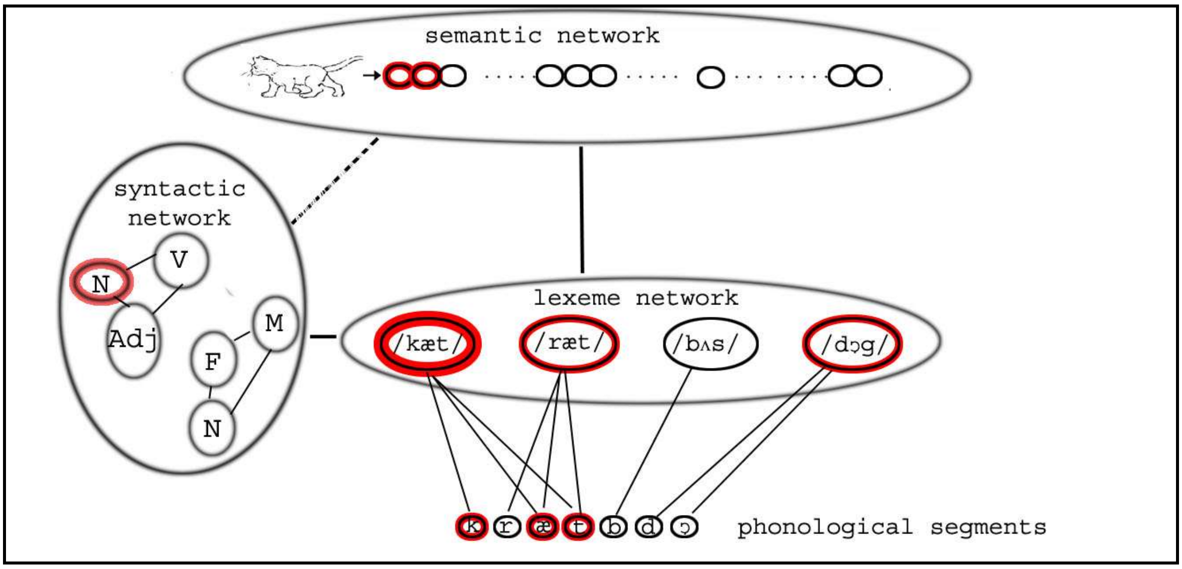


Figure 4. A representation of the network structure in Caramazza's Independent Network model.

The activation flow is, as said, feed-forward and when activating a lexical semantic/mental concept, all other concepts that share features with the target concept become weakly activated. After selection of the lexical-semantic representation, activation spreads independently and simultaneously to the syntactic network and the lexeme network. The other concepts that were activated because they share features with the target concept, also activate their associated lexeme. Grammatical category and verb tense for example can also become weakly activated through the links between the lexical-semantic network and the lexical-syntactic network. This activation is normally not sufficient for a grammatical feature to reach the selection threshold.

> "Not all grammatical features can be activated by the semantic network. For example, with the exception of natural, gender-marked words (e.g. uomo [man] in Italian), gender features do not receive activation from the semantic network. However, grammatical category and verb tense features, for example, do receive activation from the semantic network..."(Caramazza, 1997:195).

Unlike the serial discrete two-step model and the interactive two-step model, the IN model does not incorporate the notion of lemma. Strong evidence in support of the lemma as an intermediate level between the lexical-semantic and phonological form (or lexeme) level, has been the findings that grammatical gender and other grammatical information can be retrieved in a TOT-state. However, Caramazza and Miozzo (1997) found that Italian speakers can retrieve partial phonological information of a word while in a TOT condition, but not grammatical gender and vice versa. This is in conflict with one interpretation of the serial discrete two-step model, where the lemma (with all its syntactic information) has to be selected before its phonological form becomes available.

Like Levelt et al.'s model, all nouns of same gender are connected to a gender node which specifies that gender. But as mentioned before, the link goes from the phonological lexeme network to the gender node and not from lemma to gender node.

In chapter 4, we will look more into the ongoing debate concerning the selection of grammatical gender. Is a noun's grammatical gender selected through activation and competition (e.g. Schriefers, 1993) or is grammatical features, like gender, an automatic, non-competitive consequence of selecting a lemma or lexical node (e.g. Schiller and Caramazza, 2003)? Before we move on to chapter 4, a method often used in studies of grammatical gender in speech production, the picture-word interference paradigm, will be described.

CHAPTER 3, THE PICTURE-WORD INTERFERENCE PARADIGM

The picture-word interference paradigm (PWI) is a modified version of the original "Stroop-task". In Stroop's original experiment, subjects were asked to name the colors in a series of color patches and a series of colored color words. It took significant longer to name the colors of incongruent color words, than it took to name the color of colored patches (Stroop, 1935). The effect of increase in reaction time is known as the "Stroop interference". This effect is perhaps the best known example of word interference. An example of a typical Stroop-task, to name the color of the ink the color word is written in, is shown in Figure 5a (color congruent) and 5b (color incongruent).

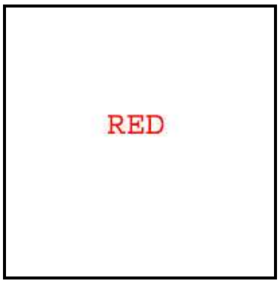


Figure 5a. An example of the Stroop-task where red is the target word. (The word is written in red).

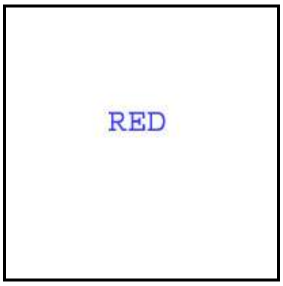


Figure 5b. An example of the Stroop-task where blue is the target word. (The word is written in blue).

Word interference can occur for all sorts of words and all sorts of pictures, not only colors (see MacLeod, 1991 for a review on the literature).

> “The interest in the effects of (...) word interference originates in part from the view that these effects involve the selection of the pictures' names rather than other stages of picture naming (e.g., picture recognition or meaning retrieval), and, therefore, they could be used to inform theories of lexical access...”(Miozzo & Caramazza, 2003:228).

In the picture-word interference paradigm subjects have to respond to a picture (e.g. a line drawing) presented to them while ignoring a superimposed distractor word. The reaction time, i.e., the time it takes to name the picture or push a button, is recorded. It is possible to vary the time interval between when the picture is shown and when the word is superimposed, this time relationship between target and distractor is called stimulus onset asynchrony (SOA). If picture and word are displayed at the same time, the SOA=0. The distractor word does not necessarily have to be a written word, it can also be an auditory distractor. That is, the respondents hear a distractor word instead of reading it. If the distractor word is displayed, for example, 100 ms before the target picture, the SOA= -100 ms. If the distractor word is displayed, for example, 100 ms after the target picture, the SOA= +100 ms. With different SOAs it is possible to investigate the time course aspect of certain effects. Many studies have been conducted with this method and there are two well-established findings that stand out in the literature.

The first well-established finding I will describe is the effect of semantic interference (e.g., Lupker 1979, Glaser & Düngelhoff, 1984; Roelofs, 1992; inter al.). It takes longer to name a picture of a car if a semantically related distractor word like *bus* is superimposed than it would if a semantically unrelated word, such as *glove*, appeared as distractor word (see Figure 6a and 6 b).

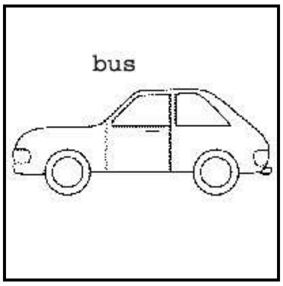


Figure 6a. An example of semantically related picture-word pair in the PWI

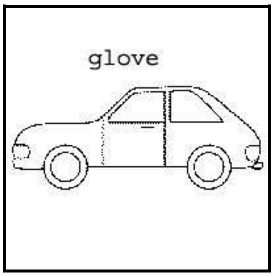


Figure 6b. An example of semantically unrelated picture-word pair in the PWI

To explain this effect, we first have to assume two things. One, that when a lemma (or lexical node in Caramazza's independent network model in Figure 4, chapter 2)[4] is activated, it spreads some activation to other semantically related lemmas through the connections they share at the semantic level. The lemmas can be thought of as nodes in a network as we have seen in chapter 2, Figure 2, 3, and 4. For example, if the speaker wants to say the word *cat*, the lemma nodes of *dog, rat, animal, hamster,* etc. will also become activated. The second assumption is that the selection of a lemma node is sensitive to the activation level of the other activated lemma nodes. In other words, for a node to be the one that is actually selected among all the other alternatives (that is the other activated nodes), it has to exceed the other nodes activation level with a "certain amount".

---

[4] I will use the notion lemma which is called lexical node by Caramazza.

With this in mind, the semantic interference effect found in picture-word interference tasks can be explained as follows: The target picture activates its lexical concept and lemma. In addition it activates semantically related concepts which in turn activate their lemmas. So, if the target picture is one of a cat, then the lemmas of dog, rat, hamster, tiger etc are also activated. If the distractor word is *dog,* the lemma and lexical concept DOG will become activated. Through the connections it shares with a cat at the semantic level, the concept and lemma node of CAT will receive activation not only from the target picture, but also from the distractor word. And the same goes for *dog's* lemma node which not only receives activation from the distractor word but also from the target picture of the cat. However, if the distractor word is semantically unrelated to the target picture, for example the word *glove,* the lemma of *glove* will not receive activation from *cat*, only from the distractor word itself. This means that the lemma node of *glove* will have a lower activation level than that of *dog* (which receives activation from two sources). Since the difference in activation levels between *cat* and *glove* is bigger than between *cat* and *dog*, that "certain amount" that the target node has to exceed the other nodes' activation level in order to be selected for further processing will be reached faster with a semantically unrelated distractor word.

The semantic interference effect usually appears within a limited time range and is usually detected with stimulus onset asynchronies (SOAs) of about -200 ms (distractor word presented 200 ms before the target picture) to +200 ms (distractor word presented 200 ms after the target picture) (e.g., Glaser and Düngelhoff, 1984; La Heij, Dirkx, & Kramer, 1990 cited in Starreveld & La Heij, 1995).

The second finding is the phonological or orthographical facilitation effect (e.g., Underwood & Briggs, 1984; Starreveld & La Heij, 1995). The naming latencies are reduced if the superimposed distractor word is orthographically or phonologically related to the picture name compared to a word that is phonologically or orthographically unrelated to the picture name. Underwood and Briggs (1984) reported that they found a facilitation effect in picture naming with orthographically related distractor words. For example: The picture of a *comb* was named faster when the distractor word was orthographically related to the word like *bomb* than when the distractor word was *pace,* which is orthographically unrelated.

> “These interference and facilitation effects are assumed to reflect processes at different levels of lexical access. The semantic interference effect is commonly thought to reflect competition at the level of lexical node selection, and the phonological facilitation effect is thought to reflect priming of the phonological content of the lexical node selected for production. Therefore the investigation of these effects could reveal properties of the lexical access system” (Caramazza et al., 2001:212).

It usually takes about 150 ms to process a picture visually and activate the appropriate concept. Selecting the lemma takes around 125 ms. Phonological encoding takes place around 275 ms and the utterance of the word starts from around 600 ms (Harley, 2001). The time intervals for these processes are shown in Figure 7. However, these times are estimates and can vary. It is possible to name a picture faster than 600 ms as can be seen in the response times reported in chapter 5.

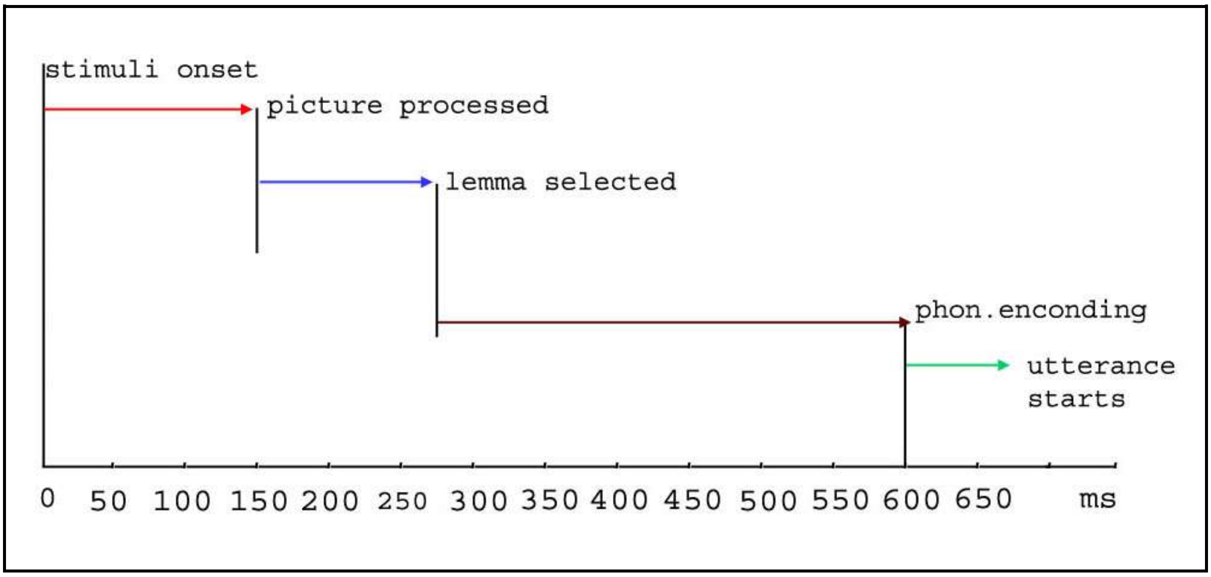


Figure 7. A representation of the time interval for the different steps in the picture naming process.

Reading a word aloud is a faster process. For example, Cattell (1885, cited in Glaser, 1992) found that reading a list in his particular experiment of 100 nouns took about 25-35 seconds, while naming a comparable list with line drawings or colored dots took about 50-60 seconds. This difference in speed between picture naming and word naming has been replicated many times since then (e.g., Glaser & Düngelhoff, 1984, experiment 1; inter al. via Glaser, 1992). Fraisse (1967, 1969 cited in Glaser, 1992) performed an experiment, where the identical symbol O was named as *circle* in 619 ms, as *zero* in 514 ms and read *oh* in 453 ms.

In the next chapter we will see how the PWI-paradigm has been used to investigate the processes involved in grammatical feature selection, more specifically the process of grammatical gender retrieval.

CHAPTER 4, STUDIES ON GENDER PRODUCTION

Schriefers (1993) extended the use of the picture-word interference paradigm to investigate the mechanisms that govern syntactic processes in Dutch noun phrases. He varied whether the grammatical gender of the distractor word was congruent (i.e. both target picture name and the distractor word had the same gender) or incongruent (i.e. the target picture's gender was different than that of the distractor word) to that of the target picture.

> “Schriefers reasoned that if grammatical feature selection functions with principles similar to those involved in the selection of lexical nodes and phonological segments (i.e., graded activation and selection competition), the manipulation of gender relatedness should produce measurable effects” (Caramazza et al. 2001:212).

Varying the grammatical gender in this manner did produce measurable effects (see section 4.1 for more details on the effects obtained in Schriefers' experiment from 1993) and the picture-word paradigm has been used to investigate the processes of gender selection in various languages (e.g., Schriefers 1993; Miozzo & Caramazza, 1999; Schriefers & Teruel, 2000; Van Berkum, 1997; La Heij et al., 1998; Alario & Caramazza, 2002; Schiller & Caramazza, 2003).

4.1 Previous studies on gender production

Schriefers (1993) conducted an experiment with Dutch speakers in order to study the syntactic processes involved in determining the definite article and the adjective inflections in production of Dutch noun phrases. In Dutch, the definite article can either be *de* or *het. De* is the definite article for masculine (or common) nouns and *het* for neuter nouns. (I will use the common gender term for the rest of the text). In Dutch both the definite article and the adjective is placed in front of the noun in a NP. Examples of Dutch NPs with gender-marked definite article is shown in sentence (6) and (7).

(6) *de groene stoel* [the green chair, common]
(7) *het groene bed* [the green bed, neuter]

NPs without definite article, but with adjective, are gender-marked by an inflection suffix added to the adjective's stem as shown in sentence (8). For neuter gender the adjective is identical to the stem as in sentence (9).

(8) *groene stoel* [green chair, common]
(9) *groen bed* [green bed, neuter]

Subjects were asked to name colored line drawings producing NPs on the form Det + ADJ + N (het groene bed / de groene stoel) or ADJ + N (groen bed /groene stoel). The SOA was varied in three steps with SOA = -200, SOA = 0, and SOA = +450. The results showed (for SOA = -200 and SOA = 0) that it took longer to name the picture if the distractor word had another gender than the target word. Schriefers interpreted this finding to be a result of competition between the target's gender information and the distractor word's gender information when selecting the appropriate gender node. This effect was referred to as the *gender congruency effect* by Schriefers.

The gender congruency effect has later been found in several experiments for Dutch when subjects were to produce a gender marked NP (e.g. Van Berkum, 1997; La Heij et al., 1998; Schiller & Caramazza, 2003), German (e.g. Schriefers & Teruel, 2000; Schiller & Caramazza, 2003) and Croatian (Costa et al., 2003). However, La Heij et al, (1998) found that when subjects were to produce only a noun without a gender-marked element, there was no gender congruency effect. This indicates that gender is only selected when needed.

Miozzo & Caramazza (1999) failed to replicate the gender congruency effect found by Schriefers (1993) with Italian Speakers. As they pointed out, there is an interesting difference between the Italian gender system and the German and Dutch gender system. Consider the examples from (10 a) to (12 c) from Caramazza et al. (2001:217)

(10a) *Il treno/ i treni* [the train/ the trains]
(10b) *Il piccolo treno* [the small train]
(10c) *Il treno piccolo* [literally, the train small]
(11a) *Lo sgabello/ gli sgabelli* [the stool/ the stools]
(11b) *Il piccolo sgabello* [the small stool]
(11c) *Lo sgabello piccolo* [literally, the stool small]
(12a) *La forchetta/ le forchette* [the fork/ the forks]
(12b) *La piccola forchetta* [the small fork]

(12c) *La forchetta piccola* [literally, the fork small]

In Italian the proper determiner is not only dependent on the noun's gender (as it is for Dutch and German) but also on the local phonological context in which the determiner occurs. In Italian feminine nouns take *la* as the definite article in singular as shown in sentence (12a-c) and *le* as the definite article in plural as shown in sentence (12a). Masculine nouns can take two different definite articles. The determiners *lo* in singular and *gli* in plural are selected if the immediately following phonological context is a vowel, a consonant cluster of the form "s+consonant" or "gn", or with an affricate. Examples of that are shown in sentence (11a) and (11c). For the remaining cases the determiners *il/i* are selected (sentence (11b) and (10a-c)). This means that in order for the proper determiner to be selected the immediately following phonological context has to be specified first.

Miozzo & Caramazza (1999) and Caramazza et al. (2001) interpreted the absence of a gender congruency effect in Italian speakers to be caused by the fact that the proper determiner cannot be selected before the noun's phonological word form has been accessed. Retrieval of phonological form is a rather late process in word production, so any possible interference from the distractor word might have died out by this time and would not interfere with the selection of the target word's determiner.

Languages with this property (that is, the appropriate determiner is not immediately available) were referred to as *Late selection languages*. In Dutch and German sufficient information for selecting the proper determiner is available immediately after retrieving the noun's gender feature. Languages with this property were referred to as *Early selection languages* by Caramazza and colleagues.

According to Caramazza et al. (2001), the gender congruency effect has not been replicated in Romance languages where the selection of the appropriate determiner is depended on a combination of grammatical and phonological properties (e.g. Miozzo & Caramazza (1999) with Italian speakers; Alario & Caramazza (2002) with French speakers). Therefore they questioned Schriefers' interpretation of the gender congruency effect being a result of competition at the level of gender feature selection. Instead they suggested that the competition might be between the different determiners rather than the abstract gender features, and only found for early selection languages. Determiner competition, rather than gender feature competition, could also account for Schriefers' results, assuming that the distractor word activates its determiner.

Schiller and Caramazza (2003) conducted a series of experiments with German and Dutch speakers to test these two different accounts for the gender congruency effect. German has three genders, masculine, feminine and neuter. In nominative case, singular, masculine nouns take the determiner *der,* feminine nouns take *die* and neuter nouns take *das*. An example from each gender is shown in (13), (14) and (15).

(13) *der Tisch* [the table, masculine]
(14) *die Wand* [the wall, feminine]
(15) *das Buch* [the book, neuter]

Both Dutch and German have different determiners in singular, but they only have one determiner in plural. In nominative case the plural determiner in German is *die* (*die Tische* [the tables, masc], *die Wände* [the walls, fem], *die Bücher* [the books, neu]). The plural determiner in Dutch is *de* irrespective of gender (*de tafels* [the tables, com], *de boeken* [the books, neu]).

This property made it possible for Schiller and Caramazza to distinguish between the two potential causes for the gender congruency effect; the gender selection interference hypothesis (GSIH) proposed by Schriefers and the determiner selection interference hypothesis (DSIH) (Schiller & Caramazza, 2003). They predicted that if the gender congruency effect is caused by interference during grammatical feature selection, this effect would also appear in production of plural NPs. According to the determiner selection interference hypothesis on the other hand, a gender congruency effect should not appear during the production of plural NPs since all the nouns have the same determiner and therefore there should not be a conflict during the process of determiner selection.

In a series of experiments using the PWI-paradigm they investigated these different accounts for the gender congruency effect in German and Dutch. The gender congruency effect was found in the singular condition but not in the plural condition. This effect is predicted by the DSIH because the plural determiner is identical for all genders, therefore there should not be any competition. However, they did not manage to replicate Schriefers' gender congruency effect found in the Adj + N naming task. They interpreted the absence of this effect to be caused by the fact that “selection interference effect only occurs for free-standing morphemes (such as determiners), but not bound morphemes (such as inflections).“ (Schiller & Caramazza, 2003:188). Another possibility is that:

> “(...) since inflectional affixes must be "attached" to the end of adjective stems, they would only be needed fairly late in the process of NP production. On this account, any competition between inflectional forms would be resolved before they would be needed for attachment to the adjective stem, and therefore invisible in the type of experiments we have carried out” (Schiller & Caramazza, 2003:188).

They interpreted the effect found for singular conditions, but not for plural, such that retrieval of grammatical features, like gender, is an automatic process in lexical node selection and not exposed to competition from the distractor word's activated gender node, as Schriefers (1993) originally suggested. If a noun's gender feature was up for competition, the interference from the incongruent noun's gender information should lead to longer naming latencies also in the plural conditions. Further support for the DSIH comes from Schriefers, Jescheniak and Hantsch (2002).

4.2 This study on gender production, the hypothesis.

There have not been conducted similar experiments in Norwegian with the focus on processing of gender in speech. My hypothesis is:

“Norwegian is an early selection language, therefore, some sort of gender congruency effect should also be obtained for Norwegian speakers.“

The definite article in Norwegian (Bokmål) is a gender-marked suffix attached to the noun. Following Schiller and Caramazza's (2003) thoughts on gender-congruency effect only appearing for free standing morphemes (such as the determiners) and not inflectional forms, we would not expect a gender congruency effect to appear for the definite form of the noun. Therefore, the task was to produce a demonstrative + definite noun phrase like *det huset* [that house, directly translated: that $_{\text{neu}}$ house_the $_{\text{neu}}$] where both the determiner in front of the noun and the noun itself are definite.

Both DSIH and GSIH would predict a gender congruency effect for Norwegian neuter target pictures when paired with a gender incongruent word. The DSIH would not predict a gender interference effect between feminine and masculine picture-word pairs because the demonstrative is *den* for both genders. The GSIH would predict a gender interference effect for these picture-word pairs. However, at this point the experiment had a more exploring approach as to what kind of gender congruency/interference effect we would anticipate.

Based on the experimental set-up described in the next chapter, the hypothesis would predict a delay in response latencies for the gender incongruent trials compared to the gender congruent trials, i.e., a gender/determiner congruency effect. The null hypothesis would be that there is no significant difference in reaction times between the gender congruent condition and the gender incongruent condition, in other words; there is no effect of gender congruency. As for the semantic control condition, I expect a delay in response time for the semantically related trials compared to the semantically unrelated trials, i.e., the semantic interference effect.

CHAPTER 5, METHOD

In two experiments it was attempted to obtain a gender congruency effect in Norwegian Bokmål.

Native Norwegian speakers with Bokmål as their first written language were told to name a set of pictures. Each picture was shown with either a congruent or incongruent distractor word.

Experiment 1 (Dem + N + Def. suffix naming task) (30 subjects)
The respondents were told to name the picture using the appropriate demonstrative, producing gender-marked Demonstrative + Noun + gender-marked definite suffix singular NPs e.g., *det huset* [that house, directly translated: that_{neu} $\text{house\_the}_{\text{neu}}$] or *den armen* [that arm, directly translated: $\text{that}_{\text{masc}}$ $\text{arm\_the}_{\text{masc}}$] in Experiment 1. The data set for Experiment 1 is found in Appendix A.

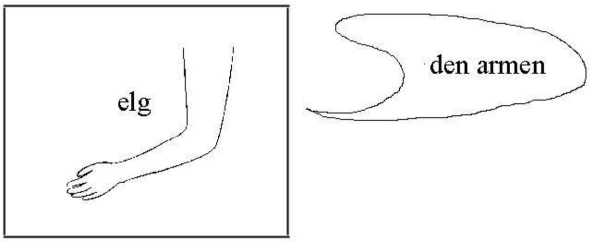


Figure 8. An example of a trial in Experiment 1 with the gender congruent distractor word elg *[moose] presented in its base form singular. The target utterance is "den armen" [that arm], as shown in the "cartoon bubble".*

Experiment 3 [5] (Dem+N+Def. suffix naming task with gender-marked distractor word) (20 subjects)
In this experiment, there were no feminine words in the data set, only masculine and neuter words. The naming task was the same as in Experiment 1. The difference from the first experiment was that the distractor word appeared in definite form singular (*bil**en*** [the_{masc} car]),

---

[5] Experiment 2 is not reported in this paper, as noted in the preface, because it did not produce any informative results.

and not indefinite form singular (*bil* [car]) as in the previous experiments. The motivation for this was to get maximum gender interference from the distractor word. The data set used in Experiment 3 can be found in Appendix B.

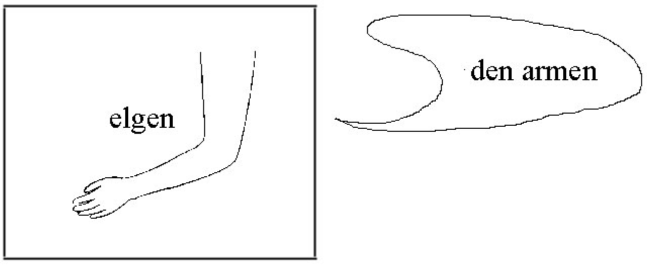


Figure 9. An example of a trial in Experiment 3 with the gender congruent distractor word elgen *[the moose] presented in its definite form singular. The target utterance is "den armen" [that arm], as shown in the "cartoon bubble".*

Selection criteria for participants

30 Norwegians participated in Experiment 1 and 20 in Experiment 3. One criterion for participating in the experiment was that Norwegian was their mother tongue. The age selection criterion was based on findings from research on the Stroop task, which has shown that there is greater interference for young children and adults over 60 (e.g., Comalli et.al, 1962 via MacLeod, 1991; Panek, Rush, & Slade, 1984 via MacLeod, 1991). Therefore the participants' age varied between 17 and 62 (where 3 respondents were older than 60). Another selection criterion was written norm. Since a noun may have different gender in Bokmål and Nynorsk, only participants with Bokmål as their written norm could participate. However, one subject used both Bokmål and Nynorsk as written norm just as often and got to participate in Experiment 1, but was afterward excluded because the dialect was too close to Nynorsk and therefore differed on gender in some of the target pictures.

5.1 Experiment 1 (Dem + N + Def suffix naming task) (30 subjects)

The respondents were told to name the picture using the appropriate demonstrative, producing gender-marked Demonstrative + Noun + gender-marked definite suffix singular NPs e.g., *det* *huset* [that house, that$_{neu}$ house_the$_{neu}$] or *den tomaten* [that tomato, that$_{masc}$ tomato_the$_{masc}$].

Participants

Experiment 1 had 30 participants who satisfied the selection criteria mentioned above. All were offered payment for participating. None of them participated in the other experiments.

Material

The material consisted of 60 mono-morphemic target pictures for naming, 20 feminine nouns, 20 masculine nouns and 20 neuter nouns in Experiment 1. Each target picture was paired with a gender incongruent distractor word and a gender congruent distractor word. In the incongruent condition there were equally many distractor words from the two incongruent genders (e.g. half of the neuter target pictures were paired with masculine distractor words and the other half with feminine distractor words, etc.).

Target picture names and their distractor words were chosen so that they did not start with the same phoneme or grapheme to avoid a phonological/orthographic facilitation effect, and the length in syllables were similar for the target words and distractor words. The mean frequencies for the different targets (the feminine, the masculine and the neuter picture names) are similar per 18.3 million words and were tested against the Oslo Corpus of tagged Norwegian texts, the Bokmål part. The frequency is based on all inflectional forms of the noun. For words with alternative spellings (e.g., ferge/ferje [ferry]) their mean frequency was taken. The congruent and incongruent distractor words to an individual target picture were also matched for frequency. In some cases, however, it was impossible to make a semantic category without violating this criterion. Since the corpus is a collection of written text (which varies quite a lot from spoken language), I did not follow the frequency numbers from the corpus strictly, but used my own knowledge of the spoken language in cases where the written word frequency seem to deviate from the spoken word frequency. For example, an everyday object which is often spoken of but little written of, is a tomato.

Semantically related conditions were used as a control measure because a semantic relation between target picture and distractor word causes an interference effect and delays the response time (as we have seen in chapter 3). This control measure was included in the test set to be sure that the distractor words were read by the participants. The incongruent distractor words in the semantically related condition were taken from the two other gender's target words in the semantically related condition. For example, half of the neuter semantically related target words were used as incongruent distractor words for the feminine and the other half for the masculine semantically related target word condition. In the semantically unrelated condition, the target words from the semantically related condition were used as distractor words. The reason for re-using the pictures and words from the semantically related conditions in the semantically unrelated conditions, is to make sure that any difference between the two conditions is not caused by the difference between the distractor words themselves, but rather between the picture-distractor word relation. The distractor words were shuffled around so that there was no semantic or phonological relationship between the target words and their distractor words. In addition there were ten practice pictures paired with distractor words according to the same criteria as mentioned above.

The pictures were simple black line drawings presented on a white background. Most of the pictures were taken from the picture database of the Max Planck Institute for Psycholinguistics in Nijmegen. The rest were downloaded from the Internet or made by me in Photoshop 6.0.

Distractor words were presented in their singular form in black characters (font type and size: Times New Roman, 30 points) across or around the picture object. Pictures appeared in the center of the screen with the distractor words appearing at slightly different positions around the fixation point to prevent the participants from ignoring the distractor word. For an individual picture the distractor words appeared in the same position. The complete sets of words are to be found in Appendix A and B.

Problems with the material in Experiment 1

Since there are no pure feminine nouns in Bokmål, nouns that could be either feminine or masculine were regarded as feminine. This is unfortunately not consistent throughout the data set. During the first look ups in a dictionary, I used the paper version of Norsk Ordbok (Kunnskapsforlaget, 1998). After some look-ups, I changed to the electronic version of Bokmålsordboka (developed by the University of Oslo). However, these two dictionaries differ when it comes to which words are strictly masculine. This unfortunate use of two dictionaries was the reason that the word *bombe* [bomb] has been marked as masculine in the test set, though it should be marked feminine according to the selection criteria I have used on the rest of the test set. The same goes for the word *ape* [monkey].

Another error has slipped through, the word *ball* [ball] which has two meanings in Norwegian. *Ball* (as in football) is masculine and the other sense (dance party) is neuter. The word was not excluded from the analysis because of the so-called frequency effect. The frequency effect is that high frequent words are named faster than low frequent words (e.g., Oldfield & Wingfield, 1965 via Miozzo and Caramazza, 2003). However, since I did not find literature on my case exactly, which consisted of ambiguous words with difference in gender and frequency, I made a survey where the results are found in Appendix C. The results from the survey showed that all of the respondents retrieved the intended sense (the football sense) immediately.

The first 9 subjects had three errors in the trial files. One incongruent distractor word appeared two times instead of one time (the unrelated, incongruent distractor word *sekk* [sack] was shown instead of *katt* [cat]). Two other target words were paired with the wrong word file in the semantically unrelated condition (*øye* and *nøtt* instead of *øye* and *gaupe* which was the intention, and *glass* and *dyne* instead of *glass* and *nøtt*). These trials are coded with error type 2 and excluded from the analyses for the first nine subjects.

Other problems with Experiment 1

Another problem with the experimental set up was that there were both subjects with a two gender system and subjects with a three gender system. Subjects with a two gender system do not use feminine gender. The original idea was to exclude all the feminine trials. That way, both two gender participants and three gender participants could take part in the experiment. However, the test set's architecture does not reflect this intention at all. The feminine words are woven into the test set instead of being an independent module, functioning as “fillers” only, that could be removed without affecting the relevant trials.

Procedure

Participants were tested individually in a quiet room. They sat in front of a computer screen at a viewing distance of approximately 65 cm. The computer screen was a 15.4 inch WSXGA+ monitor, on an Inspiron 8500 Dell laptop. Each trial started with a fixation point which appeared for 300 ms. 200 ms (at 500 ms) after it had disappeared, it was followed by a picture and a distractor word that was present for a 1000 ms. Participants were instructed to name the picture as ACCURATELY and as fast as possible with the appropriate demonstrative. At picture and word onset a voice key connected to a microphone was activated to measure the naming latencies. The trial timed out after 2000 ms. If a response was not given within the 2000 ms, the trial was counted as invalid. The next trial started immediately after the previous trial was finished. The trial sequence presentation and the reaction time measurements were controlled by NESU (Nijmegen Experiment Set-Up).

Design

The experiment was divided into three parts. First, participants saw each picture on the computer screen to become familiar with the pictures and learn their designated picture names. A fixation point appeared in the middle of the screen for 300 ms followed by a picture that appeared on the computer screen for two seconds as a simple black line-drawing on white background. Then its designated name appeared under the picture (in 48 point Times New Roman) and they were both present on the screen for another three seconds. Participants were asked to use the designated names for the pictures. This familiarization period was done to make sure that the participants did not prefer alternative names to the pictures. After the familiarization phase, participants went through a practice phase. Each picture was presented in the center of the screen, preceded by a fixation point. Participants were asked to name the picture using the appropriate demonstrative and the designated picture name, e.g., *den stolen* [that chair]. This was done to make sure that the participant knew the correct demonstrative for each picture name. The designated picture name appeared under the picture two seconds after the picture was presented on the screen. Participants saw the target picture and its designated name twice before the real experiment started. The naming phase started immediately after the practice phase. The stimuli was presented in four blocks of seventy trials each. The stimulus-onset asynchrony (SOA) was 0 ms because this is where Schriefers (1993) obtained the biggest effect with the SOAs he tested, and this has also been replicated by Schiller & Caramazza (2003). The first ten trials in each block were taken from the set of practice pictures and functioned as warm-up trials. These trials were not included in the analyses. In each block, the targets and distractor words of each of the three genders were presented approximately equally many times. Blocks were randomized so that: a) before the same object was presented again, at least four other objects appeared in between, and b) targets could have the same gender on no more than two consecutive trials. The order of the blocks was varied across participants. The experiment lasted approximately 50 minutes.

Results and discussion

Naming latencies shorter than 350 ms and longer than 1500 ms were regarded as outliers and were excluded from the analyses (in Exp.1: 7.24% of the valid replies and in Exp.3: 6.65% of the valid replies). A trial was excluded from the analyses if the respondent used a wrong target word (error type 5), repaired an utterance (error type 5 or 6 depending on the repair), produced an incorrect demonstrative (error type 6), hesitated between demonstrative and target word (error type 7), only produced a demonstrative (error type 8), other sounds triggered the microphone (error type 9) or did not reply before the deadline of two seconds had past. One participant was not included in the analyses because the person used another gender than the intended one on several of the target words due to a Nynorsk influenced dialect. One target word, *kanon* [canon] was excluded from the analyses because it was shown with two distractor words that were marked with incorrect gender in the test set (bombe [*bomb*] and *ape* [monkey]).

Two ANOVAs were conducted on the mean reaction times (RTs), with semantic relatedness (related or unrelated) and gender condition (congruent or incongruent) as independent variables. Separate analyses were carried out with participants (F_1) and items (F_2) as random variables. In the subject/participant analysis (F_1) target gender (feminine, masculine and neuter) is also an independent variable, but it is a between-subjects factor in the items analysis (F_2). In the participant analysis, the mean RT for each condition was calculated for each participant and treated as a single observation. In the items analysis, the mean RT for each target across participants was treated as a single score, for each of the conditions.

According to Clark (1973), the reason for doing one analysis by subjects (participants as random variable) and one by items (target words as random variable) is as follows: If the F_1 analysis shows that the difference between the gender congruent condition and the gender incongruent condition is significant, it roughly tells us that the effect of gender congruency will be replicated if we give the same dataset to a new, equally large, sample of participants. However, we have not proven yet that the difference between the gender congruent and the gender incongruent condition can be generalized beyond the specific sample of words in the dataset. If we do a F_2 analysis and find a significant difference between the gender congruent condition and the gender incongruent condition, then this result from the F_2 analysis will roughly tell us that we would find a gender congruency effect with a new, similar dataset if we run the experiment on the same sample of participants. If the F_2 analysis is not significant, that implies that the gender congruency effect cannot necessarily be replicated on a new sample of words.

Rejecting the null hypothesis if both the F_1 analysis and the F_2 analysis are significant has been common among many researchers after Clark's (1973) influential article. However, Raaijmakers, Schrijnemakers and Gremmen (1999) claim that this is an almost twenty year old misconception of Clark's article, and in addition, they argue against the need of performing a separate item analysis "(...) since the traditional F_1 is the correct test statistic. In particular this is the case when item variability is experimentally controlled by matching or by counterbalancing." (Raaijmakers et al. 1999:416). However, since my material is matched, for example, for frequency, but not matched, for example, for picture recognition times (which in turn might affect the naming latencies) and since there are other words that could have been used instead, I will treat item as a random variable (following Clark, 1973). Therefore, I performed a F_2 analysis in addition to the F_1 analysis, and report both results.

Two ANOVAs were also conducted on the errors in the same manner as in the RT analyses, except that the means of each condition were based on the number of errors, not the mean RT. In the error analyses only types of errors that might contain relevant information were included. The alpha level was set at .05. In the remaining experiments, the same statistical tests as reported here are performed.

The RT analysis by subjects showed a 7.97 ms advantage for the gender congruent condition over the gender incongruent condition in Experiment 1. This difference reached significance ($F_1(1,29) = 4.77$, $MS_e = 1200.26$, $p < .05$), and approached significance in the items analysis ($F_2(1,56) = 3.14$, $MS_e = 1764.13$, $p = .08$) revealing that naming latencies were faster for trials when distractor and target word had the same gender than when they had different gender. The error analyses showed that more errors were made when the distractor word's gender was incongruent to that of the target word, reflecting interference from the distractor word's gender information. This difference reached significance in both error analyses ($F_1(1,29) = 13.67$, $MS_e = 1.3$, $p < .05$; $F_2 (1,56) = 12.07$, $MS_e = 2.20$, $p < .05$). The mean RTs are displayed in Table 4 broken down by semantic condition and target gender.

An unexpected finding was that the effect of target gender reached significance with the neuter targets (592.34 ms) being named about 20 ms slower than the masculine (570.92 ms) and feminine targets (572.07 ms) in the RT analysis by subjects ($F_1(2,28) = 7.74$, $MS_e = 1860.81$, $p < .05$). This was not supported by the items analysis ($F_2(2,56) = 1.17$, $MS_e = 12527.87$, $p = .32$). In the error analyses, the effect of target gender reached significance in the analysis by items ($F_2(2,56) = 3.88$, $MS_e = 10.29$, $p <.05$) and also in the analysis by subjects ($F_1(2,28) = 17.10$, $MS_e = 1.24$, $p < .01$) where the feminine targets had a mean error of 0.7, the masculine targets had a mean error of 0.84 and the neuter targets had a mean error of 1.6.

The semantically related trials were replied to 22.36 ms slower than the semantically unrelated trials in the RT analysis by subjects. This time difference was significant with $F_1(1,29) = 26.06$, $MS_e = 1726.74$, $p < .01$. The effect of semantic interference was supported by the results from the items analysis ($F_2(1,56) = 21.51$, $MS_e = 1733.86$, $p < .01$) and by both error analyses. In the error analysis by subjects, the effect of semantic interference reached significance with $F_1(1,29) = 12.71$, $MS_e = 1.78$, $p < 05$, where the semantic related condition had a mean error of 1.3 while the the semantic unrelated condition had a mean error of 0.8. The effect of semantic interference was also obtained in the error analysis by items ($F_2(1,56) = 18.84$, $MS_e = 1.80$, $p < .01$)

An interaction effect of semantic relatedness and target gender was found in the RT analysis by subjects ($F_1(2,28) = 6.9$, $MS_e = 1011.07$, $p < .05$.). The feminine targets were replied to 21.05 ms faster in the semantically unrelated conditions (561.54 ms) than in the semantically related conditions (582.59 ms), the masculine targets had a 36.2 ms advantage in the semantically unrelated conditions (552.82 ms) over the related ones (589.02 ms), whereas the neuter targets had only a 9.82 ms advantage in the semantically unrelated conditions (587.43 ms) over the semantically related conditions (597.25 ms). However, this interaction effect was not supported by the results from the analysis by items ($F_2(2,56) = 1.27$, $p = .29$)

In the error analyses, an interaction effect between semantic relatedness and gender condition was found ($F_1(1,29) = 8.38$, $MS_e = .9$, $p < .05$; $F_2 (1,56) = 7.68$, $MS_e = 83.18$, $p < .05$). The semantically related gender incongruent conditions yielded more errors than the other conditions reflecting interference through both intended channels. No other main effects or interaction effects were significant.

Table 4. Mean naming latencies and percentage errors[6] (in parenthesis) in Exp.1 (Dem+N+def.suff.naming) from the RT analysis by subjects.

Condition	Target gender			Mean
	feminine	masculine	neuter	
Congruent	561(1.5)	552(3.7)	581(5.7)	565(3.6)
Incongruent	562(2.8)	554(3.3)	594(7)	570(4.4)
Semantically related (congr)	573(3.7)	584(4)	597(6.5)	585(4.7)
Semantically related (incongr)	592(6.2)	594(6.2)	597(12.8)	594(8.4)

The semantic interference effect obtained in all the analyses shows that the distractor words were processed by the subjects**.** The analyses also show that there is a trend towards a gender interference effect (though not statistically significant in the RT analysis by items) with delayed RTs when the distractor have different gender than the target, compared to when they have the same gender. This indicates that conflicting gender information do interfere when naming the target picture. That gender incongruent distractors interfere is supported by the error analyses, where the gender incongruent trials yielded significantly more errors than the gender congruent trials.

The reason for the slow response time for the neuter targets is unknown. It could be that the pictures of the neuter target words took longer time to recognize than the masculine or feminine target pictures. Another explanation could be frequency of usage. Neuter words make up only about 25% of Norwegian nouns. However, the material is matched (though not very strictly) for frequency, so that the feminine, masculine and neuter targets had similar mean frequency. A third explanation could be that 2/3 of the material should be responded to with the demonstrative *den* and only 1/3 with *det*. This could have led the respondents to make the demonstrative *den* the default demonstrative after a while, leading to greater interference and longer response times in the cases where the demonstrative should be *det*.

---

[6] The error percentage includes only those errors with linguistic information used in the error analysis.

The 7.97 ms advantage for the gender congruent conditions over the gender incongruent conditions reached significance in the RT analysis by subjects and in both error analyses and it approached significance in the RT analysis by items. However, both the RT analysis by subjects and the RT analysis by items should reach significance in order to reject the null hypothesis.

With the test set used in Experiment 1, the effect of gender congruency obtained in three of the analyses was rather unexpected. Recall that feminine pictures shown with masculine words (and vice versa) were considered as incongruent trials. Almost half of the respondents had a two gender system and therefore no distinction between feminine and masculine nouns. For these subjects, we would not expect an effect of gender interference in the trials where feminine words are shown with masculine pictures (or vice versa). In addition, if the interference effect is caused by competition between determiners and not between gender features, then we would not expect a gender congruency effect in these trials for subjects with a three gender dialect either (since both feminine and masculine nouns take the demonstrative *den*). To get a clearer picture of the results obtained, the respondents with a three gender system and those with a two gender system were divided into two groups and their results reanalyzed.

5.2 Reanalysis of the results from Experiment 1

With the test set used in Experiment 1, we would expect a visible effect of gender interference only in the trials with neuter targets for all the respondents in Experiment 1, regardless of whether the competition occurs between determiners or between gender features. However, if the effect is caused by competition between gender features, we would expect an overall effect of gender congruency (i.e. no interaction effect between target gender and gender condition), but only for the three gender subjects.

The respondents that had a three gender / two and a half-system and those which had a strict two gender system were divided into two groups and their results were reanalyzed. The two-gender group consisted of subjects with Bergen city dialect and the rest were considered to belong to the three / two and a half gender-group (from now on referred to as three-gender group for simplicity reasons). The two-gender group consisted of 14 subjects and the three-gender group, 16 subjects.

5.2.1 Analysis of the three-gender group

The RT analysis by subjects showed a 14.41 ms advantage for the gender congruent conditions over the gender incongruent conditions for the three-gender group. The effect of gender condition reached significance in the RT analyses ($F_1(1,15) = 15.32$, $MS_e = 650.5$, $p < .05$; $F_2(1,56) = 6.65$, $MS_e = 2175.63$, $p < .05$) and in the error analyses ($F_1(1,15) = 6.64$, $MS_e = 1.02$, $p < .05$; $F_2(1,56) = 5.92$, $MS_e = .93$, $p < .05$) with more errors being made if the distractor word and target picture had different gender than if they had the same gender. The mean RTs for the three-gender group are displayed in Table 5 broken down by semantic condition and target gender.

Table 5. Mean naming latencies for the three/two and a half gender group (16 subjects) and percentage errors (in parenthesis) in Exp.1 (Dem+N+definite suffix naming).

Condition	Target gender			Mean
	feminine	masculine	neuter	
Congruent	576(2.2)	560(4.1)	590(5.9)	575(4.1)
Incongruent	581(2.5)	573(4.1)	615(6.3)	590(4.3)
Semantically related (congr)	599(4.4)	604(3.1)	605(6.6)	603(4.7)
Semantically related (incongr)	616(5.9)	612(7.5)	622(11.3)	617(8.2)

The effect of semantic relatedness reached significance in the RT analyses ($F_1(1,15) = 4.44$, $MS_e = 2446.3$, $p < .05$; $F_2(1,56) = 22.37$, $MS_e = 2356.17$, $p < .01$) with the semantically related trials taking longer to name than the semantically unrelated trials. The effect of semantic relatedness was also obtained in the error analyses ($F_1(1,15) = 10.76$, $MS_e = .85$, p

$< .05$; $F_2(1,56) = 6.97$, $MS_e = 1.05$, $p < .05$) where more errors were made in the semantically related conditions than in the semantically unrelated conditions.

The RT analysis by subjects also showed an interaction effect between semantic condition and target gender ($F_1(2,14) = 4.16$, $MS_e = 1078,37$, $p < .05$). The feminine targets were replied to 29.03 ms faster in the semantically unrelated conditions (578.44 ms) than in the semantically related conditions (607.47 ms), the masculine targets had a 41.51 ms advantage in the semantically unrelated conditions (566.70 ms) over the related ones (608.21 ms), whereas the neuter targets had only a 10.83 ms advantage in the semantically unrelated conditions (602.52 ms) compared to the semantically related conditions (613.35 ms). This interaction effect was not significant in the analysis by items however ($F_2(2,56) = 1.34$, $MS_e = 1713.04$, $p = 27$).

The error analyses revealed an interaction effect between semantic conditions and gender conditions, with more errors being made in the semantically related gender incongruent conditions than the other conditions, reflecting interference from the semantically related distractor word and also interference from conflicting gender information from the distractor word. The difference in errors reached significance with $F_1(1,15) = 6.86$, $MS_e = .78$, $p < .05$ and $F_2(1,56) = 7.17$, $MS_e = .61$, $p < .05$. No other main effects or interaction effects were significant.

The results from the analyses are clear: Distractor words with different gender than the target word interfere more than disctractor words with the same gender as the target. Furthermore, the semantic condition showed that semantically related distractor words interfere more than semantically unrelated distractors words and they cause more errors than semantically unrelated distractor words.

The time difference between the gender congruent condition and the gender incongruent condition reached significance in both RT analyses, indicating an effect of gender congruency. Since the results from both RT analyses are also supported by the error analyses, the null hypothesis can safely be rejected.

5.2.2 Analysis of the two-gender group

The two-gender group (14 subjects) showed a rather different result than the three-gender group. Gender condition was not significant ($F_1<1$; $F_2<1$). The gender congruent trials had a mean RT of 557.93 ms and the incongruent trials had a mean RT of 558.55 ms in the RT analysis by subjects yielding only 0.62 ms advantage for the gender congruent trials. However, recall that the feminine words are not incongruent with the masculine words for these subjects even though they are coded as incongruent in the analysis. Therefore the results should not be taken into account since they do not reflect what they are supposed to reflect if the gender interference effect is caused by competition between gender features. I will still report the main findings from the analyses though because they have some information value if contrasted with the results from the three-gender group. The mean RTs for the two-gender group broken down by semantic relatedness, gender conditions and target gender, are shown in Table 6.

Table 6. Mean naming latencies for the two-gender group [7]*(14subjects) and percentage errors (in parenthesis) in Exp.1 (Dem+N+definite suffix naming).*

Condition	Target gender			Mean
	feminine	masculine	neuter	
Congruent	543(0.7)	542(3.2)	570(5.4)	552(3.1)
Incongruent	541(3.2)	532(2.5)	570(7.9)	548(4.5)
Semantically related (congr)	543(2.9)	561(5)	589(6.4)	564(4.8)
Semantically related (incongr)	565(6.4)	574(4.6)	569(14.6)	569(8.5)

Though there was no sign of a gender congruency effect in the RT analyses, the effect of gender condition did reach significance in the error analyses with more errors being made in

---

[7] The grayed out numbers show conditions with "doubtful" values.

the gender incongruent conditions than the gender congruent conditions ($F_1(1,13) = 6.82$, $MS_e = 1.69$, $p < .05$; $F_2(1,56) = 10.88$, $MS_e = .72$, $p < .05$).

The effect of target gender reached significance for the two-gender group with the neuter targets taking longer to name ($F_1(2,12) = 5.53$, $MS_e = 2077.17$, $p < .05$). However, this was not supported by the items analysis ($F_2(2,56) = 1.36$, $MS_e = 16340.98$, $p = .27$).

Semantic conditions were as in the other analyses significant in the RT analyses ($F_1(1,13) = 14.15$, $MS_e = 847.80$, $p < .05$; $F_2(1,56) = 9.53$, $MS_e = 2251.25$, $p < .05$) with the semantically related trials having longer naming latencies than the semantically unrelated trials. This effect was also obtained in the error analyses ($F_1(1,13) = 4.68$, $MS_e = 2.93$, $p = .05$; $F_2(1,56) = 11.75$, $MS_e = .82$, $p < .05$) where the semantically related conditions caused more errors than the semantically unrelated conditions.

However, these results have limited information value when it comes to revealing an effect of gender congruency for the two-gender group. Further analyses need to be done in order to reject or keep the null hypothesis. A t-test of the neuter targets could give some answers, because the incongruent trials for the neuter targets is always incongruent whether the competition is between determiners or gender features.

t-test, neuter targets

Two-tailed, paired sample t-tests were performed with pairwise comparison between semantic related gender congruent (sr_gc_9) and semantically related gender incongruent (sr_gi_9), between semantically unrelated gender congruent (su_gc_9) and semantically unrelated gender incongruent conditions (su_gi_9). The semantic conditions were merged together and a pairwise comparison between the means of the gender congruent conditions (gc_9) and gender incongruent conditions (gi_9) was also performed. These tests were done to see if the neuter targets for the two-gender group showed a move towards a gender congruency effect. The descriptive statistics for the different pairs for the two-gender group is shown in Table 7.

Table 7.The descriptive statistics from SPSS output file for the two-gender group. Neuter targets only.

Paired Samples Statistics

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	*sr_gc_9*	*588,5499*	*14*	*64,86459*	*17,33579*
	sr_gi_9	*569,1690*	*14*	*64,61141*	*17,26813*
Pair 2	*su_gc_9*	*569,9630*	*14*	*54,97921*	*14,69381*
	su_gi_9	*570,3895*	*14*	*76,17026*	*20,35736*
Pair 3	*gc_9*	*579,2565*	*14*	*56,67861*	*15,14800*
	gi_9	*569,7792*	*14*	*60,78681*	*16,24596*

As can be seen in Table 7, the two-gender group actually used 19.4 ms more to name pictures in the semantically related gender congruent trials than in the semantically related gender incongruent trials, contrary to what we expected. The results of the analysis (output from SPSS) are shown in Table 8. The comparison of the means between the semantically related gender congruent and incongruent trials in Pair 1 did not show a significant difference between these two conditions (t (13) = 1.23, p = .24). Nor was the 0.4 ms advantage for the gender congruent condition in Pair 2 statistically significant (t (13) = -.03, p = .98). The pairwise comparison between the means of all the gender congruent and gender incongruent trials in Pair 3, where the gender incongruent trials actually had a 9.48 ms advantage over the gender congruent trials, did not show a significant difference either (t (13) = 1.06, p = .31).

Table 8. The results of the analysis of neuter targets only for the two-gender group in the different gender conditions, pairwise comparisons.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sr_gc_9 - sr_gi_9	19,38090	58,79563	15,71379	-14,56668	53,32849	1,233	13	,239
Pair 2	su_gc_9 - su_gi_9	-,42646	57,07770	15,25466	-33,38215	32,52922	-,028	13	,978
Pair 3	gc_9 - gi_9	9,47722	33,62579	8,98687	-9,93773	28,89217	1,055	13	,311

The three-gender group had a somewhat different result when looking at the neuter targets only. The descriptive statistics are shown in Table 9.

Table 9. The descriptive statistics from SPSS output file for the three-gender group. Neuter targets only.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sr_gc_9	604,5598	16	75,98724	18,99681
	sr_gi_9	622,1349	16	83,31650	20,82912
Pair 2	su_gc_9	589,7827	16	76,36597	19,09149
	su_gi_9	615,2557	16	75,66306	18,91576
Pair 3	gc_9	597,1713	16	72,39460	18,09865
	gi_9	618,6953	16	75,22360	18,80590

Unlike the two-gender group, the three-gender group showed an advantage for the gender congruent conditions over the gender incongruent conditions in all of the three pairs. In Pair 1 the semantically related gender congruent trials were named 17.58 ms faster than the semantically related gender incongruent trials. However, this difference was not significant (*t* (15) = -1.36, *p* = .20). The 25.48 ms time advantage for the semantically unrelated gender congruent condition compared to the semantically unrelated gender incongruent condition found in Pair 2, reached significance with *t* (15) = -3.23, *p* < .05. The 21.52 ms time advantage for all the gender congruent trials over the gender incongruent trial (Pair 3) was also statistically significant with *t* (15) = -2.69, *p* < .05.

While the three-gender group show the expected results, i.e. an increase in naming latencies when the distractor word have different gender than the target word compared to when they have the same gender, the two-gender group show a different result. None of the RT analyses indicate an effect of gender congruency, but the error analyses do. The control condition, the presence of a semantic interference effect, reached significance, reflecting that the distractor words were read by the participants with a two gender system also.

However, rejecting the null hypothesis based on the error analyses alone seems careless. After all, the results from the t-test of the neuter targets alone, do not support a presence of a gender congruency effect for the two-gender group.

5.2.3 Discussion, reanalysis of Experiment 1

As mentioned before, it is important not to put too much emphasis on the results from the two-gender group in Experiment 1 because of the unfortunate test set architecture mentioned earlier in the text. That said, the difference between the two-gender group and the three-gender group is interesting because of some properties of the "unfortunate" test set architecture which allow us to test between the DSIH and GSIH without involving an externally set factor like singular and plural. An overview of how the different conditions in the test set really are for the two different groups of respondents are shown in Table 10, where the trials that were supposed to be incongruent but might actually be congruent, are marked with red letters.

Table 10. A representation of the different distractor conditions for the two different gender system and their relation with target gender. The table is made by Niels Schiller and Herbert Schriefers.

Target	*Incongruent distractors*	*Gender system*			
		2 genders		*3 genders*	
		Target-distractor relation for abstract gender (neu, fem, masc)	*Target-distractor relation for concrete gender (det, den)*	*Target-distractor relation for abstract gender (neu, fem, masc)*	*Target-distractor relation for concrete gender (det, den)*
Neuter (det)					
	Fem (den)	*Incongruent*	*Incongruent*	*Incongruent*	*Incongruent*
	Masc (den)	*Incongruent*	*Incongruent*	*Incongruent*	*Incongruent*
Feminine (den)					
	Neu (det)	*Incongruent*	*Incongruent*	*Incongruent*	*Incongruent*
	Masc (den)	*Congruent*	*Congruent*	*Incongruent*	*Congruent*
Masculine (den)					
	Neu (det)	*Incongruent*	*Incongruent*	*Incongruent*	*Incongruent*
	Fem (den)	*Congruent*	*Congruent*	*Incongruent*	*Congruent*

From the table we can see that the possibility for maximum interference between distractor word and target word is found for the three-gender group at the level of abstract gender, i.e. gender features like masculine, feminine or neuter. According to the DSIH (Schiller and Caramazza, 2003) the gender interference effect should be located at the level where the different demonstratives, concrete gender, compete for selection. If that is the case, then the gender congruency/interference effect found for the three-gender group should also be found for the two-gender group since the demonstratives are the same, then following Schiller and Caramazza (2003) and assume that the gender interference effect only appears for independent word forms and not inflectional forms like the definite form suffixes. Of course, it is a possibility that interference occur for gender-marked suffixes, but this might not be possible to measure with this experimental paradigm because the respondents could start to produce the word before the gender-marked suffix is selected, thereby covering the interference which is measured in reaction time (Schiller & Caramazza, 2003). Now, the gender congruency/interference effect does not show for the two-gender group. If the interference is located at the demonstrative level, the test set architecture would not allow us to distinguish between the two different groups of respondents. The demonstrative *den* is the same for masculine and feminine, therefore the gender interference effect should show for both groups or for none. What does differ is the abstract gender node, that is the features feminine, masculine or neuter. Though realized through the same demonstrative, *den*, the feminine and masculine gender do differ in other cases, such as in indefinite articles (or else there would just be a *common* gender).

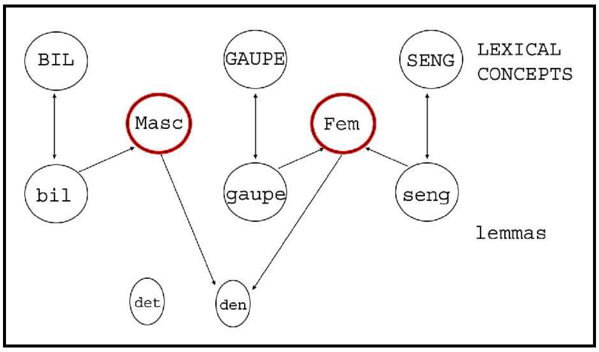


Figure 10 a. Competition represented with red circles at abstract gender node level, i.e. between the gender nodes feminine and masculine.

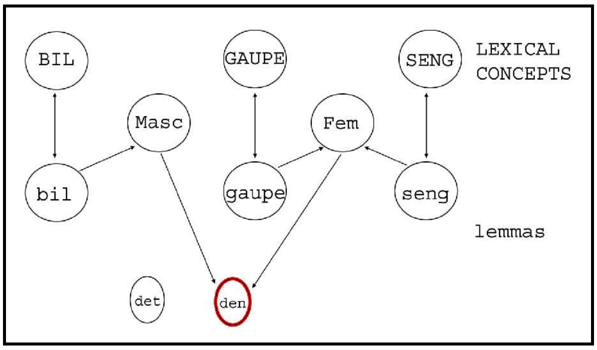


Figure 10 b. Competition (or rather lack of) represented with red circle(s) at concrete gender node level, i.e. between the demonstrative nodes.

There are several arguments for the interference being located at the abstract gender level. In terms of a model where abstract gender is represented by a node, we can think of the abstract gender as a gender node, where we have a masculine gender node, a feminine gender node and a neuter gender node. The interference effect would occur because of competition between the gender nodes when selecting the noun's gender feature. Figure 10 a shows the gender interference effect located at gender node level between the feminine gender node and the masculine gender node. The red circles represent the nodes that are in competition with each other. The first argument for the gender congruency effect occurring at level of gender node selection rather than the level of determiner node selection, is the mere presence of a gender congruency effect for the three-gender group with the test set that is used. In the test set a feminine target picture can be shown with a masculine distractor word and this combination would be a gender incongruent condition, in accordance with the results obtained. If the locus for interference was at the level where concrete gender nodes, that is the demonstrative nodes, are selected, then a feminine picture with masculine distractor words would be a gender congruent condition, as we can see from Figure 10 b, and should therefore not lead to the observed delayed response latencies. The second argument for a gender interference effect rather than a determiner interference effect is that the same picture-word combinations would always be gender congruent for the two-gender group. An interference effect should therefore not be present for the two-gender group whether the locus is at abstract gender node level or at demonstrative level (except for the neuter target conditions). Therefore, the two-gender group's results do not say anything about the locus of interference, but contrasted with the results from the analyses of the three-gender group, the lack of a gender congruency/interference effect for this group and the presence of a gender congruency/interference effect for the three-gender group support the hypothesis that the gender interference effect is caused by competition between gender nodes. Third, the two-gender group had much smaller naming latencies than the three-gender group as can be seen in Table 11.

Table 11. Mean naming latencies in ms from the reanalysis of Exp.1 (Dem+N+definite suffix naming).

Three gender subjects		Two gender subjects	
Target gender	Mean RT	Target gender	Mean RT
Feminine	593	Feminine[8]	548
Masculine	587	Masculine	553
Neuter	608	Neuter	575

The reason for the smaller naming latencies could be that the respondents that have a two gender system, were exposed to gender priming. If a respondent in an experiment is “over-exposed” to one gender, repeated activation of a gender node will speed-up its retrieval time (Levelt, 2001) because the link between lemma and gender node is recency sensitive. This

---

[8] Feminine is grayed out because it does not display the mean RTs for the feminine nouns since there are no feminine nouns for the two gender subjects.

effect is one of the two effects referred to as gender priming (Schriefers & Jescheniak, 1999). However, van Berkum (1997) failed to find a gender priming effect (or more exactly a gender recency effect) and "(...) evidence for a gender recency has been obtained in the context of a metalinguistic gender decision task only, but not in genuine production tasks." (Schriefers & Jescheniak, 1999:582). The faster naming latencies all in all could be due to less interference instead, if we assume that the interference is between gender nodes. The intended incongruent conditions involving feminine/masculine pictures paired with masculine/feminine words would not cause the intended interference, and thereby longer reaction times, as opposed to the three-gender group that are exposed to interference in all the intended trials. Both explanations above might explain the faster naming latencies for the two-gender group for the masculine and feminine target trials to some degree, but it does not explain why the neuter targets, and the gender congruent conditions, are replied to faster by the two-gender group than the three-gender group. However, a plausible explanation for the difference in response times can be again found at the abstract gender node level. If we assume that selecting a lemma or gender node is depended on exceeding the activation level of all the lemmas or gender nodes in the experiment with some critical difference; the more competitive nodes, the longer it will take to exceed the activation level of the other nodes with the necessary amount (Roelofs, 1992). A speaker that evokes two gender nodes will be able to select the appropriate gender feature faster than a speaker that evokes three gender nodes. This explanation works well if we assume that the competition occurs at abstract gender node level. If the competition was located at the demonstrative level, according to assumption above, we should not observe longer naming latencies for the three-gender group than the two-gender group.

The gender interference effect can also be explained by another difference in concrete gender, the suffixes, rather than competition between gender nodes (at least for the three-gender group). As mentioned before, the demonstrative is *den* for both feminine and masculine nouns. But the definite form suffixes differ. The definite form suffix for masculine nouns are -*(e)n*, and for feminine it is -*a*.

The idea that the gender congruency effect is located at suffix-level is not in accordance with Schiller and Caramazza's (2003) suggestion that the gender congruency effect only occurs for free-standing morphemes. It is not completely in accordance with what we know about incremental sentence production/planning either. We do not have to plan the entire sentence before we start uttering it. Rather, we can start speaking immediately after the first words of the sentence is retrieved, and do the rest of the planning as we speak (De Smedt, 1996). In Norwegian the proper demonstrative can be retrieved immediately after the noun's gender has been selected because sufficient information for selecting the appropriate demonstrative is available. The definite suffix, *-en* or *-a* in the feminine and masculine cases, do not have an impact on which demonstrative that should be selected. Whether the suffix is *-en* or *-a* the demonstrative is still *den* for both genders and the correct form is only effected by the noun's gender feature. Therefore, it seems unlikely that the gender congruency effect is caused by interference from incongruent suffix-information from the distractor word.

To sum it up so far, the overall results from the analyses indicate that the competition from the gender incongruent distractors occur between gender features and not determiners. Under the DSIH we would expect the three-gender group and two-gender group to produce the same results, because their demonstrative systems are identical. This is not in accordance with the results from the analyses of the three-gender group only and the two-gender group only. Under the GSIH we would predict a difference in results between the two groups of respondents, as observed, because they differ in their gender systems and thereby, with the given test set, the two-gender group will be exposed to fewer gender incongruent trials than the three-gender group. This prediction is in accordance with the results obtained. Under the DSIH we would expect to obtain an effect of gender congruency for the neuter targets only, and for both the two-gender group and the three-gender group. Under the GSIH we would expect a gender congruency effect, with the test set used in Experiment 1, for the three-gender group only, independent of target gender. This is in accordance with the results obtained in reanalysis of Experiment 1 where no interaction effect between target gender and gender condition was found for the three-gender group (both Fs < 1 in the RT analyses). However, the reason for lack of gender congruency effect for the two-gender group for the neuter targets (contrary to the predictions made by both the GSIH and DSIH), is still unknown, but might be a result of an inappropriate test set architecture for obtaining a gender congruency effect for this group of respondents.

A graphic illustration of some of the processes involved, according to the gender node selection interference account, is shown in Figure 11. The lexical concept will activate its lemma, which in turn will activate its gender node. The findings from the three-gender group, and the basis for the model here, do not allow us to distinguish between a cascading or serial discrete model. That is why the arrows from lemma to lexeme nodes are without arrowheads.

The motivation for a unidirectional arrow from lemma to gender nodes is that evidence for a link from gender node to lemma and back is, at most, very weak, if present at all (Schriefers and Jescheniak, 1999).

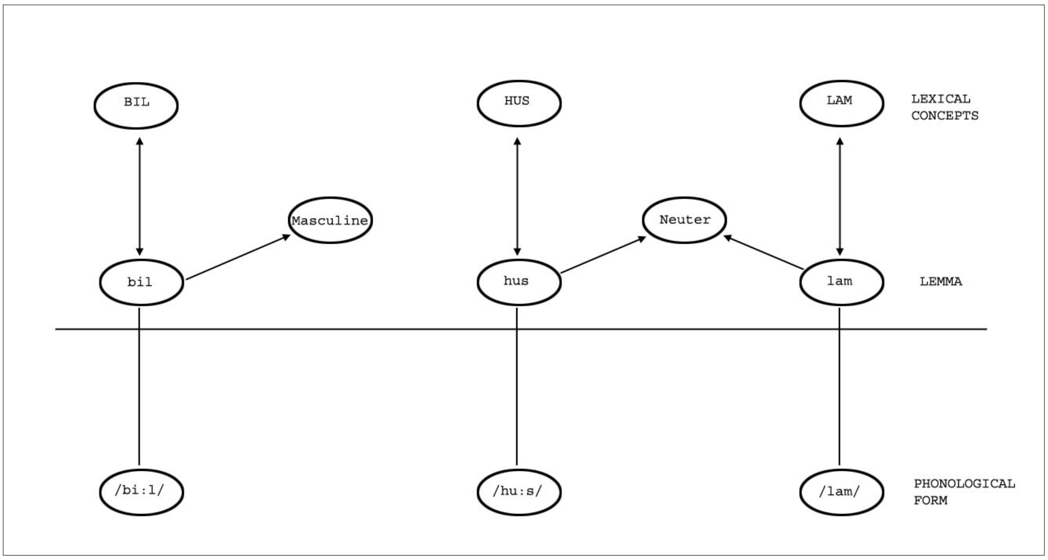


Figure 11. A simplified model, based on the findings from Experiment 1, showing fragments of the Norwegian lexicon.

5.3 Experiment 3 **(Dem + N + Def suffix naming task) (20 subjects)**

In this experiment the distractor words were presented in gender-marked definite form (e.g., *bil**en*** [the car] instead of *bil* [car] as in Experiment 1) in order to maximize the gender interference in the gender incongruent conditions.[9] Since the results from Experiment 1 indicated some sort of interference in the gender incongruent conditions, but the feminine words made it an inappropriate test set for respondents without feminine gender, the material used in Experiment 3 only contained masculine and neuter words.

Participants

20 subjects participated in the experiment. None of them had participated in the previous experiments. All were offered payment for participating.

---

[9] The motivation for doing Experiment 3 was because the first analyses of Experiment 1 with the target word *kanon* [canon] showed a trend towards a gender congruency effect, but it did not reach significance in the RT analyses. The analyses of Experiment 1 without the target *kanon* [canon] were done after Experiment 3 was conducted.

Material

The feminine words in the material used in Experiment 1 were excluded so that the material only consisted of masculine and neuter words. In addition, the target word *hare* [hare] was changed to *kanin* [rabbit] because the picture of the hare was confusingly similar to the picture of the squirrel which lead to unnecessary many mistakes in naming the target picture of the squirrel. The target word *tempel* [temple] was exchanged with *hus* [house] because the picture turned out to be tricky for the participants. Some incongruent distractor words were changed in order to match better on frequency with the congruent distractor words since some of the congruent distractors were shuffled around due to the exclusion of the feminine words. Some other improvements from the material used in Experiment 1 are that *ball* [ball] was exchanged with *løk* [onion], *ape* [monkey] with *gaffel* [fork] and *bombe* [bomb] with *pistol* [pistol]. The complete list of words can be found in Appendix B

Procedure

The procedure was the same as in Experiment 1.

Design

The design was the same as in Experiment 1 except that the stimuli were presented in four blocks of 50 trials each, 40 ordinary trials and 10 warm-up trials. The experiment lasted approximately 25 minutes.

Results and discussion

Naming latencies shorter than 350 ms and longer than 1500 ms were regarded as outliers and were excluded from the analyses, i.e. 6.65% of the valid replies in Experiment 3. A trial was excluded from the analyses if the respondent used a wrong target word, repaired an utterance, produced an incorrect demonstrative, hesitated between demonstrative and target word, other sounds triggered the microphone or did not reply before the deadline of two seconds had past. ANOVAs, described in Experiment 1, were conducted.

The RT analysis by subjects showed that the gender congruent conditions (571.79) were named 7.52 ms faster than the gender incongruent conditions (579.31). This difference did not reach significance ($F_1(1,19) = 1.59$, $MS_e = 1420.44$, $p = .22$). However, the effect of gender congruency did reach significance in the RT analysis by items with $F_2(1,38) = 6.38$, $MS_e = 1231.72$, $p < .05$. and in the error analyses ($F_1(1,19) = 32.88$, $MS_e = 1.041$, $p < .01$; $F_2(1,38) =$

14.28, MS_e = 2.40, *p < .05)* where more errors were made in the gender incongruent conditions than in the the gender congruent conditions, reflecting interference from the distractor word's incongruent gender information. The mean RTs broken down by semantic condition and target gender are shown in Table 12.

Semantic relatedness was significant in all the four analyses. The RT analysis by subjects shows a 14.8 ms advantage for the semantically unrelated conditions compared to the semantically related conditions. The effect of semantic condition reached significance in both RT analyses ($F_1(1,19) = 7.08$, $MS_e = 1239.2$, $p < .05$; $F_2(1,38) = 8.5$, $MS_e = 1776.82$, $p < .05$) and in the error analyses where more errors were made in the semantically related trials than in the unrelated ones ($F_1(1,19) = 45.0$, $MS_e = .5$, $p < .01$; $F_2(1,38) = 19.26$, $MS_e = 1.17$, $p < .01$). No other main effects or interaction effects were significant.

Table 12. Mean naming latencies and percentage errors (in parenthesis) in Exp.3, 20 subjects, taken from the RT analysis by subjects.

Condition	Target gender		Mean
	masculine	neuter	
Congruent	557(2)	568(2.5)	563(2.3)
Incongruent	563(5.8)	585(4.5)	574(5.2)
Semantically related (congr)	575(5)	587(3.5)	581(4.3)
Semantically related (incongr)	582(12.3)	588(9)	585(10.7)

The group of twenty respondents who participated in the experiment consisted of fifteen subjects with a two gender dialect and five with a three/two and a half gender dialect. Would the results from the analysis of Experiment 3 change if we removed the five subjects with a three gender system?

The difference between the gender congruent condition and the gender incongruent condition in the RT analysis by items was still significant after removing the five subjects with a three gender system ($F_2(1,38) = 5.06$, $MS_e = 1582.27$, $p < .05$). The error analyses were also still significant ($F_1(1,14) = 27.32$, $MS_e = 1.06$, $p < .01$; $F_2(1,38) = 15.12$, $MS_e = 1.44$, $p < .01$). The RT analysis by subjects showed a 10.78 ms advantage for the gender congruent condition. However, this difference did not reach significance (F_1(1,14) = 2.09, $MS_e = 1669.04$, $p = .17$).

Both the error analyses and the RT analysis by items reached significance indicating that conflicting gender information is causing more errors and longer naming latencies. Though not significant, the mean naming latencies in the gender incongruent conditions in the RT analysis by subjects also show an increase in response time for gender incongruent trials. This was also the case for the fifteen subjects with a two gender system indicating that the total lack of a gender congruency effect for the two-gender group in Experiment 1 is due to an inappropriate test set in Experiment 1 for detecting a gender congruency effect for these subjects.

Though there is a significant difference between the gender congruent condition and the gender incongruent condition in the RT analysis by item and in both the error analysis, and a descriptive move towards it in the RT analysis by subjects, it seems careless to reject the null hypothesis. After all, both RT analyses should be significant in order to reject the null hypothesis, if we want to be able to generalize the gender congruency effect to other people and other nouns.

All in all, though we cannot safely reject the null hypothesis, the overall results from Experiment 3 indicate that there is a trend towards an effect of gender interference or determiner interference. The results from the analyses of the three-gender group contrasted with the results obtained for the two-gender group in Experiment 1 indicate that the effect is a gender interference effect rather than a determiner interference effect. If the effect was a determiner effect, then we should not expect an effect independent of target gender for the three-gender group. We would only expect an effect for the neuter targets. Nor would we expect a difference in results between the two-gender group and the three-gender group. These two predictions is not in accordance with the results obtained. However, the results obtained in reanalysis of Experiment 1 is in accordance with the GSIH.

CHAPTER 6, THE COMPUTER MODEL

A computer model is a simplification of the real-world thing it is supposed to be a model of. Computer simulations allow us to test a theory without disturbance from other factors that are irrelevant for the concrete task, but is an inevitable part of the real-life task. With computer simulations we can cut away the irrelevant tasks, or factors we cannot control or measure, and focus on one aspect of the theory. In my case, leaving out the phonological and phonetic level of the production task in the computational model will allow us to focus only on the processes on the lemma level, which is where we want to test the model of gender selection. This is not possible in the real-life experiment. By simulating only the lemma and gender selection process, it is possible to test the theory of gender selection proposed in chapter 5. Comparing the simulation results with the experimental results might lead to new findings which in turn can provide new data for theory (Dijkstra & de Smedt, 1996).

The gender congruency effect obtained for the three-gender group in Experiment 1 is the foundation for the computational model. As argued for in chapter 5, the interference seems to be between different gender nodes. This is in accordance with Schriefers (1993) original interpretation of the gender congruency effect. When it comes to models of word productions, the computational model based on the results from Experiment 1 is not in conflict with any of the three models described in chapter 2. However, I assume a lemma level which is contrary to Caramazza's IN model, but that is not an essential part of this computational model. (Here, the essential part is that gender, a syntactic feature is selected through graded activation and competition.) Levelt et al.'s theory of speech production (1999) is captured in a computer model called WEAVER ++ (Roelofs, 1992, 1997, 2003), often referred to as only WEAVER. WEAVER can simulate, for example, a PWI task and it accounts well for several of the well-known findings from experimental paradigms, among them the semantic interference effect (Roelofs, 1992). Since there is nothing in the model shown in Figure 11 in chapter 5, that conflicts with the theoretical assumptions underlying WEAVER I decided to test the theory of competition between abstract gender nodes with an adaption to WEAVER, because of its convenient design for my task.

6.1 The simulations

Müller and Hagoort (submitted) developed a WEAVER-based model in their study of priming access of grammatical gender. Their model has implemented gender nodes as well as nodes for semantic fields. Similarly, my variants of Müller and Hagoort's model has nodes for two-semantic fields, animal and furniture, and nodes for each of the Norwegian genders. The input structure I have used is shown in Figure 12. As we can see, the words of the three animals *lam* [lamb], *gaupe* [lynx] and *hund* [dog] are connected through their semantic superordinate, *dyr* [animal]. These connections between semantically related concepts through a node representing the semantic superordinate, are an essential characteristic of the WEAVER for spreading activation throughout the network. Therefore the two hyperonyms *dyr* [animal] and *møbel* [furniture] are included in the computational model, even though they are not target words in the experiment.

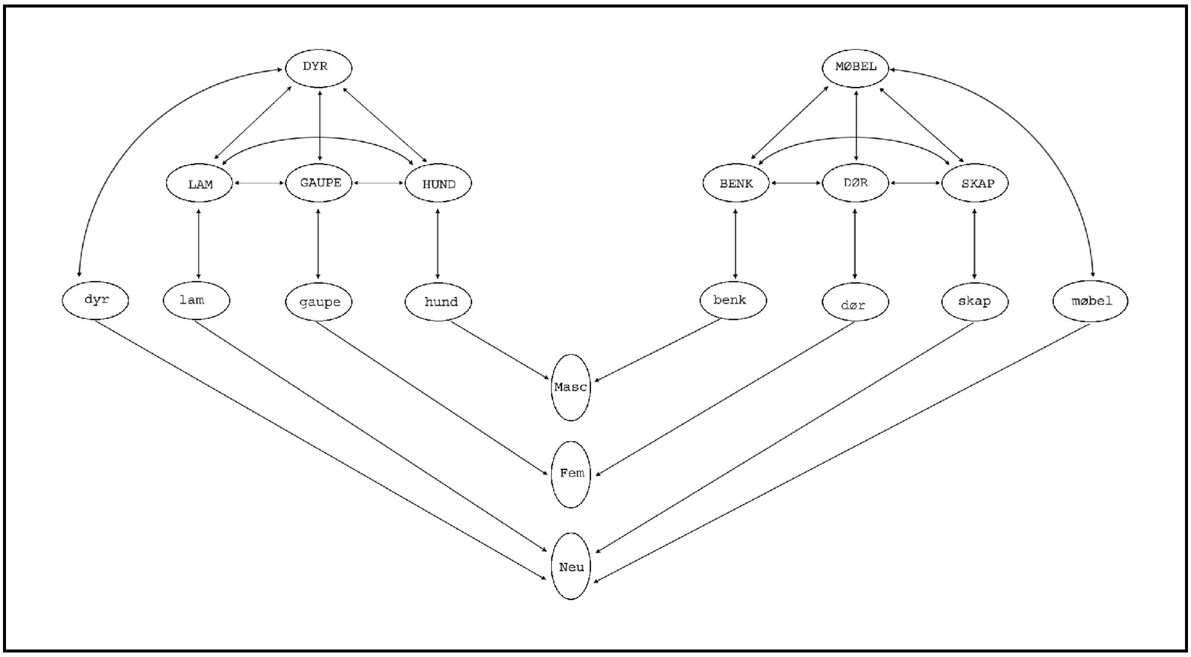


Figure 12. Representation of the computational model, input to WEAVER. Lexical concept nodes are written with capital letters. The lemma nodes are represented with small letters.

The target word in the simulations was always the masculine word *hund* [dog]. The congruent distractor word was the semantically unrelated word *benk* [bench]. The incongruent words were either the feminine word *dør* [door][10] or the neuter word *skap* [closet].

---

[10] The feminine word *seng* [bed] is actually used in the simulation instead of *dør* [door], but it does not make a difference.

Some minor code adjustments had to be done in order to run the simulations. I extended the network with an extra gender node, changed the SOA function to only include SOA=0 and implemented a network structure with Norwegian words and their connections to the gender nodes. In Simulation 3 the network was extended with lemmas for demonstrative nodes.

The simulations were run with two free parameters. The first is different distractor durations, which is a variable in the WEAVER for how long the distractor word receives external activation input. To start the simulations the target lexical concept and the distractor lemma receive external input. An extra amount of activation, called signaling activation, is added to the target picture's lexical concept node to simulate the ongoing perceptual input for the respondents from the picture displayed on the screen. Since the picture is the target, the signaling activation continuous until the target is selected. However, the distractor word must be suppressed in some sense, or else it would be selected instead, and that is simulated by shutting of the external input activation it receives after a certain time. The variable/parameter *distractor duration* (DU) determines how many milliseconds the distractor word receives external activation input. The second free parameter used in the simulation is *gender threshold*. That is the value for the activation level that the gender node has to exceed in order to be selected. In the simulations I used DU=0 ms, 75 ms, 100 ms, 125 ms, 150 ms, or 200 ms, and gender threshold=1, 2, or 3. For all the other parameters I used the ones used by Roelofs (2003, condition word reading with determiner).

I ran a simulation with three gender nodes in the network and competition implemented indirectly between gender nodes (Simulation 1). The same simulation was run one with a network with only two gender nodes (Simulation 2). I also ran two different versions where the competition was not implemented between gender nodes, but between demonstrative nodes (Simulation 3 a and b). The last simulations were run with semantically related distractor words (Simulation 4 a and b).

6.2 Simulation 1

The competition occurs indirectly between three gender nodes, feminine, masculine and neuter. The selection process goes about as follows: a production rule checks if the gender node belongs to the target lemma and that the target lemma's activation value exceeds the activation value of all the active lemma nodes that are permitted responses, and in addition that the respective gender node's activation level exceeds a certain threshold. The backward checking mechanism (that is, checking that the gender node belongs to the target lemma) is theoretically inspired by Levelt et al.'s (1999) solution to the *binding problem*. How do we make sure that we produce the retrieved lexical items in the correct way? The binding problem can be explained by following an example from Levelt et al. (1999). What prevents us from producing the sentence *Kings escort page*s if we want to produce the sentence *Pages escort kings*, or similar, what prevents us from producing the word *pings* instead of *kings* after retrieving the phonological form of the sentence? In order to avoid erroneous combinations of retrieved lexical items, there must be a mechanism that keeps track of where and who an item belongs to. This is explained by a *binding-by-checking* mechanism by Levelt et al.(1999). Each activated node has a procedure attached to it that checks if it links to an active node one level up.

The incongruent distractor words are the neuter word *skap* [closet] and the feminine word *dør* [door].

Results simulation 1

The results from WEAVER with a neuter distractor word are shown in Figure 13, 14 and 15, each with a different value for the free parameter, *gender threshold* (GT). The X-axis shows the other free parameter, distractor duration (DU). As can be seen from all the three figures, WEAVER predicts a delay in selection time for gender incongruent conditions as compared to gender congruent conditions. Recall that the simulation process is only a part of the entire speech process and therefore these results cannot be directly compared to the experimental results *per se,* but the time difference between the two conditions (gender congruent and gender incongruent) is what we should look at. Table 13 shows the predicted selection times for the different parameters.

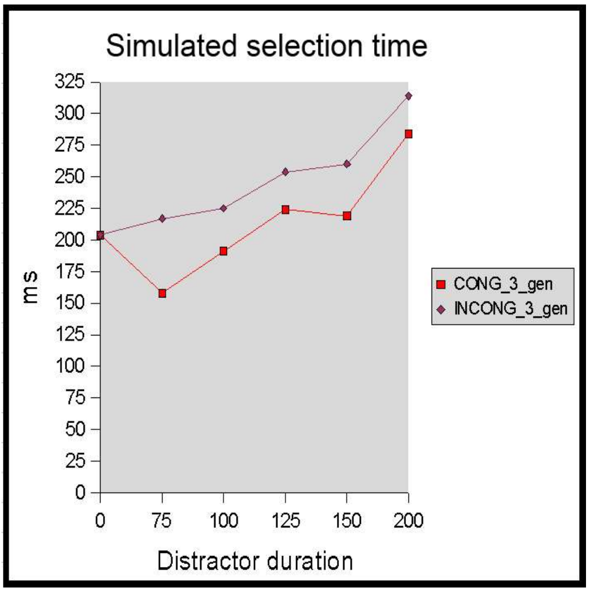


Figure 13. GT=1. Simulated selection time for abstract gender node in congruent and incongruent condition in a three gender node network, shown as a function of the parameter DU.

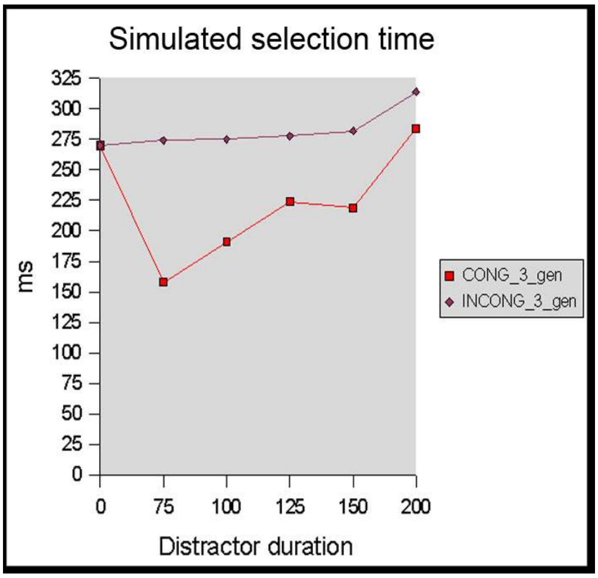


Figure 14. GT=2. Simulated selection time for abstract gender node in congruent and incongruent condition in a three gender node network, shown as a function of the parameter DU.

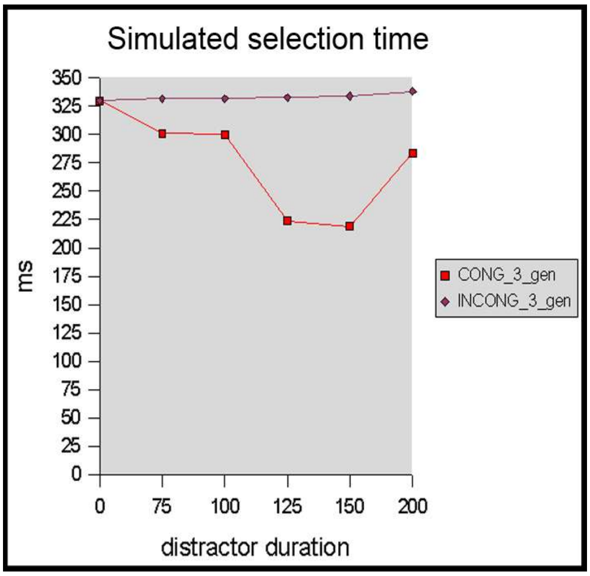
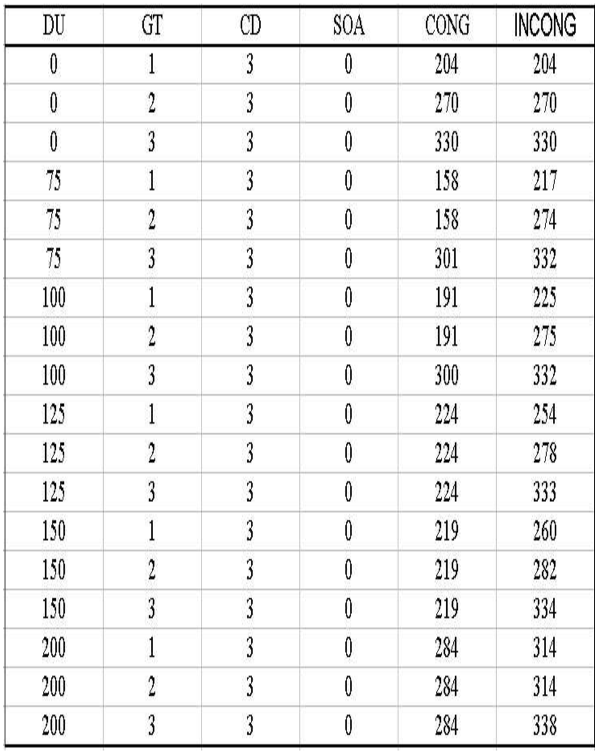


Figure 15. GT=3. Simulated selection time for abstract gender node in congruent and incongruent condition in a three gender node network, shown as a function of the parameter DU.



DU	GT	CD	SOA	CONG	INCONG
0	1	3	0	204	204
0	2	3	0	270	270
0	3	3	0	330	330
75	1	3	0	158	217
75	2	3	0	158	274
75	3	3	0	301	332
100	1	3	0	191	225
100	2	3	0	191	275
100	3	3	0	300	332
125	1	3	0	224	254
125	2	3	0	224	278
125	3	3	0	224	333
150	1	3	0	219	260
150	2	3	0	219	282
150	3	3	0	219	334
200	1	3	0	284	314
200	2	3	0	284	314
200	3	3	0	284	338

Table 13. Selected output from the WEAVER in Sim.1. GT=gender threshold, CD=critical difference, DU=distractor duration. The columns CONG and INCONG display expected selection time in ms for congruent and incongruent trials, respectively, in a three gender node network.

The three-gender group subjects in Experiment 1 had a mean RT of 575 ms for the semantically unrelated gender congruent trials and a mean RT of 590 ms for the semantically unrelated gender incongruent trials. This yields a 15 ms time difference between the means of the two conditions. The simulated results, as we can see from Figure 13, 14 and 15, varies as a function of the two parameters distractor duration (DU) and gender threshold (GT). With DU set to 75 ms and GT set to 1, the difference between the gender congruent condition and the gender incongruent condition is 59 ms. With GT set to 3, the WEAVER predicts a 143 ms increase in RT for the congruent condition, compared to GT=1 and DU=75, and a 115 ms increase for the gender incongruent condition. This difference however, is not as dramatic as it might seem. As mentioned in chapter 3, in the picture naming process we usually spend the first 150 ms to visually recognize the picture, then 125 ms or so to select the lemma. Phonological encoding starts roughly at 250 ms and articulation around 600 ms (Harley, 2001:365). Now, these are rough estimates and in my experiments, some subjects lied steadily around 500 ms, while others lied steadily around 700 ms in response time. These time differences make it difficult for me claim that one parameter is better than the other, even though one predicts a more than 100 ms longer selection time. To sum it up, there was no parameter value that was the perfect fit to my data, but all parameter settings (except distractor duration set to zero which was irrelevant for my task) in the semantically unrelated simulation, predicted a difference in RT in the intended direction.

The results from the simulation with the incongruent distractor word *dør* [door, fem] showed the same results as the incongruent distractor word *skap* [closet, neu].

6.3 Simulation 2

The procedure was the same as in simulation 1, except that there were only two gender nodes (masculine and neuter). The simulation was run with *skap* (neuter) as the incongruent distractor word. The feminine words were rewired into masculine for this simulation.

Results simulation 2

The motivation for the second simulation was to see if WEAVER predicts a different result for a network with two gender nodes versus a network with three gender nodes. Reducing the amount of gender nodes in the network architecture, led to a reduction in selection time. The predicted mean selection time for the target gender node in the two conditions (gender congruent and gender incongruent) are shown in Figure 16, 17 and 18 as a function of the two parameters, distractor duration (DU) and gender threshold (GT).

The results predict a rather clear delay in gender selection time in gender incongruent trials, also for subjects with two genders (in accordance with the gender congruency effect found for Dutch which also has a two gender system (e.g., Schriefers, 1993, inter al.)). Recall that in Experiment 1, which is the basis for the computational model simulated in Simulation 1, there was no clear effect of gender congruency/interference for the subjects with a two gender system. However, in Experiment 3 there was a descriptive finding, and a significant one in the F_2 analysis for the two gender subjects.

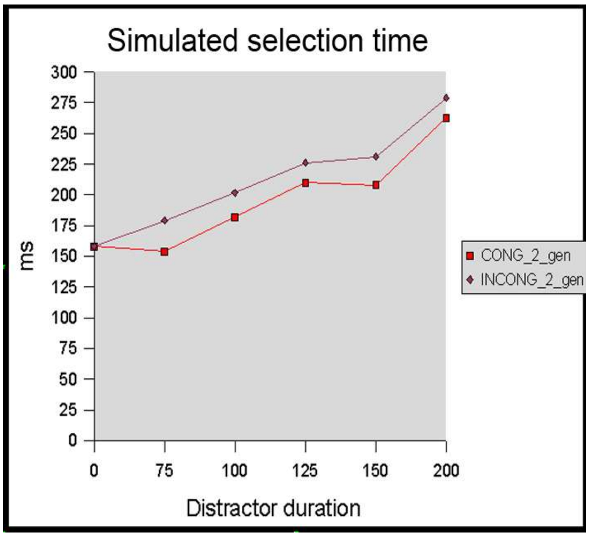


Figure 16. GT=1. Simulated selection time for abstract gender node in congruent and incongruent condition in a two gender node network, shown as a function of the parameter DU.

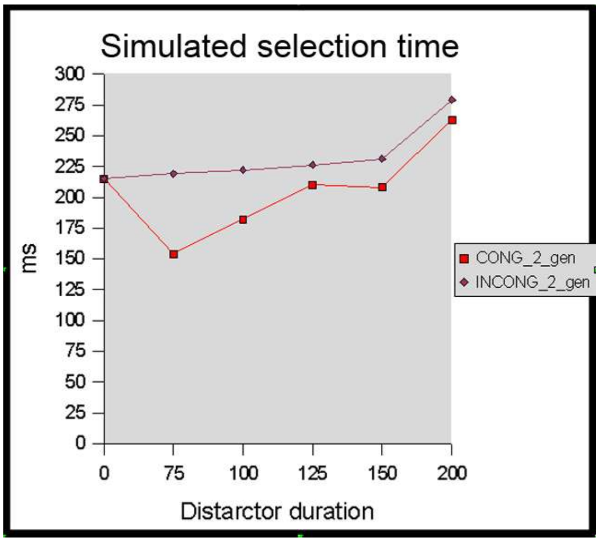


Figure 17. GT=2. Simulated selection time for abstract gender node in congruent and incongruent condition in a two gender node network, shown as a function of the parameter DU.

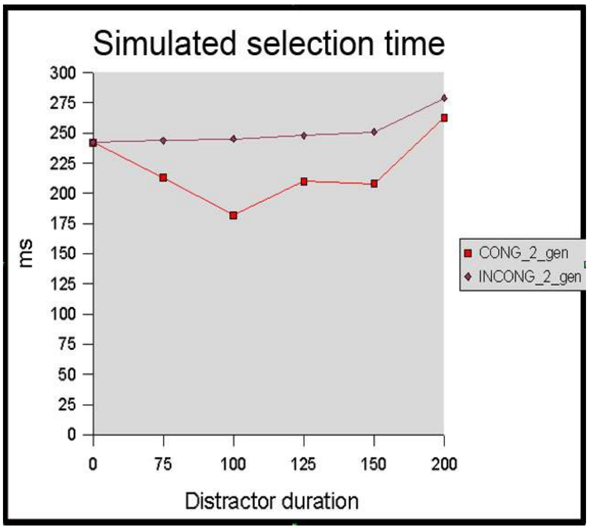
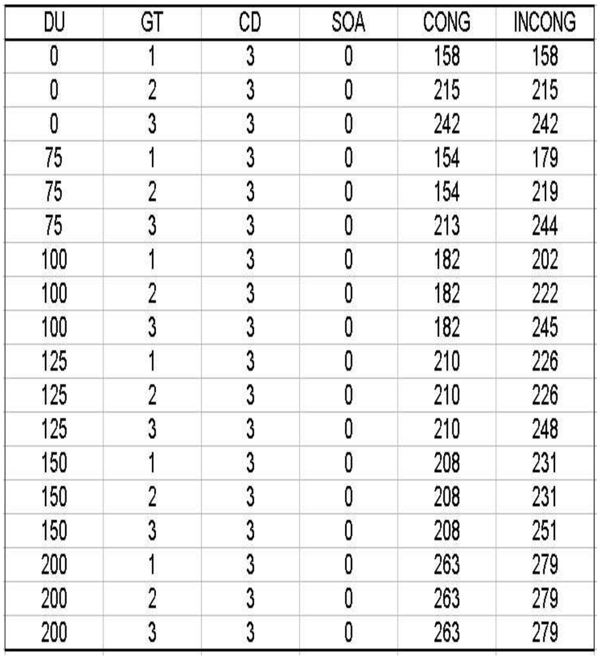


Figure 18. GT=3. Simulated selection time for abstract gender node in congruent and incongruent condition in a two gender node network, shown as a function of the parameter DU.



DU	GT	CD	SOA	CONG	INCONG
0	1	3	0	158	158
0	2	3	0	215	215
0	3	3	0	242	242
75	1	3	0	154	179
75	2	3	0	154	219
75	3	3	0	213	244
100	1	3	0	182	202
100	2	3	0	182	222
100	3	3	0	182	245
125	1	3	0	210	226
125	2	3	0	210	226
125	3	3	0	210	248
150	1	3	0	208	231
150	2	3	0	208	231
150	3	3	0	208	251
200	1	3	0	263	279
200	2	3	0	263	279
200	3	3	0	263	279

Table 14. Selected output from the WEAVER in Sim.2. GT=gender threshold, CD=critical difference, DU=distractor duration. The columns CONG and INCONG display expected selection time in ms for congruent and incongruent trials, respectively, in a two gender node network.

Running the simulation with the distractor word *dør* (which is masculine in this simulation) no gender congruency effect was found. The estimated RTs were the same as the ones predicted for the gender congruent condition. However, this is not surprising because *dør* is actually gender congruent in this simulation unlike in Simulation 1.

Another observation for the two gender subjects was that they had a faster mean RT than the three gender subjects. As can be seen from Table 13 and 14, the selection time is faster for a network with two gender nodes than the selection time for a network with three gender nodes. As mentioned earlier, one argument for interference occurring at the level where gender nodes are selected, is the delayed RT for the three-gender group compared to the two-gender group's RT claimed to be caused by the difference in the amount of gender nodes. (If it was at demonstrative level, there would be equally many demonstrative nodes for both groups). Whether this simulation result strengthens my argument for the interference occurring at the abstract gender node level rather than demonstrative level, is definitely questionable. The difference in selection time for a network consisting of three gender nodes as compared to a network with two gender nodes, is an a priori assumption implemented in the WEAVER computer program and depended on the set size of competitive nodes. That the simulation result confirms the program's algorithm does not really cast light on the abstract gender node versus demonstrative node discussion.

Another possible explanation for the time difference between the two gender network and the three gender network, can come from difference in the input networks to WEAVER. (I have later learned that WEAVER is sensitive to the distribution of gender nodes in the network). As shown in Figure 12 (section 6.1), the input network in Simulation 1 contains two feminine words. These are changed to masculine for the two gender node network used to run Simulation 2. That means that the masculine target's gender node receives activation from four nodes after a while, unlike in Simulation 1 where the masculine node only receives activation from two nodes. The extra activation to the masculine gender node (in addition to the reduced competition) might lead to a faster selection time of that gender node than it would in Simulation 1.

6.4 Simulation 3a and b

Instead of assuming that the competition occurs when selecting the abstract gender node, I assumed that the competition occurred between demonstrative nodes that are activated through a unidirectional link from the gender node to its corresponding demonstrative node lemma. Note that this is not really putting the DSIH to test, because Caramazza's IN model do not operate with lemmas, and therefore the competition should be between the demonstrative lexeme nodes and not lemma nodes then. Moreover, in the simulations where the incongruent distractor word is neuter, we would expect a gender/determiner congruency effect both under GSIH and DSIH for the two-gender group and the three-gender group, so this simulation cannot be used to differentiate between the GSIH and the DSIH.

Though I have argued against the assumption that the competition occurs between different demonstratives, it would be interesting to see what pattern we might have found if competition occurred at demonstrative level rather than that of abstract gender node. Since the relation between demonstrative and gender is, in one instance, a many-to-one relation for the three gender respondents (*den* for both feminine and masculine, but only *det* for neuter), it might yield a different result than the one-to-one relation for the respondents with two genders (*den* for masculine and *de*t for neuter). However, the competition is in both cases only between the *den* node and the *det* node, which should lead us to believe that the result would be the same for respondents with three genders and those with only two.

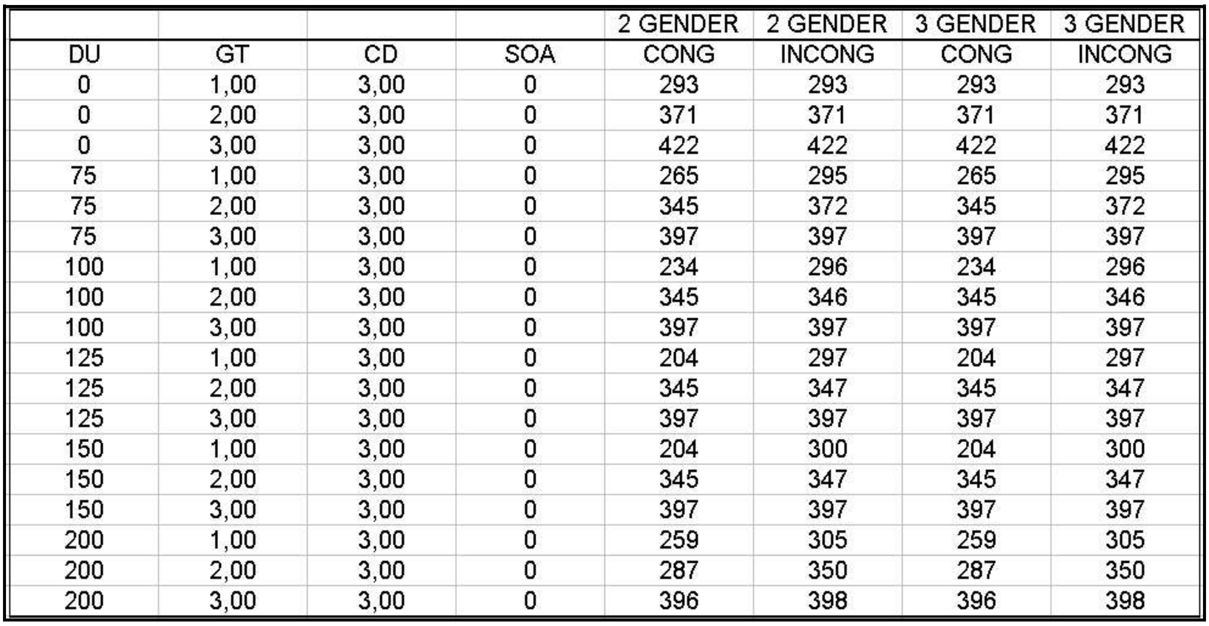
I assumed that competition occurred between the demonstrative nodes *den* and *det* instead of between masculine, neuter and feminine (if it is in the three gender network) nodes. The selection process goes about as follows for Simulation 3 a: a production rule checks that the target lemma node has been selected and that the target word's demonstrative is bigger than a threshold (set to 3). Note that this rule does not make a reference to gender node values at all. In Simulation 3 b there is a reference to the target gender node's value. The production rule checks if the target lemma has been selected and that both the target gender node's and the targets demonstrative node's activation level has passed the gender specific threshold (the same gender threshold for both the gender node and the demonstrative node).

Results Simulation 3 a and b

Running the simulations with the incongruent neuter distractor word *skap*, the results from

Simulation 3 a and 3 b were identical. The same were the results for the two gender simulations and the three gender simulations. For most of the parameter settings, WEAVER predicts a gender interference effect, or more precisely a demonstrative interference effect, for all the subjects. This is also predicted by both the DSIH and the GSIH. The expected selection times (in ms) for the demonstrative node are shown in Table 15 where DU = distractor duration, GT = gender threshold, and CD = critical difference.

Table 15. Selected output from WEAVER, Simulation 3 a and b. Predicted selection times for demonstrative node in the gender congruent (CONG) and incongruent (INCONG) condition.



				2 GENDER	2 GENDER	3 GENDER	3 GENDER
DU	GT	CD	SOA	CONG	INCONG	CONG	INCONG
0	1,00	3,00	0	293	293	293	293
0	2,00	3,00	0	371	371	371	371
0	3,00	3,00	0	422	422	422	422
75	1,00	3,00	0	265	295	265	295
75	2,00	3,00	0	345	372	345	372
75	3,00	3,00	0	397	397	397	397
100	1,00	3,00	0	234	296	234	296
100	2,00	3,00	0	345	346	345	346
100	3,00	3,00	0	397	397	397	397
125	1,00	3,00	0	204	297	204	297
125	2,00	3,00	0	345	347	345	347
125	3,00	3,00	0	397	397	397	397
150	1,00	3,00	0	204	300	204	300
150	2,00	3,00	0	345	347	345	347
150	3,00	3,00	0	397	397	397	397
200	1,00	3,00	0	259	305	259	305
200	2,00	3,00	0	287	350	287	350
200	3,00	3,00	0	396	398	396	398

According to the results from WEAVER, there should not be a difference in RT between subjects that have a two gender system and subjects that have a three gender system here as shown in Table 15. This is not in accordance with the results from Experiment 1 or Simulation 1 contrasted with Simulation 2. However, it could be that the difference in predicted selection times between Simulation 1 and Simulation 2 is caused by difference in the amount of gender nodes in the two networks which were input to WEAVER, while in Simulation 3, there are equally many determiner nodes in the two networks.

Running the same simulations with the distractor word *dør* [door] instead, yielded no gender congruency effect for neither the three gender node network nor the two gender node network as the DSIH predicts. The GSIH would predict a gender congruency effect for the three-gender group only. The results from the simulations with *dør* as distractor, showed no effect of gender congruency in any of the simulations. However, this is expected even though a congruency effect is predicted by the GSIH (which was the theoretical basis for the computational model). The only place where we would expect interference from incongruent gender information, is for the three-gender group, because the feminine gender node is competing for selection with the masculine gender node. That competition-scenario is not simulated here, but in Simulation 1. In this scenario we simulate competition between determiner lemmas. Since both target and distractor's determiner is *den* for both subjects with a two gender dialect and subjects with a three gender dialect, there is no competition from an incongruent determiner.

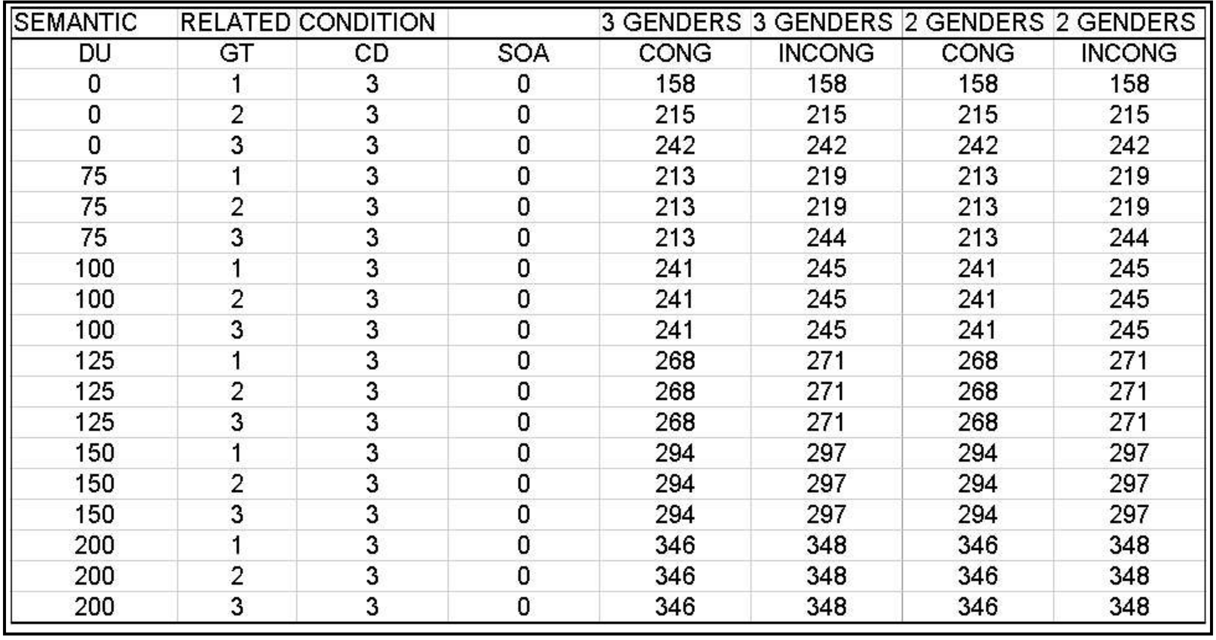
6.5 Simulation 4 a and b

Semantically related condition was also simulated with competition occurring between gender nodes. The procedure was the same as in Simulation 1, and performed on a three gender node network in 4 a, and on a two gender node network in 4 b. One modification had to be done to the input network in order to run simulation 4 a. In simulation 4 a the semantically related neuter word *lam* [lamb, neu] was replaced by the masculine word *gris* [pig] in order to run a the gender congruent condition. The target word is as always the masculine word *hund* [dog]. The semantically related gender incongruent distractor word is the feminine word *gaupe* [lynx]. Note that there is one neuter word less in this network than in the others. In Simulation 4 b (two gender network), I used *gaupe* [lynx] (which is here a masculine word) as congruent distractor and *lam* [lamb, neu] as the incongruent distractor.

Results simulation 4 a and b

WEAVER predicted identical selection times for a two gender node network and a three gender node network. The expected selection times predicted by WEAVER showed a very small advantage for the gender congruent condition compared to the gender incongruent condition, with mostly only a 3 ms difference for the different threshold values. The real data showed a difference of about 14 ms between these two conditions. The only 3 ms difference is so small that it would most likely not be significant if it was obtained in a real experiment. The predicted selection time for Simulation 4 are shown in Table 16.

Table 16. Selected output from WEAVER, Simulation 4 a and b. Predicted selection times for semantic related gender congruent condition (in the column marked CONG) and semantic related gender incongruent condition (in the column marked INCONG).



SEMANTIC	RELATED	CONDITION		3 GENDERS	3 GENDERS	2 GENDERS	2 GENDERS
DU	GT	CD	SOA	CONG	INCONG	CONG	INCONG
0	1	3	0	158	158	158	158
0	2	3	0	215	215	215	215
0	3	3	0	242	242	242	242
75	1	3	0	213	219	213	219
75	2	3	0	213	219	213	219
75	3	3	0	213	244	213	244
100	1	3	0	241	245	241	245
100	2	3	0	241	245	241	245
100	3	3	0	241	245	241	245
125	1	3	0	268	271	268	271
125	2	3	0	268	271	268	271
125	3	3	0	268	271	268	271
150	1	3	0	294	297	294	297
150	2	3	0	294	297	294	297
150	3	3	0	294	297	294	297
200	1	3	0	346	348	346	348
200	2	3	0	346	348	346	348
200	3	3	0	346	348	346	348

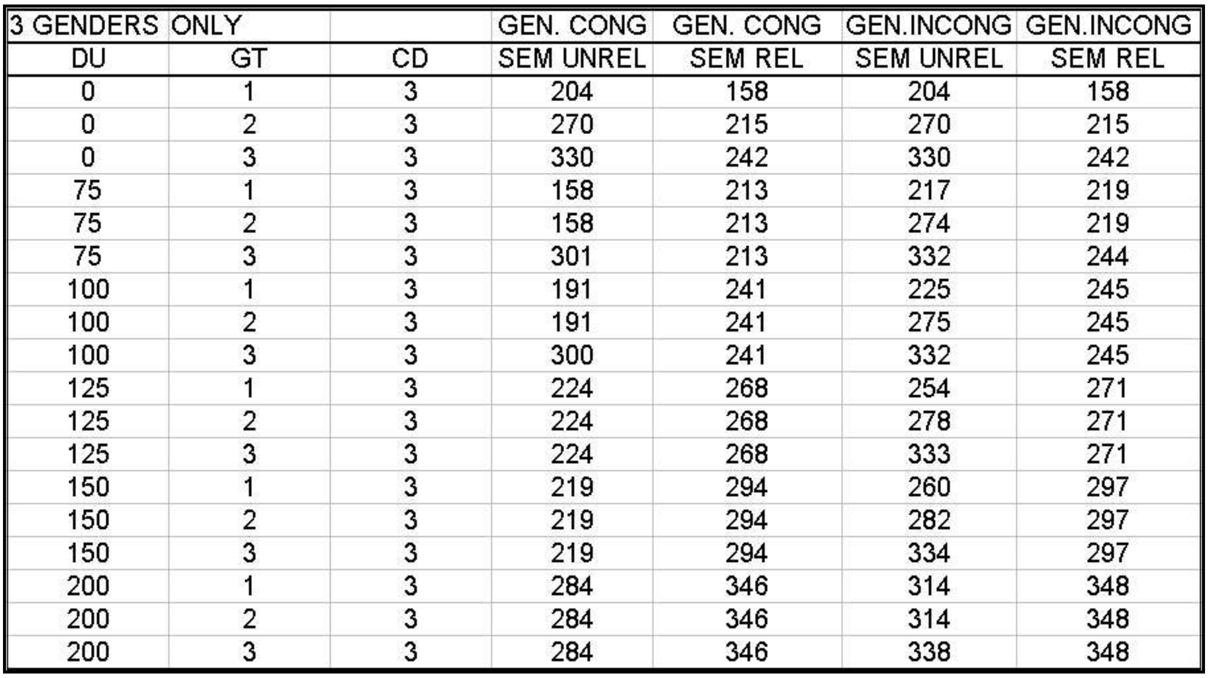
This reduction of the gender interference effect might occur because most of the parameters in WEAVER concerns lemma selection. As we know, a semantically related distractor word causes interference in lemma selection. By the time the lemma competition has finished and the lemma is selected, the interference caused by the incongruent gender information might have dissipated when the gender node is ready to be selected. However, for one parameter setting, distractor duration (DU) set to 75, gender threshold (GT) set to 3 and critical difference (CD) set to 3, WEAVER predicts a 31 ms difference between the gender congruent and gender incongruent conditions.

The additive effect of semantic interference and gender congruency obtained in Experiment 1, is only present in one parameter setting. The other parameter settings predict a reduction of the gender interference effect in the semantically related condition. The one parameter setting that does predict an additive effect, fails to account for the semantic interference effect regardless of gender condition (as can be seen in Table 17). The semantically unrelated condition had a longer predicted RT than the semantically related condition, which is opposite of what we would expect.

From Table 17 we can see three things that are opposite of what we would expect. The first is a difference in DU=0 between Simulation 1 and Simulation 4. When DU=0, that means that there is no distractor word present and both simulations should therefore have equal RT. The second thing is the reversed semantic interference effect found for GT=3 and DU=75 and for GT=3 and DU=100. The third thing is the lack of gender congruency effect for the semantic related conditions, or rather the presence of one for GT=3 and DU=75. Based on these three rather unexpected simulation results, Oliver Müller (private correspondence, 15.04.2005) made another network to compare with, where there were equally many nodes of each gender, since WEAVER is sensitive to the distribution of gender nodes in the network. When it comes to the two first findings, these abnormalities disappear in a balanced network. In other words, they are probably a result of the difference between the network used in Simulation 1 and the one used in Simulation 4. The gender congruency effect found for one parameter setting in Simulation 4 also disappeared. The main findings from Simulation 1 and 4 were still present. That means that even in a balanced network, the difference between the gender congruent and gender incongruent condition is only a couple of ms.

When it comes to the lack of gender congruency effect in the semantically related conditions, inspections of the activation files show that in these conditions, lemma selection is the latest selection criterion to be fulfilled. There are two criteria that have to be fulfilled before computation into expected selection time can start. The first criterion is that the lemma's activation node has to pass all the lemma nodes' activation value with a critical difference (CD, set to 3). The second criterion is that the target gender node has to pass the gender threshold (GT, set to 1,2 or 3). The latest criterion to be fulfilled is the criterion which determiners when the computation of selection time starts. Comparing the activation file from Simulation 4 with simulation 1 shows that in Simulation 1 (semantically unrelated condition), the lemma reaches CD at an earlier time step than in Simulation 4. In the gender congruent conditions in Simulation 1, GT is reached earlier than CD (for most of the parameter settings) as in the semantically related simulation in 4. However, since the target lemma node reaches the CD faster than in the semantically related condition, the computation of selection time can start earlier. In the gender incongruent conditions in the semantically unrelated simulation, GT is sometimes reached later than the lemma criterion. That means that GT delays the selection time compared to the gender congruent condition and hence, we get a longer selection time in the gender incongruent condition. This does not happen in the semantically related simulations (except for 1 parameter setting) because it is the lemma that delays the selection, not the gender threshold. All in all, it seems like lemma selection has more influence on the selection times in the semantically related simulations than gender threshold.

Table 17. The predicted selection times in ms, Simulation 1 (semantically unrelated, SEM UNREL) compared with simulation 4, (semantically related, SEM REL), GEN. CONG is gender congruent, GEN. INCONG is gender incongruent. The gender incongruent conditions are with a neuter distractor word.

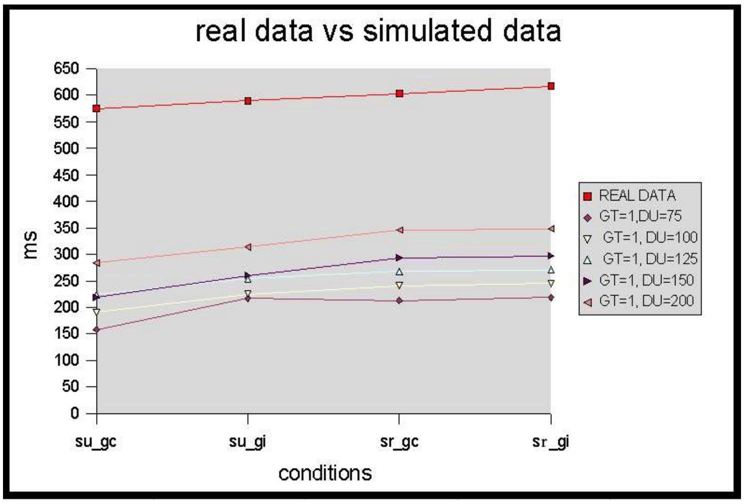


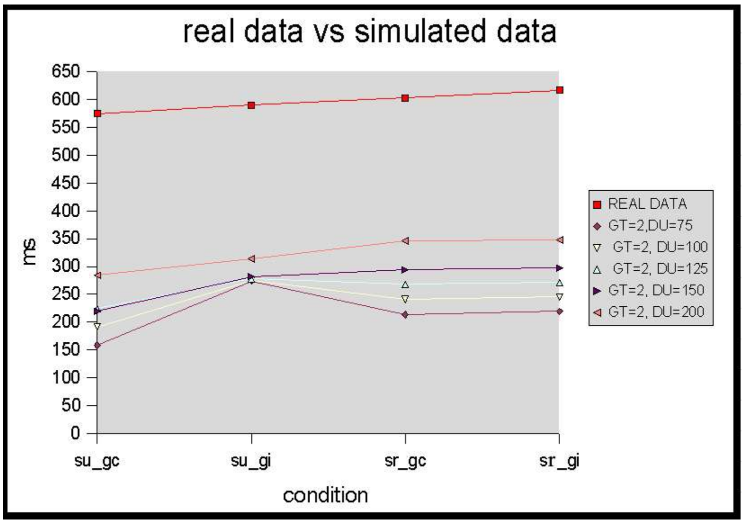
3 GENDERS	ONLY		GEN. CONG	GEN. CONG	GEN.INCONG	GEN.INCONG
DU	GT	CD	SEM UNREL	SEM REL	SEM UNREL	SEM REL
0	1	3	204	158	204	158
0	2	3	270	215	270	215
0	3	3	330	242	330	242
75	1	3	158	213	217	219
75	2	3	158	213	274	219
75	3	3	301	213	332	244
100	1	3	191	241	225	245
100	2	3	191	241	275	245
100	3	3	300	241	332	245
125	1	3	224	268	254	271
125	2	3	224	268	278	271
125	3	3	224	268	333	271
150	1	3	219	294	260	297
150	2	3	219	294	282	297
150	3	3	219	294	334	297
200	1	3	284	346	314	348
200	2	3	284	346	314	348
200	3	3	284	346	338	348

6.6 Evaluation of the simulation results

The main simulation was the first one, which was inspired by Experiment 1 where an effect of gender congruency was obtained for the three gender subjects. The input to WEAVER in Simulation 1 was a network consisting of three gender nodes, where competition is assumed to occur between different gender nodes. I will therefore use the simulations that were run on a three gender node network when evaluating the simulation results.

Since there are few data points to compare the simulation results with the real results, a "profile fit" was chosen as the evaluation method. A profile fit basically means to compare the simulated result with the real one and see if they behave in the same manner, i.e. have similar curves (Müller, private correspondence, 28.01.2005). The predicted mean selection times for the four conditions in the experiment were shown in Table 17 in section 6.5. A profile fit for the same four conditions, semantically unrelated gender congruent (su_gc), semantically unrelated gender incongruent (su_gi), semantically related gender congruent (sr_gc) and semantically related gender incongruent (sr_gi) is displayed in Figure 19 (top), 20 (middle) and 21 (below). Note that the ms-axis displays the total naming latencies for the real data, while the simulated results only show the gender node selection times.





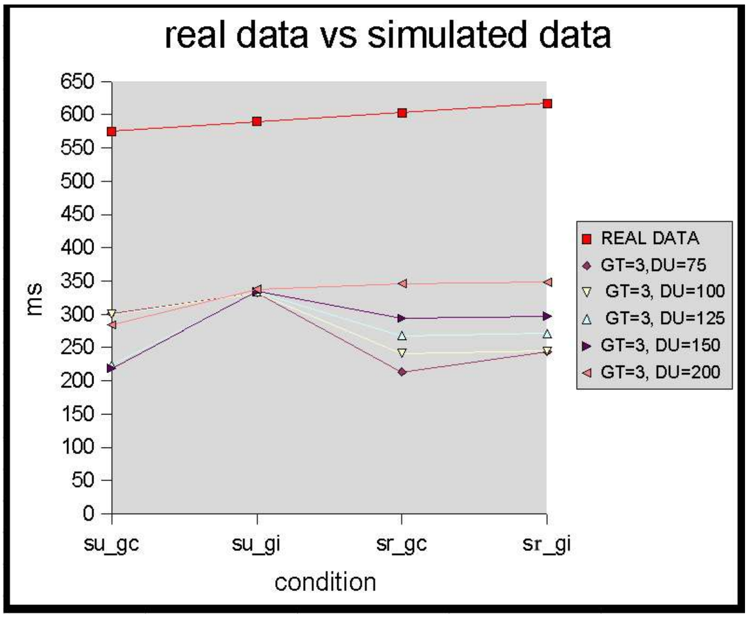


Figure 19 (top), 20 (middle) and 21 (bottom). Profile fit for real data and simulated data. Real data is from the three-gender group from Experiment 1. GT = gender threshold, DU = distractor duration.

As we can see from Figure 19, 20 and 21, long distractor durations account best for three of the four conditions. The parameters that do worst are short distractor duration and high gender threshold.

When it comes to comparing the simulation results with the results from Experiment 1, WEAVER, on one hand, does predict a slower RT for gender incongruent trials than gender congruent ones (in semantically unrelated conditions) and fits the empirical findings in that sense. On the other hand, the time difference in ms is much larger than the one obtained in Experiment 1 for the three-gender group. Most of the parameter settings are on the lemma selection task, therefore it could be that tuning on the parameter would give a better fit to the absolute time difference in ms. One could argue that tuning sufficiently on different parameters could give a fit to almost anything, and thereby losing the possible information value a simulation might have. However, WEAVER (with the present parameters) is not designed for all tasks in all languages, so tuning a bit on the parameters can be defended.

WEAVER do not predict a clear gender congruency effect in semantically related trials, though it predicts a couple of ms difference. In Experiment 1 (three-gender group), the time difference in ms between gender congruent trials and gender incongruent trials was stable across semantic condition, with about 15 ms difference in both semantically related and semantically unrelated condition. Most of the parameter settings in the semantically related condition incorrectly result in a lack of gender congruency effect (or one of an insignificant size of 3 ms). The absolute time difference predicted by WEAVER deviates from the empirical data. The real data show about 15 ms difference between gender congruent and gender incongruent conditions, in both semantically related and semantically unrelated condition.

The lack of a gender congruency effect in the semantically related simulations in WEAVER appears to caused by the cascading activation flow from lemma node to gender node. The semantic interference at the lemma level leads to a “masking” of the gender interference effect. Though my model has a seriality of selection (gender node cannot be selected before the lemma node is), it is the cascading activation flow that leads to the lack of gender congruency effect. A more strict implementation of seriality might lead to an additive effect of gender congruency and semantic interference. Changing the parameters CD or the spreading rate might also lead to an additive effect of gender congruency and semantic interference.

The parameter settings as they are, do lead to delayed RTs in the semantically related condition compared to the semantically unrelated condition, which is in accordance with the real data. In addition, delayed RTs in gender incongruent condition compared to gender congruent condition if the target and distractor are semantically unrelated, are predicted by WEAVER. This is also in accordance with the real data.

CHAPTER 7, GENERAL DISCUSSION AND CONLUSION

The starting point for the thesis was the gender congruency effect found in “early selection languages” (Miozzo & Caramazza, 1999) such as Dutch (e.g., Schriefers, 1993; La Heij et al., 1998; Van Berkum, 1997; Schiller & Caramazza, 2003) and German (e.g., Schriefers & Teruel, 2000; Schriefers et al., 2002; Schiller & Caramazza, 2003). Since Norwegian is an early selection language, a gender congruency effect should also be obtained in Norwegian.

The first experiment was more exploring in nature when it came to the question of whether the gender congruency effect occurs because of competition when selecting the appropriate abstract gender feature (e.g., Schriefers, 1993), or rather competition when selecting the appropriate determiner (Schiller & Caramazza, 2003; Schriefers et al., 2002).

The first experiment showed a trend towards a gender congruency /interference effect (though it did not reach significance in the RT analysis by items) with longer naming latencies when the picture was shown with a word with a different gender, than when the picture and word had the same gender. Another experiment was conducted, where the naming task was identical to the one in the first experiment. But this time, with a gender-marked suffix on the distractor word to “boost” the level of interference from a distractor word with conflicting gender information. This experiment showed a gender congruency effect in the RT analysis by items, supported by the error analyses. However, this finding did not reach significance in the RT analysis by subjects.

The overall results from the experiments were difficult to interpret. One reason for that was that the findings from the main experiment (Experiment 1) suffered from the unfortunate manner the feminine words were included in this experiment, since the respondents were a group consisting of people with a three gender system, two-and-a-half gender system and a two gender system. By splitting the respondents up in two groups, those with only masculine and neuter gender in one, and those who use the feminine gender also on some occasions in the other, a new pattern emerged.

The two-gender group showed no effect of gender congruency. However, the other group, those who have the gender feminine in their dialect, showed a reliable effect of gender congruency. The unintended effect of the problematic test set architecture was suddenly a good one for contrasting the Gender selection interference hypothesis (GSIH, Schriefers, 1993) and the Determiner selection interference hypothesis (DSIH, Schiller and Caramazza, 2003), without including an externally set factor, like plural, in order to contrast the determiners. With the test set architecture from the first experiment in mind, we can reason as follows: If it is true that interference occurs when selecting the appropriate determiner/demonstrative (DSIH), then the effect should be the same for all the respondents, because the demonstrative for the feminine and masculine is the same. In other words, there would not be a reason to expect a difference in results between the two groups of subjects. However, if the interference occurs while selecting the appropriate abstract gender feature (GSIH), we would expect a gender interference effect for those subjects that have the gender feature feminine (keeping the test set design in mind; normally we would expect an effect for both groups). For these subjects, the abstract features feminine and masculine would be different (as opposed to the demonstratives which are the same), and thereby interference will occur through conflicting gender feature information, and this in turn will lead to longer reaction time in the incongruent conditions. The presence of a gender congruency effect for the three/two-and-a-half gender group is in accordance with the GSIH and Schriefers' (1993) original interpretation of the gender congruency effect, and in conflict with the DISH.

This interpretation suffers from the lack of support from the two-gender group's results. If the gender congruency effect occurs, as just proposed, at the level of abstract gender node, we should find an effect for the two-gender group by looking at the neuter targets only, or by removing the unintended congruent trials (the feminine/masculine paired with masculine/feminine which would be congruent instead of incongruent for the two gender subjects) from the analyses. However, since the results from the two gender subjects in Experiment 3, show an increase in naming latencies for the gender incongruent trials for all target genders, it is more likely the test set architecture from Experiment 1 was bad for detecting a gender congruency effect for the two-gender group, than that the effect found for the three-gender group in Experiment 1 just occurred by chance.

The GSIH was the theoretical basis for the computational model for gender selection in Norwegian. The computational model was tested in WEAVER++ (Roelofs, 1992, 1997, 2003) with more or less good simulation results, though WEAVER predicted a larger difference in ms than the empirical findings. The simulations could not be used to differentiate between the DSIH and GSIH, but WEAVER accounts well for the predictions derived from the GSIH.

However, WEAVER could not account for the gender congruency effect obtained in the semantically related trials in Experiment 1. This is because, with the present parameters, lemma selection delays the selection time, not gender node selection.

Now, at least one question remains concerning the empirical findings from the experiments. Why does it seem to be a competition between abstract gender nodes in Norwegian, unlike in German (e.g., Schiller & Caramazza, 2003; Schriefers et al., 2002) and Dutch (e.g., Schiller & Caramazza, 2003) where the experimental support for the DSIH has been strong? Though German, Dutch and Norwegian are "early selection languages", there is one thing that is different: the two-and-a-half gender system in Bokmål. The "two-and-a-half" means that a feminine noun can take a masculine agreement target in some noun phrase constructions. Let us assume that the lemma of a feminine noun has two connections from itself to its gender information, one to the masculine node and one to the feminine node, in order to access both the set of masculine agreement targets and the set of feminine agreement targets. That could mean that the lemma activates the feminine gender node and the masculine gender node at the same time. Then we have the possibility of choosing between gender nodes. In German, Dutch and the two gender dialect(s) in Norwegian, where the noun's gender do not "change" depending on the noun phrase, we would find only one connection from the lemma to its gender node. Since the gender feature then is not a variable, but a constant, a selection mechanism would not be needed at that level, and therefore, gender comes as an automatic consequence of selecting the lemma. Note that these thoughts are pure speculations.

The question why competition seems to be between gender features in Norwegian and not between the determiners, is just one of many questions that remains unanswered or have come up during the experiments. Another question is why we do not find a significant gender congruency effect for the two-gender group. After all, the effect has been found in several experiments in Dutch which is practically a two gender language.

Future work

The computer simulations in WEAVER have raised some ideas about other questions that could be simulated, especially concerning the distribution of gender in the input network to WEAVER. It would be interesting to run the same simulations with an input network that reflects the real gender distribution in Experiment 1 (or in Norwegian, Bokmål), which differs between the subjects with a three gender dialect and those with a two gender dialect. For a two gender node network, would we find a longer RT for neuter targets than masculine targets if 2/3 of the network consists of masculine words? Would it have an impact on the gender congruency effect? For a three gender node network, would the predicted gender congruency effect (in ms) be reduced if the input architecture was balanced the same way the experimental test set was?

When it comes to the experiments, the gender congruency effect ought to appear for subjects with a two gender system. However, we cannot say that there is no gender congruency effect at all, because there is a descriptive trend towards the effect in Experiment 3, and most importantly, only one SOA has been tested. Schriefers and Teruel (2000) obtained a semantic interference effect at SOA=0 and a gender congruency effect at SOA = +150 in their experiment 1. This "(...) shows that the semantic interference effect and the gender interference effect can occur at different SOAs" (Schriefers & Teruel, 2000:1372). Therefore, testing with different SOAs ought to be done.

Another thing that would be interesting to test is the proposal that the gender congruency effect only occurs for free standing morphemes (Schiller & Caramazza, 2003), and not suffixes. Schiller and Caramazza failed on several occasions to replicate the gender congruency effect obtained for adjective + noun NPs for Dutch by Schriefers (1993). In Norwegian, the definite article is realized as a gender-marked suffix attached to the noun. It would be interesting to look at the suffixes' role and run an experiment where subjects produce the definite form of the noun, such as in *hus**et*** [the$_{neu}$ house] *or bil**en*** [the$_{masc}$ car]. Here we would not expect a gender congruency effect following Schiller and Caramazza. Whether we would find a gender congruency effect here or not, it would make the findings from the naming task in Experiment 1 clearer when it comes to the question concerning the suffixe's role in Experiment 1.

When it comes to the indefinite articles, the Norwegian ones, *en*, *ei* and *et,* have different stems. We would therefore expect to find a gender congruency effect. However, caution must be paid here because of the two-and-a-half gender system when choosing the words for the experiments, since the feminine words differ when it comes to how often they take a feminine form and how often they take a masculine form. A pre-study should be done before selecting the feminine words for the experiment. Another solution would be to make a new test set to run on people with Nynorsk as the written norm. Here we find a strict three gender system [11] unlike that in Bokmål, where the widely distributed two-and-a-half gender system always will contain some unknown factor when it comes to the feminine. Nynorsk would then be a "safer" testing ground for the gender congruency effect, both with an indef. Art + N naming task, and with the original Dem+ N+Def. suff. naming task. Another property with Nynorsk is that it has gender-marked definite plural form unlike German (always *die*) and Dutch (always *de*). The definite plural form is made by adding the *-(a)ne* (masculine), *-a* (neuter), and in most cases *-(e)ne* (feminine) to the end of the stem. Examples are shown in (16) to (18).

(16) *bil**ane*** [the cars, masculine]
(17) *klokke**ne*** [the bells, feminine]
(18) *hus**a*** [the houses, neuter]

However, if the gender congruency effect does not appear for suffixes, we would not expect an effect here.

The last thing listed as future work, will be to investigate the production of adjective + noun NPs in a colored naming task. A significant effect of gender congruency was obtained for Dutch by Schriefers (1993), but not replicated by Schiller and Caramazza (2003) for neither German or Dutch. This has not been tested on Norwegian yet. The gender marking on the adjectives differ a bit between Norwegian and Dutch and German. In Dutch and German, the gender-marked suffix adds an additional syllable to the stem, *rood/rode*, *groen/groene* in Dutch, and *rote, roter, rotes/grüne, grüner, grünes* in German. In Norwegian it is *rød/rødt, grønn/grønt.* This difference is another reason for testing an Adj + N naming task.

Conclusions

In conclusion, a gender congruency effect was obtained for subjects who had three genders in their dialect. The findings from Experiment 1 suggest that the gender interference effect occurs at the level of gender node selection. The findings contradict the Determiner selection interference hypothesis (DSIH) put forward by Schiller and Caramazza (2003) where retrieval of a noun's gender is thought to be an automatic consequence of selecting the noun's lexical node, and not open for competition. In addition, the results from the experiments also showed that subjects with a three gender system and those with a two gender system, should not be

---

[11] This may vary a bit in speech depending on the dialect.

mixed in an experiment incautiously.

The simulation results from the computational model showed a more or less good profile fit with the real data, though the WEAVER exaggerates the time difference compared to the experimental results form Experiment 1.

Why a clear gender congruency effect has not been found for the two-gender group is unknown. It is predicted by the WEAVER and is a robust finding in Dutch which also has a two gender system.

REFERENCES

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: evidence from French. *Cognition, 82, 179-223.*

Cacciari, C., & Cubelli, R. (2003). The neuropsychology of grammatical gender: An introduction. *Cortex, 39, Editorial.*

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14, 177-208.*

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition 64, 309-343.*

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X (2001). A cross-linguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (pp 209-226). Cambridge, MA: MIT Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A Critique of Language statistics in Psychological research. *Journal of Verbal Learning and Verbal Behavior, 12, 335-359.*

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

Costa, A., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003). The gender congruency effect and the selection of freestanding and bound morphemes: Evidence from Croatian. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 1270-1282.*

Dell,G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93, 283-321.*

Dell,G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42, 287-314.*

Dell, G. S., & Juliano, C. (1996). In Dijkstra, T., & de Smedt, K., (Eds), *Computational Psycholinguistic*s. Taylor & Francis (pp 328-359).

De Smedt, K. (1996). In Dijkstra, T., & de Smedt, K. (Eds). *Computational Psycholinguistics.* Taylor & Francis, (pp 279-307).

Dijkstra, T., & de Smedt, K. (1996). In Dijkstra, T., & de Smedt, K. (Eds). *Computational Psycholinguistics.* Taylor & Francis, (pp 23-21).

Faarlund, J. T., Lie, S., & Vannebo, K. I. (1997). *Norsk referansegrammatikk.* Universitetsforlaget AS.

Glaser, W. R. (1992). Picture naming. *Cognition, 42, 61-105.*

Glaser, W. R., & Düngelhoff, F-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10, 640-654.*

Harley, Trevor (2001). *The psychology of language: From data to theory*. Psychology Press Ltd.

Johansson, C. (2003). How is grammatical gender processed? In Tanja Gaustad (Ed.): Computational Linguistics in the Netherlands 2002, in the series *Language and Computers: Studies in practical linguistics*, 47, (pp 65-76), Rodopi, Amsterdam-New York.

La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research, 61, 209-219.*

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Schriefers, H., Meyer, A. S., Pechmann, T., Vorberg, D., & Havinga, J. (1991). The time course of lexical access in speech production: A study of Picture naming. *Psychological Review, 98, 122-142.*

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3, 223-232.*

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

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *PNAS, 98, 13464-13471.*

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition, 7, 485-495.*

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin 109, 163-203.*

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of experimental psychology: Learning, Memory and Cognition, 25, 907-922.*

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of experimental psychology; General, 132, 228-252.*

Müller, O., & Hagoort, P. (submitted). Priming access of grammatical gender in language comprehension.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41, 416-426.*

Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition, 42, 107-142.*

Roelofs, A. (1996). In Dijkstra, T., & de Smedt, K. (Eds). *Computational Psycholinguistics.* Taylor & Francis, (pp 308-325).

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling Attentional control in the Stroop Task. *Psychological Review, 110, 88-125.*

Roelfos, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64, 249-284.*

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of memory and language, 48, 169-194.*

Schwichtenberg, B., & Schiller, N.O. (2004). Semantic gender assignment regularities in German. *Brain and Language, 90, 326-337.*

Schriefers, H. (1993). Syntactic processes in the Production of Noun Phrases. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19, 813-840.*

Schriefers, H. & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of psycholinguistic research, 28, 575-600.*

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of experimental psychology: Learning, Memory and Cognition, 26, 1368-1377.*

Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner Selection in Noun Phrase Production. *Journal of experimental psychology: Learning, Memory and Cognition, 28, 941-950.*

Starreveld, P. A.,& La Heij, W. (1995). Semantic interference, Orthographic facilitation, and their interaction in naming tasks. *Journal of experimental psychology: Learning, memory and cognition, 21, 699-710.*

Stroop, J. R. (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology, 18, 1935. Reprint in Journal of Experimental Psychology: General, 1992, 121, 15-23.*

Torp, A., & Vikør, L. S. (1993). *Hovuddrag i norsk språkhistorie.* Ad Notam Gyldendal AS.

Trosterud, T. (2001). Genustilordning i norsk er regelstyrt. *Norsk Lingvistisk Tidsskrift, 19, 29-57.*

Underwood, G., & Briggs, P. (1984). The development of word recognition processes. *British journal of psychology, 75, 243-255.*

Van Berkum, J. J. A. (1997). Syntactic processes in speech production: the retrieval of grammatical gender. *Cognition, 64, 115-152.*

Appendix A: The material used in Experiment 1.

Appendix B: The material used in Experiment 3.

Appendix C: Electronic appendix, survey over *ball* as neuter or masculine word.

Appendix D: Electronic appendix, Experiment 1 and 3.

Appendix E: Electronic appendix, results from the experiments.

Appendix F: Electronic appendix, the simulation codes.

APPENDIX A

SEMANTICALLY RELATED CONDITION

Target picture		Distractor conditions			
		Congruent		Incongruent	
MASCULINE					
benk('bench')	masc	stol('chair')	masc	dør('door')	fem
hund('dog')	masc	gris('pig')	masc	mus('mouse')	fem
ost('cheese')	masc	rømme('sour cream')	masc	kake('cake')	fem
vott('mitten')	masc	sokk('sock')	masc	bukse('trousers')	fem
hatt('hat')	masc	sko('shoe')	masc	jakke('jacket')	fem
kamel('camel')	masc	sebra('zebra')	masc	geit('goat')	fem
mur('brick wall')	masc	stige('ladder')	masc	trapp('stairs')	fem
hare('hare')	masc	bjørn('bear')	masc	gaupe('lynx')	fem
fot('foot')	masc	munn('mouth')	masc	tunge('tongue')	fem
kopp('cup')	masc	asjett('tea plate')	masc	flaske('bottle')	fem
hanske('glove')	masc	knapp('button')	masc	belte('belt')	neu
kanon('canon')	masc	bombe('bomb')	masc	sverd('sward')	neu
bil('car')	masc	sykkel('bicycle')	masc	tog('train')	neu
tomat('tomato')	masc	agurk('cucumber')	masc	eple('apple')	neu
arm('arm')	masc	finger('finger')	masc	hode('head')	neu
is('ice cream')	masc	kjeks('biscuit')	masc	brød('bread')	neu
fabrikk('factory')	masc	låve('barn')	masc	slott('castle')	neu
festning('fortress')	masc	bunker('bunker')	masc	tempel('temple')	neu
behå('bra')	masc	støvel('boot')	masc	skjerf('scarf')	neu
vase('vase')	masc	duk('table cloth')	masc	speil('mirror')	neu
FEMININE					
dør('door')	fem	køye ('berth')	fem	skap('closet')	neu
mus('mouse')	fem	gås('goose')	fem	lam('lamb')	neu
kake('cake')	fem	lefse('pastry_thing')	fem	smør('butter')	neu
bukse('trousers')	fem	trøye('shirt')	fem	skjørt('skirt')	neu
jakke('jacket')	fem	hette('hood')	fem	slips('tie')	neu
geit('goat')	fem	ku('cow')	fem	esel('donkey')	neu
trapp('stairs')	fem	søyle('pillar')	fem	vindu('window')	neu
gaupe('lynx')	fem	ugle('owl')	fem	ekorn('squirrel')	neu
tunge('tongue')	fem	hånd('hand')	fem	øye('eye')	neu
flaske('bottle')	fem	gryte('casserole')	fem	glass('glass')	neu
veske('purse')	fem	kappe('cloak')	fem	hanske('glove')	masc
hagle('shotgun')	fem	lanse('lance')	fem	kanon('canon')	masc
vogn('wagon')	fem	ferge('ferry')	fem	bil('car')	masc
pære('pear')	fem	drue('grape')	fem	tomat('tomato')	masc
tå('toe')	fem	hake('chin')	fem	arm('arm')	masc
nøtt('nut')	fem	plomme('plum')	fem	is('icecream')	masc

kirke('church')	fem	rønne('shack')	fem	fabrikk('factory	masc
bru('bridge')	fem	havn('harbour')	fem	festning('fortress	masc
truse('knickers')	fem	kåpe('coat')	fem	behå('bra')	masc
lampe('lamp')	fem	hylle('shelf')	fem	vase('vase')	masc
NEUTER					
skap('closet')	neu	teppe('carpet')	neu	benk('bench')	masc
lam('lamb')	neu	føll('foal')	neu	hund('dog')	masc
smør('butter')	neu	krydder('spice')	neu	ost('cheese	masc
skjørt('skirt')	neu	forkle('apron')	neu	vott('mitten')	masc
slips('tie')	neu	tørkle('headscarf')	neu	hatt('hat')	masc
esel('donkey')	neu	rådyr('roe')	neu	kamel('camel')	masc
vindu('window')	neu	loft('attic')	neu	mur('brick wall')	masc
ekorn('squirrel')	neu	lemen('lemming')	neu	hare('hare')	masc
øye('eye')	neu	bryst('chest')	neu	fot('foot')	masc
glass('glass')	neu	beger('goblet')	neu	kopp('cup')	masc
belte('belt')	neu	sjal('scarf')	neu	veske('purse')	fem
sverd('sward')	neu	gevær('gun')	neu	hagle('shotgun')	fem
tog('train')	neu	skip('ship')	neu	vogn('wagon')	fem
eple('apple')	neu	jordbær('strawberry')	neu	pære('pear')	fem
hode('head')	neu	kne('knee')	neu	tå('toe')	fem
brød('bread')	neu	korn('grain')	neu	nøtt('nut')	fem
slott('castle')	neu	fjøs('cowshed')	neu	kirke('church')	fem
tempel('temple')	neu	naust('boat house')	neu	bru('bridge')	fem
skjerf('scarf')	neu	erme('sleeve')	neu	truse('knickers')	fem
speil('mirror')	neu	pledd('blanket')	neu	lampe('lamp')	fem

SEMANTICALLY UNRELATED CONDITION

Target picture		Distractor conditions			
		Congruent		Incongruent	
MASCULINE					
benk('bench')	masc	sjø('ocean')	masc	tog('train')	neu
hund('dog')	masc	ball('ball')	masc	glass('glass')	neu
ost('cheese')	masc	kniv('knife')	masc	trapp('stairs')	fem
vott('mitten')	masc	katt ('cat')	masc	geit('goat')	fem
hatt('hat')	masc	vind('wind')	masc	brød('bread')	neu
kamel('camel')	masc	billett('ticket')	masc	jakke('jacket')	fem
mur('brick wall')	masc	dal('valley')	masc	hode('head')	neu
hare('hare')	masc	pose('pose')	masc	veske('purse')	fem
fot('foot')	masc	ovn('oven')	masc	lampe('lamp')	fem
kopp('cup')	masc	sekk('sack')	masc	øye('eye')	neu
hanske('glove')	masc	peis('fire place')	masc	sverd('sward')	neu
kanon('canon')	masc	ape('monkey')	masc	dør('door')	fem
bil('car')	masc	rygg('back')	masc	kake('cake')	fem
tomat('tomato')	masc	koffert('suitcase')	masc	bukse('trousers')	fem

arm('arm')	masc	elg('moose')	masc	speil('mirror')	neu
is('icecream')	masc	gård('farm')	masc	flaske('bottle')	fem
fabrikk('factory')	masc	kassett('casette')	masc	tunge('tongue')	fem
festning('fortress')	masc	pensel('painting brush')	masc	ekorn('squirrel')	neu
behå('bra')	masc	hane('rooster')	masc	tempel('temple')	neu
vase('vase')	masc	kjeller('cellar')	masc	skjerf('scarf')	neu

FEMININE

dør('door')	fem	myr('marsh')	fem	skjørt('skirt')	neu
mus('mouse')	fem	sol('sun')	fem	vindu('window')	neu
kake('cake')	fem	rumpe('bottom')	fem	vase('vase')	masc
bukse('trousers')	fem	ramme('frame')	fem	smør('butter')	neu
jakke('jacket')	fem	kasse('box')	fem	eple('apple')	neu
geit('goat')	fem	bukt('bay')	fem	slips('tie')	neu
trapp('stairs')	fem	vik('cove')	fem	lam('lamb')	neu
gaupe('lynx')	fem	brygge('dock')	fem	skap('closet')	neu
tunge('tongue')	fem	hylle('shelf')	fem	ost('cheese')	masc
flaske('bottle')	fem	kiste('coffin')	fem	belte('belt')	neu
veske('purse')	fem	stang('stick')	fem	mur('brick wall')	masc
hagle('shotgun')	fem	loppe('flea')	fem	is('ice cream')	masc
vogn('wagon')	fem	ørn('eagle')	fem	slott('castle')	neu
pære('pear')	fem	snekke('fishing boat')	fem	kanon('canon')	masc
tå('toe')	fem	pil('arrow')	fem	vott('mitten')	masc
nøtt('nut')	fem	hytte('cottage')	fem	hund('dog')	masc
kirke('church')	fem	ski('ski')	fem	benk('bench')	masc
bru('bridge')	fem	strand('beach')	fem	kopp('cup')	masc
truse('knickers')	fem	bjelle('bell')	fem	esel('donkey')	neu
lampe('lamp')	fem	krabbe('crab')	fem	hanske('glove')	masc

NEUTER

skap('closet')	neu	hår('hair')	neu	bil('car')	masc
lam('lamb')	neu	språk('language')	neu	arm('arm')	masc
smør('butter')	neu	fjell('mountain')	neu	bru('bridge')	fem
skjørt('skirt')	neu	lik('corps')	neu	mus('mouse')	fem
slips('tie')	neu	dikt('poem')	neu	fabrikk('factory')	masc
esel('donkey')	neu	flagg('flag')	neu	tå('toe')	fem
vindu('window')	neu	hjul('wheel')	neu	fot('foot')	masc
ekorn('squirrel')	neu	tårn('tower')	neu	hatt('hat')	masc
øye('eye')	neu	bad('bath')	neu	gaupe('lynx')	fem
glass('glass')	neu	frø('seed')	neu	nøtt('nut')	fem
belte('belt')	neu	stempel('stamp')	neu	hare('hare')	masc
sverd('sward')	neu	plaster('band aid')	neu	kamel('camel')	masc
tog('train')	neu	jorde('field')	neu	behå('bra')	masc
eple('apple')	neu	anker('anchor')	neu	festning('fortress')	masc
hode('head')	neu	sirkus('cirkus')	neu	tomat('tomato')	masc
brød('bread')	neu	spyd('spear')	neu	kirke('church')	fem
slott('castle')	neu	nes('headland')	neu	truse('knickers')	fem
tempel('temple')	neu	gebiss('fake teeth')	neu	hagle('shotgun')	fem
skjerf('scarf')	neu	kors('cross')	neu	vogn('wagon')	fem
speil('mirror')	neu	arkiv('archive')	neu	pære('pear')	fem

APPENDIX B

SEMANTICALLY RELATED CONDITION

Target picture		Distractor conditions			
		Congruent		Incongruent	
MASCULINE					
benk('bench')	masc	stol('chair')	masc	skap('closet')	neu
hund('dog')	masc	gris('pig')	masc	lam('lamb')	neu
ost('cheese')	masc	rømme('sour cream')	masc	smør('butter')	neu
vott('mitten')	masc	sokk('sock')	masc	skjørt('skirt')')	neu
hatt('hat')	masc	sko('shoe')	masc	slips('tie')	neu
kamel('camel')	masc	sebra('zebra')	masc	esel('donkey')	neu
mur('brick wall')	masc	stige('ladder')	masc	vindu('window')	neu
kanin('rabbit')	masc	hare ('hare')	masc	ekorn('squirrel')	neu
fot('foot')	masc	munn('mouth')	masc	øye('eye')	neu
kopp('cup')	masc	asjett('tea plate')	masc	glass('glass')	neu
hanske('glove')	masc	knapp('button')	masc	belte('belt')	neu
kanon('canon')	masc	pistol('pistol')	masc	sverd('sward')	neu
bil('car')	masc	sykkel('bicyckle')	masc	tog('train')	neu
tomat('tomato')	masc	agurk('cucumber')	masc	eple('apple')	neu
arm('arm')	masc	finger('finger')	masc	hode('head')	neu
is('icecream')	masc	kjeks('biscuit')	masc	brød('bread')	neu
fabrikk('factory')	masc	låve('barn')	masc	slott('castle')	neu
festning('fortress')	mas	bunker('bunker')	masc	hus('house')	neu
behå('bra')	masc	støvel('boot')	masc	skjerf('scarf')	neu
vase('vase')	masc	duk('table cloth')	masc	speil('mirror')	neu
NEUTER					
skap('closet')	neu	teppe	neu	benk('bench')	masc
lam('lamb')	neu	føll('foal')	neu	hund('dog')	masc
smør('butter')	neu	krydder('spice')	neu	ost('cheese')	masc
skjørt('skirt')	neu	forkle('apron')	neu	vott('mitten	masc
slips('tie')	neu	tørkle('headscarf')	neu	hatt('hat')	masc
esel('donkey')	neu	rådyr('roe')	neu	kamel('camel')	masc
vindu('window')	neu	tak('roof/ceiling')	neu	mur('brick wall')	masc
ekorn('squirrel')	neu	lemen('lemming')	neu	kanin('rabbit')	masc
øye('eye')	neu	bryst('chest')	neu	fot('foot')	masc
glass('glass')	neu	beger('goblet')	neu	kopp('cup')	masc
belte('belt')	neu	sjal('scarf')	neu	hanske('glove')	masc
sverd('sward')	neu	gevær('gun')	neu	kanon('canon')	masc
tog('train')	neu	skip('ship')	neu	bil('car')	masc
eple('apple')	neu	jordbær('strawberry')	neu	tomat('tomato')	masc
hode('head')	neu	kne('knee')	neu	arm('arm')	masc
brød('bread')	neu	korn('grain')	neu	is('icecream')	masc

slott('castle')	neu	fjøs('cowshed')	neu	fabrikk('factory')	masc
hus('house')	neu	naust('boat house')	neu	festning('fortress')	masc
skjerf('scarf')	neu	erme('sleeve')	neu	behå('bra')	masc
speil('mirror')	neu	pledd('blanket')	neu	vase('vase')	masc

SEMANTICALLY UNRELATED CONDITION

Target picture		Distractor conditions			
		Congruent		Incongruent	
MASCULINE					
benk('bench')	masc	sjø('ocean')	masc	hode('head')	neu
hund('dog')	masc	løk('onion')	masc	glass('glass')	neu
ost('cheese')	masc	kniv('knife')	masc	skjørt('skirt')	neu
vott('mitten')	masc	katt ('cat')	masc	brød('bread')	neu
hatt('hat')	masc	vind('wind')	masc	vindu('window')	neu
kamel('camel')	masc	billett('ticket')	masc	smør('butter')	neu
mur('brick wall')	masc	dal('valley')	masc	tog('train')	neu
kanin('rabbit')	masc	pose('pose')	masc	eple('apple')	neu
fot('foot')	masc	ovn('oven')	masc	slips('tie')	neu
kopp('cup')	masc	rygg('back')	masc	øye('eye')	neu
hanske('glove')	masc	peis('fire place')	masc	sverd('sward')	neu
kanon('canon')	masc	gaffel('fork')	masc	lam('lamb')	neu
bil('car')	masc	sekk('sack')	masc	skap('closet')	neu
tomat('tomato')	masc	koffert('suitcase')	masc	belte('belt')	neu
arm('arm')	masc	elg('moose')	masc	speil('mirror')	neu
is('icecream')	masc	gård('farm')	masc	slott('castle')	neu
fabrikk('factory')	masc	kassett('casette')	masc	esel('donkey')	neu
festning('fortress')	masc	pensel('painting brush')	masc	ekorn('squirrel')	neu
behå('bra')	masc	vegg('wall')	masc	hus('house')	neu
vase('vase')	masc	kjeller('cellar')	masc	skjerf('scarf')	neu
NEUTER					
skap('closet')	neu	hår('hair')	neu	bil('car')	masc
lam('lamb')	neu	språk('language')	neu	arm('arm')	masc
smør('butter')	neu	fjell('mountain')	neu	fot('foot')	masc
skjørt('skirt')	neu	lik('corps')	neu	benk('bench')	masc
slips('tie')	neu	dikt('poem')	neu	fabrikk('factory')	masc
esel('donkey')	neu	flagg('flag')	neu	mur('brick wall')	masc
vindu('window')	neu	hjul('wheel')	neu	ost('cheese')	masc
ekorn('squirrel')	neu	tårn('tower')	neu	hatt('hat')	masc
øye('eye')	neu	bad('bath')	neu	is('ice cream')	masc
glass('glass')	neu	frø('seed')	neu	kanon('canon')	masc
belte('belt')	neu	stempel('stamp')	neu	vase('vase')	masc
sverd('sward')	neu	plaster('band aid')	neu	kamel('camel')	masc
tog('train')	neu	jorde('field')	neu	behå('bra')	masc

eple('apple')	neu	anker('anchor')	neu	festning('fortress')	masc
hode('head')	neu	sirkus('cirkus')	neu	tomat('tomato')	masc
brød('bread')	neu	spyd('spear')	neu	vott('mitten')	masc
slott('castle')	neu	fat('plate')	neu	hund('dog')	masc
hus('house')	neu	gebiss('fake teeth')	neu	kanin('rabbit')	masc
skjerf('scarf')	neu	kors('cross')	neu	kopp('cup')	masc
speil('mirror')	neu	arkiv('archive')	neu	hanske('glove')	masc