Uncommon gender

Loerts, Hanneke

Publication date:
2012

UNCOMMON GENDER

Eyes and brains, native and second language learners,
& grammatical gender

Hanneke Loerts

bcn

university of groningen
faculty of arts
CLCG

The work presented in this thesis has been carried out under the auspices of the School for Behavioral and Cognitive Neurosciences (BCN) and the Center for Language and Cognition Groningen (CLCG).

Groningen Dissertations in Linguistics 110
ISBN: 978-90-367-5882-6 (printed version)
ISBN: 978-90-367-5883-3 (electronic version)

Printed by: Off Page, Amsterdam, The Netherlands – www.offpage.nl

RIJKSUNIVERSITEIT GRONINGEN

UNCOMMON GENDER

Eyes and brains, native and second language learners, & grammatical gender

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
donderdag 6 december 2012
om 11:00 uur

door

Hanneke Loerts

geboren op 10 december 1983
te Leeuwarden

Promotor: Prof. dr. M.S. Schmid

Copromotor: Dr. L.A. Stowe

Beoordelingscommissie: Prof. dr. P.E. Dussias
Prof. dr. C. Frenck-Mestre
Prof. dr. P. Hendriks

ISBN: 978-90-367-5882-6

CONTENTS

List of ABBREVIATIONS

AJT	Acceptability Judgement Task
AoA	Age of Acquisition
AOI	Area of Interest
COM	Common gender
CP(H)	Critical Period (Hypothesis)
DP	Declarative/procedural
EEG	Electroencephalography
ERP	Event-related potential(s)
FEM	Feminine gender
FFFH	Failed Functional Features Hypothesis
FTFA	Full Transfer Full Access
GJT	Grammaticality Judgement Task
ICC	Intraclass Correlation Coefficient
L1	First Language
L2	Second Language
L3	Third Language
LAD	Language Acquisition Device
LoR	Length of Residence
MASC	Masculine gender
ms	Milliseconds
NA	Not available
NEU	Neuter gender
NP	Noun phrase
PL	Plural
ROI	Region of Interest
RT	Reaction times
SD	Standard deviation
SSH	Shallow Structures Hypothesis
SVO	Subject Verb Object
UG	Universal Grammar

List of TABLES

List of
FIGURES

Chapters to be published

IN PEER REVIEWED JOURNALS

Chapter 4
Loerts, H., Wieling, M., & Schmid, M.S. (Under Review). Neuter is not Common in Dutch: Eye Movements reveal Asymmetrical Gender Processing. Submitted (February 29, 2012) to *Journal of Psycholinguistic Research.*

Chapter 5
Loerts, H., Stowe, L.A., & Schmid, M.S. (Under Review). Predictability Speeds up the Re-analysis Process: An ERP investigation of gender agreement and cloze probability. Submitted (June 4, 2012) to *Journal of Neurolinguistics.*

Chapter 6
Loerts, H. (Under Review). Can Late L2 learners of Dutch use Gender Cues during Spoken Word Recognition? Evidence from Eye Movements. Submitted (July 11, 2012) to *Applied Psycholinguistics.*

Chapter 7
Loerts, H., Stowe, L.A., & Schmid, M.S. (Under Review). Enough Time for Gender: Morphological Processing in Late Second Language Learners of Dutch. Submitted (July 27, 2012) to *Bilingualism: Language and Cognition.*

Chapter 1

GENERAL INTRODUCTION

Anyone who has ever started learning a new language at a later stage in life knows that it is a difficult and frustrating process that stands in stark contrast to the ease with which they have acquired their mother tongue. Given enough time and exposure, most late learners will eventually become quite fluent in their second language (henceforth L2), but they hardly ever become indistinguishable from native speakers. One linguistic feature that appears to remain problematic, even after years of exposure, is grammatical gender. When acquiring a gendered language, the learner is faced with the task of assigning a gender category to a noun that often does not have any gender characteristics. An L2 learner of Dutch has to grasp the fact that the neuter determiner *'het'* is used when referring to a knife in Dutch ('het$_{NEU}$ mes$_{NEU}$'), while a spoon has to be preceded by the common determiner *'de'* ('de$_{COM}$ lepel$_{COM}$'). Even if the person's mother tongue has gender, this might initially not be of any help at all: this exact same spoon is masculine in German ('der$_{MASC}$ Löffel$_{MASC}$') while it is assigned feminine gender in Spanish ('la$_{FEM}$ cuchara$_{FEM}$'). Is it possible to ever fully acquire such an arbitrary system in a language other than your own?

The answer to this question might be highly dependent on the age at which a person started learning the L2. In immigrant families, for example, children are often observed to acquire the new language faster than their parents. One explanation for this difference is that we are born with an innate capacity to learn, or maybe even an innate capacity to learn language. This ability declines as we grow older and hence the learner has to rely on different, compensatory strategies when acquiring a new language at a later age. Consequently, the later learned language might be represented and/or processed in a different way by the late L2 learner than a language that has been acquired from birth or at an early stage in life. Age is, however, not the only crucial difference between the immigrant child and his or her parents. The child is often immersed in the language in school and forced to use the new language with his or her peers, while the parents continue using their mother tongue at home. Some researchers suggest that adult L2 learners, such as these immigrant parents, lag behind their children mainly due to the reduced amount of input and experience with the new language. Additionally, late L2 learners have already fully acquired a language and this mother tongue is likely to become more deeply entrenched as we grow older. The late L2 learner thus has to suppress his or her mother tongue while producing or comprehending the new language. The increased strain on resources and/or the interference from the native language causes late L2 learners to produce speech errors and respond and process the L2 more slowly as compared to native speakers, but the underlying representations and/or processing mechanisms might be fundamentally the same as those employed by native speakers. More on the different views on why and how L1 and L2 acquisition differ follows in §1.2. below.

The influence of the native language and the amount of input from the second language is especially interesting when it comes to the acquisition of abstract grammatical gender systems, such as the Dutch two-way gender system. As compared to many Romance languages, the Dutch gender system is relatively arbitrary and hardly any cues are available to determine whether a noun has common or neuter gender (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997). The nouns *'kluis'* and *'huis'* ('safe' and 'house'), for example, are almost identical except for their initial phonemes. Nevertheless, the word *'kluis'* is common and *'huis'* is neuter and hence these have to be preceded by the common and neuter definite determiners *'de'* and *'het'* respectively. The definite determiners are, along with adjectives and pronouns, the only clear sign to the (late L2) learner that nouns have different genders in the Dutch language. To further complicate the learning process, these signs are not consistently present in the input. First of all, common gender evolved from the collapse of the original masculine and feminine genders and consequently comprises around 75% of all Dutch nouns (Van Berkum, 1996). Furthermore, while the bare Dutch adjective *'mooi'* ('beautiful') has to be used when modifying an indefinite neuter noun, as in *'een mooi huis'*, the adjective is followed by a non-salient schwa-ending when modifying an indefinite common noun, as in *'een mooie kluis'*. The latter form of the adjective, however, can precede both common and neuter nouns when definite: *'het mooie huis'* and *'de mooie kluis'*. The non-transparency, asymmetry and relatively scarce evidence in the input for neuter gender causes monolingual as well as bilingual children to overgeneralize the definite, common determiner *'de'* until a relatively late age. Native Dutch children have been reported to overuse *'de'* until the age of 6 (Blom, Polisenskà, & Weerman, 2008; Van der Velde, 2004) and bilingual children, even when exposed to Dutch from birth, continue to make such errors beyond this age (Blom et al., 2008). Dutch appears to be a special case in this respect, as will be further explained in §1.1. The main question addressed in the present dissertation is whether late L2 learners, who started learning the new language after puberty, can ever reach native-like knowledge of the Dutch gender system as well as whether they are able to use and process gender similar to native speakers.

Research concerning (late) L2 acquisition of grammatical gender has thus far yielded inconclusive results. Researchers are still not out on whether late L2 learners are essentially the same or fundamentally different from native speakers. Moreover, they do not agree on whether it is the knowledge that differs (fundamentally or partially) or whether it is the processing that differs (fundamentally or partially). In other words, is the grammatical (gender) system differently represented in the brain or is it simply not used in the same way during production and comprehension. The variation in results is, at least partly, caused by the different measures used in these studies. Furthermore, many studies have focussed on measures of competence and/or knowledge, such as, e.g. reaction times related to acceptability judgements in response to sentences that are grammatical or ungrammatical. While these studies have increased our understanding of the process of language comprehension, they have not been able to accurately answer questions about the underlying processes of language comprehension and the relation between the different components that can be related to these processes. The elapsed time between listening to and deciding whether a sentence contains an error of some sort tells us how long it took a person to decide whether a sentence contains an error. It does not tell us anything about what kinds of information in the input were used at what point in time to come to this conclusion, let alone how long these possibly separable processes took and whether they influenced each other in any way. Moreover, if a late L2 learner responds later as compared to a native speaker in response to the ungrammatical phrase *'*de huis'* ('the$_{\text{COM}}$ house$_{\text{NEU}}$'), this does not necessarily mean that the underlying processes preceding the response differ from those processes of the native speaker.

To be able to examine what types of information are used at what point in time during language comprehension, online measures are needed such as eye tracking and event-related potentials. Since our eyes tend to move towards objects that we think correspond most closely to what we hear in the speech stream, eye movements can be a powerful tool to examine what kind of (linguistic) information is used to comprehend speech as it unfolds in time. By recording brain activity while subjects are listening to sentences that are either correct or incorrect, we can gather information about how people process a specific linguistic feature as well as when this is being processed.

The present thesis focusses on the real-time processes that occur while people are listening to their native or second language and, in particular, to the subtle grammatical gender cues that are hidden in the speech stream. The experiments presented in this dissertation are an attempt to combine behavioural and online data to come to a deeper understanding of the incremental nature of language processing in Dutch natives and late L2 learners of Dutch. In several experiments, late L2 learners of Dutch from a gendered first language (i.e. L1 Polish), will be compared to native Dutch speakers to examine whether (a) native Dutch speakers are able to use gender to facilitate language comprehension and whether they process gender in the same way as other morphological structures and whether (b) late, advanced learners of Dutch process gender in the L2 in a similar way as compared to natives and/or if they tend to use gender information from their mother tongue instead.[1] This introductory chapter provides the theoretical and methodological basis for the experiments which will be explained in more detail in Chapter 2. The next section will start out by explaining the notion of grammatical gender and its curious manifestations in the Dutch language (§1.1) followed by an outline of the most important findings concerning the acquisition of gender in the native and in the second language. In section 1.2, relevant linguistic theories on (second) language acquisition will be discussed and related to the phenomenon of grammatical gender after which section 1.3 focusses on the rationale behind the measures used in the experiments. The theoretical and methodological outline forms the basis for the main research questions and hypotheses as discussed in §1.4. Finally, section 1.5 provides an overview of the remainder of this dissertation.

1.1. Grammatical gender and its curious manifestations in Dutch

What is it about a spoon that makes the people in Germany regard it as masculine while Spanish people refer to it as if it was feminine? The short and simple answer to this question is that there is no connection between the word for spoon and any male or female characteristics of its referent in the real world. Gender is simply a term that derives from the Latin word *la genus*, originally meaning class, kind, variety, or sort. According to Aristotle, we can blame Protagoras for this bad choice in labels, because in the fifth century BC, he was said to have divided Greek nouns into masculine, feminine, and neuter categories (Aristotle, trans 1991, 1407b). Most researchers have a hard time believing that this man actually sat down to determine the gender class of each and every noun in the language, but linguists from all disciplines have ever since struggled, and mostly failed, to explain the relationship between grammatical gender and the world surrounding us.

Researchers do agree that grammatical gender is a noun classification system that is present is many, though not all, languages of the world. Depending on the language, this classification system divides nouns into several different categories, such as, e.g. masculine, feminine, and neuter. Most speakers (i.e. non-linguists) of a gendered language do not consciously know the genders of the words in their language, but can automatically assign the correct gender to the noun. Thus, Dutch natives effortlessly talk about '*het huis*' ('the_{FEM} $\text{house}_{\text{FEM}}$') and '*de auto*' ('$\text{the}_{\text{COM}}$ car_{COM}'), often without realizing that the determiner carries gender information that matches the gender of the following noun. Often only when confronted with misuse of these articles, as happens while listening to an L2 speaker of Dutch, do natives realize that there is some form of gender agreement in their language.

---

[1] The main focus throughout the dissertation will be on late L2 learners. Data from simultaneous bilinguals was also gathered to compare early and late acquisition of L2 Dutch gender. Due to several insurmountable reasons, however, too few simultaneous bilinguals were tested in order to allow sound statistical analyses as reported in Chapters 6 and 7. Nevertheless, the data from the group of simultaneous bilinguals was used whenever that had added value, as, e.g. in Chapter 3.

The difficulty of acquiring an L2 gender system can partly be related to the absence of gender systems in some languages as well as the difference between gender systems across languages. When an English native speaker starts learning French, he or she has to add a grammatical category, i.e. the fact that some nouns are masculine and require *'le'* as an article and other nouns require the feminine article *'la'* (instead of all nouns receiving *'the'*). In addition, if a native French person is learning Dutch, he or she cannot simply copy the gender characteristics present in the L1 onto the L2, because '*la maison*' (the house) is feminine in French, but neuter in Dutch ('het_{NEU} huis_{NEU}').

In linguistic terms, (grammatical) gender can be defined as 'a system in which the class to which a noun is assigned is reflected in the forms that are taken by other elements syntactically related to it' (Matthews, 1997:36). This definition brings us to another important distinction to be made: gender assignment versus gender agreement. Gender is a specific inherent feature of a noun (e.g. '*huis'* (house) is assigned neuter gender in Dutch), but the words surrounding the noun have to agree in gender with that noun (e.g. '*huis'* receives the neuter article *'het'*). The second part of Matthews' definition (1997), stating that the noun's gender 'is reflected in the forms that are taken by other elements syntactically related to it', refers to this gender agreement system, which is governed by rules. All neuter words in Dutch, for example, require the neuter definite determiner *'het'*. Depending on the language, words that refer to the noun may receive suffixes or affixes agreeing with the noun's gender. These agreement targets can include determiners, adjectives, relative pronouns and even verbs, and agreement can occur within as well as outside of the noun phrase (NP). This rule-system, which is also called agreement or concord, differs per language in that languages can have different agreement targets as well as different realizations of the agreement relationship. The relationship between the gender-carrying noun and its syntactically related words comprises a syntactic phenomenon, because the agreement targets through which it shows up can only be defined in terms of their syntactic category and their syntactic relationship to the controller noun (Van Berkum, 1996). Gender is also a morphological phenomenon, because the agreement is marked by inflectional devices, such as suffixes, prefixes, infixes, and/or suppletion (Van Berkum, 1996). Before being able to apply these agreement rules, however, one has to know to which of the genders a noun belongs. A problem here might be that the classes or genders that nouns are assigned to, the first part of Matthews' definition, does not seem to be straightforward in all gendered languages.

Of course, the classification of genders sometimes corresponds to the biological gender: e.g. '$\textit{die}_{\text{FEM}}$ $\textit{Frau}_{\text{FEM}}$' (the woman) and '$\textit{der}_{\text{MASC}}$ $\textit{Mann}_{\text{MASC}}$' (the man) in German. This type of gender is often referred to as natural or biological gender. Biological sex, however, seems completely unrelated to the linguistic gender of other nouns, such as inanimate objects: *'die Gabel'* (the fork) is feminine in German, whereas *'der Loffel'* (the spoon) is a masculine word that even has a feminine counterpart, i.e. *'die Kelle'* (the ladle). Scientists have long tried to determine whether gender assignment is arbitrary or, to a certain degree, rule-governed. Van Berkum (1996) rightly points out the importance of this question within psycholinguistics: if it is completely arbitrary, people would have to memorize the gender of each and every noun in a language, but if there is some systematicity present, we may be able to exploit it. Many researchers initially agreed with Bloomfield's well-known quote (1933) stating that gender categories 'do not agree with anything in the practical world' and that 'there seems to be no practical criterion by which the gender of a noun in German, French, or Latin could be determined'. Recent investigations have shown, however, that gender assignment does not seem to be a completely arbitrary feature.

Corbett (1991) studied a variety of assignment systems of languages around the world and concluded that gender assignment is essentially systematic in every language that has a gender system. An important distinction, however, is made between languages with an overt and languages with a covert gender system. In an overt gender system, the noun's gender is reflected in the morphology of its agreement targets as well as in the form of the noun itself. In many Romance languages, this is based on phonological aspects, e.g. Spanish nouns ending with a suffix '*–a*' are mostly feminine (*'la señora'*: madam), whereas Spanish nouns ending in '*–o*' are usually masculine (*'el marido'*: husband). In other languages, such as French, this relationship is less obvious and based on more abstract rules: e.g. nouns that end with the suffix '*–ion*' generally receive the feminine gender (Foucart, 2008). When acquiring more or less overt gender systems, one can first acquire the assignment system, i.e. learn which nouns belong to which category, and then apply the appropriate rules for agreement. In a covert gender system, however, gender has to be marked on the agreement targets, but it cannot be seen from the noun itself. Gender assignment in a covert system can thus only be acquired via the gender agreement system. This might be one of the main reasons why the covert Dutch gender system is so extremely difficult to acquire.

1.1.1. Grammatical gender in Dutch

Dutch nouns have one of two grammatical genders: common and neuter (Van Berkum, 1996). Common gender has evolved from the original distinction of masculine and feminine, which has now almost completely disappeared and is only made when referring back to a noun by using an anaphoric pronoun. In this case, gender is based on the apparent sex of the noun (e.g. *'De vrouw en <u>haar</u> man'*: The woman and <u>her</u> husband) or the suffix that indicates the original gender (e.g. *'-aard'* being masculine and '*-heid'* being feminine: this will be discussed in more detail below). As a result of common nouns (*de*-words) being a combination of feminine and masculine words, these nouns make up around 75% of all nouns (Van Berkum, 1996: 11). Not all dictionaries still make the distinction between masculine and feminine for common nouns, but the Taalunie (i.e. the organisation in which the Netherlands, Flanders and Suriname cooperate to, e.g. regulate the Dutch spelling) does indicate this with *'de (m.)'* for masculine (*'mannelijke'*) words and *'de (v.)'* for feminine (*'vrouwelijke'*) words. The Taalunie has published 'Het groene boekje' (2005, 'the Green Booklet') in which most rules concerning the Dutch gender assignment system are outlined. The most important ones are depicted in Table 1-1.

Overall, linguists do agree that assignment systems are, at least to a certain degree, based on regularities that tend to be related either to semantic principles (e.g. biological gender) or formal principles related to the morphology of the words. Examples of these formal principles in Dutch can be found in Table 1-1, e.g. the fact that words ending in *'–heid'* are feminine in Dutch. The rules underlying the systems, however, can only be applied to a subset of the nouns (i.e. the nouns that have suffixes) and the rules often appear to contain many exceptions (as can also be seen in Table 1-1). Moreover, many short nouns in the Dutch language, such as monosyllabic words, do not carry any suffixes that can help determine its gender. Nevertheless, knowing a noun's gender category, the gender assignment system, is a prerequisite for using the gender agreement system.

Grammatical gender in Dutch is expressed by the article, the adjective and by pronouns, but not by verbs and/or cases (as in many other languages). Table 1-2 shows the main agreement targets in the Dutch gender system. Only some of the pronouns are mentioned in this table. It should be noted that, in their attributive use, demonstrative, possessive, interrogative, and a number of indefinite pronouns all show gender agreement with the noun (Van Berkum, 1996). The general rule behind the Dutch gender system is that a root singular noun receives the definite article *'het'* when it is neuter, whereas *'de'* is used in all other definite cases. As can be seen in Table 1-2, the singular neuter case is the only deviant one in the Dutch gender system, in which there is a clear distinction from the general default form. In a singular definite noun phrase, the adjective never shows a gender distinction: all adjectives receive the suffix *–e* regardless of the gender of the noun they modify. In the singular indefinite case, however, the article is always *'een'*, with no gender distinction, and the adjective marks the gender in these phrases: the suffix *–e* is added for the common nouns and the bare adjective is used when preceding a neuter noun. Attributive adjectives, i.e. adjectives appearing in front of the noun they qualify, thus only decline when they modify nouns preceded by an indefinite article or no article at all (or by indefinite and/or interrogative pronouns like '*geen*' (none) and '*welk(e)*' (which)).

No distinct gender-marking is used in the plural where all nouns are preceded by the common definite determiner *'de',* all adjectives are marked with the suffix *–e*, and the same pronominal form is used, which is furthermore identical to the singular common pronominal form, regardless of the gender of the plural noun they refer to. In sum, the Dutch gender system can be covered by a rule for assigning gender (root noun receives *'het'* when singular and neuter, whereas all other nouns receive '*de'*) and a rule for gender agreement (all adjectives receive the suffix *'-e'*, unless the modified noun is an indefinite singular neuter noun).

Table 1-1: Overview of Dutch nouns and their division across gender categories. Adapted from the Woordenlijst van de Nederlandse taal (Word List of Dutch: 2005) as collected on behalf of the Belgian and Dutch governments. This list, also known as 'Het groene boekje' (The green booklet), has received a lot of criticism over the past few years, mainly because of its complicated rules and many exceptions to these rules. Nevertheless, the table does summarize the most important rules governing the Dutch gender assignment system. Common gender was originally divided into masculine and feminine gender. Some words can be used with both masculine and feminine gender; the double-gender words.

Common Nouns (*de*-words)			**Neuter Nouns (*het*-words)**
Masculine ***'de (m.)'***	**Feminine** ***'de (v.)'***	**Double-gender** ***'de (m/v)'***	**Neuter** ***'het'***
Words ending with: *-aar*, *-aard*, *-er*, and *-erd*. ***Exceptions:*** *'bakker'* (baker) is masculine, whereas *'baker'* (midwife) is feminine.	Words ending with: *-heid, -nis, -schap*; *-de* or *-te*; *-ij, -erij, -arij;* *-e(r)nij*; *-ing* or *-st* (as a suffix of a verb's stem) ***Exception:*** *'dienst'* (service) is masculine.	Most of the nouns for objects that used to be feminine, such as *'bank'* (couch).	Diminutives, characterized by the suffix *-(t)je*
Verb stems used as Nouns.	Words ending with: *-ie, -tie, -logie;* *-sofie, - agogie;* *-iek, -ica;* *-theek, -teit, -iteit;* *-tuur, -suur;* *-ade, -ide, ode, -ude; -age, -ine, -se;* *-sis, -xis, -tis.* ***Exceptions:*** *'kanarie'* (canary) is masculine and *'lambiek'* (lambic) is masculine.	General geographical names as well as names for stars and other celestial bodies.	Stems of verbs that begin with the following prefixes: *-be, -ge,* and *–ont*
Nouns referring to masculine persons or animals.	Nouns referring to feminine persons or animals.	Adjectives that are used as nouns, such as *'zieke'* (ill person), and*'blinde'* (blind person).	Names referring to countries and cities.
		Nouns referring to persons that can be either masculine or feminine, such as *'baby'* (baby) and *'arts'* (doctor).	

Table 1-2: Overview of the main agreement targets in Dutch. English equivalents of the phrases are printed in italic below every phrase. Note that the indefinite article is always *'een'* in the singular and no equivalent indefinite article exists for the plural.

	Gender	Definite Article	Adjectives in Definite NPs	Adjectives in Indefinite NPs	Demonstrative pronouns	Relative Pronouns
Single	Common	De …	De mooi***e***...	Een mooi***e***...	Deze/die/onze/elke	De/een ... die
	Neuter	Het ...	Het mooi***e***...	Een mooi...	Dit/dat/ons/elk ...	Het/een ... dat
Gloss		*The ...*	*The beautiful..*	*A beautiful...*	*This/that/our/every*	*The/a ... that*
Plural	Common	De ...	De mooie ...	x	Deze/die/onze ...	De ... die
	Neuter	De ...	De mooie ...	x	Deze/die/onze …	De ... die
Gloss		*The ...*	*The beautiful..*	x	*These/those/our …*	*The …that*

To summarize the situation for the language learner, the covert Dutch gender system provides only a few cues to gender and those few cues in the input only show up in the gender agreement system. To complicate matters even further, the agreement system is not consistent and the most clearly matched case, the singular neuter, is only a minority group. The challenges of acquiring such an asymmetrical and non-transparent system are further outlined in §1.2. The next section will first describe the gender system in Polish, the native language of the late L2 learners tested in the present dissertation.

1.1.2. Grammatical gender in Polish

The Polish language has a highly complex morphological system (Dąbrowska, 2004). The canonical word order in Polish is SVO, but this word order is flexible due to the presence of a case system, consisting of seven forms in both the singular and the plural. Grammatical gender in Polish is quite complex because it combines three categories: gender (masculine, feminine and neuter), personhood (personal and non-personal), and animacy (animate and inanimate). Animacy only affects the masculine gender in the singular and not the feminine and neuter genders. In the plural, a division is drawn between masculine personal gender and non-masculine personal gender, which is also called feminine object gender (Koniuszaniec & Błaszkowska, 2003). This non-masculine personal gender comprises words that are feminine and neuter in the singular as well as masculine non-personal words. In other words, the distinction is made between masculine personal and a common plural gender for all other categories. Schematically, the Polish grammatical gender system looks like this:

Single case
- Masculine
  - Animate
  - Inanimate
- Feminine
- Neuter

Plural case
- Masculine Personal
- Non-masculine Personal

In Polish, gender is a fixed property of every noun and besides gender, nouns are also marked for case and number and the inflectional features consist of a suffix being either -Ø or a vowel (Perlak & Jarema, 2003). As opposed to the Dutch gender system, gender in Polish is overt and can be extracted from the phonological and morphological characteristics of the noun itself. For example, as can also be seen in the examples below, masculine nouns generally end in consonants (1-1a), feminine nouns in *'–a'* (1-1b) and neuter nouns generally end in *'–o'* or *'–e'* (1-1c).

1-1a *Ten jeden mały chłopiec*
'this$_{MASC}$ one$_{MASC}$ little$_{MASC}$ boy$_{MASC}$'
1-1b *Ta jedna mała dziewczynka*
'this$_{FEM}$ one$_{FEM}$ little$_{FEM}$ girl$_{FEM}$'
1-1c *To jedno małe dziecko*
'this$_{NEU}$ one$_{NEU}$ little$_{NEU}$ child$_{NEU}$'

Examples from Koniuszaniec and Błaszkowska (2003)

Important for the purpose of the present dissertation is that the Polish language has no articles, but nouns can be modified by other determiners, numerals or adjectives that must agree in case, number and gender (Perlak & Jarema, 2003). Consequently, Polish-Dutch L2 learners need to learn to express gender on determiners. On the other hand, Polish adjectives precede the noun, as they do in Dutch. Hence, Polish learners of Dutch will be familiar with the combinations of determiners and adjectives preceding the noun they modify. The present project focusses on the acquisition of the determiner-noun and adjectival-noun system in Dutch, but it should be noted that Polish verbs are also inflected for gender (as well as person, number, aspect, mood, and voice). At least partly due to its prominence, transparency, and systematicity, the Polish gender system has been found to be mastered by native Polish children as early as age 2 (Smoczyńska, 1986), which contrasts with the overgeneralizations of Dutch common gender-marking by native Dutch children until the age of 6 (Blom et al., 2008). The next section will focus more on the relative difficulty of the Dutch gender system.

1.1.3. The challenge of acquiring Dutch gender

When acquiring a gender system, one can use semantic cues (i.e. natural or biological gender), morphophonological cues (i.e. a noun's suffixes, such as *'-heid'*), and syntactic cues (i.e. the form of the syntactically related words, such as determiners, adjectives, and pronouns) that are present in the language (Franceschina, 2005). Given the fact that most nouns refer to objects that do not have biological gender and that many nouns in the Dutch vocabulary do not have any morphophonological cues to determine the noun's gender (Haeseryn et al., 1997), learners of Dutch have to rely mainly on syntactic cues. These syntactic cues, however, are not always reliably linked to one of the two genders. The Dutch gender is not only covert and non-transparent, but is also characterized by the asymmetrical distribution of the two genders (75% common versus 25% neuter). Common gender-marking is also used for all nouns, regardless of their gender, in the plural. On the other hand, singular Dutch nouns become neuter when used in the diminutive form.

Due to the absence of clear cues in the input, acquisition of a noun's gender in Dutch has to occur mainly on an item by item basis (Unsworth, 2008). The Dutch gender system is therefore not only challenging for L2 learners, but even Dutch native children have been reported to have problems acquiring their own gender system. Van der Velde (2004) demonstrated that Dutch monolinguals overgeneralize the definite, common determiner *'de'* until the age of 6. He hypothesized that children first adopt the default value for the gender feature, which in Dutch is the common determiner *'de'*. Only at a later stage are children able to use the specific value, i.e. the neuter definite determiner *'het'*, in addition to the default value. Van der Velde's argument has been supported by production data from monolingual Dutch children (see, e.g. Hulk & Cornips 2006a,b; Cornips, van der Hoek, and Verwer, 2006, and Blom, Polišenskà, and Weerman, 2008), showing that children almost exclusively use the common definite determiner *'de'* and overgeneralize *'de'* with both neuter and common nouns. Their production appears target-like only after the age of 6, but even then children produce some gender errors until a more advanced age.

Dutch appears to be an exceptional case in this respect. German monolingual children, for example, show correct usage of masculine, feminine and neuter articles around the age of 3 (Mills, 1986; Szagun, Stumper, Sondag, & Franik, 2007) and Polish monolingual children have even been reported to show correct gender usage around the age of 2 (Smoczyńska, 1986). The same holds true for children acquiring French and Spanish as mother tongues and a systematic overgeneralization to a single default gender (as seen in Dutch) is not present in monolingual speakers of these languages (Franceschina, 2005). These studies focussing on L1 acquisition of Dutch gender additionally show that children acquire gender agreement between the definite article and the noun (i.e. common *'de'* versus neuter *'het'*) before agreement between the adjective and the noun (i.e. the common schwa-ending versus the bare adjective for indefinite singular neuter nouns). This might reflect the fact that, before being able to apply the rules of the gender agreement system one has to know the gender of the noun, i.e. the gender assignment system. Moreover, the determiner system in Dutch is more saliently present and more reliable than the adjectival system in that the schwa-ending on the adjective is also used to precede neuter nouns in definite NPs. Children might also initially acquire the combinations of definite determiners and nouns as whole units and only later do they realize that the unit can be separated into two detached items of which the determiner carries gender information.

This latter idea would correspond to the finding that learning new words and the overall meaning in linguistic structures (semantics) often precedes the acquisition of the structural relationships between these meaningful units in a language (morphology and syntax) (see, e.g. Slabakova, 2006). Especially grammatical gender appears to be one of the most difficult aspects to master and specifically so when L2 learners start learning at a later stage in life (Carroll, 1989). Most L2 learners have been shown to fossilize in their acquisition of Dutch gender and continue to overgeneralize either common or neuter gender (e.g. Hulk & Cornips, 2006a, b; Unsworth, 2008). What makes grammatical gender and especially the Dutch gender system so extremely difficult to acquire? Do late L2 learners mainly have difficulty with an L2 gender system, because it does not resemble their own gender system? Or is it mainly the asymmetry and covertness of the Dutch gender system that causes problems? The next section will provide an overview of the similarities and differences between native and L2 acquisition, and, more specifically, the various explanations that have been put forward to account for these similarities and/or differences.

1.2. Theoretical approaches to Second Language Acquisition

Overall, linguists agree that L2 acquisition differs from L1 acquisition. They do not agree, however, on the causes and manifestations of this difference. The various accounts differ as to whether (second) language acquisition involves an inborn capacity that can only be used before puberty or whether it is mainly input and amount of exposure that determines the ultimate level of proficiency. Secondly, the approaches differ as to whether it is mainly the linguistic representation or mainly the use of this linguistic information that distinguishes L2 learners from native speakers. Finally, researchers do not agree on the potential influence of the native language on L2 knowledge and/or processing. Ultimately, the most important distinction between these various approaches concerns the question of whether L2 learners are fundamentally different or essentially similar to native speakers. Although there are various gradations along these spectrums, this section will provide an outline of three of the main views on second language acquisition (SLA) and how proponents of these views explain the difficulties in the acquisition of gender systems in particular.

1.2.1. Explanations for L1-L2 differences based on Universal Grammar

The observation that children appear to be able to produce completely new structures that are perfectly correct without ever encountering these structures in the input was formulated by Noam Chomsky (1995) as the poverty of the stimulus hypothesis or the logical problem of language acquisition. He argued that, if children can acquire the complex rules of grammar without having received explicit instructions, they must possess some innate 'language acquisition device' (LAD) containing a set of basic principles of grammar. The specific settings or parameters of this Universal Grammar (UG) are then set to the specific language conforming to the input we receive as a child. To explain the differences between L1 and L2 acquisition, some researchers claim that this UG needs to be stimulated within the Critical Period (CP), i.e. a ‘window of opportunity’ that is present until around puberty. During this period, children are hypothesized to be specifically sensitive to linguistic input, allowing them to acquire a language with remarkable ease. After puberty, the window closes due to maturational changes after which native-like performance in the target language is not feasible. In line with this reasoning, it has been suggested that Chomsky's Universal Grammar (Chomsky, 1995), i.e. the set of rules that all languages share as a common structural basis, is only partially available to late L2 learners. One version of this hypothesis is known as the Failed Functional Features Hypothesis (FFFH: Hawkins and Chan, 1997) and states that post-puberty learners are unable to acquire functional features that are not represented in their L1. More specifically, this hypothesis assumes that late L2 learners are able to acquire interpretable features such as new isolated words. The ‘representational deficit’, however, concerns uninterpretable features related to functional categories that do not substantially contribute to meaning such as agreement.

The FFFH predicts that only learners from a gendered L1 can acquire an L2 gender system, while learners from a non-gendered language, such as English, cannot acquire L2 gender because it is not instantiated in the L1 (and UG can only be accessed via the mother tongue). The present dissertation only tested the acquisition of L2 gender by Polish-Dutch L2 learners and hence cannot test the assumption that having L1 gender is a requisite for being able to acquire an L2 gender system. It is interesting to investigate, however, whether Polish-Dutch L2 learners indeed show native-like behaviour and have the appropriate grammatical representations, despite the typological differences between the languages in general and the gender systems in particular (also see §1.2.4. below for more specific hypotheses of the FFFH).

Opposing the view that the functional feature make-up of the L1 prevents many L2 learners from becoming fully target-like (Hawkins & Franceschina, 2004) are scientists who believe that learners can acquire grammatical gender in their L2 regardless of whether this feature is present in their L1 (Bruhn de Garavito & White, 2000). The Full-Transfer Full-Access model (FTFA) assumes, like FFFH, that the initial representation of grammatical features is based on features that are present in the L1. Proponents of this model also believe, however, that L2 learners can become native-like regardless of their age of acquisition because they can still access the parameters provided by UG when they are needed in the second language.

Contrary to the FFFH, the FTFA model predicts full access to UG and hence predicts the possibility of L2 acquisition by advanced learners from both gendered as well as non-gendered languages. Most important to the present study, the FTFA predicts that L2 learners rely on their L1 grammar in initial stages of L2 acquisition. The Polish-Dutch L2 learners in the present study might thus be expected to show effects of L1 transfer in initial, but not in more advanced stages of L2 acquisition.

1.2.2. Usage-based approaches and Emergentism

Diametrically opposite the generativists are proponents of a usage-based account who believe that (children's) language development is based on a derivate rather than an innate grammar (Tomasello, 2003). Usage-based accounts argue that language is primarily learned through input and use in combination with the specific human learning capacities. Developmental psychologist Tomasello (2003) claims that the social and cognitive need to understand others is what drives language learning and, by evolution, we possess inherited cognitive skills to understand and imitate adult's communicative intentions and analyse the input to find patterns and create categories as well as new structures.

Less extreme than these empiricists are the emergentists who argue that neither nature nor nurture would suffice for adequate language acquisition. Language learning is a cognitive process that emerges from the combination of input and communicative processes. The Competition Model proposed by Bates and MacWhinney (1987) is one of the most influential theories holding an emergentist view on language acquisition. This approach acknowledges that we use semantic, syntactic, and morphological cues in the input to comprehend language. Both the availability and frequency of a cue as well as its reliability, i.e. how often it points towards a particular interpretation, are crucial in learning language. With respect to second language acquisition, this model posits that features that are similar in the L1 and the L2 are easy to acquire while features that are different in the two languages are difficult to acquire due to competition from the mother tongue. Additionally, features that are unique to the L2 are also easy to acquire due to the absence of competition between the native and the second language.

The Competition Model is especially interesting when looking at the non-transparent and asymmetrical Dutch gender system. Cues for gender agreement in Dutch are highly available, especially in the form of determiner-noun combinations, but the cues are not reliable in that the schwa-ending on the adjective can precede both common and neuter nouns (also see Table 1-2). From this point of view, it is definitely not surprising that children acquire frequent and reliable morphological structures in the input (e.g. a tensed auxiliary followed by a past participle) well before grammatical gender around the age of 2;5 years (De Houwer & Gillis, 1998). Additionally, the overgeneralization of common gender can, from a Competition Model perspective, easily be explained by the increased availability and reliability of common gender-marking in the input.

With respect to the late Polish-Dutch learners tested in the experiments in the present dissertation, the Competition Model would thus expect differences in the acquisition and processing of common and neuter gender, but also differences in the acquisition of structures that are similar and/or unique to the L2 as opposed to structures that are different in the two languages. These assumptions were tested in the experiment presented in Chapter 7 and more details concerning the hypotheses will follow in section 1.4 of the present chapter.

1.2.3. Psycho- and neurolinguistic approaches

The learning strategies of late L2 learners, who have already successfully acquired a language, may well be based on more meta-linguistic knowledge. A model that approaches L2 processing in terms of memory systems has been developed by Paradis (2004) and Ullman (2001, 2004). Ullman's declarative/procedural model (DP model) claims that there are two different neural systems: the declarative and the procedural memory system that distinguish between the lexicon and grammar. The declarative system is believed to be specialized in associative learning while the procedural system is thought to be specialized in computing sequences. Hence, the declarative system appears to be very good at learning arbitrarily related information, such as words and their meanings. The acquisition and use of grammatical rules, on the other hand, involves more implicit (non-conscious) learning as performed by the procedural system. Ullman hypothesizes that the procedural memory system weakens when we get older and therefore late L2 learners have to rely mainly on the declarative system for both lexical and syntactic processing. More specifically related to grammatical gender, the DP model assumes that L2 learners differ from native speakers in that they fail to decompose morphologically complex words.

A focus on processing was also stimulated by the development and fast progress of online methods, such as event-related brain potentials (ERP) and eye movement research (see §1.3 for more detail on these approaches). The evidence provided by such processing measures led Clahsen and Felser (2006a) to assume that late L2 learners might use processing strategies that are completely different from those in native speakers. Their 'Shallow Structure Hypothesis' (SSH) assumes that language learners have difficulty integrating different information sources online. L2 learners, as opposed to native speakers, mainly rely on associative and lexico-semantic information rather than syntax and are hence engaged in less detailed syntactic processing while comprehending the L2. As a consequence, late L2 learners are able to process semantics and morphology in a native-like manner, but not syntax. Additionally, adult L2 learners may process the target language less rapidly as compared to natives, which might reflect a lack of automaticity. The authors assume that the L2 learners' native language can influence lexical processing, but the L1 grammar cannot influence the real-time processing of grammatical structures in the L2 due to the unavailability of an L2 grammar system to transfer the L1 grammar to.

Like the FFFH, both the DP model and the SSH hypothesize late L2 learners of Dutch, regardless of their mother tongue, to be fundamentally different from native speakers. As opposed to the FFFH, however, the difference is more procedurally defined and late L2 learners are hypothesized not to be able to use procedural memory or to engage in deep processing of syntactic structures in a native-like manner. The specific predictions constitute differences in processing and will be further specified in section 1.4, after explaining the online methods used to compare L1 and L2 processing in section 1.3.

1.2.4. Nature, nurture, declarative/procedural or shallow processing?

The models and hypotheses outlined above differ with respect to whether late L2 learners' knowledge and/or processing of the L2 are fundamentally different or fundamentally similar as compared to native speakers of the language. Fundamental differences are predicted by both the Failed Functional Features Hypothesis and the Shallow Structure Hypothesis, but while the latter suggests processing differences based on a general impairment within late L2 learners the FFFH suggests the difference to reflect underlying knowledge representations that are largely influenced by the L1. These differences in representations and/or processing would lead to qualitative differences between native speakers and late L2 learners, such as incomparable error patterns and different processing strategies that may or may not be related to the mother tongue. The Full-Transfer Full-Access model and the Competition Model, on the other hand, suggest that L2 learners rely initially on their L1 grammar and may eventually, after increased practice and exposure, share the same representations and processing resources as native speakers. Proficiency, length of exposure, and the degree of similarity and/or competition between the two languages, however, may cause L2 learners to perform quantitatively different from native speakers. Such quantitative differences may result in higher error rates (but comparable errors) and delays in L2 processing.

Age of acquisition and the impact of the mother tongue, two of the main potential causes for the difference between native and second language learners, have been investigated specifically with respect to the acquisition of grammatical gender. To test the assumption of both the FFFH and the DP Model that age only affects the acquisition of functional features or the procedural system, Blom, Polišenskà, and Weerman (2008) directly compared Dutch gender acquisition of both children and adults with Moroccan (a gendered language) as L1 background to native Dutch children. They hypothesized that, as predicted by the FFFH and the DP model, age only affects grammar-driven gender representations and not lexical-driven gender representations, because the declarative system remains strong throughout our lives. In other words, adult L2 speakers of Dutch may show native-like performance on morphophonological gender correlates, but their knowledge never exceeds the lexical level and hence they will show problems at the grammar level of gender which is an uninterpretable feature. Children, on the other hand, can use their procedural system to acquire and process uninterpretable features, so they will not show differences in performance on lexical-driven and grammar-driven gender representations.

Within the Dutch gender system, both the lexical and grammatical levels of representation are effective for definite articles, but only the grammatical level of representation is more adequate for attributive adjectives. The authors thus tried to elicit speech containing definite determiners and attributive adjectives by means of a sentence completion task in which subjects were asked to describe pictures. Their results showed that, with respect to determiners, all groups made few errors with common and considerably more with neuter nouns, concurrent with the frequency distribution of the two genders. Both children groups made few errors with common nouns and showed an overuse of the schwa-suffix in the attributive adjectives, whereas adults performed significantly worse with common nouns and made less errors with neuter nouns. The error pattern of the children can thus be interpreted as overgeneralizations of the common gender feature with both determiners and adjectives, which coincides with a grammar-based representation. The inconsistency of the errors made by adults with respect to determiners and adjectives as well as the fact that the errors made with attributive adjectives occur in two directions is, on the other hand, more compatible with the relatively more equal input distributions of the schwa-ending and the bare adjective and are hence more likely to reflect lexicon-based input-driven learning.

The results of the above experiment suggest that late L2 learners, despite the presence of a gender system in the L1, have fundamentally different representations and underlying knowledge of the L2 gender system as compared to native speakers of the language. Although the main focus of the present dissertation is the acquisition of Dutch gender by Polish-Dutch bilinguals, i.e. L2 learners who have gender in their mother tongue, it is important to acknowledge the possible impact of the first language, especially because this variable, whether or not it is kept constant within experiments, differs between experiments. Research has thus far led most scientists to agree that there is transfer in acquiring an L2, but the degree of transfer is still under debate. Either L2 learners can acquire grammatical gender in their L2 regardless of whether this feature is present in their L1 (Bruhn de Garavito & White, 2000; White, Valenzuela, Macgregor, Leung, & Ben Ayed, 2001), or the functional feature make-up of the L1 prevents them from becoming fully target-like (Hawkins & Franceschina, 2004; Franceschina, 2005). An interesting additional position has been proposed by Sabourin (2003), who argued that just having gender in the L1 may not suffice and that the systems should be closely related in order for the L2 learners to be able to become native-like.

The role of the native language and the issue of overgeneralization of common gender in Dutch was the focus of a study conducted by Cornips, van der Hoek, & Verwer (2006), who tested older children (between 10-12 years of age) with either Turkish [-gender] or Moroccan Arabic [+gender] as L1 background. These children were compared to Dutch natives on their performance on a sentence completion task. The main focus of the study was to look at the acquisition of attributive adjectives and relative pronouns to see if these aspects reveal the same difficulties as definite determiners, i.e. overgeneralization of the common definite determiner *'de'*. With the exception of the common relative pronoun *'die'*, Moroccan and Turkish bilinguals revealed a serious delay in the acquisition of the Dutch gender system as compared to their Dutch peers, as reflected by overall lower accuracy scores. Cornips et al. (2006) do not report overall significant effects of language background, but they do report that the Moroccan children performed significantly better than the Turkish on the correct form of the relative pronoun in the neuter gender condition (*'dat'*). In line with the results reported by Blom et al. (2008), the acquisition of neuter gender was found to be delayed in child L2 learners, and a decline in target-like performance was found across the constructions tested with highest accuracy on determiners, and increasing difficulty in the use of attributive adjectives and relative pronouns respectively. Interestingly, even very early bilinguals could not perform at the same level as monolinguals with respect to the use of the neuter definite determiner, whether their L1 has grammatical gender or not.

The presence or absence of a gender system in the native language might only impact late L2 acquisition as opposed to child L2 acquisition as suggested by the results reported by Franceschina (2005). She compared native speakers of Spanish to highly proficient non-native speakers of Spanish with first languages that either had gender or not.[2] L2 learner's performance on a combination of off-line production and comprehension tasks showed that learners from a gendered language performed significantly better than L2 learners who do not have gender in their L1. She concludes that, as predicted by the Failed Functional Features Hypothesis, only speakers of languages with a gender system can achieve native-like accuracy in handling L2 syntactic gender. This conclusion should be viewed with some caution, because the people in the [+gender] group could have performed better due to the sometimes clear resemblances of the words used. Italian natives, for example, were confronted with words of which the morphological features can be directly copied from their L1 onto the L2: e.g. *'el problema'* compared to *'il problema'* (the problem).

Results contradicting Franceschina (2005) were found by White, Valenzuela, Kozlowska-MacGregor and Leung (2004) who aimed at testing the 'no parameter resetting hypothesis' (which is similar to the FFFH) against 'full access theories' (FTFA). Therefore, the authors examined the acquisition of Spanish gender and number features and their agreement properties in French [+gender] and English [-gender] learners of Spanish. As opposed to the late L2 learners of Spanish tested by Franceschina, who had all received multiple years of naturalistic exposure to Spanish, the participants tested by White and her colleagues were students acquiring Spanish as an L2 or an L3 in an educational setting. These L2 learners were divided into low, intermediate, and advanced proficiency groups and tested on both their comprehension and production of gender and number agreement in Spanish. Interestingly, while the authors do not report significant differences based on the L1 of the participants, i.e. the presence or absence of a gender system in the mother tongue, they do report significant differences based on proficiency level with advanced, but even intermediate, learners of Spanish performing similar to native Spanish speakers. The absence of a difference between French [+gender] and English [-gender] learners of Spanish supports the idea that the parameters of UG can be reset. In other words, overall proficiency level is a

[2] It is important to note that the groups Franceschina compared were of different sizes. The native control group contained 42 subjects and the experimental group with [+gender] languages as L1 contained 53 participants. On the contrary, only 15 people were part of the English experimental group [-gender]. Additionally, the [-gender] group was homogeneous with respect to L1 background, whereas the [+gender] group contained participants with various L1 backgrounds: Italian, Portuguese, French, Greek, German, and Arabic.

crucial factor in explaining differences in the acquisition of gender and the native language has no influence in L2 acquisition, because learners are able to fully access their UG to reset the parameters of gender, as predicted by the Full-Access Full-Transfer model.

In addition to distinguishing between having or not having a gender system in the mother tongue, Sabourin, Stowe, and de Haan (2006) hypothesized that there can either be no transfer, partial transfer or full transfer from the L1. The no transfer position would predict L2 learners with different L1 backgrounds to behave similarly. Partial and full transfer positions, on the other hand, would predict an influence from the native language with L1 grammar either being partially or fully transferred, at least during initial stages of L2 acquisition. The authors tested German [+gender], Romance [+gender], and English [-gender] late learners of Dutch to assess the additional potential impact of the similarity of the gender systems in the L1 and the L2. German is very closely related to Dutch, but Romance languages (in this case French, Italian and Spanish) have a grammatical gender system that is not congruent with the Dutch system. The first experiment focussed on gender assignment and subjects were asked to make a *'de'*/*'het'* decision for each noun. All groups appeared to be very good at assigning gender with an overall average performance of above 80%, but they did show an overuse of the default common definite determiner *'de'* with less frequent nouns. In the second experiment, focussing on gender agreement, participants were asked to judge every sentence on grammaticality as well as correcting those sentences that they marked as ungrammatical. An obvious effect of L1 was found for this task with Germans performing the best, followed by the Romance group and finally the English group (who actually performed at chance level). These results suggest that L2 acquisition of grammatical gender is additionally affected by the morphological similarity of gender marking in the L1 and L2 (i.e. surface transfer) as opposed to the presence of abstract syntactic gender features in the L1 per se (i.e. deep transfer). The results presented by Sabourin et al. (2006) are thus in line with Franceschina (2005) and contradictory to the results found by White et al. (2004).

The contradictory results concerning L1 transfer and L2 performance presented thus far might have to do with the fact that the above mentioned studies mainly used behavioural and/or offline tasks. These results only tell us something about the performance and the end-product of a certain task. Moreover, grammaticality judgement tasks, for example, might only provide us with information concerning meta-linguistic knowledge without revealing direct evidence of language processing. The next section will provide an outline of the online methods used to assess L2 language acquisition and processing and, more specifically, the advantages and disadvantages of these methods.

1.3. Experimental approaches to Second Language Acquisition

Over the years, scientists have studied language in many different ways using many different methodologies. Broadly speaking these different methodologies can be divided into offline and online measures and behavioural and physiological measures of language comprehension. Behavioural measures mostly concern reaction times and accuracy scores, such as grammaticality judgements, self-paced reading, and word monitoring. These measures can be very useful for certain purposes, but they only allow us to draw conclusions about the end-state of a particular action, e.g. the time it took to read a sentence or a word. Fortunately, relatively new online methods are now available to study the time-course and the underlying, unconscious processes involved in language comprehension.

A distinction often made in second language research is between offline and online measures of comprehension. Offline tasks of language comprehension typically involve the need of a participant to respond to a certain stimulus, such as in grammaticality judgement tasks (GJT). These tasks measure sentence comprehension and interpretation after the sentence has been heard and hence these measures tap into explicit knowledge about the language being tested (for an overview of the distinction between offline and online tasks, also see Marinis, 2010). Online comprehension tasks, on the other hand, measure the participant's performance while the linguistic information is being processed and hence rely mostly on implicit knowledge of the language user or learner and less so on meta-linguistic knowledge.

The experiments used in the present dissertation exploit both offline and online measures, but the focus lies on real-time language processing. According to Clahsen (2008), there are 'two basic types of time-sensitive measures available to examine language processing: behavioural measures (e.g. comprehension response times and production latencies) and physiological measures (e.g. event-related brain potentials (ERPs) and eye movements)'. These latter two techniques will be discussed in the next sections.

1.3.1. Eye movements

Both offline and online measures have shown that people are able to understand language in a very rapid fashion. This rapid process of understanding language is often linked to what we see in our surrounding environment. The visual world allows us, and helps us, to focus on objects already before the speech stream has completely unfolded. Vision helps us to disambiguate linguistic information and to relate single utterances to objects in the world surrounding us. Visual perception or the study of eye movements has recently gained more and more attention from psycholinguists because it can help us understand and investigate the language comprehension system.

Our eyes constantly move, trying to capture the overwhelming amount of visual information surrounding us. Of course, we are not able to capture everything, but our eyes are able to inspect small portions of the real world in a very rapid sequence. And whereas we can see around 200°, our eyes only receive detailed information from 2° (Levi, Klein, & Aitsebaomo, 1985). The region that is responsible for this sharp central vision is called the fovea, or fovea centralis. This fovea can make extremely rapid movements, but when it rests it gives us highly detailed pictures of the surrounding visual world. The movements and resting states of the eyes, also known as saccades and fixations, are an important distinction for eye tracking research. Saccades refer to the ballistic jerks of the eyeball from one position to the other, whereas fixations occur 'between any two saccades, during which the eyes are relatively stationary and all visual input occurs' (Martin, 1974). A person can thus only process new visual input during fixations while no new visual input is processed during saccades. It is assumed that these scanning movements are programmed during fixations, when the eyes rest on a certain object in a scene. As Richardson & Spivey (2004) point out, both fixations and saccades are closely related to attentional mechanisms, thus providing insight into cognitive processes such as language comprehension, mental imagery and decision making. With increasingly new and advanced technology, eye movements can be tracked with great speed and precision and this relatively new technique has proven to be very useful to the study of language.[3]

Eye tracking can be defined as the technique of measuring the eye's gaze (i.e. the area the eyes are looking at) or the movement of the eyes. When the relevance of eye movements first became noticed by scientists, they started using mirrors, telescopes, and peep holes to study the eye's movements. Currently, scientists are using either one of three techniques: head-mounted eye trackers, remote eye trackers, or video-based eye trackers (the 'poor man's eye tracker'). The present project has the advantage of using a remote eye tracker, with which the subject is freely sitting in a chair while looking at a computer screen. Nevertheless, this section will provide a short description of all three techniques.

The 'poor man's eye tracker' earns its name from the fact that anyone with a video camera and enough knowledge about eye tracking can use this technique to study eye movements. A camera trained on the participant's face and eyes records the eye's gaze while the subject is looking at a played-out scene or objects on a platform. The direction of gaze is then analysed by coding these as quadrants in the scene. This technique, and in particular the analysis, is extremely time-consuming, but it does allow for a relatively cheap option to track the eyes. It also allows participants to move their heads during experiments, which is not an option when using a head-mounted eye tracker. Head-mounted eye trackers make use of very small cameras and optics mounted on a visor. There are two of these small cameras, one focussing on the eye and the other on the surrounding visual world. The video data from the camera that follows the eye is analysed in real time to calculate the location of the pupil as well as the corneal reflection, as also depicted in Figure 1-1. This cornea, the transparent layer covering the

[3] For a more detailed overview of the history of eye tracking, the reader is referred to Richardson and Spivey (2004).

front part of the eye, transmits and focusses light into the eye. For analysis of the data recorded by head-mounted eye trackers, the coordinates of the eye's gaze are sampled and paired with the coordinate position in the scene camera (Duchowski, 2003).

Remote eye trackers work in a similar way as head-mounted eye trackers, except for the fact that no visor is needed. While subjects are reading sentences or looking at a scene of pictures on a screen, infrared light is reflected from the eye, which in turn is recorded by an optical sensor. The corneal reflection of the eye is then used to extract and analyse gaze positions and eye rotations.

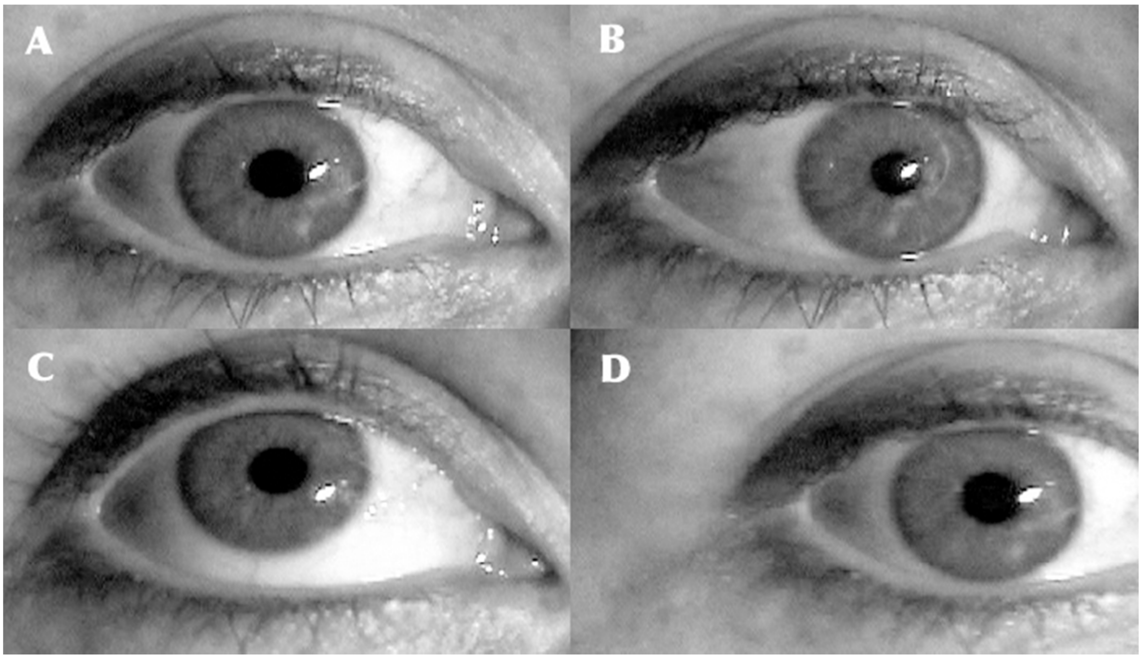


Figure 1-1: Taken from Richardson and Spivey (2004). Corneal reflections produced by different eye-head positions. The corneal reflection appears as a bright white dot that is visible to the right of the pupil (A). When the eye moves horizontally, the relative position of the pupil and the corneal reflection changes accordingly with, in this case, a smaller distance between the pupil and the bright white dot due to movement to the left (B). Similarly, when the eye makes vertical movements, the relative position of the pupil and corneal reflection changes along the vertical axis (C). The relationship between pupil and corneal reflection does not change when the head moves, but the eye is kept stable (D).

Researchers have been using these different methods for the same purpose: measuring positions and latencies of the subjects' eyes on pictures or sentences and sometimes while listening to linguistic input. Within the field of (second) language acquisition, two main eye tracking techniques have been used. First of all, researchers have looked at eye movements while reading to find out how people's eyes move back and forth across a sentence in order to understand the meaning (see, e.g. Rayner, 1998). The other eye tracking technique, which will be used in the studies presented in this dissertation, measures movements and fixations of the eyes to certain regions of interest (ROIs) in a scene while people listen to sentences or instructions (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

This latter technique, better known as the visual-world paradigm, relies on an old finding by Cooper (1974) that people strongly tend to look at the referents of the words they hear. Cooper presented four pictures to his subjects while they were listening to stories about these pictures. His results showed that the eye's gaze can be closely time-locked to the words that are heard in the speech stream. Two decades later, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) re-introduced this procedure, which resulted in the now well-known eye tracking methodology of the visual-world paradigm. Tanenhaus and his colleagues recorded eye movements while participants listened to instructions telling them to manipulate objects in the scene they observed. Their studies showed that reference is established incrementally by the listener and the movements of the eye appear to be closely time-locked to referents in the spoken input. The most important finding concerning this visual-world paradigm with respect to grammatical gender is that people tend to look at objects even before the spoken referent of the object has been fully pronounced. The speed with which people can use linguistic as well as non-linguistic information tells us something about how they use phonological information as well as how they create structures. The visual-world paradigm relies on this real-time interdependence between spoken input and the surrounding visual world. If, for example, a subject is instructed to '*pick up the candy*' when seeing four pictures consisting of a candy, a candle, a dog and a spoon, the subjects sometimes look at the picture depicting the candle first (Allopenna, Magnuson, & Tanenhaus, 1998). The fact that people initially fixate on the two referents that correspond to the first few phonemes in the input (*'can-'* in this case) and immediately use the following disambiguating information to choose the correct picture (*'-dy'* in this case), led scientists to believe that morphological information might be used in a similar manner to facilitate language comprehension. In particular when concerned with grammatical gender, the question is whether people can use morphological cues in the speech stream, such as definite determiners marked for gender, to decide which object to focus on. The literature with respect to using morphological and, in particular, grammatical gender cues during language comprehension will be discussed in detail in Chapter 4 (native processing) and Chapter 6 (second language processing).

1.3.2. Event-related brain potentials

The event-related potential technique (ERP) uses electroencephalography, or EEG, to measure the electrical activity in the brain during (language) tasks. This is done by placing electrodes on the scalp, amplifying the signal, and plotting the changes in voltage over time. As outlined in Luck (2005), ERPs are positive or negative voltage changes in the brain time-locked to specific (linguistic) events or stimuli. Events and functions, such as storing and retrieving information from the lexicon or focussing on an object, cause neurons in the brain to 'communicate' with each other. This communicative process takes place by firing of neurons, which in turn results in an electrical current that flows in and out of cells creating dipoles of negative and positive electrical charges separated by a small distance. When many of these neuronal dipoles receive the same type of input and have a similar orientation (i.e. either positive or negative) they will summate and then they may become measurable outside the skull. EEG allows us to measure this activity, which results in an overall picture of brain activity from underneath the scalp. ERPs are the changes in electrical current within this overall picture that have resulted from a specific stimulus or activity, hence the name event-related potential.

In order to get exactly that part of the EEG that represents the activity resulting from the stimulus (which is the ERP), we need repeated measures. Repeated measures are needed because the EEG also includes lots of brain activity that is irrelevant for the purpose of ERP research. The EEG must thus be time-locked to the specific stimulus and averaged after which the event-related potential becomes apparent, because the non-time-locked, irrelevant brain activity is averaged out. In sum, an ERP component can thus be defined as a 'scalp-recorded neural activity that is generated in a given neuroanatomical module when a specific computational operation is performed' (Luck, 2005:59). Different ERP effects can be characterized and described in terms of polarity (positive or negative), amplitude (smaller or larger), latency (timing of the effects), and (to a certain degree) scalp distribution.

ERPs have poor spatial resolution, but they do have an excellent temporal resolution of milliseconds. Therefore, the ERP technique has proven to be a sensitive measure of real-time language processing and is extremely suitable for the purpose of this study, because this study aims to examine the fast online processing of spoken sentences as they unfold. Previous ERP experiments have already found components that can be related to different semantic and syntactic processes. The most important effects and components in relation to the studies presented in the present dissertation (Chapters 5 and 7) are the N400, the LAN and the P600.

The N400 is a negative-going wave with a peak at 400 ms post-onset and reliably seen with semantic violations (Kutas & Hillyard, 1980). The N400 is often largest over centroparietal areas and usually larger over the right hemisphere than the left hemisphere (Kutas, Van Petten, & Besson, 1988). All content words are known to elicit a negative wave around 400 ms post-onset due to semantic integration, but this N400 is larger for words that are difficult to integrate into a sentence. The N400 is known to be affected by several semantic features, such as context, word position in a sentence, word class, and also by semantic expectancy or cloze probability, with greater negativities for words that are not preceded by a constraining context (Kutas & Hillyard, 1980).

By directly comparing semantic and syntactic violations, Kutas and Hillyard (1983) were the first to demonstrate that morphosyntactic violations elicit different ERP components as compared to semantic violations. Their study included semantic anomalies as well as number agreement discrepancies and misplaced finite and nonfinite verbs. The semantic anomalies elicited an N400, while all morphosyntactic anomalies elicited a fronto-central negativity around 300-400 ms and posterior positivities starting around 300 ms after the onset of the violation. The authors argued that semantics and syntax might have separate underlying neural processing systems.

The frontal negativity found by Kutas and Hillyard (1983) has become known as the LAN, the left anterior negativity. The LAN, named after its scalp distribution, occurs around 300 to 500 milliseconds after the onset of the stimulus and is most pronounced at anterior sites and mostly left lateralized. The LAN has been found to be correlated persistently with syntactic processing and is thought to reflect a detection process of syntactic violations. The idea that the LAN is correlated specifically to syntax has been supported by studies showing that the amplitude of the LAN did not reduce when semantic content was reduced, such as in studies in which content words were replaced by pseudo words (Yamada & Neville, 2004). The LAN has been found to precede the P600 effect, which will be explained below, in sentences with morphosyntactic violations such as subject-verb agreement violations (Coulson, King & Kutas, 1998; Hagoort & Brown, 2000), but it has also been found for word-category violations in Dutch (Hagoort, Wassenaar, & Brown, 2003). The LAN in response to agreement violations and verb inflection errors has been shown for English (Osterhout & Mobley, 1995), Spanish (Barber & Carreiras, 2005), German (Rossi, Gugler, Friederici, & Hahne, 2006), and Dutch (Gunter, Stowe, & Mulder, 1997).

The late positivity found for syntactic violations peaks at 600 ms post-onset, beginning at 500 ms and often persisting up to 1000 and even 1500 ms and is most pronounced on the back of the scalp. This effect was first reported as being a syntactic effect by Osterhout and Holcomb (1992), who labelled it the P600 effect.[4] This P600 has been found in response to a broad range of syntactic violations, such as morpho-syntactic violations, violations of category expectancy, and gender agreement violations and is thought of to reflect a repair or re-analysis process (for a recent review, see Gouvea, Phillips, Kazanina, & Poeppel, 2010).

The effects and components mentioned above have been put into a neurocognitive model of sentence comprehension by Friederici (2002), who separates comprehension into three consecutive phases. Like syntax-first models, the first phase of Friederici's model assumes an autonomous role for syntax and involves initial syntactic structure-building. The second phase (where both the N400 and the LAN can be found) begins after the initial syntactic structure has been processed and involves processes of assigning thematic roles by using semantic as well as syntactic information (thus allowing parallel processing of syntax and semantics). The final and last phase (where the P600 can be found) constitutes an interactive phase in which all preceding information is integrated, like in the interactive models. The different phases not only represent different aspects of language comprehension, but every consecutive phase also

[4] Note that other researchers have labelled this late positive ERP component the SPS, or Syntactic Positive Shift (Hagoort, Brown, and Groothusen, 1993). For the sake of consistency and clarity, I will refer to the late positivity as the P600 throughout this dissertation.

becomes more controlled. Early processes, such as distinguishing phonemes in auditory processing, are acquired at a very early age. These early acquired processes occur extremely often, which causes them to become automatic, i.e. not requiring attention or conscious analyses anymore. Consequently, these early processes are fast and unavoidable and cannot be manipulated by factors from within ourselves. The partially automatic stage requires a little more conscious awareness and late processes are thought to require full attention. Late controlled processes, as reflected by the P600, are thus used in new situations involving decision-making. Furthermore, they are flexible, meaning that controlled processes can be manipulated by factors from within the subject, such as task demands, and not by external factors such as differences in the visual contrast provided. Very early ERP responses (i.e. elicited before 100 ms post stimulus), on the contrary, are known to change in a predictable way when the physical parameters of the stimulus are varied in terms of, e.g. frequency, intensity, and duration (Kutas, Van Petten, & Kluender, 2006). These components are therefore called automatic or ‘exogenous’, i.e. rather impervious to the state of alertness of the subject. Later ERP effects such as the P600, on the other hand, are seen as controlled or ‘endogenous’ components, i.e. relatively insensitive to physical aspects of the stimuli, but generally sensitive to aspects from within the subject reflecting his or her choice of how to process the stimulus.

A P600 in response to sentences containing word-category violations, for example, has been shown to disappear when the subjects were instructed to ignore the violations (Hahne & Friederici, 2002). Additionally, the P600 has been shown to be influenced by the proportion of experimental sentences (Gunter et al., 1997; Coulson et al., 1998) and it has not been found for morphosyntactic violations on pseudo words (Münte, Matzke, & Johannes, 1997). A recent experiment by Hanulíková, van Alphen, van Goch, & Weber (2012) even showed that the P600 in response to gender violations in Dutch disappeared when the same sentences containing gender violations were pronounced by a speaker with a Turkish accent. Turkish lacks any form of gender concord and gender agreement errors are a very common phenomenon among Turkish learners of Dutch (Cornips & Hulk, 2008). The absence of a repair process in response to ungrammatical sentences that are uttered by someone who is expected to make these mistakes shows that the P600 is indeed susceptible to probability manipulations such as error probability. These results indicate that the P600 appears to be only influenced by factors within the subject’s control and hence this component is thought to reflect controlled processes.

1.4. Challenges and research questions

The main overarching question addressed in this dissertation is whether late L2 learners can process the Dutch language in a similar way as natives. The native language of the L2 learners (i.e. Polish) has been kept constant, but age of acquisition, length of residence and several different measures of proficiency varied in order to assess their impact on L2 knowledge and processing.

Thus far, scientists have mostly looked at late L2 learners from similar language backgrounds, such as Romance languages and/or Germanic languages and the gender systems under investigation often contained more cues than the Dutch language (e.g. morphophonological cues such as the suffix '–*a*' often denoting feminine gender in Spanish). The present study attempts to provide a contribution to the field of SLA by looking at the L2 acquisition of the non-transparent and asymmetrical Dutch gender system by learners from a typologically different language background, i.e. Polish. Moreover, previous research has mostly used behavioural and/or offline data to investigate L2 acquisition and these measures cannot adequately tap into the processes that are going on as speech unfolds over time. The present study additionally investigated a non-university population in order to provide evidence for similarities and differences between natives and L2 speakers that are generalizable outside a population of college students. Many of the research has thus far used visual stimuli, such as written sentences, to investigate L2 knowledge and processing. Since the late L2 learners in the present study have primarily acquired the L2 through listening and written comprehension might underestimate the actual abilities of the L2 learners, auditory stimuli were used in both the eye tracking and the ERP experiment. The next chapter provides a clear outline of the experiments and the experimental design that was chosen to come to a deeper understanding of language processing in both native speakers and L2 learners of Dutch.

The main question addressed in this dissertation is whether late L2 learners from a typologically different language are able to acquire and process the Dutch language in a similar way as native speakers. More specifically, the aim was to investigate whether the differences between native and L2 learners of Dutch concern fundamental or partial differences in either the representations and/or the processing of the gendered language. Before being able to answer this question, however, it was examined how native Dutch speakers use and process Dutch and its grammatical gender system.

The first question that was investigated was whether a non-transparent and asymmetrical gender system is used by native speakers to facilitate comprehension. In other words, are native Dutch speakers able to use gender-marked cues that precede the noun to guide their looks towards the correct referent depicted on the screen (see also Chapter 4). The second question addressed was whether late Polish-Dutch L2 learners behave native-like in their use of Dutch gender-marking. Assuming that Dutch natives can use gender-marking cues, the experiment outlined in Chapter 6 examined whether late L2 learners are also able to use Dutch gender cues to facilitate comprehension in the second language. Additionally, the eye tracking experiment aimed at examining whether, as predicted for example by the FTFA, late L2 learners activate and use lexico-syntactic knowledge such as gender from their native language, especially in earlier stages of L2 acquisition.

The event-related potential (ERP) technique was used to examine whether native Dutch speakers process grammatical gender in a similar way as compared to other morphological structures as well as whether the processing of gender can additionally be speeded up by a predictable context (see also Chapter 5). The fourth and final question was whether late L2 learners can process morphological structures in a native-like manner and aimed at investigating the influence of L2 proficiency and the availability and reliability of morphological structures in the input (see also Chapter 7). According to Ullman's (2001, 2004) Declarative/Procedural model, adult L2 learners might, in earlier stages of learning, show an N400 in response to syntactic violations, because they process it through the declarative memory system which instantiates lexical features. The processing measures might, however, also reveal mostly quantitative differences as in absent and/or delayed P600 effects in response to grammatical gender violations, instead of qualitative differences with the absence of early, automatic processing as would be predicted on the basis of the Shallow Structure Hypothesis (Clahsen & Felser, 2006a).

Polish-Dutch bilinguals do have a gender system in the native language, but as outlined in section 1.1.2, this gender system is not at all compatible with the Dutch gender system and Dutch gender largely has to be acquired on an item-by-item basis. If late post-puberty L2 learners have full access to UG, they might possess enough knowledge to perform accurately on both gender assignment and gender agreement and they might process gender in a similar way as native speakers do. If, however, acquisition of gender occurs mainly on the basis of input, the lack of consistent input due to the asymmetrical Dutch gender system and its abstract rules might cause L2 learners to have severe difficulty in the acquisition of the Dutch gender system. In terms of the Competition Model, one could argue that although Polish-Dutch learners possess a gender system in their native language, this knowledge can only partially be transferred to the L2 due to lack of overlap between categories and rules.

1.5. Outline of the dissertation

The next chapter will provide an overview of the experimental design, the tasks and measures used to asses L2 proficiency and the rationale behind these different tasks. Chapter 3 will provide an overview of the participants' performance on the different proficiency tasks outlined in Chapter 2. Chapters 4 and 5 will provide a detailed overview of native processing of gender as provided by evidence from eye tracking and event-related potentials respectively. Late L2 learners processing of grammatical gender will be dealt with in detail in relation to the eye tracking experiment in Chapter 6 and the ERP experiment in Chapter 7. Finally, Chapter 8 summarizes and discusses the combined results to answer the questions outlined in the present Chapter.

Chapter 2

EXPERIMENTAL DESIGN

The goal of this dissertation, in its broadest sense, is to compare grammatical gender processing in Dutch natives and late L2 learners of Dutch. Such a general question demands a combination of behavioural and online measures to examine how and when gender in Dutch is processed and whether these processes differ between these groups. In order to achieve this goal, two main experiments were devised that primarily focussed on the incremental nature of (second) language processing. The first study (as described in Chapter 4 and Chapter 6) examined whether prenominal Dutch gender-marking is used to facilitate language comprehension or whether, in the case of the L2 learners, the gender system from the mother tongue interferes. Section 2.3 of the present chapter will elaborate on the experimental design and rationale of this eye tracking experiment. In the second experiment described in this dissertation (Chapters 5 and 7), event-related potentials (ERPs) were recorded to examine the timing of gender processing and the potential differences and/or similarities between processing Dutch gender in the native and second language. The details of this second experiment are outlined in section 2.4 of the present chapter.

Before being able to compare the native and L2 group on how they process grammatical gender in Dutch, however, it is essential to gather independent measures on the basis of which these subjects can be matched and/or compared. The next section (§2.1) will discuss all pre-experimental details of how participants were selected and section 2.2 elaborates on how (L2) proficiency was measured.

2.1. Subjects and data collection procedure

Two main groups of subjects were included in the present dissertation: a native control group and a group of late L2 learners of Dutch with L1 Polish. Initially, one of the main goals was to investigate the impact of the mother tongue, i.e. the presence or absence of an L1 gender system. Therefore, the initial intention was to compare Polish and Turkish learners of Dutch. Turkish is not only a language that does not have any form of gender, but, like Polish, it also lacks determiners and originates from a different language family than the Dutch language. In other words, the presence or absence of an L1 gender system would have been likely to be the cause of any potential differences in the acquisition and processing of Dutch gender between the two groups. Unfortunately, it proved to be impossible to attract enough late L2 learners of Dutch with Turkish as a mother tongue due to religious, cultural, and work-related issues. It was therefore decided to compare early and late Polish learners of Dutch to investigate the potential influence of age of acquisition. The decision to compare late L2 learners to native Dutch speakers as well as simultaneous bilinguals was, however, taken at a later stage and hence this latter group was too small to be included in the present dissertation.

The decision to collect a wide variety of proficiency measures, as outlined in the present and next chapter, was also only taken at a later stage and hence only a small portion of participants from the native control group performed these different tasks. Fortunately, data from the new group of participants had started being collected and these simultaneous Polish-Dutch bilinguals could be used to compare performance on behavioural tasks as described in the present and next chapter.

2.1.1. Participant selection

Simultaneous bilingualism was defined as having acquired two languages, i.e. Polish and Dutch, from birth or before one year of age (De Houwer, 2005). Late L2 learners were defined as those individuals who started learning L2 Dutch after the age of 16. The overall characteristics of the three groups are presented in Table 2-1.

Participants were recruited by means of advertisements in the local newspaper, the online Polish newspaper Polonia, Humanitas Groningen and by means of social media such as Hyves and Facebook. In most, if not all studies that are dependent on the response rate of potential participants, women tend to be overrepresented because (among other reasons) they are often still more likely to work part-time jobs than men and (based on my own experience) they are often more willing to spend their free time in a lab to make their own contribution to science. Table 2-1 reveals that women are overrepresented in all the subjects groups, but particularly so in the two L2 groups. Of course, we had more than enough responses from potential Dutch participants and hence had the luxury to choose our participants on the basis of factors such as age and education. The minimal amount of Polish-Dutch late acquirers, and especially the Polish-Dutch bilinguals, required us to accept any registration without being overly critical. For this same reason, our first group (Native Control Group) contained a considerably wide variety of ages, ranging from 34 to 60. Data collection for the native control group had started well before the collection of the L2 learners' data and, since the main aim was to examine late, but advanced learners of Dutch, a similar broad range of ages in the late L2 learners' group was initially expected. Unfortunately, we did not manage to achieve the same age range in the other two groups, but all late L2 learners had acquired Dutch after the age of 20 and hence they all fall well into the category of late acquirers. The late L2 learners did differ considerably in their Length of Residence (LoR) in the Netherlands, which ranged from 2 to 27 years.

Table 2-1: Overview of the three groups of participants used in the present study. The table provides information on the division of genders, ages, ages of acquisition (AoA), and lengths of residence in the Netherlands (LoR) in the different groups.

Group	**Female/Male**	**Mean Age** (range)	**AoA** (range)	**LoR** (range)
Native Control Group (*n* = 28)	16/12	48 (34-60)	-	-
Late L2 learners (*n* = 27)	21/6	34 (24-52)	25 (20-34)	9 (2-27)
Simultaneous bilinguals (*n* = 6)	4/2	24 (19-30)	0	-

Due to the lack of choice of L2 learners, most of the participants in the L2 groups also turned out to be highly educated with more than three-quarter of the groups having a university degree. This is also depicted in Table 2-2.

Table 2-2: Overview of the distribution of educational levels among the different groups of participants.

Group	**Secondary School**	**Secondary Vocational Education**	**Higher Professional Education**	**University**
Native Control Group	6 (21.43%)	7 (25.00%)	6 (21.43%)	9 (32.14%)
Late L2 learners	0 (0.00%)	3 (11.11%)	3 (11.11%)	21 (77.78%)
Simultaneous bilinguals	0 (0.00%)	0 (0.00%)	1 (16.67%)	5 (83.33%)

Although the distribution of educational levels in between the groups is not optimal, it does not endanger our intention to directly compare the different groups. If, as would be expected, the native control group outperforms the late L2 learners of Dutch, this result could not be attributed to an effect of education, since such an effect should only benefit the latter group.

2.1.2. Collection of subject proficiency and experimental data

The data collection procedure started with an online questionnaire to ensure suitability of the subject and this was sent to all interested participants. The first appointment was made in the eye tracking lab in Groningen, where we first expanded our knowledge of the participant by administering a ***sociolinguistic questionnaire*** (§2.1.3). After that, the participant performed a ***film retelling task*** (§2.2.1) and filled out a ***C-Test*** (§2.2.2) to assess an overall level of L2 proficiency. Finally, the subject was seated to perform the eye tracking experiment explained in §2.3 (see also Chapters 4 and 6). The second appointment took place at the NeuroImaging Center on a different day.[5] During this second meeting, the participant performed the EEG experiment which is covered in Chapters 5 and 7. During this experiment, they performed an online version of the ***acceptability judgement task*** which will be explained in section 2.2.3. The rationale behind the event-related potentials experiment is further dealt with in detail in §2.4 of the present chapter. After the EEG experiment, participants in the L2 groups were asked to perform an extra ***gender assignment task*** (§2.2.4) to test their conscious knowledge of the Dutch gender system.

The scores on all these different tasks were analysed separately (as presented in Chapter 3) and an average L2 proficiency score based on these different tasks was calculated and taken along in the analyses presented in Chapters 5 and 7.

2.1.3. Sociolinguistic questionnaire

A large sociolinguistic questionnaire was administered to collect personal information about the subjects as well as how they assessed their own language use and proficiency. The questionnaire consisted of two parts: an online questionnaire that was used to screen and select the participants and one additional questionnaire containing several questions concerning their use of the native and, for L2 learners, second language. The online Recruitment and Selection questionnaire can be found in Appendix A-1. This online questionnaire ensured that all participants were right-handed as assessed by an adapted version of the Edinburgh Handedness Inventory (Oldfield, 1971, see also Appendix A-1) and that none of them suffered from colour-blindness. All participants had normal or corrected-to-normal vision and hearing and none of them had ever suffered from any neurological complaints or deficits.

The sociolinguistic questionnaire, adapted from the test battery developed by Schmid (2011), contained many general subject-related questions to be included in later analyses, such as, e.g. age, sex, place of birth, level of education, and current job position. Subjects were also asked about their linguistic history: whether they had ever

[5] In most cases, the second appointment was scheduled a week after the first appointment to reduce any effects of fatigue. Unfortunately, some participants had to travel quite a distance to come to Groningen and therefore we decided to schedule both appointments on the same day, but with a rather long break in between.

lived in other countries and what languages they had learned before they went to school. They were asked to make specific lists containing the languages they learned before and at school and to fill out a Likert-scale on how well they were at speaking, listening, reading, and writing in the specific language.

Besides age of acquisition (AoA) and length of residence (LoR) as important variables, L2 participants were asked to rate themselves on their Dutch and Polish as well as to provide percentages on how often they spoke Dutch with several groups of people. They were also asked questions about Dutch courses and how often they still visited their home country. The results from this questionnaire will be discussed in detail in section §3.1 of Chapter 3 and the full questionnaire can be found in Appendix A-2.

2.2. Measuring L2 proficiency

Previous investigations comparing native and second language processing have mostly used one particular measure to provide information on how well the participants are in producing or comprehending the second language. One problem is that researchers do not agree on a single appropriate standardized language assessment test. Furthermore, learners are not always equally proficient in the different domains (speaking, reading, listening, and writing) nor does everybody have equal skills that are required for certain tasks (more details will follow below). Most important to the present study is that even advanced L2 learners have been shown to have severe difficulty with L2 grammatical gender and the acquisition of gender or other linguistic aspects can thus not always predict the overall proficiency level of the L2 learner. Since L2 proficiency level is highly likely to influence L2 learners' performance, this variable was measured extensively across linguistic aspects and across domains. By averaging over these various measures, a general proficiency measure was calculated in which potential variability between skills is averaged out. Additionally, the separate measures could be used to more precisely examine which facets contribute to native-like processing of grammatical gender in Dutch.

2.2.1. Film retelling task

Although the main aim of the studies presented in this dissertation is to investigate spoken language comprehension, a test to elicit semi-spontaneous speech was also included as a measure of proficiency. Free speech requires the person to think of the correct words to use to refer to the objects and concepts of interest, but also to put them into a meaningful and grammatical sentence, and finally pronounce this sentence so that it is understandable to the interlocutor. In order to produce speech, we thus need to incorporate conceptual-semantic, morphological, syntactic and phonological/phonetic information and, if, for example, the L2 learner has not yet acquired certain aspects of the language, he or she is likely to have trouble producing speech containing this particular aspect. This makes free speech a very important factor in determining one's general and specific proficiency level.

An excerpt of about 10 minutes from the 1936 Charlie Chaplin film *Modern Times* was used to elicit semi-spontaneous speech. The procedure used was based on the original description of the task by Clive Perdue (1993), which has been used extensively in Second Language Acquisition (SLA) research. The silent movie has the particular advantage that it contains a story-line with a prominent role for a poor girl that steals a piece of bread. The words 'girl' and 'bread' are neuter in Dutch, so the participants are more or less pushed to use neuter nouns in their narrative, which are less common in the Dutch language. Furthermore, participants tend to have few problems following the story-line and tend to find it rather amusing to watch (also see Schmid, 2011).

After completion of the sociolinguistic questionnaire, the subject was told that they would be watching a 10-minute excerpt of a Charlie Chaplin movie after which they would be asked to tell what they had seen and what had happened to Charlie and the poor girl. They were not interrupted during the narrative of their version except for a couple of times when the participant only told a very brief summary of the story. In that case they were asked to elaborate in order to provide more data to base the proficiency score on.

The semi-spontaneous speech data could have been analysed extensively on several aspects, but for the present studies we only incorporated a general, holistic proficiency score and a specific analysis of the use of grammatical gender in production. ***Holistic proficiency*** was assessed by asking three linguists to listen to the free speech of each participant and rate them on fluency, pronunciation, intonation, syntax, and lexicon on a scale from 1 (very basic) to 6 (native-like). Fluency referred to the amount of hesitations, corrections, pauses and silences, false starts and stuttering. Pronunciation concerns the 'thickness' of the accent: the more Polish sounding the pronunciation, the lower the score. The native-likeness of the melodic pattern of the utterances (pitch and/or tone) was translated into a score for Intonation. Each subject was rated on their syntactic skills: the less accurate the grammar, the lower the score for Syntax. A larger amount of basic frequent words in the utterances was translated into a lower score on Vocabulary, while the use of more advanced, complicated, and less frequent words was associated with a higher score on Vocabulary. The average of these scores, the overall holistic proficiency score, was included in the average proficiency score. Intraclass correlation coefficients (ICC) were calculated to assess the level of agreement or consensus between the three raters (Howell, 2009). The average ICC measure revealed excellent interrater reliability across domains (all *r*s > .836) and provided a justification to analyse the average of holistic proficiency scores across raters.

To find out whether the late L2 learners could assign the correct gender to nouns as well as whether they were accurate in applying the rules of gender agreement, ***gender in production*** was analysed to obtain a separate score for the use of gender-marking in semi-spontaneous speech. All constructions containing a noun that had to be preceded either by a gender-marked determiner or by a gender-marked article were counted and classified as being either correct or incorrect. Since the Dutch two-way gender system is asymmetrical and previous research has shown differences in the acquisition of common and neuter gender (see, e.g. Unsworth & Hulk, 2010), a gender production accuracy score was provided for common and neuter nouns separately. Table 2-3 gives an overview of the possible constructions containing gender assignment or agreement in Dutch.

Table 2-3: Overview of the possible constructions in production containing gender-marking. Note that the determiner + noun constructions concern definite noun phrases (NPs) and the adjective + noun constructions concern indefinite NPs.

Construction	Gender	Correct	Incorrect
Determiner + Noun	Common	***de*** vrouw	****het*** vrouw
	Neuter	***het*** meisje	****de*** meisje
Adjective + Noun	Common	***arme*** vrouw	****arm*** vrouw
	Neuter	***arm*** meisje	****arme*** meisje
Demonstrative Pronoun	Common	***die/deze*** vrouw	****dat/dit*** vrouw
	Neuter	***dat/dit*** meisje	****die/deze*** meisje
Relative Pronoun	Common	vrouw ***die***	* vrouw ***dat***
	Neuter	meisje ***dat***	* meisje ***die***

The analysis of gender in production will focus mainly on determiner-noun and adjective-noun constructions, as these are also used in the eye tracking and the ERP experiment (see §2.3 and §2.4). The taxonomy of errors outlined in Table 2-3 was used to asses gender in production (Montrul, Foote, & Perpiñán, 2008; Grüter, Lew-Williams, & Fernald, 2011). The decision to collect free speech data was only taken after data collection from the L2 learners had started and only 8 participants from the native control group could be tested on their free speech. Fortunately, story retellings were collected from the 6 simultaneous bilinguals who were included in these analyses. The results of these analyses are described in detail in Chapter 3 (§3.2).

2.2.2. C-Test

Another measure of overall language proficiency, and more specifically an assessment of vocabulary, was provided by the C-Test, which consists of two short texts of which parts of words were systematically deleted. The C-Test used was adapted from Keijzer (2007). The original version of the C-Test consisted of 5 texts, but due to the length of the experiments conducted for the present dissertation, we chose two of these texts instead. The first sentence of the texts was complete and the gapped words started from the second sentence onwards. Text 1 contained 18 gapped words and text 2 contained 20 gapped words and the texts can be found in Appendix A-3. Participants were given no more than five minutes to complete each text.

For the present study the more elaborate 0-9 scoring system was used (cf. Keijzer, 2007; Schmid, 2005) to obtain a score as distinctive as possible. The answers were scored using the following criteria to assign points to each item:

0 = empty
1 = incorrect lexical stem and incorrect word class
2 = incorrect lexical stem but correct word class
3 = correct lexical stem but incorrect word class
4 = correct lexical stem, correct word class, agreement error
5 = all of above correct, but still slightly wrong
6 = acceptable variant with spelling error
7 = correct word spelling error
8 = acceptable variant
9 = correct word

The proficiency score for the C-Test is the total score for all items, with a maximum of 342. These scores were used to calculate correlations with other variables of the dataset and the percentage correct for each participant was taken along in the score for the overall proficiency level.

As was the case with the film retelling, the decision to include the C-Test as a measure of general proficiency was taken once data collection had already started. Fortunately, 15 participants from the native control group were willing to participate in this C-Test allowing for sound statistical analyses to be made. The results of these analyses of the C-Test are outlined in detail in §3.2.

2.2.3. Online Acceptability Judgement Task

During the ERP experiment (also see §2.4), participants listened to sentences that were either correct or contained a grammatical violation. After participants had heard the sentence, they were asked to judge them with respect to acceptability. This procedure allowed us to check the performance of the L2 learners as well as compare their behavioural responses to the online brain activations.

The grammatical violations either consisted of rather obvious grammatical violations of finiteness that were adapted from Sabourin (2003) or violations of grammatical gender. The finiteness condition consisted of sentences in which the critical verb was a past participle or a finite form of the verb. In the ungrammatical conditions, these two forms were switched as can be seen example sentence 2-1, where the critical verb is printed in bold. The grammatical gender violations consisted of a noun that was preceded by either the wrong determiner or the incorrectly inflected adjective (example sentence 2-2). More details on the construction of these sentences can be found in §2.4.

2-1. Ze heeft alleen haar beste vriendin **uitgenodigd/*uitnodigen** voor haar verjaardag.
*She has only **invited/*invite** her best friend to her birthday.*
2-2. Vera plant rode rozen in **de$_{\mathbf{COM}}$ tuin$_{\mathbf{COM}}$/*het$_{\mathbf{NEU}}$ tuin$_{\mathbf{COM}}$** van haar ouders.
*Vera is planting red roses in **the$_{\mathbf{COM}}$ garden$_{\mathbf{COM}}$/*the$_{\mathbf{NEU}}$ garden$_{\mathbf{COM}}$** of her parents.*

In Sabourin's study (2003), all Dutch L2 learners were familiar with the finiteness construction, which was the same in their native language (i.e. English, German, and Romance), and could process finiteness in a native-like manner. Grammatical gender, on the other hand, was only present in the mother tongues of Romance and German learners of Dutch. Sabourin (2003) showed, however, that only German learners could process Dutch gender agreement violations similar to natives and concluded that having an L1 gender system is not enough: the systems additionally need to be similar. In the present dissertation, Polish-Dutch L2 learners are familiar with both constructions, but the constructions are not similar in the two languages. The two constructions were thus compared to assess whether constructions indeed need to be similar in the two languages and/or whether the frequency and reliability of the structures in the input might influence the acquisition process (see §7.1.3 for a detailed explanation of the differences and similarities of these constructions in Polish and Dutch).

Many ERP studies have used the accuracy rates on judgement tasks as a (single) measure of L2 proficiency. To avoid a focus on a single linguistic aspect, such as grammatical gender, the subjects in the present study were asked to judge the overall acceptability of the sentence. It is thus not possible to determine whether participants' judgements are based on the presence or absence of the actual grammatical violations in the sentence or whether it was something else that made them judge the sentence as incorrect. Nevertheless, accuracy and reaction times might reveal relative differences between natives and L2 learners on the one hand and judgements in response to finiteness and gender violations on the other hand. Both accuracy and reaction times were therefore analysed and the results are outlined in Chapters 5 and 7. In addition, the relationship between these acceptability judgement scores and the other proficiency measures outlined in the present chapter were investigated and the accuracy scores were taken along in the overall L2 proficiency score (see §3.4).

2.2.4. Gender assignment task

To assess whether the L2 learners actually knew the gender of the nouns they had heard in the eye tracking and the ERP experiment, the final task was a gender assignment task. Participants were asked to assign the correct article (*'de'* or *'het'*) to every noun they had previously heard in one of the experiments. Every noun was tested three times during this task to avoid chance level results.

The results of this offline gender assignment task could show whether L2 learners are indeed more proficient in assigning gender (an interpretable feature) as compared to processing grammatical gender agreement (the uninterpretable features) as would be assessed by the eye tracking and ERP experiments. Moreover, the results allowed for a distinction between nouns of which the subject knew the gender as opposed to nouns of which they either did not know the gender or acquired the incorrect article-noun combination. The gender assignment task can be found in Appendix A-4 and the analysis of the assignment scores can be found in §3.5.

2.3. Using eye tracking to look at grammatical processing

The eye tracking experiment was designed to assess the impact of the Dutch gender system on auditory sentence processing. The setup used was based on the 'visual world paradigm' (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) in which subjects are asked to manipulate the objects in their visual scene. The general idea behind this eye tracking experiment comes from the observation that people tend to guide their eyes towards objects in the visual scene that correspond to what they hear in the speech stream (Cooper, 1974; Tanenhaus et al., 1995). The use of auditory stimuli, as in the visual world paradigm, was specifically chosen for the present study because late L2 learners from a non-university population are likely to have acquired the second language through listening and hence spoken language comprehension is likely to be a better acquired skill than reading in the L2.

The experiment was designed to answer questions concerning whether natives and L2 learners with a gender system in their L1 are able to use Dutch gender concord-based cues when looking at pictures on a screen or, in the case of late L2 learners, whether they use the gender from their mother tongue instead. All participants took part in the eye tracking experiment before participating in the ERP experiment, because the ERP experiment contains a lot of input regarding gender agreement including gender agreement errors and hence provided information with respect to the goal of the experiment.

It was hypothesized that gender-marking before the noun, such as the definite determiners in definite noun phrases (*'de*$_{\text{COM}}$*'* or *'het*$_{\text{NEU}}$*'*) and the presence or absence of a schwa-ending on the adjective in indefinite noun phrases (*'rode*$_{\text{COM}}$*'* or *'rood*$_{\text{NEU}}$*'*) might be used to pre-select the upcoming noun from a set of pictures on the screen (e.g. Dahan, Swingley, Tanenhaus, and Magnuson, 2000). For a more detailed overview of the literature concerning native and bilinguals processing of gender cues, see Chapters 4 and 6 respectively. To investigate this hypothesis, participants were instructed to look at a screen containing four coloured pictures while listening to instructions to click on one of the four objects displayed on the screen: e.g. *Klik op de rode appel* ('Click on the$_{\text{COM}}$ red$_{\text{COM}}$ apple'). The nouns depicted varied among two dimensions: syntactic gender and colour and the screen contained four pictures: the target, a competitor, and two unrelated distractors. The competitor's role differed across 4 conditions in which the competitor and the target either:

(1) matched in gender and in colour;
(2) matched in gender, but not in colour;
(3) matched in colour, but not in gender;
(4) matched neither in colour, nor in gender.

The two distractors on the screen never matched in colour nor gender with the target, as also in condition (1) mentioned above. One would expect that, if people use gender-marked information in the speech stream to facilitate comprehension, eye movements towards the target would occur earlier if the gender of the target is unique as compared to situations when more pictures on the screen share gender with the target. In the example display below, for example, participants would be asked to, e.g. *Klik op de rode appel* ('Click on the$_{COM}$ red$_{COM}$ apple'). While colour information is not disambiguating in this specific example, participants might use the common definite determiner *'de'* to pre-select the apple since common gender is unique to the target.

Since the adjective receives no gender-marking in the definite case, an indefinite trial was included in which participants were asked to *Klik op het plaatje met een rode appel* ('Click on the picture with a red$_{COM}$ apple'). Inclusion of these two constructions allows for an investigation of the potential difference in use of gender-marking on definite determiners and the less salient gender-marking (i.e. the presence or absence of the schwa-ending) on attributive adjectives. Additionally, inclusion of these two constructions allows examining whether gender-marking is used to pre-select the upcoming noun regardless of the reduced amount of time available to use these cues in the indefinite as compared to the definite condition.

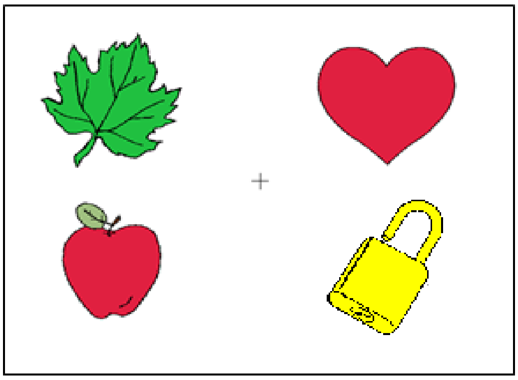


Figure 2-1: Visual display of condition (3) during the eye tracking experiment. Participants were asked to, e.g. *Klik op de rode appel* ('Click on the$_{COM}$ red$_{COM}$ apple'). In this specific case, it was hypothesized that, if people use gender-marked cues to pre-select the upcoming referent, participants would look less towards the picture of a red heart even before hearing the word *'apple'*, since heart is a neuter noun that could not have been preceded by the common definite determiner *'de'*.

All nouns used were translated into Polish to make sure no cognates were used as well as to include information with respect to the gender of the noun in Polish. In order to be able to investigate whether Polish learners of Dutch activate and/or use knowledge about their L1 gender system while comprehending in their second language, half of the nouns/pictures used overlapped in gender between Dutch and Polish (i.e. the nouns were neuter in both languages or common in Dutch and feminine or masculine in Polish) while the other half of the items used did not overlap in gender in Polish and Dutch. The exact target-competitor pairs used in the experiment can be found in Appendix B-1. The experiment was set up using E-prime extensions for Tobii software and a description of how this was programmed can be found in Appendix B-2.

A detailed description of the literature, methodology, and results of this eye tracking experiment are outlined in Chapter 4 (native Dutch speakers) and Chapter 6 (late L2 leaners).

2.4. Making use of brain potentials to examine grammatical processing

ERP studies to date have mostly used visual stimuli such as written sentences to investigate the processes involved in language comprehension. However, as Müller, King, and Kutas (1997) have pointed out, listening is just as important to investigate as reading if we want to know more about how people process linguistic features. Müller et al. (1997) rightly address the fact that auditory language comprehension is more natural than reading. Not only is reading a multimodal process that has taken the human species quite some time to create, but it is also a rather complex skill. People are able to listen and discriminate sounds at a very early age, whereas reading has to be learned explicitly at a later stage in life. The reason that most researchers do not use auditory stimuli in ERP studies is simply because it is a very complex and long process and auditory stimuli are more difficult to create than written material. Many visual ERP experiments have suggested that L2 learners are not able to process L2 grammars similarly to native speakers of the language. The learners' overall ability in the L2 might, however, be highly underestimated when looking solely at the complex processes involved in reading. An additional important reason for using auditory stimuli in the present study concerns the fact that L2 learners (in this case Polish learners of Dutch) mostly receive auditory input when acquiring their second language. And, above all, investigating more natural processes of language comprehension, as involved in listening, are worth the extra work.

The primary objective of the ERP experiment was to investigate whether (1) native speakers process grammatical gender similarly as compared to other morphological structures and (2) whether L2 learners are able to process morphological information in the L2 that is either frequently, noticeably and reliably present in the input or not: finiteness and grammatical gender. In addition, as gender-marking might help native speakers to anticipate upcoming words (as tested in the eye tracking experiment), semantic expectancy of the noun might also facilitate language comprehension. Therefore, cloze probability or semantic expectancy was added as a semantic variable to investigate potential interactions between gender agreement and cloze probability in native speakers (see Chapter 5 for more details).

Differences in morphological processing between natives and late L2 learners might constitute qualitative differences such as the absence of a P600 effect reflecting re-analyses of the morphological violation, or the presence of an N400 instead of the native-like P600 reflecting initial stages of learning in which the specific construction has been acquired as one chunk of information instead of a morphologically combined set of words (as predicted by Ullman's DP Model, 2004). One might also expect L2 learners to process morphological information fundamentally similar to natives, resulting in a native-like re-analysis process that only differs between the groups on a quantitative level, such as later or smaller effects in L2 learners as compared to native speakers. For details concerning previous research on gender processing and event-related potentials and the results of these experiments, see Chapters 5 (native processing) and 7 (L2 processing).

2.4.1. Pre-testing of stimuli

Since brain activation is influenced by many factors, such as frequency, cloze probability and plausibility of stimuli, these variables were tested and controlled for in the ERP experiment. Note that this is more important for the analyses of the ERP data than the eye tracking data, since the sentences used in the eye tracking study were equal up to the noun phrase. Moreover, eye data were analysed using mixed-effects regression models in which the frequency of the individual items could be taken into account by putting it in the model (see §4.2.5 for more detail on these statistical analyses). Since ERP data has to be analysed using repeated measures ANOVA, in which data is averaged across items and subjects, a balanced design is required. A sentence completion task was administered to assess the variable cloze probability, which was operationalized as the percentage of respondents who filled out the target word in a sentence completion task (for more details about this test, see section 5.2.2). Additionaly, a plausibility test, asking subjects to rate the experimental sentences on their acceptability and plausibility, was administered to make sure all sentences were equally plausible. This plausibility test consisted of all experimental sentences randomly intermixed with implausible filler sentences. Subjects were asked to judge each sentence on a Likert scale from 1 (very implausible) to 5 (completely normal).

Experimental sentences in the grammatical gender condition were divided into three different noun phrase constructions: determiner-noun (DN), definite determiner-adjective-noun (DaN), and indefinite determiner- adjective-noun (IaN). More information on these constructions can be found in §5.2.2. Cloze probability (percentage of expectancy) of the target nouns was manipulated. A word's cloze probability is based on the percentage of individuals that continue a sentence fragment with that word. For every sentence, a high and low cloze variant were created that were as similar as possible and matched on sentence length, sentence construction and frequency of the words used. In most cases, only one or two words differed between the two conditions.

Items were regarded as high cloze when cloze probability was above 70 %. Low cloze items always had a cloze probability below 30 %. Only high frequency nouns were selected as critical items to make sure that the L2 learners of Dutch were familiar with the words. The attributive adjectives used were also high frequent ones and they were used equally often across conditions. Initially, the goal was to use a different adjective for each item, but this seemed impossible due to the high co-occurence of adjective-noun items, which made them too predictable and hence impossible to include in the low cloze conditions. Table 2-4 below shows the average percentage and confidence intervals of the items in high- and low cloze conditions respectively. The average relative frequency of the critical nouns in each condition is presented in the leftmost column of the table.

Table 2-4: Average relative frequency as well as the average cloze probability scores for high and low cloze conditions and the corresponding 95% confidence intervals (CI). Note that the same nouns were used in both the high and the low cloze sentences.

Type of NP Construction	Gender	Average Relative Frequency of items	Average High Cloze (%)	Average Low Cloze (%)
Definite determiner	Common	0.32	89.5 (CI=[66,111])	3.6 (CI=[-1,23])
+ Noun	Neuter	0.34	87.3 (CI=[68,109])	8.7 (CI=[-10,32])
Definite + adjective	Common	0.57	85.6 (CI=[65,112])	9.2 (CI=[-14,36])
+ Noun	Neuter	0.70	92.2 (CI=[72,105])	17.3 (CI=[-20,41])
Indefinite + adjective	Common	0.72	85.8 (CI=[65,112])	9.3 (CI=[-15,37])
+ Noun	Neuter	0.43	91.2 (CI=[69,108])	17.2 (CI=[-17,38])

To make sure that the sentences were equally plausible across conditions, an additional test was administered in which subjects were asked to rate the sentences on a Likert-scale from 1 (very implausible) to 5 (completely normal). 68 Native Dutch volunteers (mean age: 21.4, range 18-33; 49 female) participated in this test. Sentences were only used when they were scored above 3.5. The table below shows the means per condition as well as the corresponding confidence intervals.

Table 2-5: Average plausibility scores across conditions for high and low cloze items and the corresponding 95% confidence intervals (CI).

Type of NP Construction	Gender	Average High Cloze Plausibility (1-5)	Average Low Cloze Plausibility (1-5)
Definite determiner	Common	4.7 (CI=[4.1,4.9])	4.5 (CI=[3.8,5.2])
+ Noun	Neuter	4.7 (CI=[4.0,5.0])	4.6 (CI=[3.6,5.4])
Definite + adjective	Common	4.5 (CI=[3.6,5.4])	4.4 (CI=[3.8,5.2])
+ Noun	Neuter	4.7 (CI=[4.1,4.9])	4.1 (CI=[3.4,5.5])
Indefinite + adjective	Common	4.6 (CI=[4.1,4.9])	4.4 (CI=[3.6,5.4])
+ Noun	Neuter	4.6 (CI=[3.7, 5.3])	4.3 (CI=[3.6,5.4])

2.5. Auditory stimuli preparation

For both experiments, sentences were digitally recorded in a soundproof booth by a female native speaker of Dutch with natural intonation and spoken at a normal rate. The sampling rate used was 44100 Hz. After the recordings, every sentence was checked for outliers in intensity. In language production, the beginning of the sentence is often pronounced with more volume which decreases over time (also see Figure 2-2). For these experiments, the sentences were required to be of the same volume throughout the sentence to avoid artifacts caused by (sudden) changes in volume. Therefore, volume within the sentences was adjusted, which can be seen in the Figure 2.2 below.

In addition to adjustments of the volume of the soudwaves within sentences, soundwaves were adjusted to make sure that they all had the same volume within and across conditions. After adjusting the volume within each sentence, all sentences were normalized to a mean intensity level of 70 dB SPL.

The results of the eye tracking and ERP experiment are outlined in detail in Chapters 4, 5, 6, and 7.

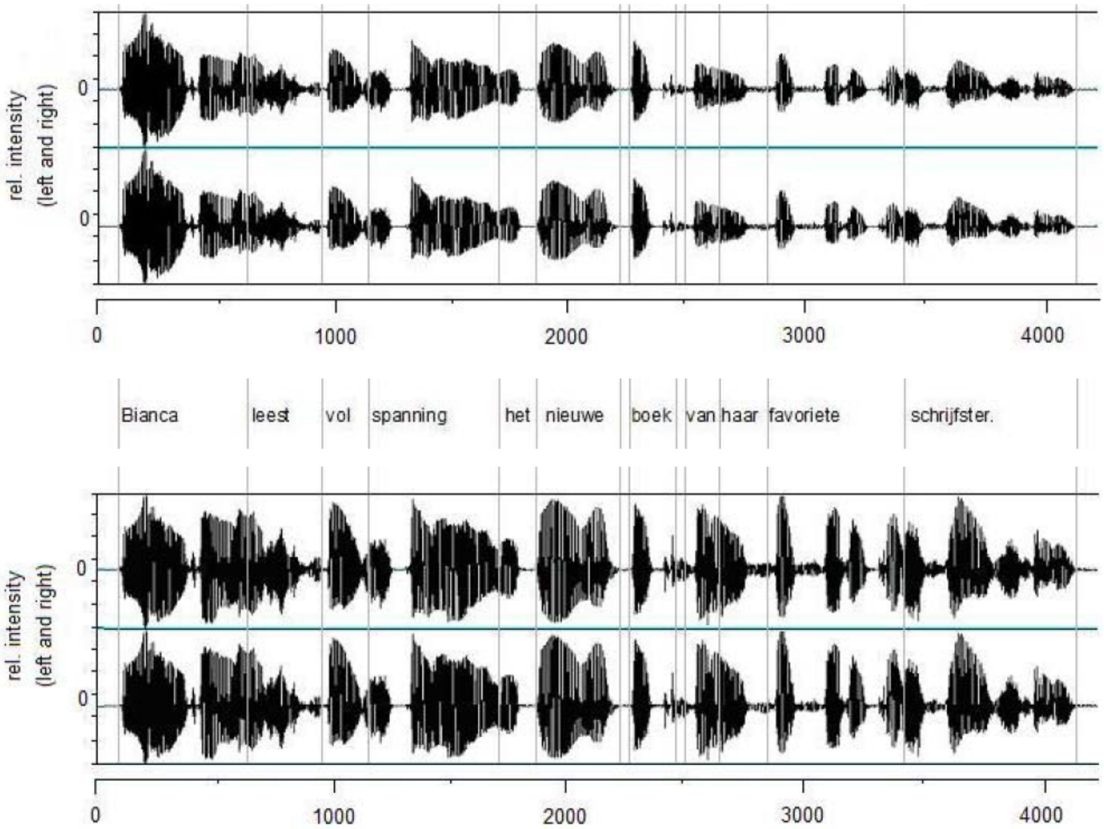


Figure 2-2: From Loerts (2009). An example of a soundwave of which the volume has been adjusted, in this case a sentence from the ERP experiment. As can be seen in the upper panel, the raw soundwave, the beginning of the sentence contains an outlier in volume. By decreasing the volume of this outlier, the volume of the soundwave was adjusted in such a way that the volume became equal across the whole sentence, as can be seen in the lower panel (the soundwave after volume adjust).

Although the main aim of the present dissertation is to investigate the online processes of (second) language comprehension while speech unfolds, the outcomes of the different proficiency measures outlined in the next chapter will provide information as to whether late L2 learners have acquired the Dutch language in general and the Dutch grammatical gender system in particular. The chapter provides an overview of these different proficiency scores as well as how they correlate to one another and to other personal characteristics of the subjects, such as age of acquisition and length of residence.

Chapter 3

BEHAVIOURAL MEASURES & L2 PROFICIENCY

The subjects who participated in the experiments on which the present dissertation is based performed several behavioural tasks (see also Chapter 2). Taken together, these tasks were used to assess the L2 proficiency of the late L2 learners of Dutch as compared to those participants who were born in the Netherlands. Since not all of the subjects from the native control group had performed all behavioural tasks presented here, this latter group also contained Polish-Dutch simultaneous bilinguals (also see §2.1.1). The present chapter provides an overview of these behavioural measures and the relationship between the different task outcomes.

Important variables for all three groups contain sex, age, and educational background. For each score that was analysed, the influence of these factors was also assessed and reported if significant.

3.1. Sociolinguistic Background Questionnaire

The sociolinguistic questions that were posed to all the participants were whether they had ever lived in another country and whether they had learned any other languages before attending primary school. Four out of the 28 native Dutch people had lived in another country for over 6 months and 10 out of 27 of the late L2 learners reported to have lived in another country (besides Poland and The Netherlands) for over 6 months. Only one simultaneous bilingual had lived abroad and, in this case, she had lived in Poland for 6 months. Only three out of 28 Dutch natives reported having learned another language or dialect (i.e. Frisian or the Groningen dialect) before primary school. For the late L2 learners, 7 out of 27 had learned a dialect next to Polish at a young age.

Of course, all 6 simultaneous bilinguals had learned both Dutch and Polish before primary school.

Almost all of the late learners of Dutch (26 out of 27) reported that they visit their home country once or twice every year. Only one person visited Poland less often: once every 3 to 5 years. Their visits normally lasted less than two weeks (24 of 27) or between 2 and 4 weeks (3 out of 27). In the group of simultaneous bilinguals, there was also only one person who went to visit Poland once every 3 to 5 years. The other 5 reported to visit relatives in Poland once or twice a year. Half of them stayed less than 2 weeks and the other half stayed a bit longer (between 2 and 4 weeks).

The simultaneous bilinguals and the late L2 learners were asked to rate themselves on both their proficiency in Dutch and their proficiency in Polish on a Likert-scale from 1 (very bad) to 5 (excellent). All simultaneous bilinguals regarded themselves as being completely native-like in Dutch, with only half of them (3 out of 6) reporting to be equally native-like in Polish. The other half reported to be either fairly proficient in Polish (3 and 3.5) or almost native-like (4.5). Almost all late L2 learners (24 out of 27) regarded their L1 proficiency as excellent, but three of them rated it as pretty good with a score of 4. When rating their overall proficiency in Dutch, one person gave himself 2 out of 5 points. Five people regarded their L2 proficiency as reasonable (3 out of 5) and 4 of them a little better than reasonable (3.5 out of 5). Eleven people thought their Dutch was good (4 out of 5) and 6 even reported to have native-like proficiency in their L2 (5 out of 5). Not surprisingly, all Dutch natives rated their Dutch as native-like with a score of 5 out of 5.

An average Dutch ***Self-rating*** score was calculated from this overall score and the scores on the different modules, i.e. reading, writing, speaking, and listening, since they all correlated with each other. Dutch natives and simultaneous bilinguals were not analysed with respect to their self-rating since they all scored at ceiling (with only a few of them rating some of the module(s) at 4).

Within the group of late L2 Learners, a marginally significant negative correlation was found between *Self-rating* and Age of Acquisition (AoA: $r_s(26) = -.38$, $p = .052$), showing that the later people had started learning Dutch, the lower they tended to rate their skills in Dutch. Length of Residence (LoR) was not significantly correlated with *self-rating* ($r_s(26) = 0.32$, $p = 0.11$) and the correlation between AoA and LoR did not reach significance ($r_s(26) = -0.18$, $p = 0.37$).[6] *Self-rating* was added to the overall L2 proficiency score that was used in subsequent analyses.

Most participants had been taught about 3 languages at school; those people born in the Netherlands mostly learned English, German, and French. The Polish participants mostly learned Russian, German and English in high school. When asked which language was easier for them at this point in their life, 15 out of 27 late L2 learners said it was still their mother tongue, Polish. Eight of them said that it did not

[6] The Spearman's rho correlation coefficient was used in all correlational analyses presented in the present chapter, because the dependent variables under investigation violated the parametric assumptions (Field, 2005).

really matter, and 4 of them regarded Dutch as the easier language. All except one simultaneous bilingual, to whom both languages were equal, said that Dutch was easier than Polish. When asked how often they spoke Polish, 15 out of 27 late L2 learners said they spoke Polish on a daily basis, 11 spoke their mother tongue weekly, and 1 reported speaking Polish only monthly. The simultaneous bilinguals had a similar division with 2 out of 6 speaking Polish daily, 3 of them weekly, and one monthly.

On average, the late Polish-Dutch L2 learners said that about half of their friends were Dutch (range: 10% – 98%) and about 35% were Polish (range: 2% - 80%). The simultaneous bilinguals reported to have a majority of Dutch-speaking friends with an average of 97% (range: 92% - 99%) and only some of them reported to have Polish friends (average 4%; range 0% - 8%).

Of those late L2 learners with a partner, 18 out of 27 had a Dutch spouse or partner. The majority of this group, 14 out of 18, reported to speak almost exclusively Dutch with their partner and a minority, 4 out of 18, mostly spoke English with their significant other. One Polish participant, who was married to a Kurdish person, reported to also speak Dutch at home. A minority of 5 out of 27 participants was married to a Polish-speaking person and all of them reported to speak mostly Polish to each other. None of the simultaneous bilinguals were in an inter-language relationship.

Almost all of the late L2 learners (23 to 27) had taken a Dutch course that lasted for about 10 months on average (range: 2 to 24 months), but most, if not all of them, reported to have acquired Dutch outside of the classroom.

3.2. Film retelling task

As outlined in the previous chapter, the semi-spontaneous speech data was collected by means of a story retelling task in which people viewed and afterwards narrated a 10-minute excerpt of a Charlie Chaplin movie. The decision to include free speech as a measure of L2 proficiency had only been taken after the Dutch control group had already been tested. Fortunately, 8 of the 28 Dutch natives could be tracked down and were willing to cooperate with an extra test.

Holistic proficiency was assessed by three linguists and experienced raters, who rated the free speech of each participant on fluency, pronunciation, intonation, syntax, and lexicon (see §2.2.1. for a more detailed description of the rating procedure). The overall *holistic proficiency* did not differ significantly between the native control group and the simultaneous bilinguals, as revealed by pairwise Wilcoxon rank-sum tests with Bonferroni corrections ($W_s(n_1 = 8, n_2 = 6) = 32.5, p = 0.294$) and hence they were analysed together as one group ($n = 14$) to (partly) overcome the problem of small sample sizes.

The late L2 learners differed significantly from the natives and simultaneous bilinguals ($W_s(n_1 = 27, n_2 = 14) = 378, p < 0.001$), as can also be seen in Figure 3-1. When comparing individual scores of the late L2 learners to the other groups as a whole, it appeared that none of the L2 learners scored within 2 standard deviations below the native mean, a measure that has previously been used to asses native-likeness (Bongaerts, 1999; Ritchie & Bhatia, 2009). Moreover, as can also be seen in Figure 3-1, the native control and simultaneous bilingual groups performed at ceiling (for reasons of consistency and completeness, the groups are plotted separately). Hence, subsequent analyses focussed solely on the late L2 learners and factors that might contribute to their *holistic proficiency* score.

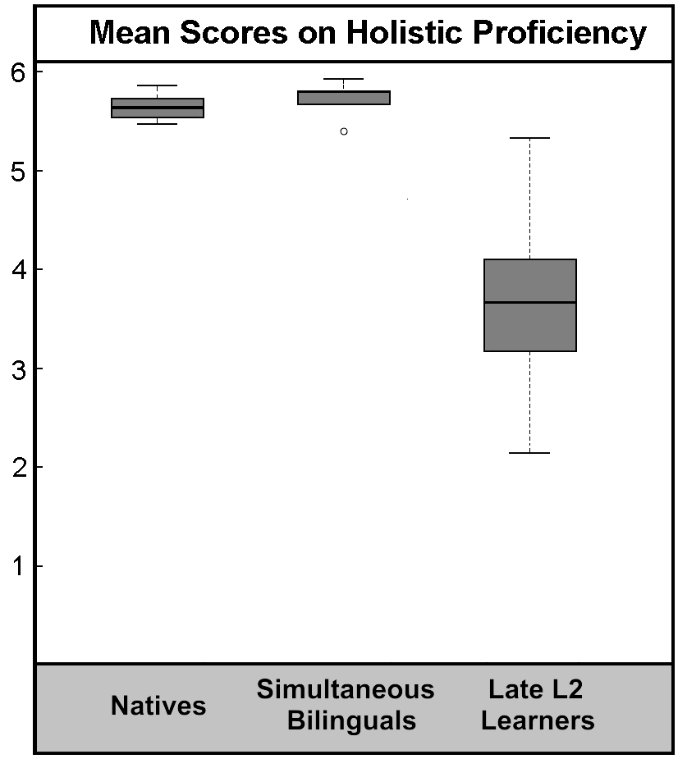


Figure 3-1: Mean scores on *holistic proficiency*, a measure based on the semi-spontaneous speech data that was gathered by means of a story retelling task and rated on fluency, pronunciation, intonation, syntax, and lexicon.

Within the late L2 learner group, a strong negative correlation was found between *holistic proficiency* score and Age of Acquisition (AoA: $r_s(26) = -0.59$, $p = 0.001$), again revealing higher proficiency scores for those who had acquired Dutch earlier in life. Length of Residence (LoR) also correlated with *holistic proficiency* ($r_s(26) = 0.41$, $p < 0.05$), but here the relationship was positive showing that higher scores can be related to longer residences in the Netherlands. The negative impact of AoA and the positive impact of LoR on *holistic proficiency* are graphically presented in Figure 3-2. The overall *holistic proficiency* score was included in the average L2 proficiency score.

The data from the film retelling task was also used to asses proficiency specifically related to the Dutch gender system, henceforth referred to as ***gender in production***. As explained in §2.2.1, all instances of gender assignment and gender agreement in the semi-spontaneous speech were counted as either correct or incorrect.

Late L2 learners appeared to make an additional mistake that may be related to their mother tongue, which was the omission of determiners. Since Polish does not have determiners such as *'een'*, *'de'*, and *'het'*, this might well constitute a form of negative L1 transfer. Due to this potentially interesting nature, these errors were counted separately.

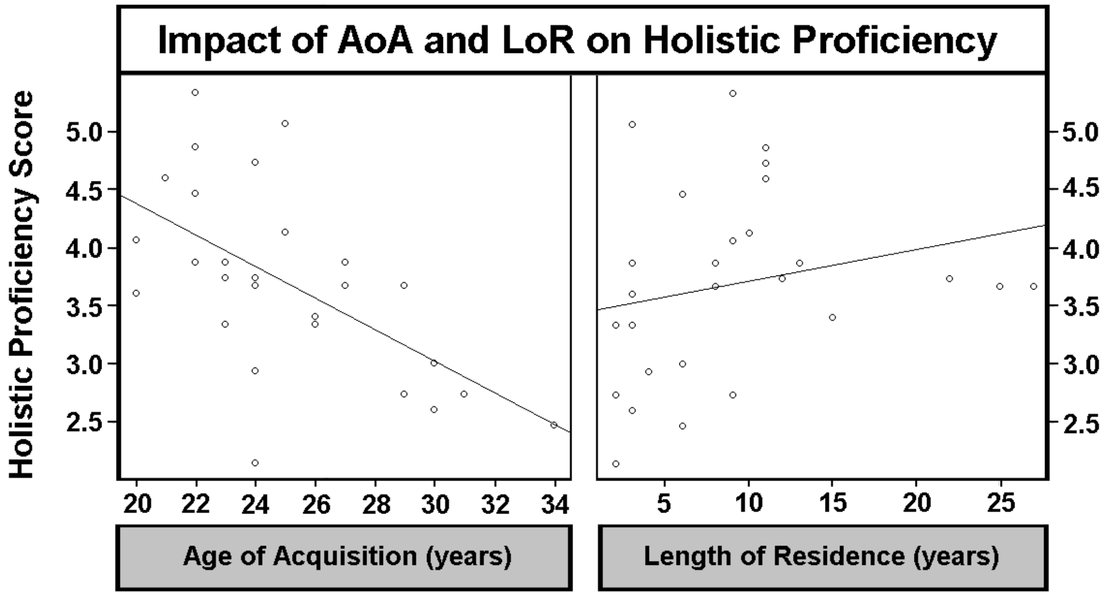


Figure 3-2: The correlational structures between Age of Acquisition (AoA) and *holistic proficiency* (left panel) and Length of Residence (LoR) and *holistic proficiency* (right panel) for the group of late L2 learners.

A first comparison of the production of noun phrases (Pairwise Wilcoxon rank-sum test with Bonferroni corrections and preceded by Kruskal-Wallis tests) revealed that the native control group produced significantly more NPs than both the simultaneous bilinguals ($W_s(n_1 = 8, n_2 = 6) = 2.5$, $p < 0.05$) and the late L2 learners ($W_s(n_1 = 8, n_2 = 27) = 40$, $p < 0.01$).[7] The two latter groups behaved similarly with respect to the amount of NPs produced ($p = 0.907$). A critical look at the data suggested that age of the participants was positively correlated with the number of NP's produced. Unfortunately, no conclusions can be drawn on the basis of such a correlation, since age is confounded with group in the present data set, i.e. participants in the native control group are overall older as compared to the late L2 learners and both these groups are older than the simultaneous bilinguals (also see section 2.1.).

On average, as would be predicted by the 75:25 ratio difference between common and neuter nouns in Dutch (Van Berkum, 1996), 63% of the produced noun

[7] Behavioural data from all three groups showed violations of normality. Fortunately, all three groups shared the same shape of distribution (a negative, left-band skew) allowing for a comparison of the data with the Kruskal-Wallis test for non-parametric data (Fagerland & Sandvik, 2009).

phrases were built around common nouns and about 37% contained neuter nouns. Participants of all groups made significantly more errors on neuter as compared to common gender in production (Wilcoxon Signed Rank Test: all *p*s < 0.001), as can also be seen in Figure 3-3. Previous research has also suggested differences in the acquisition of common and neuter (Unsworth & Hulk, 2010) and therefore *gender in production* was analysed for common and neuter nouns separately.

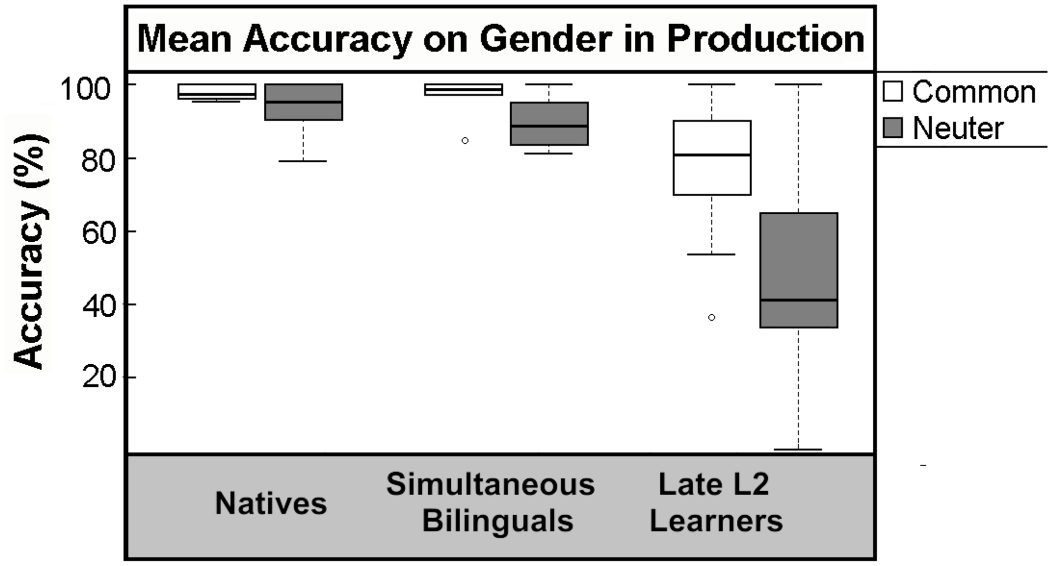


Figure 3-3: Mean percentage of accuracy on common and neuter nouns in the semi-spontaneous speech data.

As is clearly visible in Figure 3-3, late L2 learners are significantly less accurate in producing both common and neuter noun phrases as compared to both natives (*p*s $<$ 0.001) and simultaneous bilinguals (*p*s < 0.01). These two latter groups did not differ significantly on common or neuter gender in production ($W_s(n_1 = 8, n_2 = 6) = 26.5, p = 0.7881$; $W_s(n_1 = 8, n_2 = 6) = 15.5, p = 0.2963$) and could hence again be regarded as one native control group. Surprisingly, within the late L2 learners, it was found that there is no one-to-one relationship between the accuracy rates of common and neuter nouns in production ($r_s(26) = 0.28, p = 0.1526$). This lack of a correspondence also becomes apparent when looking at Figure 3-4.

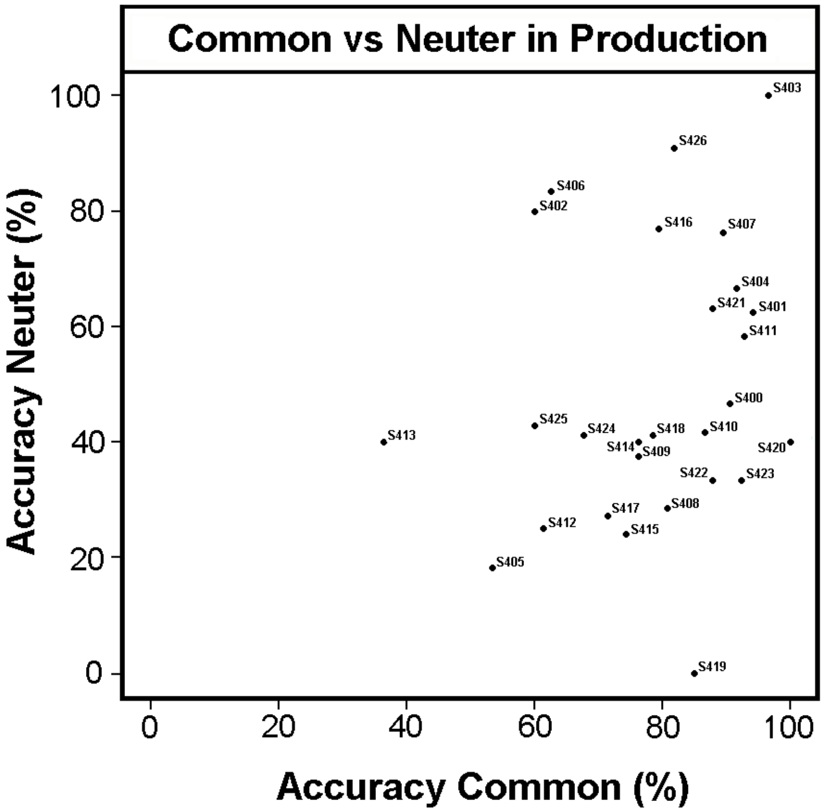


Figure 3-4: Relation between the individual scores on common and neuter nouns in production.

Figure 3-4 shows that whereas some people score below chance on both common and neuter (e.g. S413), others score native-like on both genders (e.g. S403). The majority of late L2 learners, however, appear to score pretty high on producing common nouns and they diverge in their accuracy rates on neuter nouns. Most learners are slightly better in producing common as compared to neuter gender-marking (e.g. S420), but some of them are better in neuter as compared to common nouns (e.g. S402). The general tendency, however, does appear to be that people overgeneralize common gender.

When looking solely at scores related to accuracy in gender assignment, i.e. determiner and noun congruency, this pattern also emerges. Only the late L2 learners differ from the other two groups (native controls: $W_s(n_1 = 8, n_2 = 27) = 28.5, p < 0.01$; simultaneous bilinguals: $W_s(n_1 = 6, n_2 = 27) = 134, p = 0.01$) and are significantly less accurate in producing both common and neuter noun phrases as compared to both native groups.

Gender agreement, i.e. congruency between adjective and noun, was also analysed separately. Statistical analyses using Kruskal-Wallis tests revealed that the average accuracy scores for common nouns did not differ significantly between the groups ($\chi^2(2, N = 38) = 1.45, p = 0.4843$). The data revealed that, although not many NPs containing adjectives were produced; those adjectives that were used to modify common nouns were correctly inflected. The three groups did differ with respect to neuter-marked adjectives ($\chi^2(2, N = 38) = 16.9547, p < 0.001$). Subsequent pairwise

Wilcoxon rank-sum tests with Bonferroni correction showed that the late L2 learners of Dutch differ significantly from both the native control group ($p < 0.001$) and the simultaneous bilinguals ($p < 0.01$), but these two groups do not differ from each other ($p = 0.1489$) with respect to neuter gender in production.

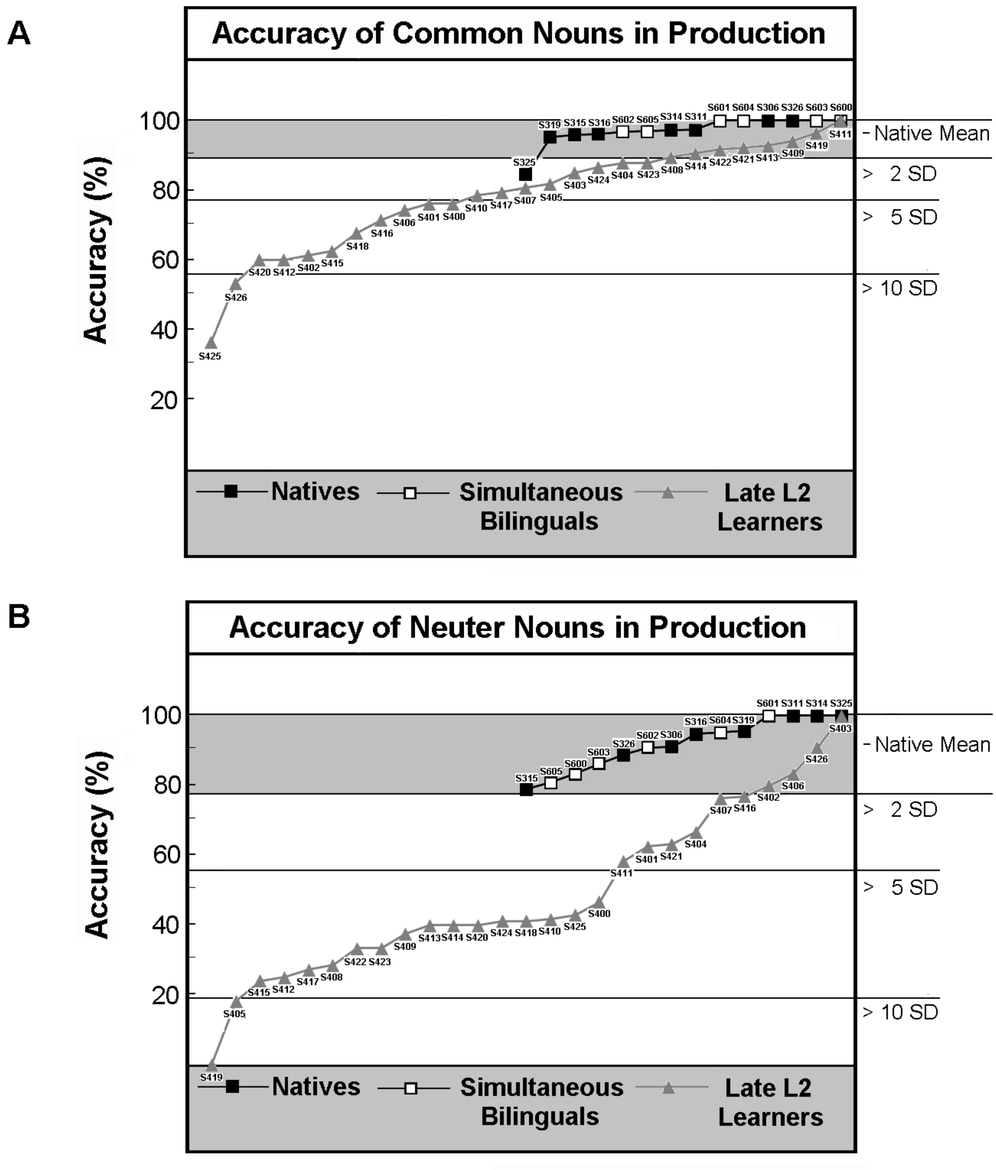


Figure 3-5: Individual accuracy scores for A) common and B) neuter nouns in production. Grey areas represent the native range, i.e. scores within 2 standard deviations from the native mean.

Figure 3-5 clearly shows the difference in performance between common and neuter nouns. The range of accuracy for natives is clearly larger for neuter nouns (3-5. B) as compared to common nouns (3-5. A). Accuracy scores within 2 standard deviations from the native means were taken as a measure of native-like performance (Ritchie & Bhatia, 2009). As can also be seen in Figure 3-5, 8 late L2 learners are native-like in common gender production and, despite the broader range, only 4 perform within native range on neuter gender in production. More than half of the participants from the late L2 learners group, 17 out of 27, scored below chance level on neuter nouns, which corresponds to scores more than 5 standard deviations below the native mean.

The special status of neuter nouns is present in all aspects of gender in production. For example, Kruskal-Wallis tests revealed that the groups did not differ in the accuracy of use of the common demonstrative pronoun *'die'* ($\chi^2(2, N = 32) = 2.0885$, $p = 0.3520$). Significant differences were found, however, with respect to the use of the neuter demonstrative pronoun *'dat'* ($\chi^2(2, N = 31) = 6.7664$, $p < 0.05$), for which late L2 learners were less accurate while the other two groups performed at ceiling. With respect to the production of relative pronouns, the groups also did not differ significantly on the production of the common relative pronoun *'die'* ($\chi^2(2, N = 31) = 1.7187$, $p = 0.4234$) where they all score at ceiling, but significant differences between the groups arise when comparing them on the production of the relative pronoun *'dat'* ($\chi^2(2, N = 29) = 11.3905$, $p < 0.01$). Strikingly, as can also be seen in Table 3-1, none of the three groups score at ceiling on producing the neuter relative pronoun *'dat'* accurately.

Table 3-1: Overview of the distribution of errors in common and neuter gender in production made by the native control group (NCG), the simultaneous bilinguals (SB), and the late L2 learners (LL2) respectively.

Domain of Agreement		**Common**	**Neuter**
Determiner + Noun	NCG:	0/154 incorrect (5/154 omitted)	1/95 incorrect (4/95 omitted)
	SB:	0/72 incorrect (4/72 omitted)	1/40 incorrect (3/40 omitted)
	LL2:	4/369 incorrect (75/369 omitted)	59/146 incorrect (34/146 omitted)
Adjective + Noun	NCG:	1/20 incorrect	0/11 incorrect
	SB:	0/3 incorrect	0/7 incorrect
	LL2:	3/38 incorrect	26/42 incorrect
Demonstrative Pronoun	NCG:	0/75 incorrect	3/37 incorrect
	SB:	0/12 incorrect	0/10 incorrect
	LL2:	6/118 incorrect	43/86 incorrect
Relative Pronoun	NCG:	0/44 incorrect	4/17 incorrect
	SB:	0/17 incorrect	3/7 incorrect
	LL2:	1/66 incorrect	31/39 incorrect

Table 3-1 also reveals a typical error for the late L2 learners, which constitutes omission of determiners. Pairwise Wilcoxon rank-sum tests with Bonferroni corrections (again preceded by Kruskal-Wallis tests) revealed that late L2 learners omit definite and indefinite determiners significantly more often than the other two groups for both common and neuter nouns (native control: $Ws(n_1 = 8, n_2 = 27) > 192$, $ps < 0.001$; simultaneous bilinguals: $Ws(n_1 = 6, n_2 = 27) > 15$, $ps < 0.05$), who barely make such mistakes. This type of error reflects a form of negative transfer, since the Polish language does not have any determiners (Perlak & Jarema, 2003).

The overall scores on common and neuter *gender in production* did not correlate significantly with Age of Acquisition. There was, however, a small positive correlation between Length of Residence and common gender in production ($r_s(26) = 0.34$, $p = 0.07984$), which was absent for neuter nouns ($r_s(26) = 0.10$, $p = 0.6191$). These overall scores for common and neuter were taken along in an average proficiency score for the two genders respectively. Omission of determiners did not correlate with AoA, but there was a small negative correlation between Length of Residence and omission of common determiners ($r_s(26) = -0.38$, $p = 0.05$). No such correlation was found for the omission of neuter determiners ($r_s(26) = -0.05$, $p = 0.8233$). It thus appears that AoA does not impact performance on either common and neuter nouns in production, but the longer the LoR the better late L2 learners are in using common gender and the less they tend to omit common determiners.

As can also be seen from the data in Table 3-1, natives and simultaneous bilinguals hardly make any errors except that they sometimes use the common relative pronoun *'die'* to refer back to a neuter noun. This occurred 7 times (out of the 24 neuter pronouns used) and in all instances this concerned reference to a person (i.e. *'het meisje'*, the girl). This is a relatively frequent error made by native Dutch speakers probably resulting from the fact that persons are more likely to be referred to by common gender. Recent investigations on the pronominal gender system of modern spoken Dutch indeed suggest that new semantic foundations are being used in pronominal use, most likely resulting from the loss of the feminine-masculine distinction (Audring, 2009). For this same reason, the studies presented in the present dissertation have focussed on the processing and acquisition of determiners and gender-marked adjectives and did not include measures addressing the use or comprehension of relative pronouns.

Both the accuracy scores of common and neuter *gender in production* were taken along in a proficiency score specifically related to grammatical gender. Omissions of determiners were not taken along in the proficiency score.

3.3. The C-Test

Statistical analyses using Kruskal-Wallis tests revealed a significant difference between the three groups with respect to their C-Test scores ($\chi^2(2, N = 47) = 24.262$, $p < 0.001$).

Two-sample Wilcoxon rank-sum tests with Bonferroni correction showed that the late L2 learners of Dutch differed significantly from both the ‘monolingual’ native speakers ($p < 0.05$) and the early simultaneous bilinguals ($p < 0.001$). As one might expect, the latter group did not differ in performance on the C-Test from the native control group ($p = 0.9377$). These groups can thus be regarded and analysed as one group of native Dutch speakers ($n = 21$) to overcome the relatively small sample sizes of the groups.

As Figure 3-6 shows, the natives and simultaneous bilinguals appear to perform at ceiling, while there is a lot of variability in the late L2 learners group.

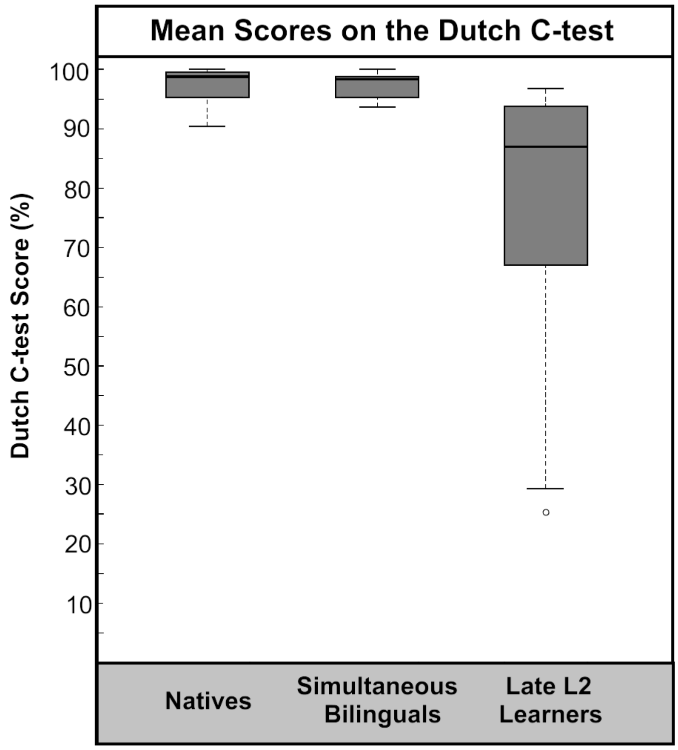


Figure 3-6: The mean scores of the native control group, the simultaneous bilinguals, and the late L2 learners of Dutch on the C-Test.

The heterogeneity within the late L2 learners is also clearly visible in Figure 3-7, where the individual scores on the C-Test are plotted. With respect to the native Dutch speakers, all except for one scored within 2 standard deviations below the native mean and the maximum score that is possible on this C-Test (i.e. the grey area in Figure 3-7, which depicts C-Test scores between *322* and *342*). Interestingly, Figure 3-7 clearly reveals that there are late L2 learners of Dutch who do not deviate from the native Dutch controls and simultaneous bilinguals in that they also score within 2 standard deviations from the native mean (Bongaerts, 1999; Ritchie & Bhatia, 2009). The graph also shows some extreme outliers within this latter group with scores of more than 20 standard deviations away from the native mean. When taken together, no significant correlations were found between age and C-Test score ($r_s(47) = 0.18$, $p = 0.226$) nor between level of education and C-Test score ($r_s(47) = -0.11$, $p = 0.451$). A Wilcoxon rank-sum test with Bonferroni correction showed no difference in performance on the C-Test between men and women ($p = 0.9242$).

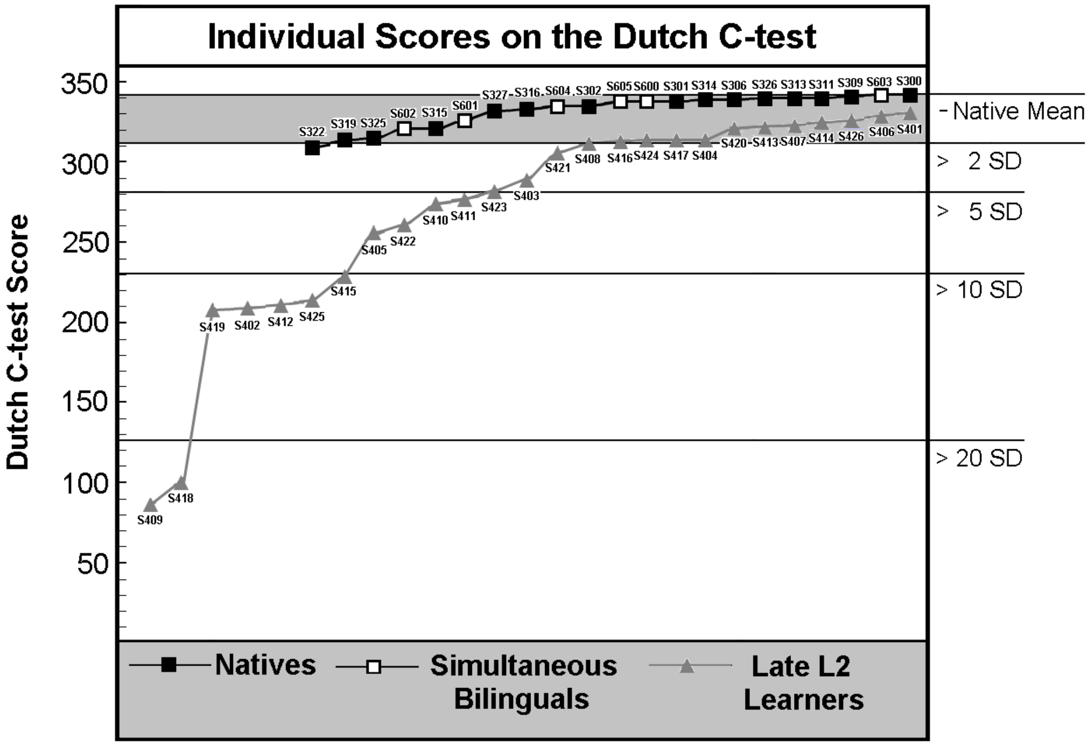


Figure 3-7: The individual scores of the native control group and the simultaneous bilinguals are plotted against the late L2 learners of Dutch. A score in between 2 standard deviations below the native mean and the maximum score that is possible on this C-Test (i.e. a score between *322* and *342*) can be regarded as a native-like score. The graph shows that some late L2 learners fall well within the native range.

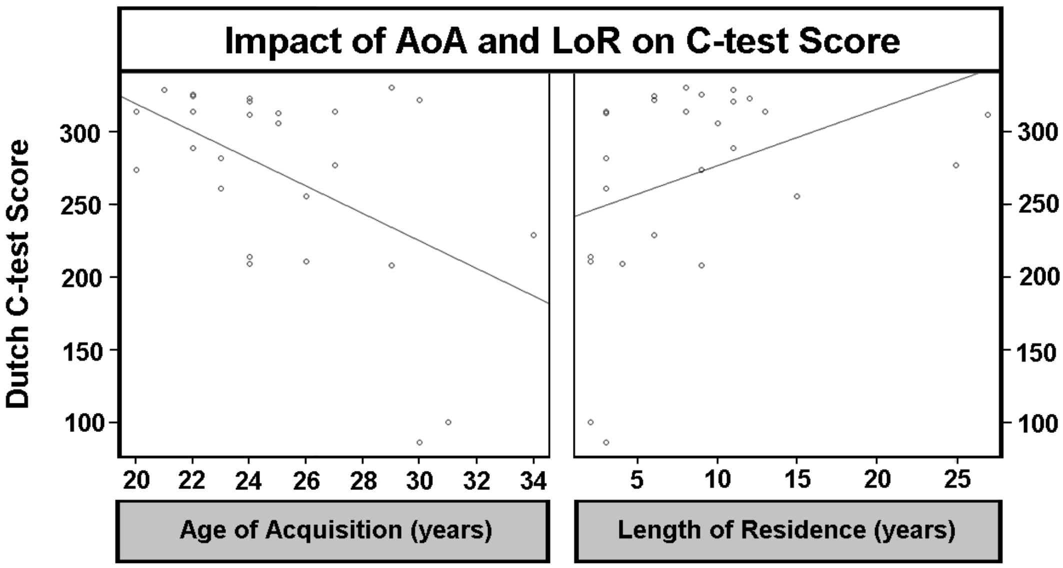


Figure 3-8: The correlational structures between Age of Acquisition (AoA) and C-Test score (left panel) and Length of Residence (LoR) and C-Test score (right panel) for the group of late L2 learners.

Within the late L2 learners, C-Test scores correlated negatively with Age of Acquisition ($r_s(26) = -0.40$, $p < 0.05$), revealing that the later people had started learning Dutch, the lower the general level of proficiency as revealed by the C-Test score. Similarly, Length of Residence (LoR) was positively correlated with C-Test score ($r_s(26) = 0.39$, $p < 0.05$), showing that the longer the participant had lived in the Netherlands, the higher his or her score on the C-Test was. These two correlations are also depicted in Figure 3-8.

3.4. Acceptability judgement task

To create an overall proficiency score, the overall percentage of accurate responses for sentences that contained correct constructions and violations of finiteness (*AJT Finiteness*) and gender agreement (*AJT Common* and *AJT Neuter*), as administered during the ERP experiment, was taken into account. Detailed analyses of the accuracy as well as the reaction time data of this grammaticality judgement task are discussed in detail in §5.3.1 and §7.3.1.

3.5. Gender assignment task

The offline gender assignment task was administered to check participant's knowledge of Dutch nouns and the appropriate determiners belonging to these nouns. Due to the relatively long experimental sessions, native speakers were not asked to assign the gender to each noun and, in line with the results on the previously mentioned proficiency measures, it was assumed that these groups would have no difficulty in assigning the correct gender to the high frequency Dutch nouns used in the present experiments. Late L2 learners' performance on this task could thus only be compared to the performance of the 6 Simultaneous Bilinguals.

Wilcoxon rank-sum tests with Bonferroni correction showed that the late L2 learners of Dutch differed significantly from the simultaneous bilinguals ($p < 0.001$). The latter group scored at ceiling with only 1 out of 6 bilinguals assigning the incorrect gender to one of the nouns used. Therefore, further analyses included only late L2 learners.

As was also found for the accuracy on common and neuter gender in production, there was again no clear one-to-one relationship between the accuracy rates of common and neuter nouns in the gender assignment task ($r_s(26) = -0.34$, $p = 0.09229$). No correlations were found between *gender assignment* score and age, education or sex. A significant correlation, however, was found between AoA and gender assignment score ($r_s(26) = -0.42$, $p < 0.05$), revealing that lower AoAs correspond to a more accurate assignment score. Separate analyses of the genders revealed that this was mainly driven by a significant positive interaction between AoA and the ability to correctly assign neuter gender ($r_s(26) = -0.39$, $p < 0.05$). The correlation between AoA and common gender assignment did not reach significance ($r_s(26) = -0.30$, p = 0.1413).

Interestingly, Length of Residence appeared to be an even better predictor of the gender assignment score ($r_s(26) = 0.67$, $p < 0.001$) and separate analyses for the genders revealed that this strong positive correlation holds for both common ($r_s(26) = 0.52$, $p < 0.01$) and neuter nouns separately ($r_s(26) = 0.42$, $p < 0.05$).

3.6. Overall proficiency level

As the results show, most of the late L2 learners of Dutch were fairly proficient and the scores on the different tasks were not always equally affected by Age of Acquisition and/or Length of Residence of the late L2 learners of Dutch.

Table 3-2 provides an overview of the correlations between all separate proficiency measures and AoA and LoR.

Table 3-2: Overview of the correlations between the proficiency measures, age of acquisition (AoA), and length of residence (LoR). Significant levels are conveyed by asterisks and clarified below the table.

Domain	AoA	LoR
1. C-Test	**-.41****	**.41***
2. Self-rating	**-.38***	.33
3. Holistic Proficiency	**-.60****	**.43***
4. AJT finiteness	-.10	.34
5a. Production Common	-.06	.36
5b. Production Neuter	-.31	.19
6a. Assignment Common	-.30	**.52****
6b. Assignment Neuter	**-.39***	**.42***
7a. AJT Common	-0.19	0.18
7b. AJT Neuter	-0.34	0.18
8a. Omission Common	.10	**-.38***
8b. Omission Neuter	-.01	-.05

* 0.05
** 0.01
*** 0.001

As can be seen from Table 3-2, several general proficiency scores, such as the scores on the C-Test, the self-ratings, and the score they received with respect to their free speech data (*holistic proficiency*), correlated significantly with Age of Acquisition. These negative correlations reveal that, even in late L2 learners who all acquired the second language after the age of 20, an earlier onset of acquisition is still beneficial in that it results in higher scores on these general proficiency measures. Similarly, Length of Residence positively contributed to these scores with longer stays in the Netherlands resulting in higher scores on both the C-Test and *holistic proficiency*. The correlation between LoR and self-rating, however, did not reach significance.

In addition, a later AoA also resulted in a significantly lower accuracy in assigning the correct gender to neuter nouns (*6b. Assignment neuter*) and strong positive correlations between Length of Residence and both the accuracy of assigning gender to common and to neuter nouns suggest an important role of input quantity in the acquisition of grammatical gender. Longer LoR's were additionally correlated with a decrease in the number of omissions of common determiners, which suggests that longer exposure also reduced the amount of negative transfer from the mother tongue.

Strikingly, there was not always a clear one-to-one relationship between the proficiency on common and neuter nouns in production as well as comprehension, as can also be seen in Table 3-3. Therefore, all measures were entered separately in the analyses of the eye tracking data (Chapter 6) as well as in the analyses of the ERP amplitudes of the effects found in the group of late L2 learners (Chapters 7).

Table 3-3: Correlations between the different proficiency scores for late L2 learners. Significant levels are conveyed by asterisks and clarified below the table.

Domain	1	2	3	4	5a	5b	6a	6b	7a	7b	8a	8b
1. C-Test	-											
2. Self-rating	**.72*****	-										
3. Holistic Proficiency	**.67*****	**.69*****	-									
4. GJT finiteness	**.62****	**.56****	**.56****	-								
5a. Production Common	.23	.30	**.44***	.11	-							
5b. Production Neuter	**.42***	**.42***	**.52****	.31	28	-						
6a. Assignment Common	**.74*****	**.42***	**.39***	.34	.36	**.45***	-					
6b. Assignment Neuter	**.42***	**.50****	**.46****	**.43***	**.43***	**.59****	.33	-				
7a. GJT Common	**.75*****	**.49***	**.50****	**.48***	**.40***	**.51****	**.65****	**.63****	-			
7b. GJT Neuter	**.76*****	**.51****	**.54****	.30	.18	**.58****	**.60****	.33	**.63****	-		
8a. Omission Common	-.17	-.23	**-.46***	.04	**-.92*****	-.30	-.36	-.34	-.25	-.15	-	
8b. Omission Neuter	-.17	-.11	-.25	-.08	-.09	**-.64*****	-.29	-.19	-.16	-.24	.25	-

* 0.05
** 0.01
*** 0.001

For all global analyses, the L2 learners were divided into a group of low and high proficiency learners. The division of these groups was based on the overall proficiency level based on all separable scores on the proficiency tests as is visualized in Table 3-4.

Table 3-4: Overview of the scores on the different proficiency measures per proficiency group (as used for the ERP experiment). Subject S408 did not participate in the second part of the study and did not perform the C-Test, the acceptability judgement task, and the offline gender assignment task.

Subject	Proficiency	Overall score (%)	1. C-Test	2. Self-rating	3. Holistic Proficiency	4. AJT Finiteness	5a. Production Common	5b. Production Neuter	6a. Assignment Common	6b. Assignment Neuter	7a. AJT Common	7b. AJT Neuter
S418	low	45.5										
S409	low	48.7										
S419	low	49.0										
S415	low	51.0										
S412	low	52.0										
S402	low	53.6										
S425	low	53.7										
S405	low	56.2										
S422	low	60.7										
S413	low	61.1										
S423	low	62.1										
S410	low	63.7										
S417	low	63.7										
S400	high	66.5										
S424	high	66.5										
S408	high	67.5	NA			NA			NA	NA	NA	NA
S416	high	69.9										
S411	high	70.0										
S414	high	71.9										
S420	high	72.6										
S421	high	75.7										
S404	high	77.4										
S401	high	77.6										
S406	high	78.7										
S403	high	82.4										
S407	high	83.9										
S426	high	88.3										

- Score within 2 sd from native mean / score above 90%
- Score within 5 sd from native mean / score above 80 %
- Score within 10 sd from native mean / score above 65 %
- Score more than 10 sd from native mean / score below 60 %
- Score more than 15 sd from native mean / score below 50 %

The difficulty with producing and comprehending gender in general, and neuter gender in particular, is also visible in this Table 3-4. The white and lighter grey colours reflect more native-like scores and the darker grey colours reflect scores that deviated more from the native mean. L2 learners within the group of highly proficient learners, presented in the lower rows, still reveal difficulty with the use, assignment and comprehension of neuter nouns and especially the use of neuter gender-marking in production (*5b. Production Neuter*), which remains difficult for L2 learners even with increasing proficiency.

Similarly, Table 3-4 provides information concerning the lack of interactions between several scores. Subject number S402, for example, scored low on almost all proficiency measures, but produced native-like accuracy in neuter gender-marking (*5b. Production Neuter*). Subject S416, on the other hand, was highly proficient and scored relatively high on both the production and comprehension of neuter nouns, but appeared to still have problems correctly assigning common gender (*6a. Assignment Common*) and accurately judge the acceptability of sentences containing common nouns (*7a. AJT Common*).

As can be seen in Table 3-3, most scores correlate with one another to a certain degree. Table 3-4 clearly shows, however, that the different proficiency measures do not suggest the same median split for the same people. Therefore, it was decided to take along all separate measures of proficiency in addition to the overall proficiency score used to divide the L2 learners into two groups.

Chapter 4

NEUTER IS NOT COMMON IN DUTCH:

Eye Movements Reveal Asymmetrical Gender Processing

Native speakers of languages with transparent gender systems can use gender cues to anticipate upcoming words. To examine whether this also holds true for a non-transparent two-way gender system, i.e. Dutch, eye movements were monitored as participants followed spoken instructions to click on one of four displayed items on a screen (e.g. *Klik op de*$_{COM}$ *rode appel*$_{COM}$, 'Click on the$_{COM}$ red apple$_{COM}$'). The items contained the target, a colour- and/or gender-matching competitor, and two unrelated distractors. A mixed-effects regression analysis revealed that the presence of a colour-matching and/or gender-matching competitor significantly slowed the process of finding the target. The gender effect, however, was only observed for common nouns, reflecting the fact that neuter gender-marking cannot disambiguate as all Dutch nouns become neuter when used as diminutives. The gender effect for common nouns occurred before noun onset, suggesting that gender information is, at least partially, activated automatically before encountering the noun.

This chapter has been submitted for publication: Loerts, H., Wieling, M., & Schmid, M.S. (Under Review). Neuter is not Common in Dutch: Eye Movements reveal Asymmetrical Gender Processing. Submitted (February 29, 2012) to *Journal of Psycholinguistic Research.*

4.1. Introduction

While listening to language, people are not passively interpreting each word separately. Instead, our brain is actively trying to predict what the next word will be. People are therefore faster in responding to words that are predictable on the basis of semantic context (e.g. Stanovich & West, 1983) and can decide what word will be pronounced after only hearing the first few phonemes (Grosjean, 1980). The speed with which certain (non-)linguistic cues in the speech stream are used to anticipate and hence facilitate language comprehension has increasingly been investigated by following people's eye movements while they receive auditory input. These eye tracking experiments are based on the initial finding by Cooper (1974), who showed that while people are listening to short narratives, they strongly tend to look at those objects in the visual field that are most closely related to the words they hear in the speech stream. Over two decades after this discovery, Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) re-introduced and further developed Cooper's (1974) methodology into what is now known as the 'visual world paradigm'. By measuring participants' eye movements while they are following instructions to manipulate objects, the studies by Tanenhaus and his colleagues revealed a real-time interdependence between spoken input and the surrounding visual world. If, for example, subjects were instructed to '*pick up the can...*' they fixated equally often on a picture of a candle and a picture of a candy (Allopenna, Magnuson, & Tanenhaus, 1998). This effect illustrates the incremental nature of language comprehension: the onset of a word activates a cohort, a relevant set of lexical candidates that compete for recognition until the uniqueness point of the word is heard (Marslen-Wilson, 1987).

More recent models argue that competition occurs on the basis of all acoustic input (e.g. the TRACE model (McClelland & Elman, 1986), Shortlist (Norris, 1994), and the neighbourhood activation model (NAM; Luce & Pisoni, 1998)). Ambiguities in the speech stream may therefore not only concern phonological overlap between words, but items may also compete due to ambiguous morphological cues in the input. Accordingly, gender-marked items preceding the noun might facilitate access by reducing the set of possible lexical candidates to gender-matching nouns. This hypothesis was first tested in French by Dahan, Swingley, Tanenhaus, and Magnuson (2000). Their participants were instructed to click on one of four displayed pictures (e.g. *Cliquez sur le bouton,* 'Click on the$_{\text{MASC}}$ button$_{\text{MASC}}$'). Every screen of four pictures contained a target, a cohort competitor (i.e. a word sharing its initial phonemes with the target, e.g. *bouteille*, 'bottle$_{\text{FEM}}$'), and two unrelated distractors. When the auditory input contained no gender-marking cues, as in the baseline condition containing French plural '*les*', the same cohort competitor effect as reported by Allopenna et al. (1998) was found: people looked equally often at a picture of a bottle and a picture of a button upon hearing '*cliquez sur les bou...*'. Interestingly, this cohort effect disappeared when singular gender-marked determiners preceded the noun in the speech stream. Upon hearing '*cliquez sur le*$_{\text{MASC}}$ *bou…*', natives never fixated on the picture of a feminine referent (*la*$_{FEM}$ *bouteille*$_{FEM}$) more often than on the unrelated distractors, indicating that gender cues facilitate language processing by minimizing the set of possible lexical candidates. Dahan et al. (2000) also found that, by itself, a gender-marked article does not seem to prime all gender-matching nouns. Upon hearing the feminine article, e.g. in *la*$_{FEM}$ *louche*$_{FEM}$ (‘the$_{FEM}$ ladle$_{FEM}$’), participants did not fixate more on a picture of a gender-matching noun that did not overlap in phonemic onset (e.g. a sock: *chaussette*$_{FEM}$). These additional results suggest that gender information, as present on the definite determiner preceding the noun, might be used post-lexically, i.e. requiring the output of the lexical processing system for its operation. The gender effect could then be defined as an inhibitory rather than a facilitatory effect.

The absence of a gender effect in Dahan et al. (2000) with regard to phonemically non-overlapping article-noun combinations might be a matter of locality (Paris, Weber, and Crocker, 2006). Subjects may have had too little time to pre-activate and in turn use gender information on the article, because the short French articles are immediately followed by the target nouns. To increase the time people have to possibly pre-activate gender congruent nouns, Paris et al. (2006) studied the phenomenon in German which allows adjectives to intervene between determiner and noun. Confirming their hypothesis, they found that participants began fixating targets and gender-matching pictures significantly more than gender-mismatching pictures immediately after having heard the gender-marked article in a sentence such as, e.g. *Wo befindet sich die*$_{FEM}$ *schwere*$_{FEM}$ *Melone*$_{FEM}$ (‘Where is the$_{FEM}$ heavy$_{FEM}$ melon$_{FEM}$?’). This effect only decreased when subjects heard the onset of the target noun (which was unique to the target). The authors do mention that the effect they found is smaller than the effect found with phonemically overlapping nouns (Dahan et al., 2000), but the presence of a gender effect with non-overlapping nouns suggests that gender cues on articles influence referential processes before encountering the target noun. This result provides at least some support for a pre-activation account in which gender cues facilitate language comprehension in a more automatic fashion.

Until now, a gender effect (either facilitatory or inhibitory) has been found in eye tracking studies in German (e.g. Paris, Weber, & Crocker, 2006), French (Dahan et al., 2000), Russian (Sekerina, Brooks, & Kempe, 2006), and Spanish (Lew-Williams and Fernald, 2010). Gendered languages, however, differ greatly with respect to the amount and type of gender categories they have as well as the extent to which a noun’s gender can be interpreted on the basis of phonological and/or morphological features (Corbett, 1991). Most researchers believe that the assignment of nouns to a particular gender category is a rather arbitrary process and in most instances, e.g. for inanimate objects, there is no clear relationship between the noun’s gender and any male or female characteristics of its referent in the real world. In languages with an overt gender system the noun’s gender is reflected in the morphology of its agreement targets as well as in the form of the noun itself. In many Romance languages, gender has a phonological correlate. For example, Spanish nouns ending with the suffix ‘*–a*’ are mostly feminine (‘*la senora*’: madam), whereas nouns ending in ‘*–o*’ are usually masculine (‘*el marido*’:

husband), although there are exceptions to this principle. Other languages, such as French and German, contain more abstract, morphological cues that help to determine the gender of a noun. For example, French nouns that end with the suffix '*-ion*' generally receive the feminine gender (Foucart, 2008) and German nouns are assigned feminine gender when ending in the feminine suffix '*-keit*' (Zubin & Köpcke, 1984).

In a largely covert gender system, such as Dutch, gender has to be marked on the agreement targets, but is usually not obvious from the shape or meaning of the noun (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997). Dutch has two genders: common and neuter. Common nouns receive the definite article '*de*' while neuter nouns (as well as diminutives) take '*het*'. Attributive adjectives always receive a schwa-ending, except when modifying a neuter noun in an indefinite noun phrase (NP; see also Table 4-1). The special status of the neuter gender in indefinite NPs is accompanied by an asymmetrical distribution of the two gender categories. Common gender evolved from the collapse of the original masculine and feminine genders and consequently comprises around 75% of all Dutch nouns (Van Berkum, 1996). This overrepresentation of nouns taking '*de*' is further reinforced by the fact that '*de*' is also the definite plural article, irrespective of the gender of the noun. The relatively scarce evidence in the input for neuter gender causes monolinguals to overgeneralize the definite, common determiner '*de*' until the age of 6 (Blom, Polisenskà, & Weerman, 2008; Van der Velde, 2004). This is rather late in comparison to French native children, who have been reported to have acquired their gender system at the age of 4 (Van der Velde, 2004) as well as German monolinguals, who appear to master their system by age 3 (Müller, 1990; Szagun, Stumper, Sondag, & Franik, 2007).

Table 4-1: Overview of the main agreement targets in Dutch. Adapted from Loerts, Stowe, & Schmid (Under Review). English equivalents of the phrases are printed in italic below each phrase. Note that the indefinite article is always *'een'* in the singular and no equivalent indefinite article exists for the plural case.

	Gender	**Definite Article**	**Adjectives in Definite NPs**	**Adjectives in Indefinite NPs**
Single	common	***de*** appel	***de*** rode appel	een ***rode*** appel
	neuter	***het*** huis	***het*** rode huis	een ***rood*** huis
English equivalent		*the apple/house*	*The red apple/house*	*a red apple/house*
Plural	common	de appels	de rode appels	--
	neuter	de huizen	de rode huizen	--
English equivalent		*the apples/houses*	*the red apples/houses*	--

The Dutch gender system with its asymmetrical distribution of gender classes and agreement markers thus largely has to be learned on an item-by-item basis. An interesting question, therefore, is if gender-marking in such an asymmetrical and non-transparent gender system is also used to facilitate and/or inhibit spoken word recognition?

Brouwer, Unsworth, and Mak (2010) recently investigated this question using the eye tracking methodology. They recorded eye movements from monolingual Dutch natives (both children and adults) while they looked at two pictures on a computer screen. The two nouns on the screen always shared the same colour, but the competitor either had a different gender or shared the gender with the target. Participants listened to speech stimuli containing gender-marked articles such as '*Can you see the$_{NEU}$ yellow house$_{NEU}$?*'. As a baseline, the non-gender-marked possessive adjective '*mijn*' (my) preceded the adjective and the noun. The authors reported that adults are marginally sensitive to gender-marking, as revealed by earlier detection of the target picture when the item was preceded by a gender-marked article as opposed to the possessive '*mijn*'. It should be pointed out that a potentially confounding factor in this study is the fact that the non-gender-marked adjective '*mijn*' might be a problematic baseline as it is somewhat longer than both gender-marked determiners (*'de'* or *'het'*). Consequently, the earlier detection of the target picture in the gender-marked conditions might, at least to some extent, be caused by the shorter duration of these determiners as opposed to the possessive '*mijn*'. The present study improves on Brouwer et al.'s approach by using similar durations throughout the auditory stimuli and therefore attempts to investigate more clearly whether, despite the non-transparency of the system, Dutch natives are able to use prenominal gender cues during language comprehension to pre-select upcoming nouns. The setup used was similar to the visual world paradigm of Tanenhaus et al. (1995), asking subjects to manipulate the objects in their visual scene. The nouns depicted varied along two dimensions: syntactic gender and colour. Reasoning in line with Paris et al. (2006), the Dutch language also allows for the placement of adjectives in between articles and nouns, possibly reducing the high co-occurrence of definite articles and nouns and allowing more time for gender information to be processed.

If gender-marking on an item immediately preceding the noun does not allow the listener enough time to process and use the information, then a possible gender effect is only expected in the definite condition, i.e. the condition in which gender-marking occurs on the determiner that is followed by an intervening adjective allowing for more processing time. This definite condition was directly compared to the indefinite 'baseline' condition, in which gender information only becomes apparent at the offset of the adjective and might thus not allow the listener enough time to process and use the information. The present study used only phonemically non-overlapping nouns in order to examine the gender effect in more detail.

4.2. Methods

4.2.1. Participants

Twenty-eight adult native speakers of Dutch (mean age: 48, range: 34-60; 16 female) were recruited around the city of Groningen by means of advertisements in a local newspaper. All participants filled out a questionnaire to ensure that they had normal or corrected-to-normal vision, no hearing deficits and no form of colour blindness.

4.2.2. Visual stimuli

A set of 48 nouns (24 common and 24 neuter nouns) was selected based on several criteria. First of all, the nouns chosen had to refer to objects that could be depicted using images. Only nouns that could be extended by an adjective referring to the colour of the object depicted were used. As Dahan, Magnuson, & Tanenhaus (2001) have pointed out, referents with higher frequency names are more likely to be fixated on, though the fixations tend to be shorter. Therefore, only high frequency nouns were selected from the frequency list of the *Corpus Gesproken Nederlands* (corpus of spoken Dutch), containing nouns that belong to the 5000 most frequently occurring words (CGN 5000). Despite this, there were still relatively large differences in lemma frequency of the nouns in our stimuli set, ranging from 7 to 3203 (mean lemma frequency of 323.2). To correct for the skewness of this distribution, lemma frequency was log-transformed and included in the mixed-effects regression model (see below).

The visual stimuli were derived from the standardized picture set of black and white line drawings (Snodgrass and Vanderwart, 1980), but rendered in one of five different colours: red, blue, green, yellow, and brown. Initially, the goal was to use only primary and/or notable colours. Due to the large number of criteria, however, images that could not be depicted in a primary or secondary colour (such as, e.g. a dog) were used and hence coloured brown.

4.2.3. Auditory stimuli

As gender information in Dutch NPs either appears on the determiner in definite NPs (sentence 4-1a, below) or on the adjective in indefinite NPs (sentence 4-1b, below), both definite and indefinite constructions were used. Colour adjectives were placed between the determiner and the noun to reduce possible effects of high co-occurrence probabilities of determiners and nouns as well as to allow more time for gender information to be processed in the definite constructions, see 4-1a below (Dahan et al., 2000).

4-1a ‘Klik op de$_{\mathrm{COM}}$ rode appel$_{\mathrm{COM}}$’
Click on the$_{COM}$ red apple$_{\mathrm{COM}}$
4-1b ‘Klik op het plaatje met een rode$_{\mathrm{COM}}$ appel$_{\mathrm{COM}}$’
Click on the picture with a red$_{COM}$ apple$_{\mathrm{COM}}$

The auditory instructions were digitally recorded in a soundproof booth with a sampling rate of 44.1 kHz by a female native speaker of Dutch with normal intonation. A speech editor (Adobe Audition 3.0) was used to slightly adjust the auditory stimuli in such a way that all definite determiners started around 800 ms and all indefinite determiners around 2000 ms after the onset of the sentence. The mean duration of the determiners was 214 ms in the definite condition and 188 ms in the indefinite condition. On average, adjectives following a definite determiner lasted for about 532 ms and adjectives in the indefinite condition lasted about 493 ms. The nouns also lasted approximately equally long across conditions with 669 ms and 681 ms for definite and indefinite sentences, respectively. The relatively constant timing delay of 1200 ms for the indefinite NPs as compared to the definite constructions allowed for a direct comparison of the responses to these different types of sentences. Possible effects of these slight timing differences between the two conditions, however, were controlled for by including these timing variables in the mixed-effects regression analysis.

Table 4-2 shows the four conditions in the present experiment. Each trial consisted of a screen divided into four quadrants in which four pictures appeared: the target, a competitor matching in colour and/or gender with the target, and two distractors that never matched in gender or in colour with the target. The initial segments of the nouns of the four pictures never overlapped to exclude any phonological competition effects. Positions of the target, competitor and distractor objects on the screen were counterbalanced. In addition, each item occurred on the screen once as a target, once as a competitor, and twice in the distractor role. Following a Latin Square design, four different lists were created in such a way that each participant was presented only one version of each item in a certain condition.

Two blocks were created: one block containing instructions with definite NPs, and one block containing instructions with indefinite NPs. Each block consisted of 48 experimental plus 22 filler trials.

Table 4-2: Overview of the conditions in the experiment containing example stimuli while hearing the instruction to *Klik op de*$_{COM}$ *rode appel*$_{COM}$ ('Click on the$_{COM}$ red apple$_{COM}$ '). Note that in the definite conditions, gender-marking is on the definite determiner (*'de'* when referring to common and *'het'* when referring to neuter nouns) while gender-marking only appears on the adjective in indefinite NPs (*schwa*-ending when modifying a common noun and bare adjective when modifying a neuter noun).

Target	**Competitor**	**Gender Competitor**	**Colour Competitor**
De$_{COM}$ rode appel$_{COM}$ *The*$_{COM}$ *red apple*$_{COM}$	Het$_{NEU}$ groene bureau$_{NEU}$ *The*$_{NEU}$ *green desk*$_{NEU}$	different	different
	De$_{COM}$ gele zon$_{COM}$ *The*$_{COM}$ *yellow sun*$_{COM}$	same	different
	Het$_{NEU}$ rode hart$_{NEU}$ *The*$_{NEU}$ *red heart*$_{NEU}$	different	same
	De$_{COM}$ rode taart$_{COM}$ *The*$_{COM}$ *red cake*$_{COM}$	same	same

4.2.4. Procedure

Eye movements were recorded in the Eye Lab of the University of Groningen using a remote Tobii T120 eye tracker with a sampling rate of 60 Hz. Auditory and visual stimuli were presented using E-Prime (Schneider, Eschman, & Zuccolotto, 2002) and reaction times as well as eye movement data were recorded using the E-Prime Extensions for Tobii software.

Participants were seated in a comfortable chair at approximately 60 cm from the eye tracking monitor. After a standard 5-point calibration procedure, participants were instructed to click (using the mouse) on the correct picture as quickly as possible. The experiment began after a short practice session containing 6 filler items.

Each trial started with a fixation cross at the centre of the screen. To ascertain that the participants' eyes were focussed at the centre of the screen at the beginning of each trial, subjects had to fixate on the fixation cross for at least 1000 ms in order for the trial to start. Auditory instructions started as soon as the display of four pictures was presented on the screen. Immediately after clicking on a picture, the set of pictures disappeared and the next trial started.

Each subject completed two blocks: one containing instructions with definite and one with indefinite NPs. Half of the subjects received the definite instructions first and the other half received the indefinite instructions first. Each block lasted around 12 to 13 minutes and subjects were allowed a short break between the two blocks.

4.2.5. Statistical analyses

In most eye tracking studies, the dependent variable is the proportion of fixations towards the target. At present, the vast majority of studies analyse this data using subject and item ANOVAs or t-tests in order to obtain generalizable results. This approach *incorrectly* follows Clark (1973) who proposed a min-F measure based on F1 and F2 to generalize over subjects and items. As min-F is not equal to conducting a separate F1 and F2 analysis, many studies may report spurious significance (type-I errors). Additionally, it is not always necessary to take F2 into account, especially when the research design is counterbalanced (Raaijmakers, Schrijnemakers, & Gremmen, 1999). Using ANOVA to analyse categorical data (or proportional data, which is bounded by 0 and 1) is even more problematic as it violates the assumptions necessary to use an ANOVA and can lead to spurious significance (type-I errors) and null results (type-II errors; Jaeger, 2008). More suitable analyses have therefore been proposed, including growth curve analysis (Magnuson, Dixon, Tanenhaus, & Aslin, 2007) and mixed-effects (logistic) regression (Barr, 2008; Huettig, Rommers, & Meyer, 2011).

In this study we will apply mixed-effects regression modelling (for introductions, see, e.g. Baayen, 2008, Ch. 7 and Baayen, Davidson and Bates, 2008). In mixed-effects regression modelling fixed-effect and random-effect factors are distinguished. Fixed-effect factors have a small number of levels that exhaust all possible levels (e.g. the gender of a participant is either male or female). In contrast, random-effect factors have levels sampled from a much larger population of possible levels.

In our data, there are two random-effect factors that are likely to introduce systematic variation, namely participant and item (which has a one-to-one correspondence with a certain scene). Our observations (i.e. measurements) are specific to each participant. As these participants are a sample of a much larger set of possible participants which might have participated in our study, participant is a random effect factor. Similarly, since we could have selected different items, item is also a random effect factor.

By including random-effect factors, the structural variability associated with these factors can be taken into account. For example, some participants might focus faster on the target item than others. These adjustments to the population intercepts ('random intercepts') can be used to make the regression formula (for each participant and item) as precise as possible. Similarly, it is possible that there is variability (associated with the random-effect factors) in the effect a certain predictor has. For example, the effect (i.e. slope) of participant age might vary per item, indicating that some items might be more easily recognized by older participants than younger participants, while other items may show the opposite pattern. Together with the random intercepts, these random slopes make the regression formula as precise as possible (for each participant and item). The necessity of including these random intercepts and random slopes is verified with likelihood ratio tests evaluating whether the increase in the number of parameters is justified given the increase in goodness of fit.

The analyses used in the present paper considered such random-effect factors, as well as a contrast to distinguish the presence or absence of a colour competitor and a gender competitor in the scene (i.e. the contrast we are most interested in). In addition, we included item-related characteristics such as the gender and word frequency of the item, as well as participant-related characteristics such as gender and age. To control for a possibly imbalanced design, factors such as the position and colour of a specific item were also included in the analyses. Potential effects of fatigue and/or learning during the course of the experiment were accounted for by including predictors such as trial-number as well as dependent measures corresponding to the previous trial and testing their significance in the model.

We evaluated the significance of fixed-effect factors by means of the usual *t*-test for the coefficients. Since our data set contains a large number of observations (about one thousand items), the *t*-distribution approximates the standard normal distribution and factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

As we had two dependent variables, reaction time (henceforth RT) and eye movement data (henceforth EYE), we created two separate mixed-effects regression models. The models were fitted using a stepwise variable deletion procedure: predictors that did not contribute significantly to the model were removed. Potential nonlinear effects of each of the continuous predictor variables were also assessed. As collinearity can be a problem even in rather balanced designs, predictors were centred before analyses (Jaeger, 2008).

4.3. Results

Of all 2688 item responses (28 subjects x 48 items x 2 blocks), 2677 were answered correctly (99.6%) and only 11 items (0.4%) were answered incorrectly and therefore removed from further analysis. For 6 additional items, responses had only been given later than 1.5 seconds after the onset of the noun. These 6 extreme outliers (0.2%) were also removed. In addition to minimal a-priori data trimming, potentially influential outliers were removed after fitting a first version of the model (i.e. model criticism, see Baayen & Milin, 2010). In this study, linear mixed-effects models were fit using the *lmer* function of the lme4 package (Bates, 2005) implemented in R (version 2.11.1: The R Foundation for Statistical Computing, 2010).

The mean timing delay of 1200 ms in the indefinite constructions (compared to the definite constructions) was subtracted from the timing data in order to be able to directly compare responses from both conditions.

4.3.1. Behavioural data (reaction times)

Out of the 2671 responses of the initial data set, absolute standardized residuals exceeding 3 standard deviations were removed (0.5%). A linear mixed-effects model was fitted to the log-transformed RT data using a stepwise variable deletion procedure. The best fitting model including all necessary fixed and random effect factors (shown in Tables 4-3 and 4-4) explained approximately 67% of the variance in the RT data and the residuals of the model followed a normal distribution.

Table 4-3 shows the coefficients and the associated *t*-values of the fixed-effect factors. The random-effects structure is outlined in Table 4-4. Table 4-3 reveals an intercept of about 957 ms (6.8640 in logarithmic scale), which can be seen as the base reaction time for a given item. The coefficients of the remaining predictors or fixed factors reveal whether this base value increases or decreases when a specific variable is present.
The main question addressed in the present experiment was whether gender-marking influences the comprehension process by facilitation of gender-matching and/or inhibition of gender-mismatching nouns. As shown in Table 4-3, the presence of a competitor that shared gender with the target (a *SameGenderCompetitor*) did not by itself significantly impact reaction times ($\beta = -0.0459$, $t = -1.78$). However, when looking at the interaction with age, we observed that older people had a stronger negative impact of the presence of a *SameGenderCompetitor* than younger people (β =

0.0012, $t = 2.31$, $p < 0.05$, two-tailed test). In other words, a *SameGenderCompetitor* only increased RTs in the older participants. As consistently found across studies (see, e.g. Der & Deary, 2006), older participants were also generally slower in choosing the correct target picture (β = 0.0050, $t = 3.51$, $p < 0.01$).

The second interaction (*SameColourComp:SameGenderComp*) revealed that the effect of colour, i.e. significantly slower RTs when the competitor shared colour with the target (β = 0.1536, $t = 19.48$, $p < 0.001$), was less strong when the competitor also shared gender with the target (β = 0.1536 - 0.0152 = -.1384, $t = -2.17$, $p < 0.05$). This counterintuitive result suggests that a gender competitor speeds up the visual decision making process. In addition to the variables of interest, we tried to account for potential fatigue and/or learning effects and tested for both the effect of trial (*Sequence*) and effects of RTs on the previous trial (*PreviousRT*). An overall effect of *Sequence* was found (β = -0.0006, $t = -3.85$, $p < 0.001$), showing that participants' RT generally decreased during the course of the experiment. In addition to this learning effect, participants responded more slowly when they had a slower RT in response to the previous trial (β = 0.0744, $t = 4.92$, $p < 0.001$). By adding this control variable, temporal dependencies are not only controlled for, but inclusion also allows for a more precise estimation of the effects of the other predictors (Baayen & Milin, 2010). A second control variable, *LemmaFrequency*, also reached significance, showing that higher frequency items were responded to significantly faster (β = -0.0071, $t = -2.46$, $p < 0.05$).

Table 4-3: Fixed-effect factors in the model fitted to the reaction time (RT) data. The table provides the Estimated Coefficients, standard errors, and *t* values for the mixed-effects regression model fitted to the log-transformed reaction times (RT) in the eye tracking experiment. Positive coefficients reflect an increase in response time and negative coefficients a decrease. Factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

	Estimate	***Std. Error***	***t value***
(Intercept)	6.8640	0.1325	51.80
SameColourCompetitor	0.1536	0.0079	19.48
PreviousRT (log)	0.0744	0.0151	4.92
Age	0.0050	0.0014	3.51
Sequence	-0.0006	0.0002	-3.85
TargetIsLeft	0.0103	0.0037	2.81
TargetIsBrown	0.0297	0.0106	2.80
LemmaFrequency (log)	-0.0071	0.0029	-2.46
*n*SeenCompetitor	0.0046	0.0020	2.25
SameGenderCompetitor	-0.0459	0.0258	-1.78
Age:SameGenderCompetitor	0.0012	0.0005	2.31
SameColourComp:SameGenderComp	-0.0152	0.0070	-2.17

The analyses also included item- as well as trial-related characteristics such as the colour of the item, its competitor and the position of these pictures on the screen.

Inclusion of these predictors showed that participants were significantly slower when the target picture was coloured brown (β = 0.0297, $t = 2.80$, $p < 0.05$). In addition to the saliency of different colours, positioning of the pictures also influenced RTs: people were slower in clicking on the target when it was shown on the left side of the screen as opposed to the right (β = 0.0103, $t = 2.81$, $p < 0.05$). Although this latter result is not straight-forward to interpret, it might be related to the right-handedness of our subjects, and consequently their preference moving their hand from left to right. This might cause a delay in manual responses to objects at the left of the screen.

In order to match the targets, competitors, and distractors, the present design consisted of a set of 48 images of which each image appeared on the screen once as a target, once as a competitor, and twice as a distractor. The analyses revealed that a higher number of times a competitor had previously been on the screen in a different role, e.g. the role of distractor, significantly increased the time it took participants to select the correct referent (β = 0.0046, $t = 2.25$, $p < 0.05$, two-tailed test). This might constitute a recognition effect for the competitor, but the absence of such an effect for the amount of times the target had been on the screen reveals that this effect probably results from the tendency to pre-select the target picture, which is more likely to be a picture that has previously been on the screen in a different role, such as the role of distractor.

Table 4.4 shows the two random-effect factors (labelled as Groups) of the present experiment: *Item* and *Subject*, where *Item* refers to a specific picture or noun and *Subject* refers to the participant. The table lists the standard deviations for the adjustments to the intercept for both random effect factors and, for subject standard deviations are also listed for the adjustments to coefficients (i.e. slope) of 4 covariates: *SameColourCompetitor*, *TargetIsBrown*, *IsIndefinite*, and *Sequence*.

Table 4-4: Random-effect parameters in the model fitted to the reaction time (RT) data. The numbers in the column entitled 'Correlations with Intercept' represent correlations between the by-subject random intercept and the random slope for *SameColourCompetitor* and *TargetIsBrown*, respectively. The number in the second column represents the correlation between the by-subject random slopes for *SameColourCompetitor* and *TargetIsBrown*.

Groups	*Name*	*Standard Deviation*	*Correlations with Intercept*	*Correlations between Slopes*
Item	Intercept	0.0231		
Subject	Intercept	0.0658		
	SameColourComp	0.0324	-0.719	
	TargetIsBrown	0.0164	-0.950	0.466
Subject	IsIndefinite	0.0313		
Subject	Sequence	0.0005		
Residual		0.0903		

The population slope for *SameColourCompetitor* was 0.1536 for different gender trials and 0.1384 (0.1536 - 0.0152) for same gender trials. Figure 4-1 shows, however, that subjects not only differ with respect to the effect that a specific variable has on their RTs, but that these different variables also correlate with one another. Subjects with a more positive slope for *SameColourCompetitor* tend to be the fast responders, while slow subjects have a more negative slope for *SameColourCompetitor*. In other words, faster participants suffered more from a competitor that shared colour with the target resulting in a relatively large increase in RTs as compared to a relatively small increase in RTs for the overall slower participants.

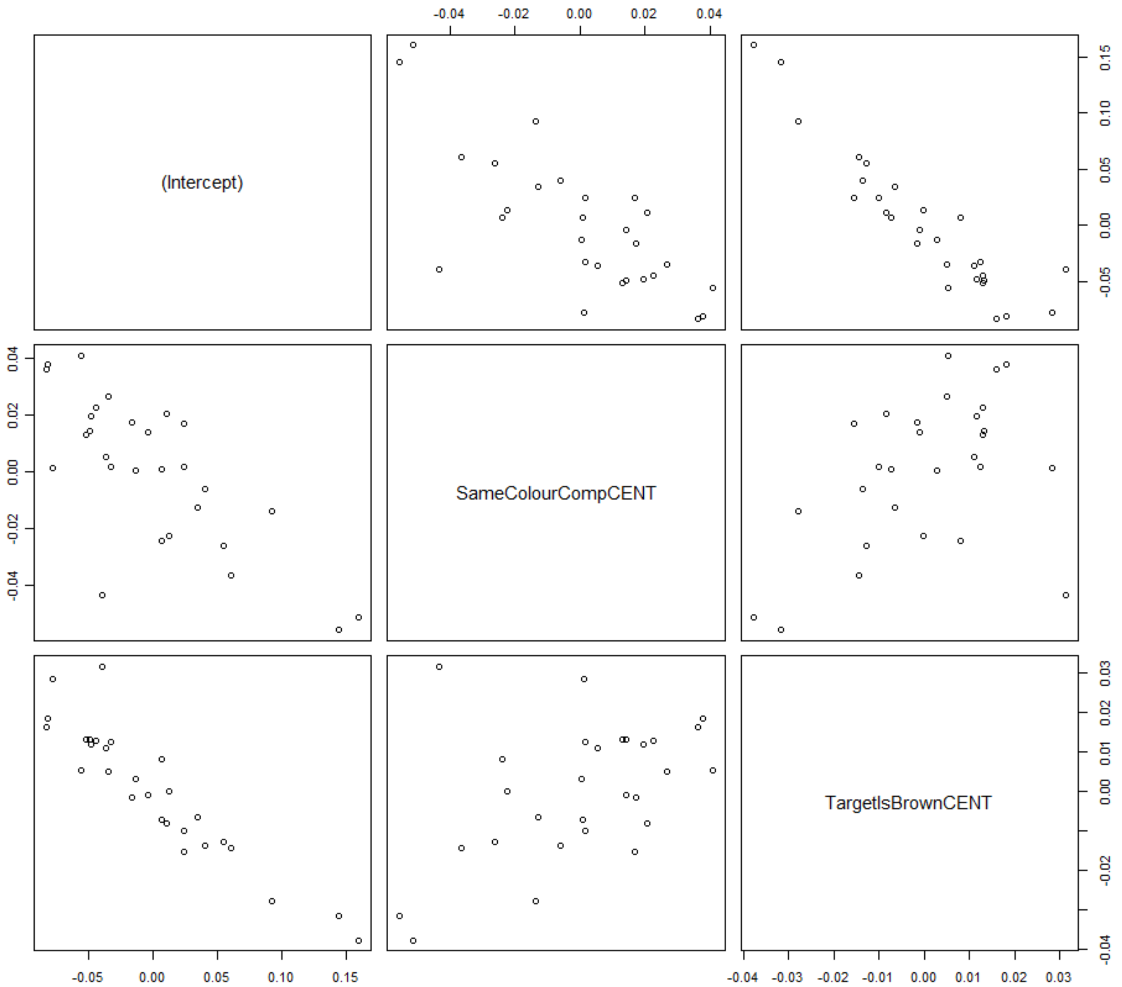


Figure 4-1: Scatterplots visualizing the correlational structures of the random intercepts and random slopes for *Subject* in the model for the reaction time (RT) data. The adjustments to the intercept position subjects with respect to the average response time. Subjects with large positive BLUPS (best linear unbiased predictions) are slow subjects; those with large negative BLUPS are fast subjects. This graph illustrates that fast subjects tend to suffer more from a *SameColourCompetitor* than slower subjects. The correlational structure between *SameColourCompetitor* and *TargetIsBrown* reveals that those subjects who suffer more from a *SameColourCompetitor* also suffer more when the *TargetIsBrown*.

Similarly, Figure 4-1 shows that the slower subjects suffered less from a brown target as compared to fast subjects. The coefficient of *TargetIsBrown* for the population is estimated to be 0.0297. Fast subjects, i.e. subjects with negative adjustments to the intercept, are characterized by upward adjustments for *TargetIsBrown* slopes. Furthermore, there is a positive correlation of the adjustments for *SameColourCompetitor* and *TargetIsBrown* (0.466), showing that subjects who suffer more from the presence of a competitor that shares colour with the target also suffer more when the target depicted is coloured brown.

In addition, by-subject random slopes for the effects of *IsIndefinite* and *Sequence* significantly improved model fit. In other words, some subjects were slower in response to indefinite constructions, while others were faster. Similarly, the learning effect that was found differed significantly among subjects, with some of them showing a large effect of *Sequence* while others only show minor or no learning effects across the duration of the experiment.

4.3.2. Eye Movement Data

Since all relevant and distinctive information that might be used to pre-select the target picture (i.e. gender-marking and colour information) occurs between the onset of the determiner and the onset of the noun, this time period was used to analyse eye movements (taking into account the 200 ms it takes to plan and launch an eye movement). Proportions of fixations are depicted in Figure 4-2.

For each item and each time point, the gaze location was calculated as being either on the Target, on the Competitor, on the Distractors, or not on an Area of Interest. For 8% of the data, the Tobii system reported to be fairly confident that the data was incorrect or corrupted (validity codes > 2). In order to retain as much data as possible, the gaze location for these specific time points was set to NA. In addition, for each individual item and subject combination an error percentage was calculated. Items for which the validity codes of the eyes were higher than 2 for more than 30% of the time were removed (9.1%). For the resulting data set, mean proportions of fixations were calculated across the entire time period (i.e. from determiner onset until noun onset plus 200 ms) per subject and per item.

It would be expected that, if participants use gender-marking cues and colour information to pre-select the correct picture, this would result in more fixations towards the target as compared to the competitor when the information is disambiguating. Therefore, the dependent variable *TargetRatio* was calculated as the percentage of looks towards the target divided by the total percentage of looks towards the target and its competitor. To avoid problems associated with performing statistical analyses on proportion data, mixed-effects models were fit to the empirical logit transformations of the variable *TargetRatio* (Barr, 2008; Jaeger, 2008).

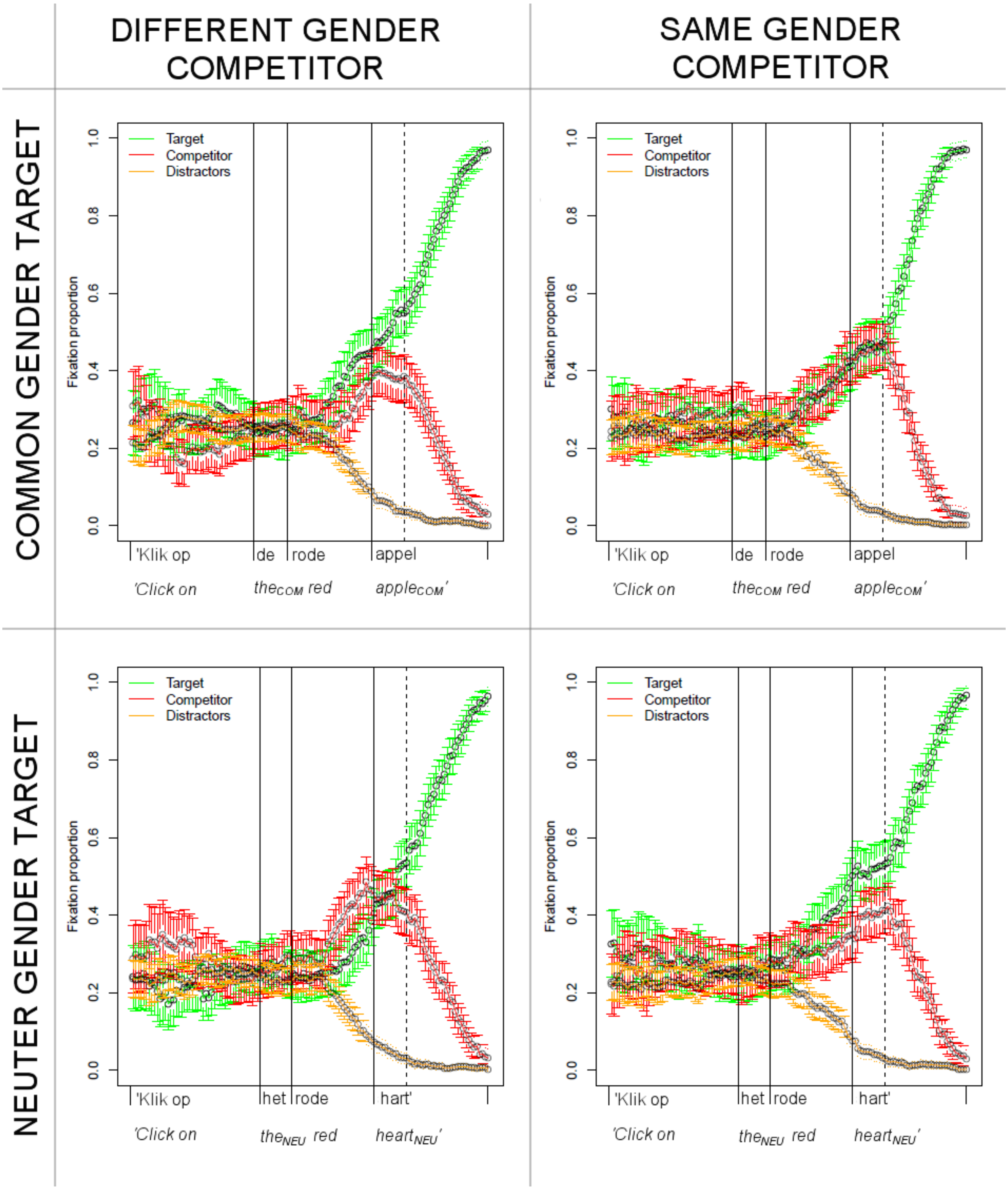


Figure 4-2: Proportions of fixations towards target, competitor, and distractors are depicted only for the conditions in which target and competitor shared colour. Trials in which the competitor did not share gender with the target are depicted in the left panel and same-gender trials are depicted on the right. The above panel shows the fixation proportions in conditions where the target was common and the lower panel reveals eye movement data in response to neuter gender targets.

A linear mixed-effects regression model was fitted to the data using a stepwise variable deletion procedure. Absolute standardized residuals exceeding 3 standard deviations were removed (1.3%) after which the model was refitted. The best model including all fixed and random effect factors (shown in Tables 4-5 and 4-6) explains approximately 32% of the variance in looking behaviour and the residuals of the model followed a normal distribution. Before explaining the random effect structure of this model, the fixed-effects will be explained in detail.

As opposed to the reaction time data, the EYE data do reveal a relative difficulty in distinguishing target and competitor when both referents shared gender. The negative coefficient for *SameGenderCompetitor* in Table 4-5 reveals that participants looked relatively more towards the competitor when it shared gender with the target ($\beta = -0.0484$, $t = -2.22$, $p < 0.05$). In addition to this general effect of gender, the analysis also revealed a difference between common and neuter nouns: when the *TargetIsNeuter*, i.e. when it is preceded by '*het*' in the definite or a bare adjective in the indefinite condition, people took significantly longer to fixate on the target ($\beta = -0.0856$, $t = -3.37$, $p < 0.01$).

Interestingly, these two predictors interacted in that the observed gender effect was present in common nouns, but absent or even reversed when both the target and the competitor were neuter ($\beta = 0.0839$, $t = 2.67$, $p < 0.05$). This difference between common and neuter nouns is also visible in Figure 4-2, where the upper panel clearly reveals an earlier selection of the target picture when it has unique gender (upper-left picture) as compared to same-gender trials (upper-right picture). The lower panel, depicting eye movements in response to neuter nouns, shows that the opposite happened in response to neuter nouns: people tended to pre-select the target when target and competitor shared gender, while relatively more looks towards the competitor occur when the gender of the target is unique. The absence of a gender effect for neuter nouns can be directly linked to the fact that all Dutch nouns, regardless of the gender they carry, are preceded by the neuter determiner '*het*' when used as a diminutive. Consequently, hearing neuter gender-marking preceding the noun does not automatically mean that all referents to common nouns can be ignored as possible targets. The fact that the gender effect appears to be reversed for neuter nouns, i.e. the fact that a competitor sharing both colour and gender with the target makes it easier as compared to a competitor that only shares colour with the target, is very difficult to explain. Although Figure 4-2 does suggest that looks towards the competitor decrease earlier in the different-gender as compared to the same-gender trials, there is an obvious preference for the competitor in the different-gender and a preference for the target in the same-gender trials.

Table 4-5: Fixed-effect factors in the model fitted to the eye movement data. The table provides the Estimated Coefficients, standard errors, and *t* values for the mixed-effects regression model fitted to the output measure *TargetRatio* (the empirical logit transformation of the relative difference between the proportion of looks towards target and competitor) in the eye tracking experiment. Positive coefficients reflect an increase in *TargetRatio* and negative coefficients a decrease. In this case, an increase in *TargetRatio* reflects relatively more fixations on the target as compared to the competitor in the relevant time-window. A decrease in *TargetRatio* signifies a difficulty in distinguishing target and competitor resulting in a more balanced distribution of fixations on the target and the competitor or relatively more looks towards the competitor.

	Estimate	***Std. Error***	***t value***
(Intercept)	-2.9161	0.3526	-8.270
SameColourCompetitor	-0.3138	0.0201	-15.61
PreviousTargetRatio	0.0361	0.0197	1.83
TimeOnsetNoun	-0.0010	0.0001	-6.67
TimeOnsetDeterminer	0.0016	0.0003	6.36
Age	-0.0045	0.0014	-3.23
IsMale	-0.0558	0.0195	-2.86
TargetIsBrown	-0.0836	0.0211	-3.97
TargetIsUp	0.0666	0.0251	2.65
CompIsRight	-0.0396	0.0159	-2.49
TargetIsNeuter	-0.0856	0.0254	-3.37
SameGenderCompetitor	-0.0485	0.0219	-2.22
SameGenderComp:TargetIsNeuter	0.0839	0.0314	2.67

Table 4-5 reveals the same substantial effect of *SameColourCompetitor* that was visible in the reaction time data ($\beta = -0.3138$, $t = -15.61$, $p < 0.001$). In this case, the presence of a competitor with the same colour as the target picture causes a significant decrease in *TargetRatio,* i.e. people fixated less on the target and relatively more on the competitor when there is a S*ameColourCompetitor* present. Reaction times already showed that people also suffered when the target picture was coloured brown: participants were significantly slower in clicking on such target pictures. The negative effect of *TargetIsBrown* is replicated in the eye data with people looking significantly less towards a brown target as compared to the competitor ($\beta = -0.0836$, $t = -3.97$, $p < 0.001$). Furthermore, looks to the target significantly increased when the target was presented in one of the two upper quadrants of the screen ($\beta = 0.0666$, $t = 2.65$, $p < 0.05$). When the competitor was placed in one of the right corners of the screen, (i.e. *CompIsRight),* looks towards the competitor also increased relative to the looks towards the Target ($\beta = -0.0396$, $t = -2.49$, $p < 0.05$). Similar to the RT data, *Age* significantly affected the dependent variable with older people taking significantly longer in choosing the correct target picture resulting in a lower *TargetRatio* ($\beta = -0.0045$, $t = -3.23$, $p < 0.01$, two-tailed test). The negative coefficient for *IsMale* shows that male participants generally showed a lower score of the dependent variable *TargetRatio*, reflecting a more equal distribution of looks towards target competitor as compared to female participants (β = -0.0558, t = -2.86, $p < 0.05$).

The auditory stimuli were recorded and manipulated in such a way that the determiner always started around 800 ms after the onset. Minor differences in timing between the items, however, could not be further reduced and the timing details were included in the model. Table 4-5 reveals that an increase in *TimeOnsetNoun,* i.e. the time (in ms) at which the noun started in the speech stream, significantly reduced the *TargetRatio* (β = -0.0010, t = -6.67, p < 0.001). A later *TimeOnsetDeterminer*, however, increased *TargetRatio* (β = 0.0016, t = 6.36, p < 0.001). These results suggest that the later the participant heard and could process the determiner, the more the target was focussed on relative to the competitor before the onset of the noun. It should be noted that the variable *TargetRatio* was calculated per item and per subject. The slight difference between the timing of the onset of the determiner and the noun thus caused a difference in the length of the time-window over which *TargetRatio* was calculated. This length was taken into account in the calculation of the empirical logit transformation of *TargetRatio.* The lower *TargetRatio* values in response to a later *TimeOnsetNoun* thus most likely reflect the fact that the information on the determiner and/or adjective had not yet been adequately processed for these items. The higher *TargetRatio* in response to items with a later *TimeOnsetDeterminer* is more difficult to explain, but might reflect that people have had more time to scan the pictures, which might help in detecting the target picture more efficiently.

Potential effects of fatigue and/or learning were again accounted for by testing both the effect of *Sequence* and the effects of the *TargetRatio* for the previous trial. *Sequence* did not affect the output measure, but subjects were found to have a higher *TargetRatio* when they had a higher *TargetRatio* on the previous trial (β = 0.0361, t = 1.83, $p < 0.05$, one-tailed test).

Table 4-6: Random-effect parameters in the model fitted to the eye movement data.

Groups	***Name***	***Standard Deviation***	***Correlations with Intercept***	***Correlations between Slopes***
Subject	Intercept	0.0262		
Subject	SameColourComp	0.0631		
Subject	TargetIsUp	0.0996		
Residual		0.3578		

In contrast to the RT data, the eye movement data did not require a random intercept for item, which was confirmed by likelihood ratio tests. This is probably related to the time-window that was analysed, from determiner onset to noun onset, in which the noun's referent has not yet been mentioned in the speech stream. In relation to this, *LemmaFrequency* had no effect on the mean proportions of fixation in the time-window analysed. As shown in Table 4-6, our model did require a random intercept for *Subject*, revealing again that some people have higher *TargetRatio* scores while other people have lower *TargetRatio* scores. Similar to the RT data, people tend to be differentially affected by the presence of a *SameColourCompetitor,* with some people suffering more than others from this predictor. Whether the target appeared in the upper or in the lower panel of the screen also impacted people's eye movements differently and inclusion of by-subject random slopes for these effects significantly improved model fit.

4.4. Discussion

Previous eye tracking experiments have shown that competition between words during language comprehension does not only occur on the basis of phonological information (e.g. Allopenna et al., 1998), but also on the basis of morphological cues such as gender-marking (e.g. Dahan et al., 2000; Paris et al., 2006). The types and amount of gender categories as well as their impact on agreement targets, however, varies between languages. The Dutch gender system, as opposed to many Romance languages for example, is relatively covert in that a noun's gender is rarely predictable on the basis of phonological and/or morphological cues in the input (Haeseryn et al., 1997). The two-way Dutch gender system is characterized by its asymmetrical distribution with about 75% of all Dutch nouns having common gender (Van Berkum, 1996). The distinct status of neuter gender is enhanced by the fact that all words use the common determiner '*de*' in the plural. On the other hand, every Dutch noun can become neuter when used in the diminutive form. The present study investigated whether gender-marking in such an asymmetrical and non-transparent gender system is used by Dutch natives during language processing to anticipate gender-congruent and/or inhibit gender-incongruent nouns.

Eye movements were recorded while people listened to sentences such as *Klik op de*$_{COM}$ *rode appel*$_{COM}$ ('Click on the$_{COM}$ red apple$_{COM}$') while looking at a screen containing the target, a competitor matching in colour and/or gender with the target, and two distractors that never matched in gender nor colour with the target. We hypothesized that, if people use prenominal gender-marking to pre-select the correct referent, this should become visible before hearing the onset of the noun (which was unique to the target), resulting in a relatively higher amount of fixations towards the target as compared to the competitor when the information in the relevant time-window contained disambiguating colour and/or gender information. Mixed-effects regression models were fitted to both the (log-transformed) reaction time (RT) data and the relative difference in fixation proportions between target and competitor in the relevant time-window, i.e. the time-window from determiner onset to noun onset in which both gender and colour information was presented.

A gender effect was found only in the eye data and, more importantly, only common gender-marking appears to affect processing of the subsequent noun by inhibiting gender-incongruent and/or anticipating gender-congruent nouns. This result is an immediate reflection of the special status of neuter gender in Dutch: all Dutch nouns can become neuter when used in the diminutive form. A common noun can thus still become a neuter noun and consequently, neuter gender-marking cues cannot disambiguate between common and neuter nouns. While some studies suggest that gender-marking on the article does not, by itself, seem to restrict the possible set of lexical candidates to gender-congruent nouns (Dahan et al., 2000), others suggest that gender cues facilitate processing immediately after hearing the gender-marked article (Paris et al., 2006). Reasoning in line with Paris et al. (2006), we used intervening adjectives in the present study to allow the participants more time to process gender cues. More importantly, since Dutch gender-marking either appears only on the determiner (definite NPs) or on the adjective (indefinite NPs), the present experiment investigated both constructions in order to disentangle this issue of locality. The fact that no difference was found between the two conditions suggests that the effect of gender is present in and similar for both constructions. In other words, gender-marking on the adjective (as in indefinite NPs) occurs immediately before the noun and the presence of a gender effect in these conditions suggests that Dutch natives do have enough time to process gender information and hence pre-select the correct referent on the screen based on the gender-marking cues present in the auditory input.

The results reported by Paris et al. (2006) were replicated in that a gender effect was found before the onset of the noun, which was unique to the target. The fixation proportions also reveal, however, that this effect occurs at the offset of the adjective as opposed to immediately after hearing the determiner. This might be the result of colour being an extremely salient feature. Both the RT data as well as the eye movement data revealed a large effect of colour: participants waited until they had heard the (onset of the) adjective before choosing the correct target picture and the fixation proportions suggest that, even if there is a disambiguating determiner present in the input, people tend to look at both the competitor and the target equally often until right before hearing the unique initial phonemes of the noun.

While the eye movement data in response to common nouns show a gender effect, i.e. a same-gender competitor slows down the process of locating the target; the opposite pattern was found for neuter nouns, i.e. a same-gender competitor appears to speed up the process of locating the target as compared to a different-gender competitor. Although the data, as presented in Figure 4-2, do suggest that looks towards the competitor decrease earlier in the different-gender as compared to the same-gender trials, there is an obvious preference for the competitor in the different-gender and a preference for the target in the same-gender trials. Similarly, the RT data suggested that the effect of colour decreased when the competitor also shared gender with the target. This result initially seems rather counterintuitive, but another recent study on gender processing yielded similar results. Cubelli, Paolieri, Lotto, and Job (2011) asked their subjects to semantically categorize two objects that either matched or did not match in gender (in Italian). The time to categorize pictures decreased when the two objects were from the same grammatical gender category. The authors hypothesized that, during the categorization task, lexical representation associated with the objects are activated and nouns within the same gender category activate each other, facilitating processing and speeding up response times. Similar results have recently also been reported in 30-months old children when confronted with two pictures from different semantic categories. These children were reported to look more towards the target picture when the competitor shared gender with the target as compared to the different-gender trials (Bobb & Mani, 2012). These results, in combination with the gender effect for common nouns occurring before noun onset, imply that gender information of the objects on the screen is made available before encountering the gender-carrying noun itself and argues in favour of a pre-activation account of gender. The fact that this facilitation of same gendered items is especially visible in the eye movement data for neuter nouns is in line with the fact that neuter Dutch nouns are relatively scarce and make up about 25% of all nouns (Van Berkum, 1996). It is thus easier to activate neuter gender nouns simply because there are relatively fewer nouns to activate and hence facilitation within this gender category speeds up the recognition process.

The RT data also reveal that the effect of gender increased with age, which corresponds to the overall age effect that was found in both the RT and the eye data. As expected, item-related characteristics such as frequency only affected the RT data with faster reaction times in response to higher frequency items (Dahan et al., 2001). Similar frequency effects were not replicated in the eye fixations, which is not surprising when considering that we analysed eye movements in the time-window before noun onset. By using mixed-effects regression analyses, the present experiment allowed to include variables in the model that are normally considered to be rather balanced. Although the present experiment also used a counter-balanced design, usage, placement and colouring of the pictures appeared to impact both reaction times and eye fixations. To create a balanced design, the present study presented every item picture once as a target, once as a competitor, and twice as a distractor. The analyses revealed that the number of times that a competitor had previously been on the screen in a different role significantly increased the relative looks towards this competitor. This effect was only found for pictures in the competitor role and not in the target role, suggesting that participants' focus was primarily on their task to select the target picture as soon as possible. The target is more likely to be a picture that had previously been on the screen in a different role, such as the role of distractor, and hence there is a greater chance that this should be the picture in the competitor role.

Not only the number of times a picture was shown affected participants' behaviour, but also the colouring and positioning of these items on the screen. When the target was brown, people were slower in locating it with their eyes, but they were also slower in clicking on the target with the mouse. It has been known for a long time that colour plays an important role in capturing a person's attention and some colours appear to attract more attention than other colours, simply because they are more salient (Osberger & Rohaly, 2001). As a colour of low intensity, brown is a tertiary colour, which means that it is a mix of three subtractive primary colours. This causes brown objects to be less noticeable and hence more difficult to find and respond to. The RT data also revealed that subjects were slower in clicking on a target when it was positioned on the left side of the screen. We hypothesized that, since all subjects were right-handed, they tended to prefer moving their hand from left to right which caused a delay in manual responses to objects on the left of the screen. Similarly, the difference in fixations towards the target and competitor positively increased when the target was positioned in one of the upper quadrants of the screen. Contrary to this, the target was more difficult to find when its competitor occurred in one of the right quadrants on the screen. The latter effect is hard to explain, but the fact that upper pictures are more easily detected most probably arises from the fact that these are the first pictures the subjects fixate on. Most participants tend to scan the screen from left to right and from the upper to the lower pane. Consequently, the concepts corresponding to the upper pictures have already been activated and hence eye movements are more easily directed towards these referents as compared to the referents in the lower pane. The fact that, as visible in the eye data, people tend to suffer from a competitor on the right side of the screen might be caused by them not having activated the referent of these competitors yet. This, however, remains speculative and it should be stressed that these predictors are not of main interest to the present study, but their inclusion in the model allows for a more precise estimation of the effects of other, more important variables. The same holds for the timing variables that significantly affected the eye movement data.

Minor differences in timing between the onset of determiners and nouns in the auditory input differentially affected participants' eye movements. A later onset of the noun was related to more equally distributed fixations on target and competitor, while a later onset of the determiner tended to increase relative fixations on the target as compared to the competitor. Although these conflicting results are difficult to explain, we hypothesized that a later onset of the noun is more likely to relate to a later onset of the gender information and this information probably had not yet adequately been processed for these items. Relatively more fixations on the target in response to later onsets of the determiner might well reflect the increase in the amount of time participants had to scan the pictures prior to receiving disambiguating information, which might help in detecting the target picture more efficiently.

Mixed-effects modelling not only allows including multiple predictors that could not have been added in an ANOVA analysis, but it also allows for the inclusion of random intercepts for both subject and item and to vary the slopes of specific predictors across subjects or items. Both the RT data as well as the eye movement data showed that a competitor sharing colour with the target differentially affected participants, with some of them suffering more from a same coloured competitor than others. The RT data additionally showed that fast responders suffered more from a competitor with the same colour than slower subjects and these subjects, likewise, also suffered more from a target being brown. Mixed-effects regression models allow the researcher to acknowledge that people respond differently to certain variables.

The present experiment found a gender effect in eye movement data, but not in reaction time data, confirming the advantages of studying language processing by investigating eye movements. Given that the gender effect is present in a non-transparent language such as Dutch and for nouns that do not share phonemic onsets, argues against the idea that previously reported gender effects, such as in French (Dahan et al., 2000), were dependent on the onset of the noun reflecting a phonological or lexical effect. Instead, the present study suggests that gender information is, at least partially, activated automatically before encountering the noun. We therefore have provided at least some evidence for a more automatic, pre-activation account on gender processing.

Chapter 5

PREDICTABILITY SPEEDS UP THE RE-ANALYSIS PROCESS:

An ERP Investigation of Gender Agreement and Cloze Probability

To investigate the timing relationship between semantic and syntactic processing, the present study compared event-related potentials (ERPs) in response to spoken Dutch sentences that were either correct or contained gender agreement violations on the article or adjective preceding the noun. The target noun was either unpredictable from the preceding sentence information (low cloze) or was preceded by a highly constraining context (high cloze) to investigate whether contextual constraints influence morphological processing of grammatical gender in real-time. In line with previous findings, gender violations elicited a P600 indicating processes of repair or re-analysis in response to the gender mismatch. Low cloze items, independent of the gender mismatch, elicited an increased N400 reflecting semantic integration difficulty. Interestingly, an interaction between gender mismatch and cloze probability occurred in the P600 time window, with the P600 starting significantly later in the low cloze conditions as compared to the high cloze conditions. These results indicate that semantic expectancy facilitates the repair process, suggesting an interplay between semantics and syntax in later stages of processing and supporting interactive accounts of language comprehension.

This chapter has been submitted for publication: Loerts, H., Stowe, L.A., & Schmid, M.S. (Under Review). Predictability Speeds up the Re-analysis Process: An ERP investigation of gender agreement and cloze probability. Submitted (June 4, 2012) to *Journal of Neurolinguistics.*

5.1. Introduction

Language comprehension relies heavily on phonological, semantic and syntactic levels of processing, and psycholinguistic research has long sought to establish the relationship between these processes. In particular, the specifics of how and when semantic and syntactic information is used are still controversial. The present study compared event-related potentials (ERPs) in response to spoken Dutch sentences that were either correct or violated article-noun or adjective-noun gender agreement. The noun, i.e. the point at which the incorrect gender-marking becomes apparent, was either preceded by a neutral context or by a highly constraining context to investigate whether and how semantic expectancy influences morphological processing of grammatical gender agreement in real-time.

Grammatical gender is a linguistic phenomenon that has recently received considerable attention with regard to the issue of serial or interactive processing. Grammatical or syntactic gender is a classification system dividing nouns into different categories or genders and it is present in many, but not all, languages of the world. The classification of genders, i.e. gender assignment, sometimes corresponds to the biological gender of the underlying concept as in '*der*$_{\text{MASC}}$ *Mann*$_{\text{MASC}}$' ('the$_{\text{MASC}}$ man$_{\text{MASC}}$' in German), but in other nouns, such as those referring to inanimate objects, there is no basis for the assignment of linguistic gender in biological sex: '*la*$_{\text{FEM}}$ *maison*$_{\text{FEM}}$' ('the$_{\text{FEM}}$ house$_{\text{FEM}}$' in French). Moreover, these classifications are not consistent across languages: a native French L2 learner of Dutch cannot simply draw upon the gender characteristics present in the first language (L1), because '*la*$_{\text{FEM}}$ *maison*$_{\text{FEM}}$' is feminine in French, but neuter in Dutch ('*het*$_{\text{NEU}}$ *huis*$_{\text{NEU}}$').

It is therefore not surprising that gender has been shown to be one of the most difficult aspects to master in acquiring language. Although Dutch has a comparatively straightforward two-way gender system (common and neuter, see below), it has to be acquired item by item on the basis of relatively sparse signals in the input (see Table 5-1 for syncretism between forms). Unlike in some Romance languages, such as, e.g. Spanish and Italian, Dutch gender assignment is hardly ever predictable on the basis of morphological or phonological features of the noun (Deutsch & Wijnen, 1985). The system thus largely has to be learned word by word, and its processing is dependent on retrieval of lexical information. Accordingly, Dutch native-speaking children have been reported to overgeneralize the definite, common determiner '*de*' until age 6 (e.g. Van der Velde, 2004; Blom, Polišenskà, & Weerman, 2008). Common gender has evolved from the collapse of the original masculine and feminine genders and consequently comprises around 75% of all Dutch nouns (Van Berkum, 1996). Therefore, common gender is often viewed as the default gender and tends to be heavily overgeneralized during acquisition.

The noun's gender triggers inflectional agreement processes which vary across languages and can affect items within or outside the noun phrase (NP). In Dutch, grammatical gender is reflected on articles, adjectives and pronouns, but not on verbs (as in other languages). Table 5-1 shows the main agreement targets in the Dutch gender system. Gender-marking on the article occurs only in the definite singular; a noun receives the definite article *'het'* when it is neuter, while *'de'* is used in all other cases. In indefinite NPs, all singular nouns receive the indefinite article *'een'* regardless of the gender of the noun they modify. Adjectives are unmarked for gender in definite NPs: all adjectives receive the suffix *–e* (schwa). In indefinite singular NPs, adjectives modifying a common noun receive a schwa-ending, whereas bare adjectives are used to modify indefinite neuter nouns. As can be seen in Table 5-1, the indefinite singular neuter context is the only exceptional form in the Dutch gender system with respect to adjectival marking. Plural nouns are unmarked for gender in Dutch and always receive the definite article *'de'* followed by adjectives containing a schwa-ending.

Table 5-1: Overview of the main agreement targets in Dutch. English equivalents of the phrases are printed in italic below each phrase. Note that the indefinite article is always *'een'* in the singular and no equivalent indefinite article exists for the plural case. Distinctive gender agreement markers are printed in bold. Only some of the pronouns are mentioned in this table. It should be noted that, in their attributive use, demonstrative, possessive, interrogative, and a number of indefinite pronouns all show gender agreement with the noun (Van Berkum, 1996).

	Gender	Definite Article	Adjectives in Definite NPs	Adjectives in Indefinite NPs	Demonstrative pronouns	Relative Pronouns
Single	Common	De …	De mooi***e***...	Een mooi***e***...	Deze/die/onze/elke	De ... die
	Neuter	Het ...	Het mooi***e***...	Een mooi...	Dit/dat/ons/elk ...	Het ... dat
English equivalent		*The ...*	*The beautiful..*	*A beautiful...*	*This/that/our/every*	*The ... that*
Plural	Common	De ...	De mooie ...	x	Deze/die/onze ...	De ... die
	Neuter	De ...	De mooie ...	x	Deze/die/onze …	De ... die
English Equivalent		*The ...*	*The beautiful..*	x	*These/those/our …*	*The …that*

The question addressed in this article is whether contextual constraints affect the processing of grammatical gender in Dutch. One semantic factor that is based on the build-up of contextual information is the probability that a specific word will be produced in a given context. The predictability of words, often operationalized as cloze probability, can facilitate language comprehension in such a way that, during discourse, context is used to disambiguate, predict, and anticipate upcoming words (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). For example, in sentence 5-1a, the word 'horse' is highly predictable on the basis of the preceding context. In sentence 5-1b, however, the word 'horse' has a relatively low cloze probability. Interactive processing models propose that sentence 5-1a is easier to process, because the preceding context can be used to predict and anticipate upcoming words and thus limit the amount of possible lexical candidates activated.

5-1a The cowboy was galloping through the desert on his *horse.*
5-1b The burglar was planning to rob the bank on his *horse.*

Linguistic theory proposes that gender assignment is done at the lexical level: The mental lexicon contains not only information on the meaning of individual words, but also syntactic information about the word such as its gender (e.g. Levelt, 1989). Grammatical gender as marked on the article or adjective can enhance the processing of an upcoming noun by increasing or decreasing the activity of possible lexical candidates (Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Gunter, Friederici, & Schriefers, 2000; also see Chapter 4 of this dissertation). This would thus result in an interaction between grammatical gender and the contextual predictability of a noun, i.e. cloze probability. Although most researchers agree that semantic and syntactic information are integrated during language comprehension, there is no consensus as to when this interaction takes place. Interactive or parallel processing models (e.g. Hagoort, 2003b) claim that all information types interact continuously at each stage of the comprehension process. Contrasting with this view is the assumption of an autonomous role for syntax, in which syntax is processed without influence of and prior to semantics (e.g. Friederici, 2002). Although this syntax-first model agrees that semantic and syntactic information are integrated during language comprehension, it would predict late interaction while parallel models assume interaction to take place during earlier stages.

5.1.1. Gender, cloze, and event-related potentials (ERPs)

Both syntax-first and fully interactive models have been tested extensively with behavioural experiments, but the results cannot provide conclusive answers to questions about the timing and independence of the different processes involved. According to Friederici (2002), behavioural data are problematic because the behavioural paradigms used cannot distinguish between the separable processing aspects, such as automatic and controlled processes of comprehension, nor can they distinguish between the separable time windows of processing. In order to provide such insights, neuroimaging methods such as EEG readings should be used to investigate the time-course of linguistic processing.

Due to their excellent temporal resolution (milliseconds), event-related potentials (ERPs) allow the researcher to characterize the detailed time-course of the different operations involved in language comprehension as the sentence unfolds. Specific and distinctive ERP effects have been consistently associated with different semantic and syntactic processes. The N400 is found to be largest for unexpected items, such as semantic anomalies, and a word's N400 amplitude has a nearly inverse relationship with its cloze probability (Kutas & Hillyard, 1983; Kutas, Lindamood, & Hillyard, 1984). That is, as a word becomes more expected in context, its N400 amplitude is reduced relative to less expected words. These findings suggest that as the prior context builds up meaning, it makes the processing of upcoming words that fit with that context easier, reducing the N400 amplitude and reflecting less processing costs. When elicited by auditory stimuli, the N400 has been found to have an earlier onset at about 200 ms post-stimulus as compared to the N400 elicited by visual stimuli (Van den Brink, Brown, & Hagoort, 2006).

Syntactic processing difficulty has frequently been found to elicit a late positivity that is sometimes preceded by an earlier left anterior negativity (LAN). The LAN occurs around 300 to 500 milliseconds after the stimulus, is most pronounced at anterior sites and mostly left lateralized. The LAN has been found to correlate consistently with syntactic processing and might reflect an automatic detection process of syntactic violations prior to later controlled processes of repair or re-analysis as reflected by the P600 (Friederici, 2002). The LAN has been found to precede the P600 in sentences with morphosyntactic violations such as incorrect subject-verb agreement (Coulson, King & Kutas, 1998; Hagoort & Brown, 2000), but it has also been found for word-category violations in Dutch (Hagoort, Wassenaar, & Brown, 2003). The LAN in response to agreement violations and verb inflection errors has been found for English (e.g. Osterhout & Mobley, 1995), German (e.g. Münte, Matzke & Johannes, 1997), and Dutch (e.g. Gunter, Stowe and Mulder, 1997). It should be noted that the LAN has not been found consistently across languages and across violations and there is no consensus as to what processes the LAN reflects. Friederici's (2002) syntax-first model states that the LAN (as well as the N400 occurring in the same time window) index a processing stage in which morphosyntactic information is used to assign thematic roles. Other researchers prefer an explanation based on working memory, with the LAN indexing a process of looking forward and looking back to integrate displaced constituents (Kutas, Van Petten, and Kluender, 2006).

A more reliable ERP effect that has been found consistently in response to a variety of violations, most of which grammatical violations and complexities, is the late positivity. This positivity occurs from about 500-1000 ms after stimulus onset, usually peaks at 600 ms post-onset and is most pronounced at posterior sites (Osterhout and Holcomb, 1992). This component, commonly referred to as the P600, is elicited by a broad range of syntactic violations and/or syntactic complexities, such as morphosyntactic violations, violations of category expectancy, garden-path sentences, gender agreement violations, and in some cases semantic anomalies (for recent reviews, see, e.g. Kaan, 2007; Steinhauer and Connolly, 2008). Similar to the discussions concerning the underlying processes of the LAN, there is still some controversy on what the P600 actually represents and whether it is purely related to syntax and language. Some claim that the P600 is a general response to stimuli that possess low probability (Coulson et.al., 1998; Gunter et al, 1997), while others claim that the P600 is a linguistic effect related to repair and re-analyses of syntactic integration and parsing.[8] A detailed discussion of whether the P600 is language-specific goes beyond the scope of this paper and even though researchers are still not out on what the LAN, N400 and P600 actually reflect, these components do reveal different stages of processing and are thus suitable to investigate time course and the interactions of different sorts of information during these stages.

Gender is an extremely interesting feature to investigate in this respect, because it comprises lexical, morphological, as well as syntactic features. The first ERP study on gender violations by Osterhout & Mobley (1995) focussed on semantic gender. Osterhout & Mobley investigated reflexive pronouns agreeing or disagreeing with an animate antecedent as in the sentence *'The woman congratulated herself/*himself on the promotion.'* Somewhat counter-intuitively, these sentences elicited a P600, reflecting re-analysis and repair at the syntactic level as opposed to semantic integration difficulties. This result is particularly interesting in light of the fact that this study concerned the English language, in which gender is solely based on semantics. Osterhout and Mobley (1995) propose that gender features, regardless of their semantic basis, may become independent of the gender carrying noun and become a syntactic feature instead. In a follow-up study, Osterhout, Bersick and McLaughlin (1997) tested sentences violating gender-related social stereotypes, such as '*The doctor prepared herself...*'. These sentences again elicited a (somewhat smaller, but reliable) P600, supporting the idea that even stereotypical gender is encoded in grammatical agreement.

Studies looking at other languages are able to distinguish between semantic and grammatical gender, a distinction that is not made in English. Grammatical gender agreement violations between article and noun have, at the sentence level, consistently been found to elicit a P600 across languages, such as Dutch (Hagoort & Brown, 2000; Sabourin & Stowe, 2008), German (Gunter, Friederici, and Schriefers, 2000; Davidson & Indefrey, 2009), French (Foucart & Frenck-Mestre, 2012), Hebrew (Deutsch and Bentin, 2001), Spanish (Barber & Carreiras, 2005) and Italian (Molinaro, Vespignani, & Job, 2008). The few studies that have investigated the processing of gender agreement violations between attributive adjectives and the noun in sentential context have also reported a P600 (Davidson & Indefrey, 2009; Sabourin & Haverkort, 2003), as well as less local grammatical gender violations extending beyond the noun phrase, such as relative clauses in Dutch (Van Berkum, Brown, & Hagoort, 1999; Sabourin, 2003). While the P600 for gender violations has been reported quite consistently across languages and across conditions, the LAN has so far been observed only for Spanish article-noun and article-adjective-noun violations of gender agreement in sentence

[8] The idea that the P600 is a general response related to infrequent stimuli is based on the fact that the P600 has been found for non-linguistic infrequent stimuli which are known to elicit a P3b (Donchin and Coles, 1988). This P3b has a similar posterior scalp distribution as compared to the P600, but peaks earlier at about 300 ms post stimulus. Since the P600 has been shown to be reduced for frequent syntactic violations as compared to infrequent ones (Coulson et.al., 1998; Gunter et al, 1997), some researchers have proposed the P600 is part of the P3 family. Others claim that the P600 is a linguistic effect related to repair and re-analyses of syntactic integration and parsing. This view is supported by studies showing that the P3b and the P600 are independent of one another (Frisch, Kotz, von Cramon, and Friederici, 2003).

processing (Barber & Carreiras, 2005) and violations of gender agreement in German (Gunter et al., 2000).

A small number of studies on gender violations have reported different results. Barber & Carreiras (2005) reported an N400 in response to gender violations in isolated noun-adjective pairs. When the gender violation occurs on the final word of the sentence, it has also been found to elicit an N400 (Hagoort, 2003a; Hagoort & Brown, 2000). In both cases, the authors hypothesize these findings to reflect integration difficulties of lexical representation either because the words occur in isolation (Barber & Carreiras, 2005) or because a violation on the final word of a sentence triggers an overall processing mechanism that checks the well-formedness of the entire sentence (end-of-sentence wrap-up effect: Hagoort, 2003a; Hagoort & Brown, 2000).

In sum, gender violations between article-noun and adjective-noun in sentence medial position have been reported to elicit a robust P600 effect which may be preceded by a LAN (e.g. Gunter et al., 2000). The LAN is generally thought of to reflect detection of the mismatch, which is followed by a repair and/or re-analysis process that is reflected by the P600.

5.1.2. ERPs and the interplay between semantics and syntax

The N400, the LAN and the P600 appear to be influenced by different factors and to be found for different experimental conditions, and might thus be linked to different neuronal substrates of the brain. Friederici (2002) has incorporated these findings into a syntax-first model in which she separates the ERP effects into a three-phase process of comprehension. The first phase occurs 100-300 ms after the critical words are presented. The model states that this phase uses category information to build an initial syntactic structure and assumes this initial phase to reflect automatic processes that are rather impervious to the state of alertness of the subject (Kutas et al., 2006). Phase two of the model (300-500 ms, the time window in which the N400 and the LAN occur) consists of the processing of morphosyntactic and lexical-semantic information to assign thematic roles and is thought of to reflect partially automatic processes. Finally, in the third phase (500-1000, where the P600 can be found) all the preceding different types of information are integrated. According to the model, this phase reflects more controlled processes, since the effects found in this time window (e.g. the P600) are relatively (though not completely, see, e.g. Allen, Badecker, & Osterhout, 2003) insensitive to physical aspects of the stimuli, but generally sensitive to aspects from within the subject. According to Friederici's model, syntax is initially autonomous (phase 1) after which semantics and syntax work in parallel (phase 2) and only interact in late stages of processing (phase 3).

An opposing, fully interactive model of syntactic processing has been proposed by Hagoort (2003b). His Unification Model is based on a computational model of parsing developed by Vosse and Kempen (2000), which postulates that every word in the lexicon is associated with a structural frame. During the unification process, structural frames are linked to one another and agreement features of the items are checked. According to Hagoort's model, (E)LANs might reflect the failure to bind two syntactic frames as in the case of agreement violations. The subsequent P600 can be related to the time the unification process takes, which is affected by competition of other possible unification links and also by semantic influences. The amount of competition is assumed to modulate the amplitude of the P600 with larger amplitudes corresponding to greater competition.

Hagoort (2003a) directly addressed the issue of serial vs. interactive processing of gender agreement and semantic factors in Dutch, using violations between the determiner and the noun (with an intervening adjective). To investigate the influence of semantics on syntax, he compared ERPs elicited by semantic violations, gender agreement violations, and a combination of the two. Gender mismatches elicited a classic P600 effect and semantic violations an N400 effect. An increased N400 amplitude preceding the P600 was observed only for sentence-final agreement violations. Sentence-internal combined violations, such as *'Het*$_{\text{NEU}}$ *eerlijke paraplu*$_{\text{COM}}$ *staat in de garage'* ('The$_{\text{NEU}}$ honest umbrella$_{\text{COM}}$ is in the garage'), showed an interaction of semantics and syntax at the level of the N400 amplitude: the N400 was larger for combined violations as compared to semantic violations without an additional gender mismatch. Since the P600 seemed not to be influenced by semantic factors, Hagoort (2003a) concluded that syntactic information influences semantic processing but not vice versa.

A different result was obtained by Gunter, Friederici, and Schriefers (2000) in an investigation of possible interactions between syntax and semantics during the processing of agreement violations in German. Similar to Hagoort (2003a), they compared native speakers' ERP readings in response to correct sentences and sentences containing a gender mismatch between the article and the noun. In contrast to Hagoort (2003a), who used semantic anomalies, Gunter et al. (2000) manipulated cloze probability as a semantic factor: half of the experimental sentences contained critical nouns that were highly constrained by the preceding context (high cloze) while the other half contained nouns that could not be predicted on the basis of the preceding sentence information (low cloze). Gunter et al. (2000) reported a main effect of cloze probability, in which low cloze conditions elicited a larger N400 reflecting higher processing costs for nouns that were unexpected given the context. Gender mismatches elicited a LAN followed by a P600 reflecting detection and re-analysis of a syntactic violation. Interestingly, there was an interaction of cloze probability and gender agreement: violations in high cloze conditions elicited a significantly larger P600 as compared to violations in low cloze conditions. Because the N400 was independent of gender mismatch and the LAN independent of cloze probability, the authors concluded that semantics and syntax work in parallel during initial phases of processing, with a LAN and an N400 in the same time window, and that interaction takes place only in a later time range, i.e. the P600 time range. The shorter and smaller P600 for sentences such as *'*Sie befährt* ***den***$_{\textbf{MASC}}$ ***Land***$_{\textbf{NEU}}$ *…'* ('She drives across **the**$_{\textbf{MASC}}$ **land**$_{\textbf{NEU}}$ …'), points towards a situation of less repair and re-analysis compared to the high cloze condition. Gunter et al. (2000) suggest that when a gender-incongruent noun occurs, it does not need deep semantic processing in the high cloze condition (because it is a highly expected word) and only requires gender repair. In the low cloze condition, however, more thorough semantic analyses are needed, causing the P600 to be reduced in amplitude. This explanation is in line with the notion that the P600 is under strategic control and, in this particular case, the strategy is to focus on meaning and syntactic re-analysis or repair consequently become less important.

Wicha, Moreno, and Kutas (2004) performed a study similar to Hagoort (2003a) and compared gender violations, semantic violations, and a combination of the two in Spanish. When subjects were confronted with double violations, such as in the sentence *'…podrıá al fin ponerse* ***el$_{MASC}$ maleta$_{FEM}$*** *por el resto de su vida' (*'... he would finally be able to wear the$_{\text{MASC}}$ suitcase$_{\text{FEM}}$ for the rest of his life'), their ERP waveforms showed both N400 enhancement as well as a larger P600 for the combined violations. This result suggests that syntax and semantics interact during the entire process of language comprehension.

As far as the question of a possible interplay between semantics and syntax in sentence processing is concerned, the findings from ERP experiments are thus conflicting. The influence of semantics on the P600 observed by Gunter et al. (2000), in particular, is in clear contrast to the Hagoort (2003a) study reporting N400-enhancement when semantic anomalies are combined with an additional gender violation. The relative non-transparency of the Dutch gender system might contribute to these different results. In Dutch, gender is primarily lexically determined, while gender in German and Spanish is more prominent due to the involvement of case in German and the presence of phonological cues in Spanish. In addition to the fact that the experiments are conducted in different languages, Gunter et al. (2000) note that the different findings between these studies may be related to the fact the two studies used different tasks: Gunter et al. (2000) used a probe detection task, Hagoort (2003a) used an acceptability judgement paradigm, and Wicha et al.'s (2004) participants read for comprehension. Yet another difference between the studies concerns the manipulation of the N400: Gunter et al. (2000) manipulated cloze probability, while Hagoort (2003a) and Wicha et al. (2004) used semantic violations per se.

Further differences between the studies include the fact that both Gunter et al. (2000) and Wicha et al. (2004) studied a more local violation, whereas Hagoort (2003a) used gender mismatches with an intervening adjective. A mismatch between the determiner and the noun, with the determiner being followed by a correctly inflected adjective, as used by Hagoort, might constitute a weaker and/or less reliable cue which, according to the Competition Model, can be expected to evoke slower reaction times as compared to stronger cues (Bates & MacWhinney, 1987). In addition, it should also be noted that Hagoort compared the P600 amplitudes of the different conditions relative to a baseline at 300 to 500 ms after noun onset instead of the more usually applied pre-stimulus baseline. Wicha et al. (2004) report that, when comparing P600 amplitudes relative to the N400 time window, as Hagoort did, the P600 is more similar for both semantically congruous and incongruous nouns (and perhaps slightly larger in the incongruous condition).

In order to shed more light on the proposed interplay between a semantic factor, cloze probability, and gender agreement in Dutch, the present study therefore attempted to eliminate the confounding factors that might have resulted in the contradicting findings by Hagoort (2003) on the one hand and Gunter et al. (2000), and Wicha et al. (2004) on the other.

5.1.3. Grammatical gender and cloze probability

The present study examined the extent to which contextually constrained semantic aspects of a noun influence the processing of grammatical gender mismatches, as reflected in the P600, by manipulating cloze probability and gender concord in spoken Dutch sentences. Local violations on determiners were embedded in a sentence context (e.g. ‘*de_{COM} $boek_{NEU}$’, *the_{COM} book_{NEU}), but sentences containing less local violations were also created with an intervening adjective (e.g. ‘*de_{COM} $mooie_{COM}$ $boek_{NEU}$’, the_{COM} $\text{beautiful}_{\text{NEU}}$ book_{NEU}). Since the saliency of the gender violations might play a role as well, violations occurred on the determiner (e.g. ‘*de_{COM} $boek_{NEU}$’) as well as on the adjective (e.g. ‘**een* $mooie_{COM}$ $boek_{NEU}$’, a $\text{beautiful}_{\text{COM}}$ book_{NEU}). Materials were designed with half of the target nouns being common and half of them being neuter. In addition, the present study only used sentence-medial gender violations to avoid any potentially confounding sentence wrap-up effects (Hagoort, 2003a).

All conditions were divided equally into sentences where the target noun was highly predictable from the preceding sentence information (high cloze) and conditions where the target could not be predicted on the basis of context (low cloze) (see below, section ‘Materials’, on how cloze probability was established). Low cloze items were expected to elicit an N400, reflecting semantic integration difficulties in response to targets with low semantic expectancy, while gender mismatches were expected to elicit a P600 response reflecting processes of repair and/or re-analysis. Assuming that both gender and context are used to anticipate upcoming words, an interaction between these two variables was expected. The contradicting findings of previous studies, i.e. the syntactic boost reported by Hagoort (2003a) as opposed to the reduced P600 reported by Gunter et al. (2000), might reflect a difference in processing gender violations in Dutch and German respectively. The opposing outcomes, however, might also be related to the different tasks they used or to the difference between processing semantic violations as compared to processing items that are unpredictable on the basis of the preceding context. The present experiment used Hagoort’s (2003a) acceptability judgement task and Gunter et al.’s (2000) manipulation of cloze probability. If the syntactic boost on the N400 for ungrammatical items results from participants judging the grammaticality of sentences, the present study should obtain the same N400-enhancement. Should the difference between the previous results be attributable to processing a semantic violation as opposed to low cloze probability nouns, an interaction of cloze probability would only be expected to occur in the P600 time window.

5.2. Methods

5.2.1. Participants

The present experiment collected data from 28 native Dutch adults (16 female, 12 male) aged 34-60 (mean age=48) who were recruited from a non-university population through advertisements in a local newspaper. All were healthy, right-handed participants (assessed according to Oldfield, 1971) with normal hearing and without history of neurological problems and/or language disorders. They were paid for their participation. ERP data from two of the participants were not analysed because of too much contamination by artefacts (blinks and/or heavy alpha activity), so that 26 participants were retained for ERP analysis.

5.2.2. Stimuli

In total, 360 experimental sentences were created using a set of 90 high frequency nouns (45 common and 45 neuter) selected from the 5000 most frequently occurring words in the Spoken Dutch Corpus (*Corpus Gesproken Nederlands*; CGN) compiled by The Centre for Language and Speech Technology (Dutch HLT Agency or TST-centrale). Since it was assumed that complexity and locality of gender violations might influence the online processing of gender mismatches, nouns were embedded in sentences using three different NP constructions: Definite determiner + Noun (DN), Definite determiner + Adjective + Noun (DaN), and Indefinite determiner + Adjective + Noun (IaN). For example sentences, see Table 5-2 below.

For each selected noun, two sentences were created: one with a constraining sentence context that was highly predictive with respect to the target noun (high cloze) and one in which the target noun could not have been easily predicted on the basis of the preceding sentence context (low cloze). In a sentence completion task administered among 24 native Dutch volunteers (mean age 26.6, range 22-50; 17 female, 7 male), subjects were presented with the sentences up to the target noun (thus including the determiner and, in DaN and IaN conditions, the adjective) and asked to fill in the first word that came to mind. Target items for which above 70% of the responses agreed were used as high cloze items in the experiment (mean average cloze probability 88.6%). Low cloze targets scored an average of 10.9% and only targets that scored below 30% were used in the experiment.

This resulted in a set of 90 high cloze sentences and 90 low cloze sentences. For each of these resulting 180 grammatical sentences, ungrammatical variants were created by violating gender-marking on the determiner or the adjective.

Table 5-2: Example sentences of the different conditions. Examples are divided into NP-construction, cloze-probability, and gender congruency. The critical NPs are printed in bold and English equivalents of the sentences are presented in italic below every sentence. Note that each condition equally contained neuter and common nouns as target words and only one example is given in the present table.

NP	Cloze	Gender	Example Sentence
DN	High	Congruent	Vera plant rode rozen in **de tuin** van haar ouders. *(Vera is planting red roses in the$_{COM}$ garden$_{COM}$ of her parents.)*
		Incongruent	Vera plant rode rozen in ***het tuin** van haar ouders. *(Vera is planting red roses in *the$_{NEU}$ garden$_{COM}$ of her parents.)*
	Low	Congruent	Vera breit een sjaal in **de tuin** van haar ouders. (*Vera is knitting a scarf in the$_{COM}$ garden$_{COM}$ of her parents.)*
		Incongruent	Vera breit een sjaal in ***het tuin** van haar ouders. (*Vera is knitting a scarf in *the$_{NEU}$ garden$_{COM}$ of her parents.)*
DaN	High	Congruent	Bianca leest (…) **het nieuwe boek** van haar favoriete schrijfster. *(Bianca is reading the$_{NEU}$ new book$_{NEU}$ of her favourite writer.)*
		Incongruent	Bianca leest (…) ***de nieuwe boek** van haar favoriete schrijfster. *(Bianca is reading the$_{COM}$ new book$_{NEU}$ of her favourite writer.)*
	Low	Congruent	Bianca zoekt (…) naar **het nieuwe boek** van haar favoriete schrijver. *(Bianca is searching for the$_{NEU}$ new book$_{NEU}$ of her favourite writer.)*
		Incongruent	Bianca zoekt (…) naar ***de nieuwe boek** van haar favoriete schrijver. *(Bianca is searching for the$_{COM}$ new book$_{NEU}$ of her favourite writer.)*
IaN	High	Congruent	Sjoerd schrijft **een lange brief** aan zijn verloofde. *(Sjoerd writes a long$_{COM}$ letter$_{COM}$ to his fiancée.)*
		Incongruent	Sjoerd schrijft **een lang brief** aan zijn verloofde. *(Sjoerd writes a long$_{NEU}$ letter$_{COM}$ to his fiancée.)*
	Low	Congruent	Sjoerd krijgt **een lange brief** van zijn verloofde. *(Sjoerd receives a long$_{COM}$ letter$_{COM}$ from his fiancée.)*
		Incongruent	Sjoerd krijgt **een lang brief** van zijn verloofde. *(Sjoerd receives a long$_{NEU}$ letter$_{COM}$ from his fiancée.)*

In Dutch, gender is marked either on the definite article (in definite NPs) or on the adjective (in indefinite NPs) (see above). In the conditions containing a definite article (the DN and DaN condition), the ungrammatical variants of the sentences were constructed by using an incorrect gender-marked determiner (i.e. replacing the common determiner *'de'* with *'het'* and vice versa). In the indefinite condition (IaN), the ungrammatical variants were created by using the incorrect form of the modifying adjective, thus either adding the common suffix *–e* in neuter conditions or omitting the schwa-ending in the common conditions.[9] By using this 2 (correct - gender mismatch) x 2 (high cloze – low cloze) design, each target noun appeared in four sentence versions varying with regard to cloze probability and grammaticality and resulting in 90 sentences per condition. Since each noun occurred once in each of the four conditions, the total set of 360 sentences was divided across 4 lists in such a way that each subject encountered a target word only once.

[9] Note that this gender feature on the adjective is only present in the indefinite condition and not in the definite condition in which attributive adjectives receive the suffix '*–e*' regardless of the gender of the noun they modify.

To make sure that the sentences were equally plausible across conditions, 68 additional native Dutch volunteers (mean age: 21.4, range 18-33; 49 female) were asked to rate the sentences on a Likert-scale from 1 (very implausible) to 5 (completely normal). Items with an average plausibility score lower than 3.5 were excluded from the material and, for the final set of experimental sentences, low cloze and high cloze items were rated as being plausible with a mean score of 4.4 and 4.6 respectively.

For the present experiment, we used auditory instead of written/visual stimuli. Sentences were digitally recorded in a soundproof booth with a sampling rate of 44100 Hz by a female native speaker of Dutch with normal intonation. All sentences were normalized to a mean intensity level of 70 dB SPL.

The order of the experimental sentences (45 grammatical gender sentences per list) was randomized with 125 plausible filler sentences. Each list was divided into a practice session followed by 4 runs each of which could be followed by a short break. The experimental sentences were randomized in such a way that the same conditions never occurred in succession of each other. Each stimulus was to be judged on its acceptability after the end of the presentation.

Markers were placed at the onset of the critical noun for later ERP analysis, since this is the point at which both the predictability of the noun as well as the gender match or mismatch become apparent.

5.2.3. Procedure

During the experiment, participants listened to correct sentences and sentences containing gender violations and finiteness violations while their continuous EEG signal was being recorded. Before each session, the sound of the speakers was adjusted to a volume of between 73dB and 77dB SPL. Participants were seated in a comfortable chair in an isolated soundproof room which was dimly illuminated. Subjects were seated in front of a screen at a distance of 60-70 cm and were instructed to minimize eye blinks and body movements when listening to the sentences. They were asked to look at a central fixation cross to fixate their eyes and instructed to blink only when a string of asterisks appeared on the screen (****). After each sentence, subjects were asked to judge the acceptability of the sentence they had heard by pressing one of two buttons using their right hand.

5.2.4. Recording and analysis

During the task, electrical brain activity was recorded with 64-channel EEG integrated in a plastic cap. Electrode positions, based on the extended 10-20 system, are shown in Figure 5-1 (nomenclature based on Sharbrough et al., 1991). Blinks and eye movements were monitored via 4 electrodes that were placed on the temples and above and below the left eye. Additionally, two electrodes were placed for reference on the mastoids and the ground electrode for common reference was placed on the sternum. To optimize the conduction between electrodes and skin, conduction gel was applied through a dull needle. Impedances of the skin were reduced to below 10 KΩ.[10] The EEG and EOG signals were recorded and digitized with a sampling frequency of 250 Hz using Brain Vision Recorder software (version 1.05).

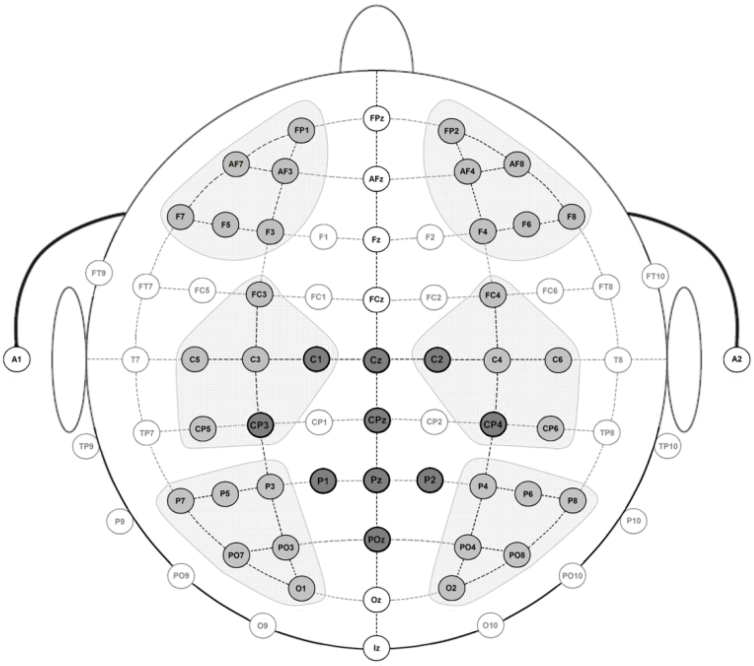


Figure 5-1: Approximate location of electrodes and the layout of the 7 regions of interest used for analyses. EEG was recorded with 64 tin electrodes integrated in a plastic cap. Electrode positions are based on the extended 10-20 system. Light grey areas represent the 6 Regions of Interest chosen to examine the distribution of the effects by dividing electrodes on the basis of the factors anterior-posterior sites and left and right hemisphere. Dark grey coloured electrodes are included in the central ROI, which was mainly used to examine onset latencies of the effects.

The EEG data time-locked to the target words were analysed with Brain Vision Analyzer 1.05. Before statistical analysis, data was re-referenced to the average of the two mastoid electrodes and filtered off-line with a 50 Hz low-pass filter and a 0.25 Hz high-pass filter. Before calculating the grand averages, the data were divided into segments from 500 ms before to 1500 ms after the critical words. All segments were screened for blinks, eye movements, and artifacts. The method of Gratton and Coles (1989) was used to correct the data for blinks and trials containing artifacts were rejected. Segments were then normalized using a baseline correction with a baseline period starting 200 ms before onset of the critical word. The averaged ERPs were calculated per subject per condition.

[10] In some instances, the temporal and frontal electrodes (Fp1 and Fp2, see figure 2 below) could only be reduced to below 20 KΩ.

ERP epochs in response to correct sentences and their equivalents containing a gender mismatch were analysed in order to determine how participants process and repair the gender mismatch as well as to explore the possible interaction between cloze probability and grammatical gender.

To examine the onsets and durations of the ERP effects, statistical analyses were carried out on 50 ms latency windows from stimulus onset until 1500 ms post stimulus. For analyses of these different time windows, the averaged data were exported from Brain Vision Analyser per subject, per time window, and per condition. In total, 6 regions of interest (ROIs) containing 6 electrodes each were selected based on anterior/posterior and hemisphere. The ROIs for statistical analysis contained left-anterior (FP1, AF3, AF7, F3, F5, F7), right-anterior (FP2, AF4, AF8, F4, F6, F8), left-central (FC3, C1, C3, C5, CP3, CP5), right-central (FC4, C2, C4, C6, CP4, CP6), left-posterior (P3, P5, P7, PO3, PO7, O1), and right-posterior (P4, P6, P8, PO4, PO8, O2). In addition, a central ROI containing 10 electrode sites (C1, Cz, C2, CP3, CPz, CP4, P1, Pz, P2, and POz) was selected to investigate onset differences and leaving out effects of distribution. These ROIs are depicted in Figure 5-1.

Reported p-values reflect the application of the Greenhouse-Geisser correction for repeated measures with more than one degree of freedom (Greenhouse & Geisser, 1959).

5.3. Results

5.3.1. Behavioural data

Overall, subjects were quite accurate with a mean score of 94% (95% confidence intervals = [92,95]) on the correct sentences and a mean score of 88% (95% CI = [85,92]) when judging incorrect sentences, as is also visible in Figure 5-2. An overall analysis of variance on accuracy scores from all conditions showed that participants were significantly more accurate in judging grammatical sentences as opposed to ungrammatical sentences, $F(1, 27) = 13.070$, $p < 0.01$. The three types of gender violations differed significantly, $F(2, 27) = 5.433$, $p < 0.05$, and showed a main effect of grammaticality, $F(1, 27) = 16.715$, $p < 0.001$, as well as an interaction of phrase type by grammaticality, $F(2, 27) = 7.532$, $p < 0.01$. A separate analysis of the three phrase types revealed that in both conditions containing an adjective (the DaN and IaN conditions), participants were significantly better in judging the grammatical sentences (DaN: $F(1, 27) = 11.011$, $p < 0.01$; IaN: $F(1, 27) = 13.034$, $p < 0.01$). Such an effect of grammaticality was absent in the DN condition (F<1).

With regard to reaction times, subjects were significantly faster when judging ungrammatical sentences as compared to grammatical ones, $F(2, 27) = 4.603$, $p < 0.05$. In accordance with the accuracy data, the DN condition caused the least difficulty with subjects being faster in this condition to rate the ungrammatical counterparts as compared to the grammatical ones, $F(1, 27) = 9.557$, $p < 0.01$. Only a marginal grammaticality effect was found in the DaN condition, $F(1, 27) = 3.996$, $p = 0.056$, and it was completely absent in the IaN condition ($F<1$). This result is in agreement with predictions made by the Competition Model (Bates & MacWhinney, 1987) which states that stronger and more reliable cues lead to faster reaction times.

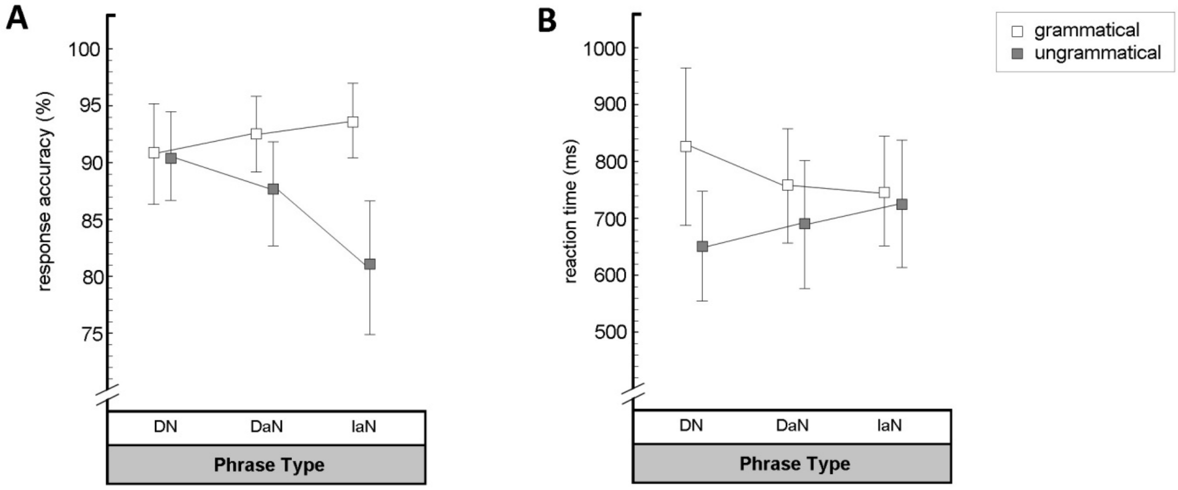


Figure 5-2: Participant's mean A) accuracy rates and B) reaction times to the correct and incorrect versions of the gender conditions. Error bars indicate 95% Confidence Intervals. The accuracy scores in the left panel reveal that participants made significantly more errors when judging ungrammatical strings that contained an adjective in between the determiner and noun (DaN and IaN). The right panel shows that participants are significantly faster in detecting an ungrammatical as opposed to a grammatical determiner-noun combination. Only a trend of this effect was found in the DaN condition and it was completely absent in the IaN condition.

Cloze probability did not affect the accuracy rates ($F<1$), nor did it impact the reaction times of the participants ($F(1,27) = 1.034$, $p = 0.318$). The effects of grammaticality on both the accuracy and reaction time data show that when participants are focussing on acceptability judgement, the article-noun mismatch thus appears to be a stronger and more reliable cue as compared to an adjective-noun mismatch. The fact that the DaN condition is in the middle confirms this: it is a more reliable cue than an adjective, but somewhat more difficult to resolve because there is a correctly inflected adjective intervening, and competing cues tend to inhibit sentence interpretation and thus lead to slower response times than cues that occur alone or converge with one another.

5.3.2. Electrophysiological Data

Grand average ERPs elicited by the critical nouns in the gender agreement conditions are presented in Figure 5-3.

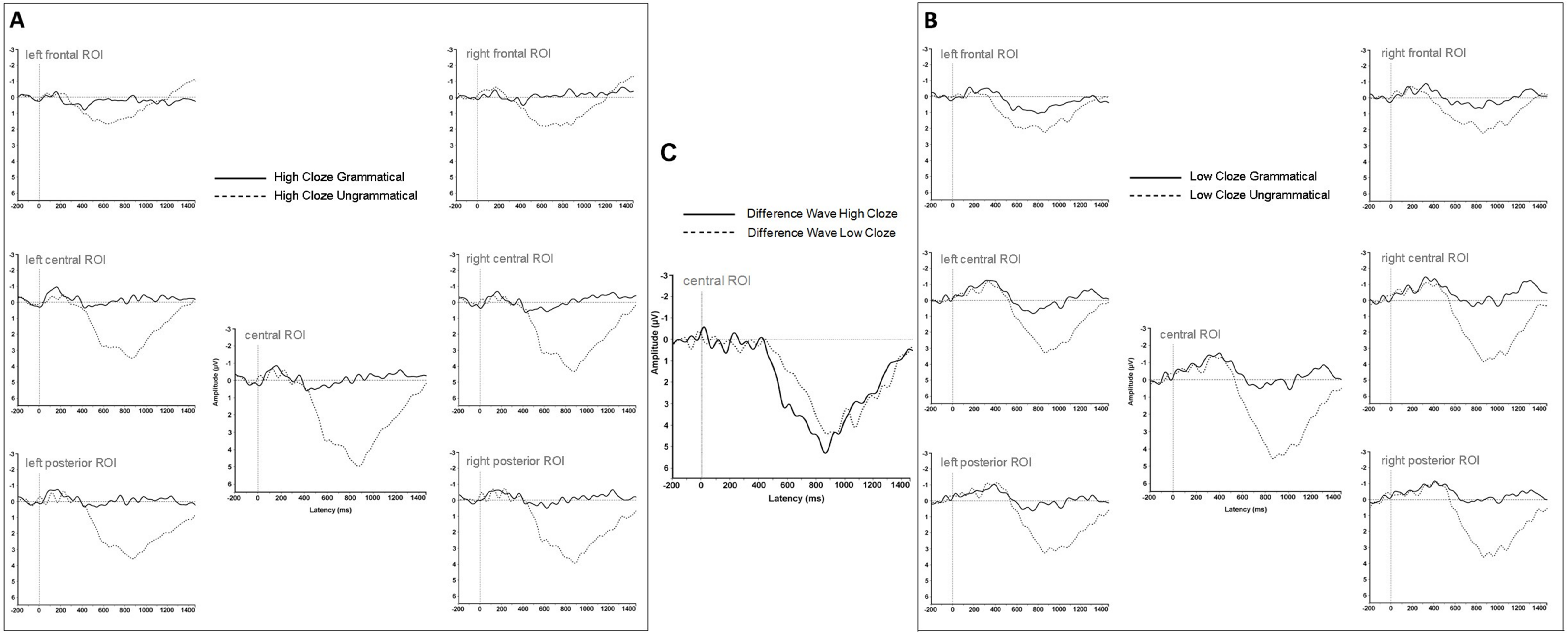


Figure 5-3: Grand average ERPs (n=26) in response to the correct gender agreement sentences (solid line) and the gender mismatch condition (dotted line) in A) the high cloze conditions and B) the low cloze conditions. ERP epochs run from 200 ms before up to 1500 ms after the critical noun. Negative voltage is plotted upwards. Figure 3C overlays the P600 for high and low cloze items and reveals the difference in onset latency between the two conditions.

The average waveforms in the low cloze condition (Figure 5-3B) revealed a clear negative-positive pattern, with an N400-like effect followed by a P600-like effect for the items containing a gender mismatch. The high cloze items (Figure 5-3A) did not show this biphasic pattern and only revealed a clear positive deflection starting around 450 ms post-onset.

ERPs elicited by the gender mismatch did not seem to contain any early anterior negativities (LAN) such as had previously been reported in response to morphological violations. This was confirmed by statistical analyses revealing no reliable differences between the conditions in the first 300 ms after word onset.

ERP amplitudes from all relevant time windows were first compared using a 2 (hemisphere) x 3 (anterior-posterior) x 2 (cloze probability) x 2 (grammaticality) repeated measures ANOVA. In subsequent analyses, focussing primarily on onset latencies of the observed effects, the central region of interest was analysed excluding the factors hemisphere and anterior-posterior. Since phrase (determiner-noun (DN), determiner-adjective-noun (DaN), and indefinite determiner-adjective-noun (IaN)) significantly affected the accuracy and reaction time data, additional analyses were performed including this factor and collapsing over cloze probability or grammaticality.

ANOVAs on the averaged ERP amplitudes with the 6 ROIs revealed a significant main effect of cloze probability from 300 until around 600 ms post stimulus, all $Fs(1,25) > 7.454$, $ps < 0.01$, reflecting larger negativities for the low cloze items as compared to the high cloze items. This result confirmed the presence of a classic N400 effect, which also interacted with the factor anterior-posterior, $Fs(2, 50) > 4.917$, $ps < 0.05$. Post hoc ANOVAs for the factor region revealed that the N400 was significantly present and bilaterally distributed in central and posterior sites (Central ROIs: $Fs(1, 25) > 16.561$, $ps < 0.001$); posterior ROIs: $F(1, 25) > 19.189$, $ps < 0.001$).

Some previous studies have reported anterior negativities (LAN) to precede a P600 in response to a gender mismatch. A hypothesis based ANOVA applied on frontal regions only, however, showed no main effects of grammaticality in the time windows from 0 to 450 ms post stimulus, all $Fs < 2.190$, all $ps > 0.151$, and a closer look at the mean amplitudes revealed that the ungrammatical conditions actually evoke a slightly more positive signal, eliminating any possibility of the presence of a LAN effect.

The more positive signal for ungrammatical items reached significance between 450 and 500 ms after noun onset and started in frontal regions only, $F(1, 25) = 4.598$, $p < 0.05$. From 500 ms onwards, this P600 became significant across the entire scalp and lasted until the end of the ERP epoch at approximately 1500 ms (Frontal ROIs: $Fs(1, 25) > 6.280$, $ps < 0.05$; Central ROIs: $Fs(1, 25) > 9.738$, $ps < 0.01$; Posterior ROIs: $Fs(1, 25) > 6.886$, $ps < 0.05$) as can also be seen in Figure 5-3. The grand-averages depicted in Figure 5-3 also suggested an earlier onset of the P600 for high cloze items and this was confirmed by the presence of an interaction between cloze probability and grammaticality starting in posterior regions around 500 ms and lasting up to approximately 800 ms after noun onset.

To investigate whether this interaction is indeed the result of an earlier onset of the P600 in response to critical nouns with high cloze probability, a repeated measures ANOVA was administered on the central ROI containing 10 electrodes (see also Figure 5-1). Post hoc ANOVAs for the factor cloze probability revealed a main effect of grammaticality for the high cloze items already starting in the 500-550 ms time window, $F(1, 25) = 6.379$, $p < 0.05$, reflecting the fact that high cloze items containing a gender mismatch are already being repaired or re-analysed in this time window. When looking at low cloze items, however, no effect of grammaticality was found in the time window from 500 to 550 ms, $F < 1$. The difference waves of the high cloze and low cloze conditions are presented in Figure 5-3C where this onset difference can clearly be seen. The interaction between cloze probability and grammaticality continues from 550 to 800 ms, Fs(1, 25) > 5.877, ps < 0.05, and is present bilaterally in central and posterior areas (Central ROIs: Fs > 4.479, ps < 0.05; Posterior ROIs: Fs > 7.554, ps < 0.05.). From 550 ms onwards, however, grammaticality significantly affects both high cloze and low cloze items (High cloze: Fs > 29.851. ps < 0.001; Low cloze: Fs > 7.965, ps < 0.01), revealing that the P600 is present in both high and low cloze conditions, but significantly smaller in the low cloze as compared to the high cloze conditions. Table 5-3 provides an overview of the significant effects and interactions of cloze probability and gender grammaticality within the Central region of interest.

Table 5-3: Overview of the significant effects found. This Table shows the effect of cloze probability (N400), the effect of grammaticality (P600) and the interaction between these two variables in the middle row. Light, medium, and dark grey colours correspond to *p*-values below 0.05, 0.01, and 0.001 respectively.

	250 - 300	300 - 350	350 - 400	400 – 450	450 - 500	500 - 550	550 - 600	600 - 650	650 - 700	700 - 750	750 - 800	800 - 850	850 - 900
Cloze													
Cloze*Grammaticality													
Grammaticality													

Although the behavioural results revealed differences between the different phrase types, no effects of or interactions with phrase were found in the electrophysiological data, suggesting a similar N400 and P600 effect regardless of the noun phrase used.

5.4. Discussion

The present study investigated the influence of a semantic factor, cloze probability, on the online processing of grammatical gender by adult Dutch natives from a non-university population. Subjects listened to correct sentences and sentences containing gender agreement violations that differed with respect to the type of NP construction. In addition, target nouns were either highly predictable from the preceding context (high cloze) or could not have been predicted from the preceding context (low cloze).

Based on previous research, we hypothesized low cloze items to elicit an enhanced N400 and gender mismatches to elicit a P600 possibly preceded by a LAN. According to interactive or parallel processing models, grammatical gender as marked on the article or adjective can enhance the processing of an upcoming noun by increasing or decreasing the activity of possible lexical candidates (Bates et al., 1996; Gunter et al., 2000). This would thus predict an interaction between the cue to grammatical gender inherent in the inflection on preceding elements and the predictability of a noun (i.e. cloze probability). We hypothesized this interaction between cloze probability and gender agreement to occur, resulting either in a 'syntactic boost' on the N400 amplitude (Hagoort, 2003a), as a delayed and/or reduced amplitude of the P600 (Gunter et al., 2000), or a combination of the two (Wicha et al., 2004). By combining Hagoort's acceptability judgement task with Gunter et al.'s semantic manipulation based on cloze probability, the present study aimed at distinguishing between the contradictory results. In other words, if task demands can explain the different findings, the present study would expect an interaction to occur on the N400 (Hagoort, 2003a). If, however, cloze probability as compared to semantic violations can explain the differences, an interaction on the P600 would be expected (Gunter et al., 2000).

From 300 to 600 ms post stimulus, a reliable N400 was found for low cloze items: the difference in cloze probability of the target nouns between conditions caused a significant increase in negativity for the low cloze items. This effect replicates previous findings (e.g. Gunter et al., 2000) and suggests that predictability facilitates processes of identification and access, with less activation and difficulties for highly predictable items (Kutas et al., 1984).

As hypothesized, subjects needed to repair and/or re-analyse when they were confronted with the ungrammatical conditions as suggested by a reliable P600 for ungrammatical items that was visible across the entire scalp. This late positivity began around 500 ms and continued for a long period up to 1500 ms. The most interesting finding concerning this P600 is the onset difference between high cloze and low cloze probability words. Analysis revealed that the repair/re-analysis process had a significantly earlier onset of about 500 ms post-stimulus in the high cloze conditions as compared to the low cloze conditions where the re-analysis process became significant around 550 ms post-stimulus.

This result, that the onset latency of the P600 is affected by semantic factors such as cloze probability, is in clear contrast with Hagoort (2003a) and Wicha et al. (2004) who report N400-enhancement when semantic anomalies are combined with an additional gender violation. Hagoort (2003a) claimed that the N400 is open to syntax, but not the other way round. The contradicting view is that the P600 is sensitive to both syntax and semantics and interaction only takes place in this later time window (Gunter et al., 2000).

Gunter et al. (2000) hypothesized that the N400-enhancement reported by Hagoort (2003a) might be affected by the task used in that study, i.e. acceptability judgement. The present study, however, also used an acceptability judgement task and reports only an N400 in response to cloze probability and no syntactic boost in response to ungrammatical sentences. Gunter et al. (2000) also suggest that the N400 found by Hagoort might be an overlapping LAN. Although we cannot determine whether or not this is the case with respect to Hagoort's findings, the present study also did not detect a LAN in response to gender mismatches and, as Hagoort (2003a) explained, such an early detection is not always expected in Dutch and German because gender is not marked as a morpheme on the noun, nor does a gender mismatch violate phrase structure.

Another difference between Gunter et al. (2000) and Hagoort (2003a) was that Gunter et al. used a more local gender violation while Hagoort used violations containing an intervening adjective. The present study shows that this difference cannot account for their conflicting results: gender mismatches between article and noun elicited a similar pattern with an N400 for low cloze and a robust P600 in response to ungrammatical variants with or without an adjective in between. The only difference found with respect to the type of NP construction was that gender mismatches between adjective and noun were more difficult to detect, as demonstrated by slower response times. This result is in line with predictions from the Competition Model, stating that less salient and/or reliable cues lead to slower response times (Bates & MacWhinney, 1987). The P600, however, was equally robust for violations between article and noun with or without an intervening adjective.

Table 5-4: Overview of the results found in different studies concerning the interplay between semantic factors and gender agreement. All studies outlined below have examined the interplay between syntax and semantics by looking at the processing of gender agreement violations. All of these studies reported to have found an N400 followed by a P600 (which was preceded by a LAN in the Gunter et al. study (2000)), but they differ with respect to the timing of the interaction between the semantic factor and the syntactic factor.

Authors	**Gender System**	**Semantic Manipulation**	**Task**	**N400 Interaction**	**P600 Interaction**
Hagoort (2003)	Dutch	Semantic violation	Acceptability Judgement	+	-
Wicha et al. (2004)	Spanish	Semantic violation	Read for Comprehension	+	+
Gunter et al. (2000)	German	Cloze probability	Probe detection	-	+
Loerts et al. (Under Review)	Dutch	Cloze probability	Acceptability Judgement	-	+

As can be seen in Table 5-4, a more plausible explanation for the conflicting results might be that this concerns a comparison of responses to anomalies and difficulties which might not undergo exactly the same processes. Hagoort (2003a) found N400-enhancement in combined violations such as *'…****het***$_{NEU}$ ***verkouden*** *droom*$_{COM}$*'* ('… **the**$_{NEU}$ **sniffling** dream$_{COM}$'). Gunter et al. (2000) report a decreased P600 for gender mismatches on a low cloze item, such as *'*Sie befährt* ***den***$_{MASC}$ ***Land***$_{NEU}$ *...'* ('She drives across **the**$_{MASC}$ **land**$_{NEU}$ …') in which no semantic violation is present. Wicha et al. (2004) reported an interaction on both the N400 and the P600 when Spanish natives were confronted with a sentence like, *'…podriá al fin ponerse* ***el***$_{MASC}$ ***maleta***$_{FEM}$ *por el resto de su vida' (*'... he would finally be able to wear **the**$_{MASC}$ **suitcase**$_{FEM}$ for the rest of his life'). The P600 was smaller for these combined violations as compared to purely syntactic violations when analysed relative to a pre-stimulus baseline. The semantic manipulation in the sentences used by Wicha et al. (2004) constitutes a violation paradigm, comparable to Hagoort (2003a). In Wicha et al.'s design, however, the semantically anomalous nouns never fit within the preceding sentence context while in Hagoort's study, adjectives were modified to manipulate semantic fit and the same high cloze nouns were used in both congruous and incongruous sentences. Based on this difference between the studies as well as the fact that the present study finds an interaction within the P600 for cloze probability and gender mismatches, there may well be a difference in processing low cloze items as compared to processing a semantic violation. When semantics is violated together with gender, a deeper and more thorough semantic analysis might indeed result in N400-enhancement as reported by Hagoort (2003a) and Wicha et al. (2004). When a gender mismatch is not combined with a semantic violation, but with a factor such as cloze probability which hampers semantic integration, the P600 might be affected the most. An interaction of both the N400 and the P600, as reported by Wicha et al. (2004), might be due to the combination of processing a noun that not only constitutes a violation, but that is also unexpected on the basis of the preceding sentence information. Future research should disentangle this issue by directly comparing semantic violations and semantic integration difficulties as well as their effect on the N400 and P600 respectively.

The results of the present study suggest that cloze probability causes an interaction with gender at the level of the P600. One question that remains is why Gunter et al. (2000) report a reduction of the P600 whereas the P600 in the present study is only delayed in the low cloze condition rather than reduced. Gunter et al. (2000) suggested that their results might reflect the fact that, in the low cloze conditions, the subjects' primary goal is the construction of meaning. The authors hypothesized that, when confronted with a high cloze noun, no thorough semantic analysis is needed, but in low cloze conditions this is the primary goal (as reflected in the N400). The present study showed a significant P600 in both conditions and the only difference found here is that the P600 started later in the low cloze conditions, i.e. only after semantic integration has been completed, whereas it is speeded up in the high cloze conditions where semantic integration is easy. The fact that the P600 in Gunter et al.'s (2000) study was delayed, reduced, and also shorter in latency for low cloze items most likely results from the fact that they used a probe detection task as opposed to the acceptability judgement task used in the present study and task effects have been shown to modulate the P600 (Coulson et al., 1998). The participants in Gunter et al.'s study were asked to recognize and hence focus on the lexical items in the sentence, while the participants in the present study were asked to focus on (overall) acceptability of the sentence. The latter task might cause participants to focus more on integration processes (as reflected in the P600), whereas Gunter et al.'s probe detection task may well have increased the participants' primary focus on meaning of the sentence, resulting in a situation of less repair and/or re-analysis. The later onset of the P600 in low cloze conditions found in the present study reflects the fact that there is no primary focus on meaning and the repair process is equally robust for both high cloze and low cloze items from 800 ms until the end of the ERP epoch around 1500 ms. In low cloze items, however, the re-analysis process can only be started once the noun has been adequately found and integrated in the mental lexicon. It makes sense that this takes longer for nouns that cannot be predicted from the preceding context, slowing down the subsequent re-analysis process.

In sum, the present study replicates previous findings showing that cloze probability facilitates sentence comprehension and that gender mismatches are processed at a syntactic level as reflected by the P600. Future research should take into account the influence of context and predictability on both semantic and syntactic processing as revealed by the earlier onset of the P600 for high cloze targets. High cloze probability speeds up the recognition process with an earlier onset of the P600. Low cloze items can only be repaired later once they have been retrieved from the mental lexicon. The fact that many experiments have shown different onsets of the P600 might well be the result of differences in predictability of the used items. The present data do not provide conclusive evidence supporting one of the two main models of sentence comprehension. On the one hand, an interaction occurring only in later stages of processing was hypothesized by Friederici's syntax first-model (2002), but the absence of any early syntactic effects again argues against an autonomous role for syntax. On the other hand, the absence of earlier interactions, such as a syntactic boost on the N400, argue against fully interactive models in which all types of information interact continuously during each step of the comprehension process. Based on the present data, we cannot draw conclusions about early processes of sentence comprehension, because the N400 appears to be the first reliable effect in the present experiment. The fact that this component is unaffected by grammar in turn argues against Hagoort's (2003a) claim that syntax can affect semantic processing and not vice versa. Our results suggest that the N400 is not influenced by the presence of a gender mismatch, but that cloze probability speeds up the re-analysis process as reflected in an earlier P600 for items that are highly expected based on the preceding context, confirming the predictions made by interactive processing models.

One could conclude that the present results contradict the idea that late processes are controlled, but it might be not as simple as that. An important characteristic of ERP effects, in this respect, is that the latency of an ERP component shows the point at which the brain reveals differences and not necessarily the onset of the underlying cognitive processes. Therefore, the P600 might have an earlier onset, which is simply not visible due to negativities in the earlier, preceding time windows (N400). This can, however, not be concluded on the basis of the present data alone and should be examined by future research. In either case, it can be concluded that cloze probability facilitates the processing and repairing of gender agreement violations supporting an interactive view of the language faculty.

Chapter 6

CAN LATE L2 LEARNERS OF DUTCH USE GENDER CUES DURING SPOKEN WORD RECOGNITION?

Evidence from Eye Movements

Dutch natives have been shown to be able to use common gender-marking preceding the noun to facilitate language comprehension. To investigate whether late advanced Polish learners of Dutch are also able to process gender cues in a native-like fashion and/or whether they use and transfer the gender information from their native language, eye movements were monitored as participants followed spoken instructions to click on one of four displayed items on a screen (e.g. *Klik op de*$_{COM}$ *rode appel*$_{COM}$: 'Click on the$_{COM}$ red apple$_{COM}$'), containing the target, a competitor, and two unrelated distractors. Half of the items shared gender with the Polish equivalent noun, while the other half did not. Dutch natives were slower in locating the target when both target and competitor shared common gender. Late L2 learners did not show this gender effect nor did they transfer gender categories from their mother tongue. Both native and L2 learners, however, showed difficulty in processing neuter as compared to common gendered nouns. Additionally, while late learners do not transfer the gender categories, they do show sensitivity to nouns that are neuter in Polish, suggesting that gender information from the L1 is activated, but not transferred, while listening to the L2.

This chapter has been submitted for publication: Loerts, H. (Under Review). Can Late L2 learners of Dutch use Gender Cues during Spoken Word Recognition? Evidence from Eye Movements. Submitted (July 11, 2012) to *Applied Psycholinguistics.*

6.1. Introduction

Speakers who acquired a second language (L2) at a later stage in life are often characterized by disfluencies in their speech as well as grammatical errors that are hardly ever made by native speakers of the language. One particularly characteristic error that is made, even by highly proficient L2 speakers, concerns grammatical gender. Learners of gendered languages have to acquire a system in which nouns are classified as belonging to a certain category, such as masculine, feminine, and neuter, and words that are syntactically related to these nouns then have to agree in gender with the noun they relate to or refer to. One obvious difficulty in acquiring the assignment system, i.e. the nouns and their corresponding genders, is that gender is a rather abstract system that has little to do with biological sex. The majority of nouns, such as nouns referring to inanimate objects, cannot be related to any biological gender characteristics of the referent and the gender of nouns can often only be extracted from more or less subtle cues in the input, such as in the morphology of the agreement targets.

In languages with overt gender systems, however, the noun itself can also carry some of these cues. In Spanish, for example, most (though not all) nouns ending with the suffix '*–a*' are feminine ('*la*$_{FEM}$ *cuchara*$_{FEM}$': the$_{FEM}$ spoon$_{FEM}$), whereas nouns ending in '*–o*' are usually masculine ('*el*$_{MASC}$ *cuchillo*$_{MASC}$': the$_{MASC}$ knife$_{MASC}$). Besides phonological cues, some languages contain more subtle morphological cues that help determine the gender of a noun. These signs often entail suffixes, such as the suffix *'-ion'* generally denoting feminine gender in French (Foucart, 2008) and the suffix *'-keit'* generally denoting feminine gender in German (Zubin & Köpcke, 1984). Dutch has a two-way gender system with nouns either being common or neuter and, while gender has to be marked on the agreement targets, it can hardly ever be distracted from the shape or meaning of the noun itself (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997). The majority of Dutch nouns are of common gender (75%), which used to contain both masculine and feminine genders but has collapsed into one common gender, and only 25% of nouns have neuter gender (Van Berkum, 1996). In addition, the agreement targets, which are determiners, adjectives, and pronouns, do not always provide unambiguous information about the Dutch gender system. For example, all common nouns are preceded by the common definite determiner *'de'* and neuter nouns take the neuter definite determiner *'het'* while adjectives are not marked for gender in definite noun-phrases (NPs) and always receive the common schwa-ending as visible in Table 6-1. In indefinite NPs, gender is only marked on the adjective with the schwa-ending used for attributive adjectives when modifying common nouns and bare adjectives are used to modify neuter nouns. Table 6-1 reveals the special status of neuter gender, with all attributive adjectives receiving a schwa-ending, except when modifying a neuter noun in an indefinite noun phrase. The default status of common gender in Dutch is amplified by the fact that common gender-marking is also used to refer to all plural nouns regardless of their gender in singular form. All nouns are treated as neuter, however, when they are used in their diminutive form. The grammatical gender system in Dutch could thus be summarized as having a default common gender for which the gender-marking is always used, except when the noun referred to is neuter or a diminutive form. The default status of common gender is reflected in the acquisition process, with children overgeneralizing the definite, determiner *'de'* until the age of 6 (Blom, Polisenskà, & Weerman, 2008; Van der Velde, 2004), which contrasts with children from other languages who have been reported to have already acquired their gender system at the age of 3 or 4 (German: Szagun, Stumper, Sondag, & Franik, 2007; French: Van der Velde, 2004).

Table 6-1: Overview of the main agreement targets in Dutch. Adapted from Loerts, Wieling, & Schmid (Under Review). English equivalents of the phrases are printed in italic below each phrase. Note that the indefinite article is always *'een'* in the singular case and no equivalent indefinite article exists for the plural case.

	Gender	Definite Article	Adjectives in Definite NPs	Adjectives in Indefinite NPs
Singular	common	***de*** appel	***de*** rode appel	een ***rode*** appel
	neuter	***het*** huis	***het*** rode huis	een ***rood*** huis
English equivalent		*the apple/house*	*The red apple/house*	*a red apple/house*
Plural	common	de appels	de rode appels	--
	neuter	de huizen	de rode huizen	--
English equivalent		*the apples/houses*	*the red apples/houses*	--

The asymmetries of both the Dutch assignment as well as the Dutch agreement system suggest that acquisition might well occur on an item-by-item basis (Unsworth, 2008). Is it possible for natives and L2 learners to use such an abstract system to facilitate language comprehension? Recent investigations using the eye tracking methodology have shown that this question can be addressed by using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This paradigm presents the participant with a display of pictures that contain a target as well as a competitor that competes with the target on a certain level, such as the initial phonemes. People then receive instructions to manipulate the target and the results have repeatedly shown that people immediately try to use the information as the speech stream unfolds to select the correct picture on the screen. When the screen contains both a picture of a candle and a picture of a candy, for example, participants fixated on both when asked to '*pick up the can...*' and only looked at the candy after hearing the disambiguating information*'...dy'* (Allopenna, Magnuson, & Tanenhaus, 1998).

This observed cohort effect illustrates how lexical candidates compete for recognition until the uniqueness point of the word is heard (Marslen-Wilson, 1987) and triggered researchers to examine whether the same holds for morphological information, such as grammatical gender. In other words, are people able to use prenominal gender-marking, such as a definite determiner, to pre-select only those referents that correspond to the gender of this determiner?

This question has been addressed in French by Dahan, Swingley, Tanenhaus, and Magnuson (2000) who asked their participants to click on one of four displayed pictures on a screen, containing the target, a cohort competitor sharing its initial phonemes with the target, and two unrelated competitors. In French, nouns are either masculine or feminine and, in the singular, the definite determiner might thus be used to pre-select only gender-matching referents and/or inhibit gender-mismatching referents. In the plural, however, no prenominal gender-marking is present in the input and hence plural items served as a baseline condition. Their results showed that, when people were asked to *'cliquez sur les bou…'* ('Click on the$_{\text{PL}}$ *bou…*'), they fixated equally often on a picture of a bottle ('bouteille$_{\text{FEM}}$') and a picture of a button ('bouton$_{\text{MASC}}$'). When asked to '*cliquez sur le*$_{\text{MASC}}$ *bou…*' ('Click on the$_{\text{MASC}}$ *bou…*'), however, natives did not fixate on the picture of a bottle (*la*$_{\text{FEM}}$ *bouteille*) more often than on the pictures containing unrelated distractors. These results indicate that prenominal gender-marking on definite determiners facilitates language processing by minimizing the set of possible lexical candidates. Similar results have also been shown for referents that do not overlap in phonemic onset when intervening adjectives are added to provide more time to process the gender information on the determiner in German (Paris, Weber, and Crocker, 2006), and more recently also in Dutch native speakers (Loerts, Wieling, & Schmid, Under Review).

The strength of the eye tracking method in general, and the visual world paradigm in particular, to expose the incremental nature of (grammatical gender) processing has increasingly led to experiments comparing native and second-language processing (for an overview including reading studies, see Dussias, 2010). Most of these studies have revealed an influence of the native language when processing the L2. When asked to pick up a marker, for example, late Russian L2 learners of English have been reported to briefly look at a picture of a stamp, which is called a 'marka' in Russian (Marian & Spivey, 2003). This suggests that, even in a complete L2 situation, lexical candidates from the native language are also activated during spoken word recognition and can act as phonemically overlapping cross-linguistic competitors. Is morphological information from the native language, such as a noun's gender, also activated and transferred during spoken word recognition in the L2?

This question was examined by Weber and Paris (2004), who asked proficient French-German bilinguals to click on one of four pictures on a computer screen containing the target, a competitor overlapping in phonemic onset in French and German, and two unrelated distractors. In half of the items, i.e. the 'same-gender' pairs, the target and competitor shared gender in both the native and second language, while target and competitor differed in gender only in French, but not in German in the 'different-gender' pairs. When participants were asked to look for a picture containing a pearl, as in '*Wo befindet sich die*$_{\text{FEM}}$ *Perle*$_{\text{FEM}}$*?'*, both natives as well as French-German L2 learners fixated more on a picture containing a wig ('die$_{\text{FEM}}$ Perücke$_{\text{FEM}}$') as compared to the unrelated distractors. When they were asked to find '*die*$_{FEM}$ *Kassette*$_{FEM}$' (the cassette), however, only German native speakers fixated more often on a picture of a canon, *'die$_{FEM}$ Kanone$_{FEM}$'*, but French-German L2 learners did not. This result suggests that the native French masculine gender of the canon, *'le$_{MASC}$ canon$_{MASC}$'*, inappropriately eliminated competitor activation in the different-gender trials. This effect of negative transfer led Weber and Paris (2004) to conclude that, at least for cognate nouns, native gender information can influence non-native spoken word recognition.

From a theoretical perspective, this result might be explained by Full-Transfer Full-Access models which assume that L2 learners will transfer all properties from their native language (Schwartz & Sprouse, 1996), or at least in initial stages of L2 acquisition (Bruhn de Garavito & White, 2000). The results, however, have only been reported in response to cognates, which are known to be accessed faster by bilinguals than non-cognates (e.g. Costa, Caramazza, & Sebastian-Galles, 2000). Non-cognates only share semantic information while cognates also share additional orthographic and phonological information and might even have shared representations within the bilingual brain (Dijkstra & Van Heuven, 2002). The results so far thus cannot answer questions about transfer of lexico-syntactic gender per se.

The present experiment aimed at examining whether gender cues in Dutch facilitate spoken word recognition in natives and L2 learners with L1 Polish, a language that differs from Dutch on many linguistic levels. The most important difference for the present study concerns the complex three-way gender system consisting of masculine, feminine, and neuter nouns that additionally divides masculine gender into animate and inanimate gender in the singular and into masculine personal and non-masculine personal in the plural (Koniuszaniec & Błaszkowska, 2003). Polish has no articles, but it does have other determiners, numerals, adjectives, pronouns, and verbs that must agree in case, number and gender with the noun (Perlak & Jarema, 2003). Most important for the present study, gender-marked colour adjectives precede the noun in Polish, as they do in Dutch.

The present experiment used the same materials as in Loerts, Wieling, and Schmid (Under Review) and compared Polish-Dutch L2 learners' eye movements to those of native speakers while they were asked to click on one of four pictures on a computer screen, containing the target, a competitor that either matched in gender and/or colour with the target, and two unrelated distractors. The auditory instructions included either a gender-marked determiner (as in definite NPs) or a gender-marked adjective (indefinite NPs) followed by the target noun to assess potential effects of locality of the gender-marking as well as saliency. With respect to locality, colour adjectives were used to allow listeners more time to process gender information in the definite condition (Dahan et al., 2000). In addition, gender-marking on the determiner constitutes two separate articles (*'de'* vs. *'het'*), while gender-marking on the adjective, as present in indefinite NPs, constitutes the less salient presence or absence of the schwa-suffix. By including both structures, we can account for potential differences in processing based on the locality and saliency of gender-marking. The target noun never overlapped phonemically with the competitor in order to examine effects of gender activation per se.

Our previous analyses of native Dutch eye movements has shown that Dutch people take longer to fixate on the target picture when both the target and the competitor share common gender, while no gender effect was found in the neuter gender trials (Loerts et al., Under Review). The absence of a gender effect for neuter nouns probably reflects the fact that all Dutch nouns, regardless of the gender they carry, are preceded by the neuter determiner *'het'* when used as a diminutive (an extremely frequent derivation in Dutch). Consequently, neuter gender-marking cannot disambiguate between common and neuter nouns.

To investigate possible effects of L1 transfer, half of the nouns used overlapped in gender with the equivalent Polish noun and the other half of the items consisted of nouns that did not overlap in gender between both languages. In addition, the Polish equivalent of the competitor either overlapped in gender in Polish with the target or not. If Polish gender of the referents is activated and used during spoken word comprehension in the L2, Polish-Dutch L2 learners should reveal difficulty in guiding their eyes towards the target when target and competitor share gender in Polish as compared to when they do not. If, however, Polish-Dutch L2 learners do not transfer gender from their mother tongue, they should not show any differences between these conditions. Ultimately, L2 learners might show sensitivity towards Dutch gender-marking, resulting in a similar pattern as native Dutch speakers, i.e. relative difficulty in guiding their eyes towards the target picture when the competitor shares common gender with the target in Dutch.

Since L2 learners' performance and, more specifically related to the present study, parallel language activation may depend on language proficiency (e.g. Weber & Cutler, 2004), levels of proficiency were tested extensively using a variety of measures, including a self-assessment score, a C-Test, a story-retelling task, and acceptability judgement tasks (see Methods section for further details of these tasks). All these proficiency scores were tested in a mixed-effects model to check their potential impact on L2 gender processing.

6.2. Methods

6.2.1. Participants

A group of 26 late advanced learners of Dutch with L1 Polish (20 female, aged 24-52, mean age=34) were compared with 28 native Dutch speaking adults (16 female, aged 34-60, mean age=48). These native speakers were the same as those tested by Loerts et al. (Under Review).

The late L2 learners had all acquired L2 Dutch after the age of 20 (range: 20-34) and the average length of exposure to the L2 was 9 years (range 2-27). All participants were recruited from a non-university population by means of advertisements in the local newspaper, the online Polish newspaper ('Polonia'), and social media. All participants had normal hearing, normal or corrected-to-normal vision and none suffered from colour blindness. They were paid for their participation.

L2 proficiency of the Polish-Dutch learners was assessed through a combination of measures that have previously successfully been used to assess proficiency including (1) a C-Test in which participants were asked to fill in the gaps, i.e. parts of words, in two texts of about 60 to 70 words each (also see Keijzer, 2007), (2) a self-assessed proficiency measure based on the scores on a 5-point Likert scale reflecting Dutch listening, reading, speaking, and writing, (3) a film retelling task (see Perdue, 1993), which was used to assess a holistic proficiency measure of fluency based on the free speech data (as rated by three experienced native Dutch linguists) and specific use of gender-marking on common and neuter nouns in definite and indefinite NPs and, (4) an online acceptability judgement task in which participants listened to correct sentences and sentences containing several grammatical violations, including violations of gender concord (for details on this online acceptability judgement task, also see Loerts, Stowe, and Schmid, Under Review). In addition, (5) participants' conscious knowledge of the Dutch gender system was assessed by asking them to circle the correct article (i.e. *'de'* or *'het'*) of a list of Dutch nouns including the ones that had been presented to them during the experiment. Each noun occurred three times in this gender assignment task to avoid performance at chance level.

Table 6-2 provides an overview of the scores on the different tests and, although most participants can be regarded as highly proficient in Dutch, the scores on the different tasks show a reasonable spread in proficiency especially when it comes to the production and comprehension of gender in general and neuter gender in particular. The potential influence of these individual continuous proficiency measures was tested in the mixed-effects model. Additionally, the influence of the conscious knowledge of the gender of nouns, as tested by the gender assignment task (5), was added per subject and item (also see 'statistical analysis' below).

6.2.2. Visual stimuli

A set of 48 high frequency nouns (24 common and 24 neuter) were selected from the CGN 5000, the frequency list of the *Corpus Gesproken Nederlands* (CGN: Corpus of Spoken Dutch) containing nouns that belong to the 5000 most frequently occurring words in the CGN. Not only are names of higher frequency more likely to attract more (though shorter) fixations (Dahan, Magnuson, & Tanenhaus, 2001), frequency effects have been shown to be larger for second language learners and have been found to be inversely related to L2 exposure during reading (Whitford & Titone, 2012). Nevertheless, there were still differences in frequency in our dataset (mean lemma frequency: 323.2; range: 7 - 3203) and this variable was therefore log-transformed and included in the mixed-effects regression model.

Table 6-2: Scores on several proficiency measures that were used to assess potential effects of proficiency on gender processing in the L2. All scores are provided in percentages and the ranges within both the group of native speakers and late L2 learners are provided in italic after each mean test score.

		Natives *(range)*	***L2 Speakers*** *(range)*
1. C-Test (%)		97.04 *(90 - 100)*	79.24 *(25 - 97)*
2. Self-rating (%)		97.57 (*84 – 100)*	78.19 *(40-100)*
3. Story-retelling (%)	*Holistic Proficiency*	94.03 *(91 - 99)*	61.48 *(36 - 89)*
	Common gender-marking	97.74 *(85 - 100)*	78.33 *(36 - 100)*
	Neuter gender-marking	93.68 *(79 - 100)*	48.85 *(0 - 100)*
4. Acceptability judgement task (% correct)	*General grammatical violations*	96.47 *(81 - 100)*	70.70 *(38 - 95)*
	Common gender violations	91.28 *(67 - 100)*	54.87 *(36 - 76)*
	Neuter gender violations	89.41 *(52 - 100)*	55.68 *(49 - 82)*
5. Offline gender assignment task (% correct)	*Common nouns*		74.41 *(23 - 100)*
	Neuter Nouns		53.22 *(9 - 85)*

All nouns chosen could be used to refer to depicturable, coloured objects in order to allow an intervening colour adjective in the auditory stimuli. The resulting list of nouns was translated into Polish in order to exclude cognates between the L1 and the L2, because we are specifically interested in the influence of having a gender system and not the influence of the native language in general. The resulting 24 Dutch common nouns and the 24 Dutch neuter nouns were equally divided according to the gender of their Polish translation equivalent to assess the potential influence of overlapping gender categories. Hence, 24 of the nouns (12 common and 12 neuter) shared the same gender as their Polish equivalent, while 24 did not. It should be noted that the neuter gender category is the only completely overlapping category. Dutch common nouns, on the other hand, overlapped with both masculine and feminine equivalents in Polish.

The experimental stimuli were derived from the standardized picture set of black and white line drawings (Snodgrass and Vanderwart, 1980), but rendered in one of five different colours: *'rood'* (red), *'blauw'* (blue), *'groen'* (green), *'geel'* (yellow), and *'bruin'* (brown).

6.2.3. Auditory stimuli

Auditory instructions were digitally recorded in a soundproof booth with a sampling rate of 44.1 kHz by a female native speaker of Dutch with normal intonation. All 48 nouns were embedded in a sentence context either containing definite NPs, in which gender-marking appears on the determiner (sentence 6-1a, below) or indefinite NPs in which gender-marking occurs only on the adjective (sentence 6-1b, below). Colour adjectives were placed between the determiner and the noun to allow the listeners more time to process gender information in the definite constructions as well as to reduce possible effects of high co-occurrence probabilities of determiners and nouns (Dahan et al., 2000). Since gender-marking on the determiner is more salient than gender-marking on the adjectives, inclusion of both constructions allowed for a comparison of potential differences in the ability to comprehend and use these structures by L2 learners.

6-1a 'Klik op de_{COM} rode $\text{appel}_{\text{COM}}$'
Click on $\textit{the}_{COM}$ red $\textit{apple}_{COM}$
6-1b 'Klik op het plaatje met een rode_{COM} $\text{appel}_{\text{COM}}$'
Click on the picture with a $\textit{red}_{COM}$ $\textit{apple}_{COM}$

Auditory stimuli were slightly adjusted using a speech editor (Adobe Audition 3.0) in order to have all definite determiners start around 800 ms and all indefinite determiners around 2000 ms. The mean duration of the determiners was 208 ms and 195 ms in the definite and indefinite condition respectively. Adjectives lasted for about 535 ms in the definite and about 525 ms in the indefinite condition. The nouns also lasted equally long across conditions with 631 ms for nouns embedded in definite sentences and 642 ms for nouns in indefinite sentences. The constant timing delay of 1200 ms for the indefinite NPs as compared to the definite NPs allowed for a direct comparison of the responses to these different types of sentences.

The four conditions in the present experiment with respect to the Dutch language and gender system are outlined in Table 6-3. Each trial consisted of a screen divided into four quadrants in which four pictures appeared: the target, a competitor matching in colour and/or gender with the target, and two distractors that never matched in gender or in colour with the target. These positions of the target, competitor and distractor objects on the screen were counterbalanced and each item occurred on the screen once as a target, once as a competitor, and twice in the distractor role in order to minimize potential effects of more interesting or notable pictures. To exclude any phonological competition effects, initial segments of the nouns of the four pictures did not overlap, as can also been seen in Table 6-3.

Table 6-3: Overview of the conditions in the experiment. Adapted from Loerts, Wieling, & Schmid (Under Review). Example stimuli used in the present experiment while hearing the instruction to *Klik op de*$_{COM}$ *rode appel*$_{COM}$ ('Click on the$_{COM}$ red apple$_{COM}$ '). Note that in the definite conditions, gender-marking is on the determiner (*'de'* when referring to a common noun and *'het'* when preceding a neuter noun) while gender-marking only appears on the adjective in indefinite NPs (*schwa*-ending when modifying a common noun and bare adjective when modifying a neuter noun).

Target	Competitor	Gender Competitor	Colour Competitor
De$_{COM}$ rode appel$_{COM}$ *The*$_{COM}$ *red apple*$_{COM}$	Het$_{NEU}$ groene bureau$_{NEU}$ *The*$_{NEU}$ *green desk*$_{NEU}$	different	different
	De$_{COM}$ gele zon$_{COM}$ *The*$_{COM}$ *yellow sun*$_{COM}$	same	different
	Het$_{NEU}$ rode hart$_{NEU}$ *The*$_{NEU}$ *red heart*$_{NEU}$	different	same
	De$_{COM}$ rode taart$_{COM}$ *The*$_{COM}$ *red cake*$_{COM}$	same	same

The same conditions as those presented in Table 6-3 were present in the experiment based on the Polish gender of the items. For half of the targets, gender overlapped in both Dutch and Polish, while the other half did not share gender in both languages. In addition, the competitor item either overlapped in Polish gender with the target or not. This set-up allows examining whether Polish gender of the items was activated and transferred during listening in the L2 when the target and competitor shared gender in both languages or not. In other words, it might be possible that a Polish neuter noun serves as a competitor for another Polish neuter noun for L2 learners, but this might also only be the case when both items are additionally neuter in Dutch.

Since both definite and indefinite NPs were tested using the same items, the experiment consisted of two separate blocks: one block containing instructions with definite NPs, and one block containing instructions with indefinite NPs. Each block contained 48 experimental and 22 filler trials. According to a Latin Square design, 4 different lists were created in such a way that each participant heard only one version of each item in a certain condition.

6.2.4. Procedure

Eye movements were recorded using a remote Tobii T120 eye tracker with a sampling rate of 60 Hz in the Eye Lab of the University of Groningen. Stimuli were presented using E-prime (Schneider, Eschman, & Zuccolotto, 2002) and data, both reaction times and eye gaze data, were recorded using the E-Prime Extensions for Tobii software.

Participants were tested individually and sat in a comfortable chair at approximately 60 cm from the eye tracking monitor during the experiment. After a standard 5-point calibration procedure, participants were instructed to use the computer mouse to click on the correct picture, as mentioned in the auditory stimulus, as soon as possible. Before the experiment began, participants performed a short practice session containing 6 filler items to get acquainted with the procedure. To make sure the participants' eyes were focussed at the centre of the screen at the beginning of each trial, subjects had to fixate on a fixation cross for at least 1000 ms in order for the trial to start. As soon as the display containing four pictures appeared on the screen, the auditory instructions started as well. Immediately after the participants had clicked on a picture, the set of pictures disappeared and the next trial started.

Each subject was presented with two blocks: one containing instructions with definite and one with indefinite NPs. Half of the subjects first received the definite instructions and the other half first completed the block containing indefinite constructions. Each block lasted around 12 to 13 minutes and, if desired, subjects could take a short break between the two blocks.

6.2.5. Statistical analyses

The main aim of the present experiment is to investigate whether Polish L2 learners of Dutch can use L2 gender-marking to facilitate language comprehension or whether they tend to transfer the gender from their mother tongue. The dependent measures, reaction times and proportions of fixations towards the different pictures, have previously often been analysed using F1 and F2 analyses in which data is averaged across subjects and items. These factorial designs require dichotomization of numerical variables, such as dividing L2 learners into a low and a high proficient group, resulting in a loss of statistical power. Additionally, categorical and proportional data violate the assumptions necessary to use ANOVA and can lead to false rejection or false acceptance of the null hypothesis (type I and type II errors; Jaeger, 2008).

Most studies using ANOVA are required to create extremely well-balanced designs, but even small differences in, for example, item frequency can already have a significant impact on reaction times (Loerts et al., Under Review). Both the items and the subjects in psycholinguistic experiments are random samples from a larger population of possible levels and can hence not be repeated. A replication of the present study would be able to find items of similar frequency, but the items themselves are highly unlikely to be the exact same sample from the population. The present study therefore uses mixed-effects regression to be able to include both fixed and random effects (Barr, 2008; for introductions, see, e.g. Baayen, 2008, Ch. 7). Fixed-effect factors are factors with a small number of levels that can exhaust all possible levels of the sampled population, such as the numerical variable word frequency or the participant's gender. Random-effect factors, on the other hand, have levels sampled from a much larger population of possible levels.

The present study contains two random-effect factors: item and participant. The items (the nouns and their corresponding picture) were chosen based on a number of criteria and were chosen from a much larger pool of items. Similarly, the participants tested are a sample of a much larger set of possible participants and since our measurements are specific to each chosen item and subject, both these factors are likely to introduce systematic variation. In mixed-effects models, these random-effect factors are viewed as sources of random noise that can be linked to the participants and items that are tested. The inclusion of random intercepts allows for a variability associated with participants or items. For example, some subjects might generally respond faster while others are relatively slow in their responses. Similarly, a certain item might evoke more fixations than another item and hence the intercept for these items may vary. Such intercept adjustments are assumed to follow a normal distribution with mean zero and a standard deviation to be estimated from the data. Once estimated from the dataset, it becomes possible to adjust the population intercept such that it becomes precise for each individual participant and each individual item. In this case, these adjusted intercepts are referred to as by-subject and by-item random intercepts.

An additional possibility would be that the variation associated with a random-effect factor affects the slopes of other predictors. For example, the slope of age of L2 acquisition (AoA) might vary per item, indicating that some items are more easily recognized by earlier acquirers, while other items are easier for learners who started learning the L2 at a later age. A mixed-effects model will estimate the by-item biases in the AoA slope and add these estimated adjustments to the general AoA slope. Consequently, this allows for a rather precise estimation of the effect of age of acquisition for each item in the experiment. For each random intercept and random slope, likelihood ratio tests are used to evaluate whether the increase in the number of parameters is justified given the increase in goodness of fit.

The analyses used in the present paper considered the random-effect factors, participant and item, as well as a contrast to distinguish the presence or absence of a colour competitor and a gender competitor in the scene (the contrast we are most interested in). Specifically related to the L2 learners, gender overlap of the target in Dutch and Polish as well as the presence or absence of a Polish gender competitor in the scene was considered. In addition, item-related characteristics such as the gender, word frequency, and the acquaintance of the L2 learners with the item were tested, as well as participant-related characteristics such as gender, age (of acquisition), length of residence, and the various proficiency measures used (see 'participants'). To control for a possibly imbalanced design, factors such as the position, colour, and previous presence of a specific item (e.g. as a distractor) were included in the analyses. Potential effects of fatigue and/or learning during the course of the experiment were accounted for by including factors such as trial, session, and responses to the previous trial and testing their significance in the model.

To test whether fixed-effect factors had to be included in the model, the usual *t*-test for the coefficients was used. The present dataset contains a large number of observations and hence the *t*-distribution approximates the standard normal distribution. Therefore, factors will be significant (i.e. $p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen, Davidson, & Bates, 2008).

Separate mixed-effects models were computed for both dependent variables: the reaction time data (henceforth RT) and the eye movement data (henceforth EYE). Models were fit to the full data set containing both natives and L2 learners as well as to the data from each group separately. To avoid redundancy and increase clarity, information about the fixed- and random-effect factors of the models fit to the separate datasets are only mentioned when they clarify effects or two- and three-way interactions in de model fitted to the full dataset.

Only predictors that led to a significant contribution to the fit of the model were retained and predictors that did not contribute significantly to the model were removed to achieve the best fit model. Potential nonlinear effects of continuous predictor variables were checked using restricted cubic splines and predictors were centred before analyses to limit the collinearity of the predictors included in the model (Jaeger, 2008).

Linear mixed-effects models were fit using the *lmer* function of the lme4 package (Bates, 2005) implemented in R (version 2.11.1: The R Foundation for Statistical Computing, 2010). Only extreme outliers in the data were removed prior to model fitting and additional potentially influential outliers were removed after fitting a first version of the model (i.e. model criticism, see Baayen & Milin, 2010). The residuals of the best fit models presented in the results were normally distributed.

6.3. Results

In accordance with the simplicity of the task, accuracy scores were very high in both natives (99.7%) as well as L2 learners (98.7%) and only 45 items (0.8%) had to be removed from further analyses due to inaccurate responses. A very small number of extreme outliers (0.44%), for which responses had only been given over 1.5 seconds after the onset of the noun, were also removed.

For the dependent measures in response to the indefinite conditions, the mean timing delay of 1200 ms was subtracted in order to be able to directly compare responses from both conditions.

6.3.1. Reaction times

Out of the 5235 responses of the initial data set, absolute standardized residuals exceeding 3 standard deviations were removed (approximately 0.5% of the data from native Dutch speakers and 2.5% from the L2 learners group) after the first best model was fit to the data. The final best fitting model including all necessary fixed and random effect factors for the entire dataset (Table 6-4 and Table 6-5) explained approximately 65% of the variance in the data.

Table 6-4: Fixed-effect factors in the best fitting mixed-effects regression model fit to the combined reaction time (RT) data including both native and second-language speakers of Dutch. Estimated Coefficients, standard errors, and *t* values are reported for the model and an increase in reaction times is reflected by positive coefficients and a decrease in RTs is reflected by negative coefficients. Factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

	Estimate	***Std. Error***	***t value***
(Intercept)	6.6080	0.1070	61.78
SameColourCompetitor	0.1401	0.0054	25.76
PreviousRT (log)	0.1024	0.0104	9.82
IsFirstSession	0.0235	0.0055	4.31
Sequence	-0.0003	0.0002	-2.15
TargetIsRed	-0.0301	0.0076	-3.96
TargetIsLeft	0.0096	0.0034	2.79
CompetitorIsRed	0.0132	0.0034	3.85
LemmaFrequency (log)	-0.0077	0.0024	-3.26
TimeOnsetAdjective (ms)	0.0001	0.0000	3.51
TimeOnsetNoun (ms)	0.0001	0.0000	2.63
IsLateLearner	0.0045	0.0184	0.25
TargetIsNeuter	-0.0002	0.0064	-0.04
IsLateLearner:TargetIsLeft	-0.0143	0.0049	-2.94
IsLateLearner:LemmaFrequency (log)	-0.0056	0.0018	-3.08
IsLateLearner: TargetIsNeuter	0.0140	0.0048	2.90

Table 6-4 shows the coefficients and *t*-values of the fixed-effect factors in the best fit model for the combined native and L2 learner reaction time data. The intercept in the top line of the table reveals a base reaction time of around 741 ms (6.6080 in logarithmic scale). The positive or negative coefficients of the fixed effects in the table reveal whether this base reaction time increases or decreases when these predictors are present or when their levels increase.

The RT data did not reveal any significant effects of the presence of a *SameGenderCompetitor*, nor was there any interaction with this main variable of interest. Interestingly, no main differences in RTs were found between native speakers and late L2 learners (*IsLateLearner:* $\beta = 0.0045$, $t = 0.25$). Differences between the groups, however, were found with respect to item-related characteristics and positioning of the pictures. When the *TargetIsLeft,* i.e. when the target is positioned in one of the left-hand quadrants of the screen, this negatively impacted reaction times ($\beta = 0.0096$, $t = 2.79$, $p < 0.05$, two-tailed test). Since these RTs constitute mouse-clicks and since all our subjects were right-handed, we hypothesized this result to reflect participants' preference to move their hand from left to right as opposed to from right to left, in turn causing a delay in manual responses to objects in the left-hand quadrants of the screen (Loerts et al., Under Review). The interaction with *IsLateLearner*, however, revealed that this effect is only present in native speakers and not in L2 learners of Dutch (β = 0.0096 - 0.0143 = -.0047, t = -2.94, $p < 0.01$). One might speculate the absence of the effect in L2 learners to reflect their potentially higher level of concentration during a simple task as compared to native speakers, arising from their drive to perform in their second language, but the result is difficult to explain on the basis of the present data alone.

Another significant control variable, *LemmaFrequency* (β = -0.0077, t = -3.26, $p < 0.01$), revealed that higher frequency items were responded to significantly faster. Interestingly, an interaction with *IsLateLearner* revealed that this frequency effect was significantly larger within the group of L2 learners (β = -0.0077 - 0.0056 = -.0133, t = -3.08, $p < 0.01$). In other words, targets of higher frequency are responded to faster as compared to targets of comparatively lower frequency and these differences are even larger in the group of L2 learners. The final interaction showed that, while native RT data are not significantly affected by the gender of the target (*TargetIsNeuter:* β = -0.0002, t = -0.04), L2 learners' response times significantly increased when the *TargetIsNeuter* (β = -0.0002 + 0.0140 = 0.0138, t = 2.90, $p < 0.01$). This result corresponds to the observed difficulty L2 learners have with both the production and comprehension of neuter nouns (also see Table 6-2).

As would be expected, RTs were, independent of the linguistic background of the participants, most severely influenced by the presence of a *SameColourCompetitor* (β = 0.1401, t = 25.76, $p < 0.001$). The large positive coefficient for this effect revealed that, when the target was accompanied by a competitor that shared colour with the target, decisions could only be made significantly later as compared to when the target's colour was unique. The specific colour of the items additionally appeared to be important. When the *TargetIsRed*, responses were significantly faster (β = -0.0301, t = -3.96, $p < 0.001$), while responses significantly slowed down when the *CompetitorIsRed* (β = 0.0132, t = 3.85, $p < 0.001$). These results reflect the fact that salient colours, such as red, attract more attention and are hence responded to faster (if *TargetIsRed*) and/or can distract attention from the target (if *CompetitorIsRed*). Although we are not specifically interested in these variables, inclusion of these and previously mentioned control variables allow for a more precise estimation of the effects of the other predictors (Baayen & Milin, 2010).

Similarly, temporal dependencies were controlled for by including potentially significant effects of fatigue and/or learning. A slower *PreviousRT*, i.e. the response time to the previous trial, significantly slowed down reaction times to the current trial (β = 0.1024, t = 9.82, $p < 0.001$). An additional learning effect revealed that participants were significantly slower overall in the first as compared to the second session (*IsFirstSession:* β = 0.0235, t = 4.31, $p < 0.001$) and became faster during the course of the experiment, as revealed by the significant negative coefficient for *Sequence* (β = -0.0003, t = -2.15, $p < 0.05$).

Table 6-4 additionally reveals that increases in both the *TimeOnsetAdjective* and the *TimeOnsetNoun,* i.e. the time (in ms) at which the adjective and the noun started in the speech stream, significantly slowed down participants' responses (β = 0.0001, t = 3.51, p < 0.01; β = 0.0001, t = 2.63, p < 0.05, respectively), reflecting that these minor differences in timing do cause the relevant information in the speech stream to be processed and hence responded to later. Again, it should be noted that these variables are not of main interest and their inclusion allows for a more precise estimation of the other (more important) effects.

The best fit model required both a random intercept for item (a specific picture-noun combination) as well as a random intercept for subject, as can be seen in Table 6-5, where the random-effect factors are labelled as Groups. The third column of the table lists the standard deviations for the adjustments to the intercept for both random-effect factors. Additionally, standard deviations are listed for the by-subject adjustments to the coefficients or random slopes of 4 covariates: *SameColourCompetitor*, *TargetIsBrown*, *IsIndefinite*, and *Sequence*.

Table 6-5: Random-effect parameters in the model fitted to the full dataset containing the reaction time (RT) data from both natives and L2 speakers. The numbers in the column entitled 'Correlations with Intercept' represent correlations between the by-subject random intercept and the random slope for *SameColourCompetitor* and *TargetIsBrown*, respectively. The number in the second column represents the correlation between the by-subject random slopes for *SameColourCompetitor* and *TargetIsBrown*.

Groups	***Name***	***Standard Deviation***	***Correlations with Intercept***	***Correlations between Slopes***
Item	Intercept	0.0183		
Subject	Intercept	0.0746		
	SameColourCompetitor	0.0363	-0.695	
	TargetIsBrown	0.0125	-0.758	0.365
Subject	IsIndefinite	0.0352		
Subject	Sequence	0.0009		
Residual		0.0848		

Both the effects of a *SameColourCompetitor* and a *TargetIsBrown* showed correlations with the intercept as well as with each other. Relatively fast subjects, i.e. those with negative adjustments to the intercept, tended to suffer more from the presence of a *SameColourCompetitor* than relatively slower subjects and the same holds when the *TargetIsBrown.* The positive correlation between these two reveals that those subjects who suffered more from a *SameColourCompetitor* also suffered more when the *TargetIsBrown*. This correlational structure is also graphically presented in Figure 6-1.

Table 6-5 additionally reveals the two by-subject random slopes for the effects of *IsIndefinite* and *Sequence* that significantly improved model fit. These random slopes reflect the fact that some subjects were slower in response to indefinite constructions as compared to definite constructions, while others were faster. Similarly, the learning effect that reflected increasing RTs during the course of the experiment appears to differ significantly among subjects, with some of them showing a large effect of *Sequence* while others only show minor or no changes in RTs during the course of the experiment.

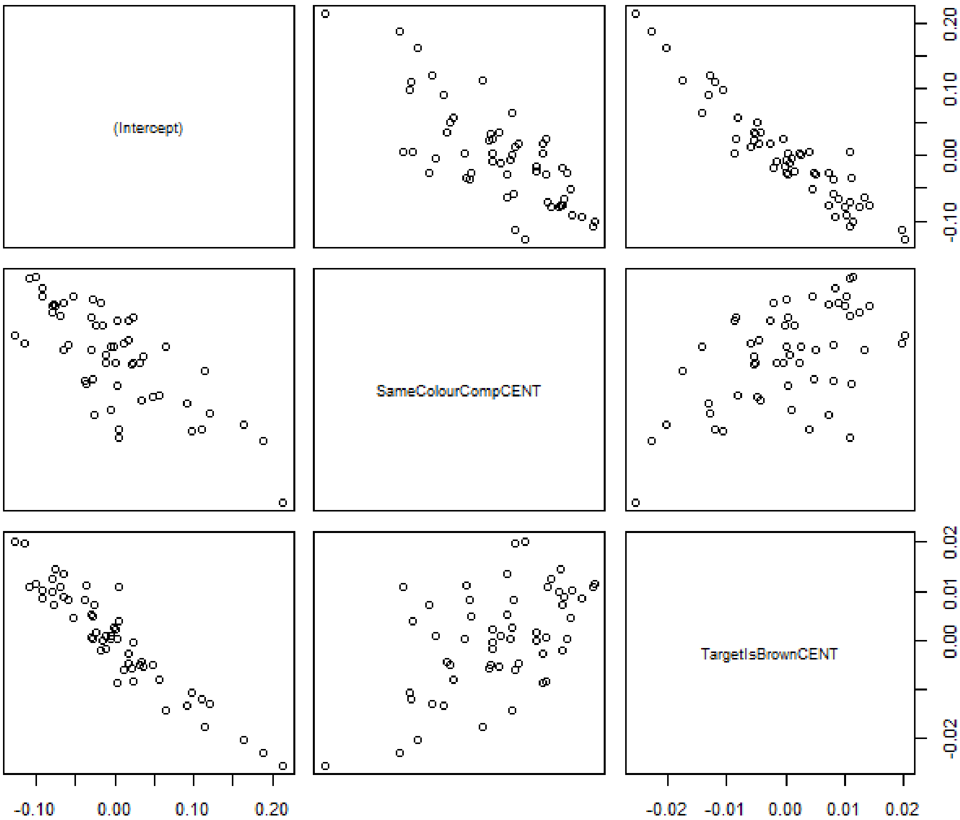


Figure 6-1: Scatterplots visualizing the correlational structures of the random intercepts and slopes for *Subject* in the model for the combined reaction time (RT) data of native and L2 speakers of Dutch. Adjustments to the intercept position subjects with respect to the average response time and subjects with large positive BLUPS (best linear unbiased predictions) are relatively slow responders while those with large negative BLUPS are relatively fast in their responses. This graph illustrates that these relatively fast subjects tend to suffer more from a *SameColourCompetitor* and also when the *TargetIsBrown*. The positive correlational structure reveals that those subjects who suffer more from a *SameColourCompetitor* also tend to suffer more when the *TargetIsBrown*.

Neither the full model nor the best model fit solely to the L2 learners' data revealed any effects of or interactions with age of acquisition, length of residence, or any of the proficiency scores also outlined in Table 6-2. In addition, no effects were found with respect to the Polish gender of the targets and/or the presence or absence of a gender competitor in Polish.

6.3.2. Eye movement data

The auditory stimuli used contain relevant and distinctive gender information on the determiner and the adjective and this information, i.e. gender-marking and colour information, might be used to pre-select the target picture. Therefore, proportions of looks from determiner onset to noun onset were analysed to examine possible gender effects (taking into account the ca. 200 ms it takes to plan and launch an eye movement).

For each item and each time point, the gaze location was calculated as being either on the Target, on the Competitor, on the Distractors, or not on an Area of Interest. When the Tobii system reported incorrect or corrupted data (corresponding to validity codes exceeding 2) the gaze location for these time points was set to NA in order to retain as much data as possible. This affected 8% of the data from the Dutch natives and 5.5% of the late L2 learners' data. Additionally, an error percentage was calculated for each individual item and subject combination and those items that contained more than 30% of incorrect or corrupted data were removed. This affected an additional 9.1% of the data in Dutch natives and 4.7% in the late L2 learners.

For the resulting data set, mean proportions of fixations were calculated across the entire time period (i.e. from determiner onset until noun onset plus 200 ms) per subject and per item. To avoid problems associated with performing statistical analyses on proportion data, a linear mixed-effects regression model was fitted to the empirical logit transformation of the dependent variable *TargetRatio*, i.e. the percentage of looks towards the target divided by the total percentage of looks towards the target and its competitor (Barr, 2008; Jaeger, 2008). If participants use gender-marking cues and colour information to pre-select the correct picture, this would lead to a higher *TargetRatio,* corresponding to relatively more fixations towards the target as compared to the competitor when the information is disambiguating.

After the first best fit model was attained, absolute standardized residuals exceeding 3 standard deviations were removed from the dataset (1.3% of the data in the native speakers and 3.2% of the data from the L2 learners) after which the model was refitted. The fixed- and random effect factors of the best model, outlined in Table 6-6 and Table 6-7, explained approximately 34% of the variance in *TargetRatio* scores.

Figures 6-2 and 6-3 show the eye movement data of the native and L2 speakers in response to conditions containing a *SameColourCompetitor.* The upper panel in Figure 6-2 shows that native Dutch speakers look at the correct common gender target earlier when the competitor does not share gender (upper-left corner) as opposed to when target and competitor share common gender (upper-right picture). The lower panel, depicting native eye movements in response to neuter nouns, shows that gender-marking on neuter nouns could not be used to facilitate the processing of neuter nouns in the different gender as opposed to the same gender condition. Figure 6-3 suggests that late L2 learners do not show differences in their fixations towards a target whether or not there is a *SameGenderCompetitor* present. The lower panels, containing L2 learners' eye movements in response to neuter gender targets, do suggest that late L2 learners require more time to select neuter as compared to common gender targets and looks towards the competitor picture appear to last slightly longer in the same gender condition (bottom-right picture) as opposed to the different gender condition (bottom-left picture).

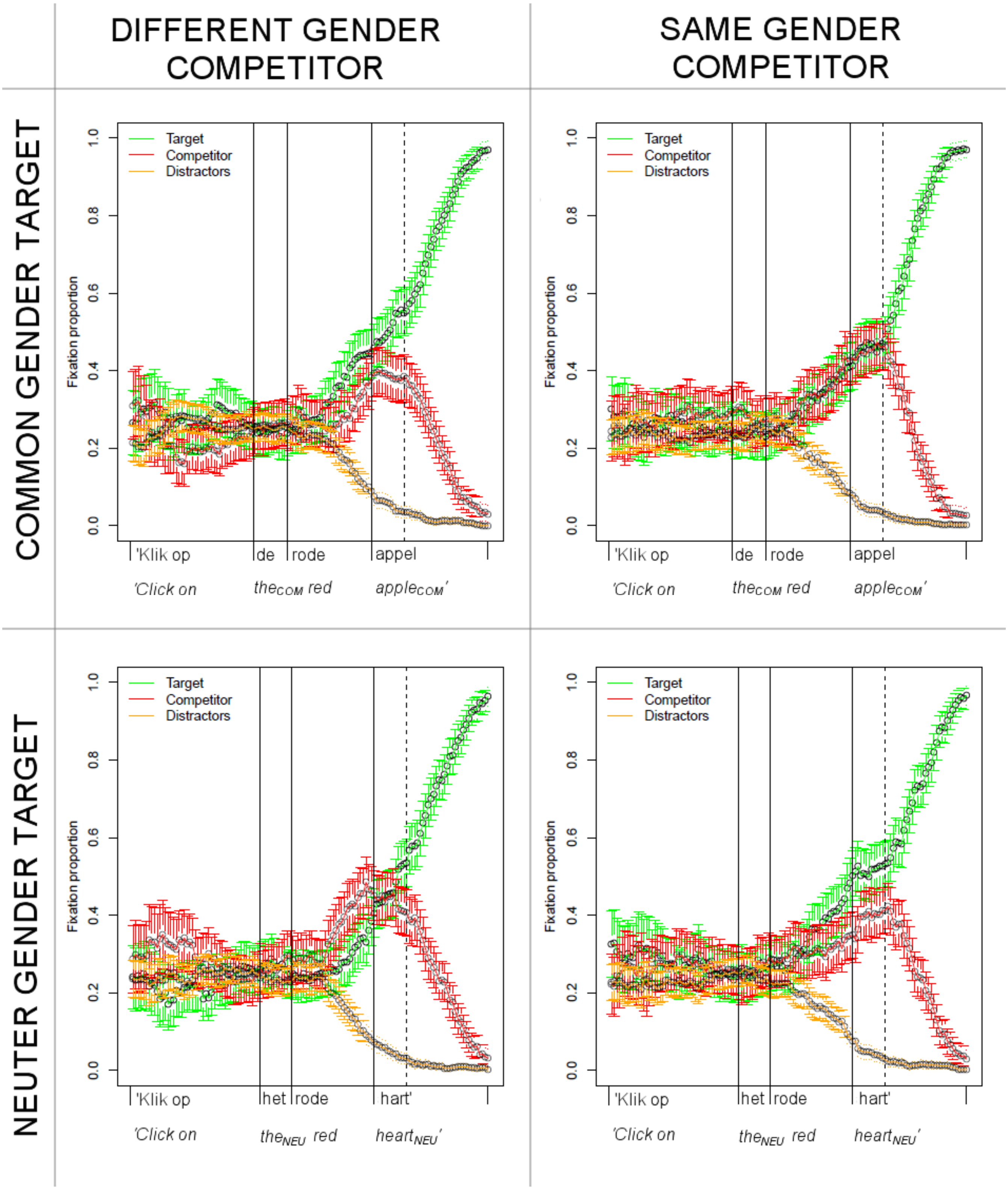


Figure 6-2: Proportions of looks within the native Dutch speakers. Adopted from Loerts, Wieling, and Schmid (Under Review). Proportions of fixations towards target (green), competitor (red), and distractors (yellow) are depicted only for the conditions in which target and competitor shared colour. Trials in which the competitor did not share gender with the target are depicted in the left panel and same-gender trials are depicted on the right. The above panel shows the fixation proportions in conditions where the target was common and the lower panel reveals eye movement data in response to neuter gender targets.

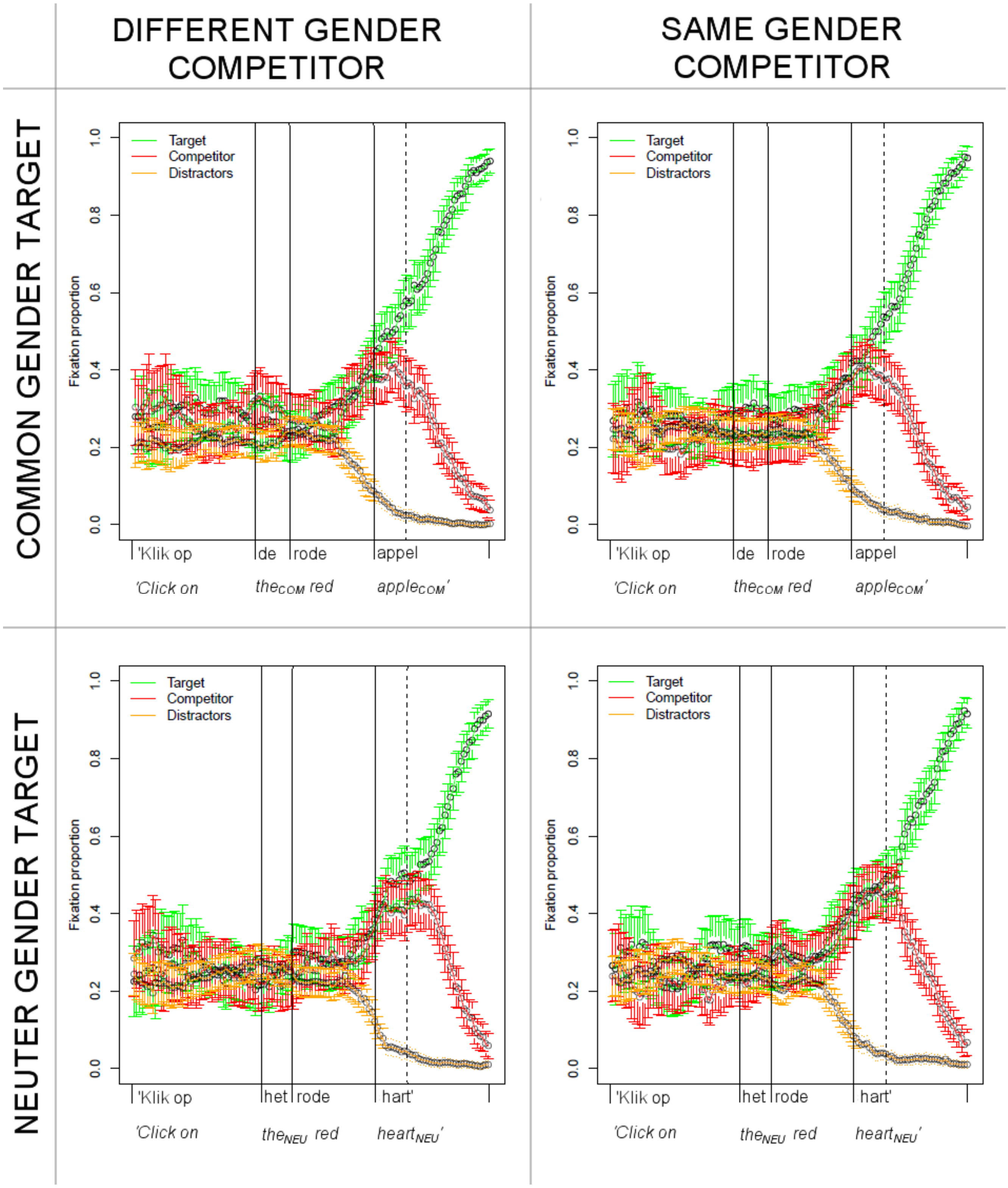


Figure 6-3: Proportions of looks within the group of late L2 learners towards the target (green), the competitor (red), and the distractors (yellow). Only conditions in which target and competitor shared colour are depicted. Trials in which the competitor did not share gender with the target are depicted in the left panel and same-gender trials are depicted on the right. The above panel shows the fixation proportions in conditions where the target was common and the lower panel reveals eye movement data in response to neuter gender targets.

As opposed to the response times, eye movements were significantly affected by the presence of a *SameGenderCompetitor* ($\beta = -0.0442$, $t = -2.23$, $p < 0.05$), as can be seen in Table 6-6. The negative coefficient reveals that if the target is accompanied by a competitor that shares gender with the target, participants looked relatively more towards the competitor as compared to trials containing different gender pairs. Additionally, when the *TargetIsNeuter*, people had significantly more difficulty in locating the target ($\beta = -0.0635$, $t = -2.91$, $p < 0.01$) and these two variables interacted showing that a *SameGenderCompetitor* is problematic when both target and competitor are common, but the gender effect disappeared and even reversed when target and competitor shared neuter gender ($\beta = -0.0442 + 0.0794 = 0.0352$, $t = 2.82$, $p < 0.01$). Separate models fitted to the data of the natives and L2 learners confirmed the presence of this interaction only in native speakers ($\beta = 0.0839$, $t = 2.67$, $p < 0.05$) and can also be seen in Figure 6-2.

Table 6-6: Fixed-effect factors in the best fitting mixed-effects regression model fit to the combined eye movement data of native and second-language speakers of Dutch. Estimated Coefficients, standard errors, and *t* values are outlined for the mixed-effects regression model fitted to the dependent measure *TargetRatio*. Increases and decreases in *TargetRatio* are reflected by positive and negative coefficients respectively. An increase reflects relatively more looks towards the target as compared to the competitor and a decrease reflects relatively more looks towards the competitor than the target. Factors are significant ($p < 0.05$) when they have an absolute value of the *t*-statistic exceeding 2 (Baayen et al., 2008).

	Estimate	***Std. Error***	***t value***
(Intercept)	-1.2608	0.3026	-4.17
SameColourCompetitor	-0.3258	0.0148	-21.96
TimeOnsetNoun	-0.0013	0.0002	-6.78
TargetIsUp	0.0664	0.0162	4.10
TargetIsBrown	-0.0574	0.0192	-2.99
TargetIsNeuter	-0.0635	0.0218	-2.91
Age	-0.0034	0.0011	-3.14
SameGenderCompetitor	-0.0442	0.0200	-2.23
IsLateLearner	-0.0374	0.0301	-1.24
TargetIsNeuterInPolish	0.0134	0.0180	0.75
SameGenderCompetitor: TargetIsNeuter	0.0794	0.0281	2.82
IsLateLearner: SameGenderCompetitor	0.0795	0.0290	2.74
IsLateLearner: TargetIsNeuter	0.0148	0.0292	0.48
IsLateLearner: TargetIsNeuterInPolish	-0.0812	0.0219	-3.71
IsLateLearner: SameGenderCompetitor: TargetIsNeuter	-0.1047	0.0413	-2.54

While the L2 learners did not differ significantly from native speakers overall (β = -0.0374, *t* = -1.24), they differed from natives in that they did not show an effect of *SameGenderCompetitor* (β = -0.0442+ 0.0795 = 0.0353, *t* = 2.74, *p* < 0.01) which was confirmed by the absence of an effect of *SameGenderCompetitor* in the best model fit to the L2 data only. They also did not show a differential effect of gender for common or neuter nouns, as revealed by the 3-way interaction at the bottom line in Table 6-6 and confirmed by the absence of this interaction in the separate L2 model. The absence of a significant interaction of *IsLateLearner* by *TargetIsNeuter* showed that L2 learners had, like native speakers, significantly more difficulty in locating the target when it was neuter as compared to when it was common. The slight difference in looking behaviour towards neuter targets that were accompanied either by a same gender or a different gender competitor, as visible in the lower panels of Figure 6-3, did not reach significance.

Interestingly, L2 learners showed difficulty in locating the target when the *TargetIsNeuterInPolish* (β = 0.0134 - 0.0812 = -0.0678, *t* = -3.71, *p* < 0.01), suggesting that the gender of the noun in the mother tongue does affect processing in the L2. This variable did not interact with factors related to the presence of a *SameGenderCompetitor* in Dutch or Polish nor did it interact with the factor of having a neuter Dutch target. These results thus suggest activation of the neuter category from the mother tongue that is unrelated to the gender of the noun in L2 Dutch and hence does not appear to result in transfer or competition effects.

Table 6-6 also clearly reveals the large impact of a *SameColourCompetitor* (β = -0.3258, *t* = -21.96, *p* < 0.001), reflecting that the presence of a competitor with the same colour as the target picture caused participants to fixate less on the target and relatively more on the competitor. In accordance with the RT data, the eye movement data showed the benefits of the target being coloured red. However, the negative impact of the *TargetIsBrown* was a significantly better predictor in the model fit to the eye data (β = -0.0574, *t* = -2.99, *p* < 0.01) and showed that people are less likely to fixate on the target when it is coloured brown. Similarly, while the mouse clicks occurred earlier when the target's position was on the right of the screen, looks towards the target occurred earlier when the *TargetIsUp,* i.e. placed in one of the upper quadrants of the screen (β = 0.0664, *t* = 4.10, *p* < 0.001). This effect most likely reflects participants' tendency to scan the screen from left to right and from the upper to the lower pane.

As was the case in the RT data, an increase in the *TimeOnsetNoun,* i.e. the time (in ms) at which the noun started in the speech stream, significantly reduced the *TargetRatio* (β = -0.0013, *t* = -6.78, *p* < 0.001) revealing relatively less looks towards the target when the noun onset started slightly later. As opposed to the RT data, *Age* significantly affected eye movements with older people taking significantly longer in choosing the correct target picture resulting in a lower *TargetRatio* (β = -0.0034, *t* = -3.14, *p* < 0.01). It should be noted that the models fit to the native and L2 learners' data separately showed that this age effect only reached significance in native speakers (β = -

0.0045, $t = -3.23$, $p < 0.01$), where the age range was relatively large. The age effect was not significant in L2 learners, but an interaction between these two variables failed to reach significance in the model fitted to the full dataset.

As opposed to the RT data, the eye movement data was not significantly affected by *Sequence* or *TargetRatio* scores for previous items, suggesting that eye movements are less susceptible to these effects of learning and/or fatigue. Additionally, no differences were found between the processing of definite and indefinite constructions in natives nor in L2 learners, suggesting that these constructions are processed similarly regardless of whether gender-marking occurs on the determiner or on the adjective.

Neither the full model nor the best model fit solely to the L2 learners' data revealed any effects of or interactions with age of acquisition, length of residence, or any of the proficiency scores outlined in Table 6-2. As mentioned, only items that had neuter gender in Polish negatively affected L2 learners' eye fixations towards the target, but no effects were found with respect to the presence or absence of a gender competitor in Polish.

The best fit model required both a random intercept for item (a specific picture-noun combination) as well as a random intercept for subject, as can be seen in Table 6-7. People tended to be differentially affected by the presence of a *SameColourCompetitor,* with some people suffering more than others from this predictor as was also visible in the reaction times. Whether the target appeared in the upper or in the lower panel of the screen (*TargetIsUp*) also impacted people's eye movements differentially and inclusion of by-subject random slopes for these effects significantly improved model fit.

Table 6-7: Random-effect parameters in the model fitted to the full eye movements' dataset containing both natives and L2 speakers.

Groups	***Name***	***Standard Deviation***	***Correlations with Intercept***	***Correlations between Slopes***
Item	Intercept	0.0311		
Subject	Intercept	0.0392		
Subject	SameColourCompetitor	0.0762		
Subject	TargetIsUp	0.0890		
Residual		0.3309		

6.4. Discussion

Previous eye tracking experiments have shown that, while second language learners are listening to their L2, they tend to activate words from their mother tongue that phonemically overlap with the words in the L2 (Marian & Spivey, 2003). Some studies have even shown that the gender of the words in the L1 might be transferred to the L2, at least when these words are cognates (Weber & Paris, 2004). The present investigation aimed at examining whether this also holds for non-cognates and for the acquisition of an extremely abstract gender system, i.e. Dutch, by late learners from a language that has a completely different gender system, i.e. Polish.

In Dutch, all nouns belong to either one of two categories: common gender (*de*-words) and neuter gender (*het*-words) and gender within a noun phrase is either marked on the article (definite NPs) or the adjective (indefinite NPs). Gender in Dutch is rarely predictable on the basis of phonological and/or morphological cues in the input (Haeseryn et al., 1997) and has an asymmetrical distribution with about 75% of all Dutch nouns having common gender (Van Berkum, 1996). Additionally, while all nouns receive neuter gender-marking when used as a diminutive, the agreement pattern used for all nouns in the plural is identical to the singular common agreement pattern. The present study investigated whether L2 learners, like native speakers, are able to use gender-marking in such an asymmetrical and non-transparent gender system to anticipate gender-congruent and/or inhibit gender-incongruent nouns. Moreover, the present study aimed at investigating the possible role of L1 transfer. To answer these questions, eye movements were monitored as participants followed spoken instructions to click on one of four displayed items on a screen (e.g. *Klik op de*$_{COM}$ *rode appel*$_{COM}$: 'Click on the$_{COM}$ red apple$_{COM}$'). The four items on the screen contained the target, a competitor that either matched in colour and/or gender with the target, and two unrelated distractors that never matched in colour nor gender with the target. To examine potential L1 transfer effects, half of the nouns used overlapped in gender with the equivalent Polish noun and the other half of the items consisted of nouns that did not overlap in gender between the two languages and, consequently, the Polish equivalent of the competitor either overlapped in gender with the Polish equivalent of the target or not. It was hypothesized that, if Polish gender of the referents is activated during spoken word comprehension in the L2, Polish-Dutch L2 learners should reveal difficulty in guiding their eyes towards the target when the competitor shares the target's gender in Polish as compared to when the competitor does not share gender with the target in Polish.

To test these hypotheses, mixed-effects models were fitted to both the (log-transformed) reaction time (RT) data and the relative difference in fixation proportions between target and competitor from determiner onset to noun onset, i.e. the time window in which both gender and colour information was presented. The results revealed a gender effect in the eye data for Dutch natives, with common nouns being fixated on relatively earlier if the gender of the target was unique. No such effect was found for neuter gendered items, probably due to the fact that neuter gender-marking does not disambiguate between common and neuter referents since all nouns can be preceded by neuter gender-marking when used in diminutive form. Late L2 learners of Dutch did not show a similar gender effect, revealing that they are not able to use common gender-marked articles or adjectives to facilitate language comprehension as native speakers do. The absence of any effects of NP construction showed that the native common gender effect is unaffected by locality and/or saliency and that natives can also use the less salient gender-marking occurring right before the onset of the noun, i.e. the presence of a schwa-suffix denoting common gender. L2 learners, however, are not able to use Dutch gender cues to facilitate spoken word comprehension regardless of whether the cues are more salient and allow for more processing time (definite determiner) or whether the cues entail less salient gender-marking on the adjective (as in indefinite NPs).

The group of late L2 learners do show a gender effect, however, in that they have significant difficulty in locating and choosing the correct target picture when this reflects a neuter noun as compared to when the target is common. This effect was also found in the native speakers' eye movement data and most likely reflects the non-disambiguating nature of neuter gender-marking as well as the less common nature of neuter nouns in Dutch as was also reflected in the extreme difficulty of L2 learners to use neuter gender-marking in production as well as their inability to recognize violations of neuter gender-marking. While late L2 learners of Dutch did not seem to transfer and use all gender categories from their mother tongue, as they did not suffer from a competitor that shared gender with the target in Polish, they did show significantly more difficulty in fixating on the target picture when its gender was neuter in Polish. Mostly, studies of co-activation of gender nodes from the L1 have shown gender congruency effects, such as faster responses when the nouns shared gender in the two languages and slower responses when gender differed between the two languages (Lemhöfer, Spalek, & Schriefers, 2008). The effect found in the present study does not entail a specific relationship between gender in the L1 and L2, but suggests a negative impact of a minority gender class in the native language: neuter gender.

Neuter gender in Polish is not only, as it is in Dutch, the smallest noun class, but Polish speakers have also been reported to be less productive with neuter gender, which can be mainly attributed to the extremely low frequency of neuter nouns as well as the more restricted phonological density of neuter nouns as compared to both masculine and feminine nouns in Polish (Dąbrowska, 2004). The fact that Polish L2 learners are slower to fixate on pictures whose referents are neuter in their mother tongue suggests co-activation of gender nodes within the two languages, even before hearing the noun in the speech stream. This activation, however, does not result in positive or negative transfer of (neuter) gender categories, possibly because the languages as well as their gender systems are relatively dissimilar. Future studies should disentangle this issue by looking at bilingual activation of gender categories across languages and the role of the status of the gender category (in this case neuter) in this process.

Although L2 learners showed no evidence of the use of L2 gender-marking, no main differences in the timing of processing were found, which is in conflict with previous research showing slower reaction times in L2 adults as opposed to natives (Lew-Williams & Fernald, 2010). The absence of a main difference between the two groups suggests no overall processing difficulty and confirms the relatively high proficiency of the L2 learners tested in the present study and/or reflects the relatively easy task that was used in the present experiment. Additionally, no effects of proficiency were found, suggesting that both lower and higher proficiency L2 learners show similar processing and similar co-activation of neuter gender while listening to L2 Dutch.

An effect of frequency of the items used in the experiment was found only in the reaction times, replicating previous findings that higher frequency items are more likely to be recognized (Dahan et al., 2001), and, as has been shown previously, this effect of frequency was larger for second language learners (Whitford & Titone, 2012). Similar frequency effects were not replicated in the eye data, which is not surprising when considering that we analysed eye movements in the time-window before noun onset. As expected, a large effect of colour was found in both the RT and eye data, showing that participants tend to wait until they had heard the (onset of the) adjective before choosing the correct target picture. The data also showed that, even if there was a disambiguating determiner present in the auditory input, people tended to wait and look at both the competitor and the target until they had heard the end of the adjective and/or the onset of the noun.

By using mixed-effects regression analyses, the present experiment allowed to include variables in the model that were initially thought of to be rather balanced, such as placement and colouring of the pictures as well as the specific timing in the auditory stimuli. Both the reaction times and the eye movement data showed that participants more easily guide their eyes towards red pictures while the colour brown does not seem to yield as much attention. Similarly, when the competitor has a salient colour, such as red, this disturbs responses towards the comparatively less salient target. These results show the importance of colour in capturing a person's attention due to their saliency (Osberger & Rohaly, 2001). Moreover, participants showed relatively more fixations towards targets that appeared in one of the upper quadrants of the screen, possibly reflecting their tendency to scan from top to bottom and from left to right. Reaction times were additionally negatively affected by a left-hand position of the target on the screen, which is likely to reflect the relative ease of right-handed participants to move their right hand from left to right as compared to from right to left. It is difficult to explain why late L2 learners did not show the same tendency, but this might be explained by their relatively higher concentration rate. This, however, remains speculative and the true nature of the (absence) of the effect cannot be answered on the basis of the present data alone.

Minor differences in the onset of the adjective and noun in the speech stream also significantly affected responses, with later onsets corresponding to less information and hence a delay in reaction times as well as fixations towards target pictures. An overall effect of age, i.e. significantly later responses for older participants, only reached significance in the eye movement data. In addition, people showed an overall learning effect with relatively faster responses as the experiment progressed as well as an influence of the previous reaction time on the current one. The fact that these latter effects only significantly impacted the reaction time data suggests that effects of learning can be obtained in controlled manual responses such as mouse clicks on the correct target picture. Eye movements, however, cannot easily be adapted by the individual since they are largely automatic and unconscious in nature. This, together with the fact that a native gender effect was only found in the eye movement data, but not in reaction time data, again confirms the advantages of studying language processing by investigating eye movements in addition to reaction times.

By analysing both RTs and eye movement data using mixed-effects modelling, the present experiment could assess the impact and importance of multiple predictors that could not have been added in an ANOVA analysis and the inclusion of random intercepts and slopes for both subjects and items allow the researcher to acknowledge that people respond differently to certain variables and that different items, no matter how balanced the design, can evoke different responses.

In sum, the present experiment reveals that native speakers use common, but not neuter gender-marking to facilitate language comprehension, agreeing with the fact that neuter gender- marking can be used for all nouns when used as diminutives and hence it has no predictive value. Difficulty in gender processing in L2 learners was reflected by the inability to use Dutch gender-marked articles and/or adjectives to facilitate processing of the subsequent noun. Considering the relatively high proficiency of the late L2 learners tested here, we can conclude that, even for advanced learners of Dutch, grammatical gender remains a difficult feature to acquire and use during spoken word recognition. Especially the processing of less frequently used and non-disambiguating neuter gender-marking cues remains problematic. The absence of L1 transfer effects suggest that, at least for non-cognates, learners do not use gender information from their mother tongue when comprehending Dutch sentences if the L2 is relatively different from the L1. They do, however, have severe difficulty with nouns that have the less frequent neuter gender in Polish, suggesting that L1 gender is activated, but not transferred during language comprehension in the L2.

Chapter 7

ENOUGH TIME FOR GENDER

Morphological Processing in Late Second Language Learners of Dutch

The present study compared Dutch natives and late Polish-Dutch L2 learners on the processing of structures that are either frequently, noticeably and reliably present in the input or not: finiteness and grammatical gender. Event-related potentials (ERPs) were recorded while participants listened to correct sentences and sentences containing morphological violations. Finiteness violations elicited an N400 – P600 pattern in natives and a delayed P600 response in L2 learners. Grammatical gender violations elicited a P600 in natives, but a delayed and reduced P600 in L2 learners was only found in highly proficient learners and only in response to violations on nouns carrying the default common gender. The availability and reliability of morphological structures in the input thus plays an important role in stages of lower proficiency. After enough exposure and with increasing levels of proficiency, however, L2 learners can process both finiteness and grammatical gender in a more native-like fashion.

This chapter has been submitted for publication: Loerts, H., Stowe, L.A., & Schmid, M.S. (Under Review). Enough Time for Gender: Morphological Processing in Late Second Language Learners of Dutch. Submitted (July 27, 2012) to *Bilingualism: Language and Cognition.*

7.1. Introduction

There is general consensus that acquiring a second language (L2) after puberty can be difficult, and mastering the L2 up to native-like level often seems almost impossible. This appears to be especially true for the acquisition of grammatical gender, a noun classification system that divides nouns into different categories, such as feminine and masculine. Languages differ as to whether they assign such nominal categories, but even if they do, there is usually little overlap. Gendered languages not only differ with respect to what gender categories they include, but also with respect to the frequency, saliency, and reliability of the gender-marking cues in the input. The Polish language, for example, has a complex gender system consisting of masculine, feminine, and neuter nouns. Determiners, numerals, adjectives, pronouns, and even verbs have to agree in gender with the noun. Dutch, on the other hand, has an asymmetrical two-way gender system consisting of the default common gender (which historically evolved as a conflation of the earlier masculine/feminine categories) and the less frequent neuter gender. Gender agreement occurs only on determiners, pronouns, and adjectives, and some other elements such as quantifiers. Additionally, while the gender of most Polish nouns can be determined from morphophonological features on the noun itself, the gender of Dutch nouns can rarely be predicted on the basis of morphophonological characteristics of the noun (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997).

The difficulty in the acquisition of a second language, and particularly an L2 gender system, has increasingly attracted interest among scientists. Although most researchers agree that L2 acquisition is different from L1 acquisition, they do not agree on the causes and manifestations of this difference. The debate centres on whether only features that are present in the L1 are available to post-puberty learners, whether the native language has an influence on the L2 acquisition process, and what the role of L2 proficiency is in these processes. The present paper aims at investigating whether the difficulty in acquiring the Dutch gender system at a later age might also be related to the frequency and reliability of the gender-marking cues in the input.

Within the generative framework, the difference between L1 and L2 acquisition is often linked to a sensitive period for language acquisition, after which native-like processing becomes impossible due to the unavailability of Universal Grammar (UG), the set of syntactic rules that are shared among the different languages in the world (Chomsky, 1995). One version of this approach, the Failed Functional Features Hypothesis (FFFH: Hawkins & Chan, 1997), states that the difficulty for late L2 learners in acquiring certain grammatical features, such as gender, is caused by the inability of post-puberty learners to acquire functional features that are not represented in their L1. More specifically, late L2 learners are able to acquire interpretable features such as new isolated words, but uninterpretable features related to functional categories such as agreement can only be acquired if instantiated by the mother tongue. A more process-oriented view on the effect of age has been proposed by Clahsen & Felser

(2006a), who claim that it is not the specifics of the L1 that hampers L2 acquisition. Instead, L2 learners are engaged in less detailed syntactic processing as compared to natives and mainly rely on associative and lexico-semantic information rather than syntax while processing the L2. As a consequence, late L2 learners are able to process semantics and morphology in a native-like manner, but not syntax.

Opposing the view that post-puberty L2 learners cannot achieve native-like processing of syntactic features in general or features that are not present in the L1 in particular, are scientists who believe that learners can acquire grammatical gender in their L2 regardless of whether this feature is present in their L1 (e.g. Bruhn de Garavito & White, 2000). The Full-Transfer Full-Access model (FTFA) assumes, like the FFFH, that the initial representation of grammatical features is based on features that are present in the L1. It is argued, however, that L2 learners can become native-like regardless of their age of acquisition as they can still access the parameters provided by UG and change the L1 parameter settings, if necessary, as they grow more proficient.

7.1.1. The difficulty of the Dutch gender system

The role of L1 transfer and L2 proficiency in late L2 acquisition is especially interesting when looking at grammatical gender, which has repeatedly been shown to be one of the most difficult aspects to master when acquiring a new language. In Dutch, the difficulty at least partly has to do with the non-transparency of the system. As can be seen in Table 7-1, Dutch nouns are either common or neuter, but bear no overt marking of their class such as the phonological '*–a*' and '*-o*' endings in Spanish mostly (though not exclusively) denoting feminine and masculine gender respectively.

Table 7-1: Overview of the main agreement targets in Dutch. In Dutch, gender-marking within the noun phrase occurs either on the determiner (definite NPs) or on the adjective (indefinite NPs). English equivalents of the phrases are printed below each phrase in italic. The indefinite article is always *'een'* in the single case and no equivalent indefinite article exists for the plural case.

	Gender	Definite Article	Adjectives in Definite NPs	Adjectives in Indefinite NPs
Single	common	***de*** hond	***de*** mooie hond	een ***mooie*** hond
	neuter	***het*** paard	***het*** mooie paard	een ***mooi*** paard
English equivalent		*the dog/horse*	*the beautiful dog/horse*	*a beautiful dog/horse*
Plural	common	de honden	de mooie honden	--
	neuter	de paarden	de mooie paarden	--
English equivalent		*the dogs/horses*	*the beautiful dogs/horses*	--

Other languages contain more subtle morphological cues that help determine the gender of a noun, such as the suffix *'-ion'* generally denoting feminine gender in French (Foucart, 2008). In a covert and non-salient system such as the Dutch one, the gender-marking that appears on determiners, attributive adjectives, and relative and demonstrative pronouns is the only clear indication of the presence of gender. Additionally, the majority of all Dutch nouns (70%) have common gender as a consequence of the collapse of the original feminine and masculine genders (Van Berkum, 1996). This asymmetrical distribution makes it difficult even for young native children to acquire, who tend to overgeneralize the common determiner *'de'* until the age of 6 (Van der Velde, 2004), while children learning German and French as mother tongues have been reported to show error-free gender use at the ages of 3 and 4 (Szagun, Stumper, Sondag, & Franik, 2007; Van der Velde, 2004). The default status of Dutch common gender is further reinforced by the fact that determiners and adjective agreement in plural noun phrases (NPs), regardless of gender, is identical to the common singular. On the other hand, all Dutch nouns become neuter when used as diminutives. Common gender can thus be seen as the default and is, consequently, often overgeneralized by learners of the Dutch language (Unsworth, 2008). The absence of systematic and salient cues in the input, specifically when referring to neuter nouns, makes Dutch gender difficult and it might well occur on an item-by-item basis that is likely to be influenced by input quantity and hence length of exposure (Gathercole & Thomas, 2005; Unsworth, 2008).

The importance of input frequency and quality is emphasized by the Competition Model (Bates & MacWhinney, 1987) which states that language acquisition occurs on the basis of lexical, morphological, and syntactic cues in the input. Consequently, the frequency as well as the reliability of a specific cue, i.e. how often it points towards a particular interpretation, is crucial in learning language. Cues for gender agreement in Dutch are highly available, especially in the form of determiner-noun combinations. The cues, however, are not reliable in that the schwa-ending on the adjective (as in 'mooi*e*') can precede both common and neuter nouns (also see Table 1). Consequently, frequent and reliable morphological structures in the input such as a tensed auxiliary followed by a past participle or a modal auxiliary plus an infinitive emerge well before grammatical gender around the age of 2;5 years in Dutch L1 acquisition (De Houwer & Gillis, 1998). Similarly, the relatively higher availability and strength of common-gender marking as opposed to neuter gender-marking causes overgeneralizations of common gender-marking in both native children and L2 learners. For (adult) L2 learners, similarity between the native and the second language is thought to play an additionally crucial role, especially in initial stages of L2 processing (MacWhinney, 2008). According to the Competition Model, both features that are similar in the two languages as well as structures that are unique to the L2 are easy to acquire because there is no competition from the L1. Structures that are not similar in the two languages, on the other hand, may cause difficulties in processing due to competition from the native language.

Questions concerning the degree of native-like processing in late L2 learners and the influence of L1 transfer and proficiency level have increasingly been examined by measuring event-related brain potentials (ERPs), which have proven to be a useful tool in comparing online processes in natives and L2 learners during language comprehension. ERPs are changes in voltage over time representing the level of activation in the brain in response to specific (linguistic) stimuli. They allow us to investigate the rapid online processes involving language comprehension by providing a detailed timeline of these processes. In other words, they provide a useful tool in assessing *when* particular aspects of language are processed. Moreover, ERPs seem to be differentially sensitive to aspects of language comprehension, such as semantics and syntax. For example, the N400, a centro-parietal negative-going wave between 300 and 500 ms post onset, has been found to be modulated predominantly by semantic factors. The P600, a long-lasting positivity starting around 500 ms post stimulus (Osterhout and Holcomb, 1992), has mostly been found in response to morphological and syntactic violations and complexities, including violations of grammatical gender (Sabourin & Stowe, 2008; Foucart & Frenck-Mestre, 2012). The P600 is thought to indicate controlled processes of repair and/or re-analysis of the grammatical violation (Friederici, 2002). In some studies, the P600 is preceded by a LAN, a left anterior negativity occurring between 300 to 500 ms post-stimulus, which is assumed to indicate a more automatic detection process prior to the repair process reflected in the P600 (e.g. Gunter, Stowe, & Mulder, 1997). Additionally, an even earlier LAN, the ELAN, has previously been reported to precede the P600 in response to word-category violations, such as '**Die Bluse wurde am gebugelt'* (The blouse was on-the ironed) (Friederici, Pfeifer, & Hahne, 1993).

Overall, studies have shown native-like processing of semantic aspects for advanced late L2 learners, while grammatical processing appears to be problematic (for recent reviews see Steinhauer, White, and Drury (2009) and Van Hell & Tokowicz (2010). The discussion of whether late L2 learners can process an L2 in a qualitatively similar manner as natives has thus far yielded inconclusive results. Researchers do not agree on whether the native-likeness of L2 processing depends mostly on the similarity between L1 and L2 grammatical features (Sabourin & Stowe, 2008), the amount of second language experience (Foucart & Frenck-Mestre, 2012), and/or competition between the native and second language (Tokowicz & MacWhinney, 2005).

7.1.2. Competition, transfer, and L2 experience

The possible influence of L1 syntactic similarity to the L2 was investigated by Tokowicz and MacWhinney (2005) in a study of ERPs of native English learners of L2 Spanish in response to syntactic structures that were similar in both languages (auxiliary omission), different in the two languages (number agreement on the determiner), or unique to the L2 (gender agreement on the determiner). Despite the lower proficiency of the subjects, the authors reported a significant P600 for the constructions that were either similar or unique to the L2. The authors concluded that, in line with the predictions made by the Competition Model and against the predictions made by the FFFH, the occurrence of implicit syntactic processing in the L2 depends on the similarity between the native and the second language. Features that are similar in the two languages or unique to the L2 are easier to acquire due to the absence of competition between the two languages. Structures that are not similar in the two languages, however, may block appropriate processing due to competition from the native language. It should be noted that Tokowicz and MacWhinney (2005) did not test a native control group. They were thus not able to directly compare processing in natives to second language learners.

Contradicting evidence has been presented by Sabourin and Stowe (2008), who used grammatical gender violations and violations of finiteness to focus on the effects of similarity between languages. To assess the impact of the first language, Sabourin and Stowe (2008) compared Dutch natives and late L2 learners with different L1 backgrounds: English, German, and a combination of Romance languages (French, Italian, and Spanish). With respect to finiteness violations consisting of past participles replaced by infinitives and vice versa, a familiar construction for all L2 learners, a P600 was present in all populations. This indicates that L2 processing can be native-like for constructions that are similar between L1 and L2. A second experiment focussed on the question of whether the presence of gender systems in the L1 is sufficient to allow native-like processing. As opposed to English, both German and Romance languages have gender, but the similarity between these and the Dutch gender system differs. German nouns are masculine, feminine, or neuter, and the categories often overlap with the Dutch common and neuter categories. Romance L2 learners, on the other hand, were only familiar with having two genders (masculine and feminine), but they could not rely on any direct mapping from the L1 onto the L2. As hypothesized, only German learners of Dutch showed a clear P600 in response to gender violations, suggesting that it is not sufficient to have gender categories in the L1, the systems additionally need to be similar, possibly at the lexical level.

The additional difficulty in acquiring and processing Dutch gender as opposed to finiteness constructions might, however, also be affected by the frequency and reliability of the structures in the input. Significant effects of input frequency were also reported in French by Foucart and Frenck-Mestre (2012). They compared grammatical gender processing in French native speakers to L2 learners of French with L1 English. Opposing the results reported by Sabourin and Stowe (2008) and in line with the results reported by Tokowicz and MacWhinney (2005), the authors reported that English-French bilinguals are sensitive to gender agreement violations as revealed by a P600 effect. Interestingly, this P600 appeared to be affected by the frequency of the structures in the input. Violations between the noun and a post-posed adjective (the canonical structure in French) elicited a native-like P600 effect. In contrast, violations between the noun and a pre-posed adjective (a less frequent construction) triggered a P600 in French speakers but an N400 in L2 learners. In line with previous research, the authors suggested this N400 to reflect initial stages of syntactic processing following stages in which L2 learners do not show neural sensitivity in response to morphological violations and preceding the stage in which L2 learners show native-like sensitivity as reflected by a (delayed and/or reduced) P600 effect (Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004).

These contrasting results might be related to the absence of clear proficiency measures in most of the ERP studies mentioned above. As noted by Clahsen and Felser (2006a), the German-speaking learners in Sabourin and Stowe's (2008) study might have been the only group showing a P600 in response to Dutch gender violations due to their higher level of proficiency (as revealed by the grammaticality judgement task) instead of the similarity between Dutch and German gender. Additionally, while the English and Romance learners of Dutch did not reveal neural sensitivity to gender violations, they largely did perform above chance in judging the violations and showed accuracy in assigning gender to nouns in an offline task (i.e. choosing the correct determiner *'de'* or *'het'*), suggesting that offline gender knowledge cannot predict online processing of gender agreement violations. On the contrary, the low proficient L2 learners of Spanish tested by Tokowicz and MacWhinney (2005) did show a P600 in response to gender violations despite the absence of a gender system in their L1, but failed to perform above chance on the grammaticality judgement task. One proficiency measure may thus not provide trustworthy results and different task demands might not only reflect language proficiency, but also an overall proficiency of the learner in coping with the specific task. Therefore, the present study used a composite measure of L2 proficiency.

The fact that the English learners tested by Tokowicz and MacWhinney (2005) and Foucart and Frenck-Mestre (2012) showed a P600 in response to a feature that is not present in their L1 is in clear contrast with Sabourin and Stowe's (2008) results showing only sensitivity to L2 Dutch gender violations by L2 learners with a similar gender system in their mother tongue. Another possible explanation for these differences might be related to differences in the saliency and transparency of the gender systems under investigation. L2 learners might be able to acquire and process L2 gender systems in which the presence of phonological (as in Spanish) or morphological (as in French) cues help determine the gender of nouns. An asymmetrical and non-transparent gender system such as the Dutch system might, on the other hand, be more difficult to acquire and hence require more language experience and a higher level of L2 proficiency.

Similarly, the fact that all learners tested by Sabourin and Stowe (2008) could process finiteness and only German L2 learners could process gender in Dutch was hypothesized to reflect L1 transfer, which is possible for finiteness in all groups but for gender only in the German group. Considering the predictions made by the Competition Model and the effect of input frequency reported by Foucart & Frenck-Mestre (2012), however, the results might also be attributed to the fact that finiteness constructions are more salient and reliably present in the input and therefore acquired prior to the Dutch gender system, which is non-transparent and non-salient and has an asymmetrical distribution with common gender-marking being more frequently and more reliably present in the input.

7.1.3. The present experiment

The present study aimed at investigating the influence of L2 proficiency on the one hand and frequency and reliability of morphological input cues in the L2 on the other hand. Event-related potentials (ERPs) were recorded while natives and late advanced L2 learners of Dutch (age of acquisition (AoA) > 20) listened to sentences that were either correct or that contained morphological violations. L2 learners differed with respect to their level of proficiency (high or low) as assessed by means of a composite measure, but age of acquisition did not differ between the low and high proficiency learners. Morphological violations consisted either of violations of finiteness or violations of gender agreement.

Finiteness

Finiteness violations consisted of tense violations containing past participles that were replaced by infinitives or vice versa. These sentences have previously been used by Sabourin & Stowe (2008) and have shown a reliable P600 in L2 learners from languages that are relatively similar to Dutch, such as German, English, and Romance languages. The present study aimed at investigating whether these constructions can also be easily acquired by learners from a language that is relatively different from Dutch, i.e. Polish. In Dutch, the perfect tense is formed by using (a tensed form of) the auxiliary verbs *'hebben'* (to have) or *'zijn'* (to be) followed by a past participle form of the verb, which is usually (though not always) recognizable by the prefix *'ge-'* as in *'geschreven'* (written) in example sentence 7-1a. Modal auxiliaries such as *'mogen'* (to be allowed to), *'moeten'* (to have to), and *'willen'* (to want to), on the other hand, have to be followed by the infinitive form of the verb (example sentence 7-2a). In Polish, verbs often occur in pairs consisting of an imperfective and a perfective variant that is usually derived by prefixation; for example, the verb *write* having the imperfective form *'pisać'* and the perfective form *'napisać'*. As can be seen in sentence 7-1b, the prefix *'na-'* denoting a perfect tense is similar to the Dutch prefix *'ge-'*. As opposed to Dutch, the auxiliary in Polish has been morphologically impoverished and only still exists in the forms of suffixes on the participle (Migdalski, 2006). Modals in Polish are, on the other hand, always followed by the infinitive of the main verb, as in example 7-2b, This construction is thus very similar in the two languages. Violations of these finiteness sentences, adopted from Sabourin & Stowe (2008), were created by replacing the finite verbs by past participles or vice versa.

7-1a. Jan heeft geschreven
Jan have$_{3SG}$ write$_{PART}$

7-1b. Jan napisał
Jan write$_{PART.M.SG}$

7-2a. Jan wil schrijven
Jan want$_{3SG}$ write$_{INF}$

7-2b. Jan chce pisać
Jan want$_{3SG}$ write$_{INF.IMPF}$

Grammatical Gender

Since all populations tested by Sabourin & Stowe (2008) showed a P600 in response to finiteness violations and only German learners of Dutch showed a P600 in response to gender violations, the authors hypothesized L2 gender acquisition to be dependent on the similarity between the L1 and L2 gender system. The results, however, might additionally also reflect differences in the frequency and reliability of the structures in the input, causing finiteness to be acquired prior to grammatical gender. To test this hypothesis, the processing of finiteness violations was compared to the processing of grammatical gender. Like Dutch, Polish contains gender, but the system in Polish is quite complex because it combines three categories: gender (masculine, feminine and neuter), personhood (personal and non-personal), and animacy (animate and inanimate). Apart from phonological and morphological cues on the noun itself, gender agreement with the noun is marked on demonstrative pronouns, numerals, adjectives, and verbs. In Dutch, gender within the NP occurs either on the determiner, as in definite NPs, or on the adjective, as in indefinite NPs (also see Table 7-2). An overview of the differences and similarities between the Dutch gender constructions used in the present experiment and their Polish equivalents are outlined in Table 7-2. For illustrative purposes, nouns with semantic gender were chosen that overlap in gender in Dutch and Polish.

As outlined in Table 7-2, gender in Polish is overt and can be extracted from the phonological and morphological characteristics of the noun itself, i.e. masculine nouns generally end in consonants, feminine nouns in *'–a'*, and neuter nouns generally end in *'–o'* or *'–e'*. As opposed to Dutch, gender agreement is always present on the adjective and, due to the absence of determiners in Polish, no distinction is made between definite or indefinite NPs. It should be noted, however, that nouns can be preceded by other determiners or numerals that must agree in case, number and gender (Perlak & Jarema, 2003). Definiteness, on the other hand, is interpreted on the basis of context in Polish. The examples in Table 7-2 illustrate that, in addition to the lack of overlap between gender categories in Dutch and Polish, gender agreement is less salient in Dutch as compared to Polish. While gender is present in Polish, definite and indefinite determiners are unique to the L2. Adjectival inflection, on the other hand, is relatively similar in Dutch and Polish. Hence, a comparison of the processing of these constructions may provide insight into the specific effects that different gradations of similarity between the mother tongue and the L2 have on potentially native-like processing of morphological information in the L2.

Table 7-2: Overview of gender agreement realization in Dutch and Polish for the constructions used in the present experiment. For illustrative purposes, nouns with semantic gender are outlined that overlap in gender category in Dutch and Polish.

Construction	**Common**		**Neuter**
	Masculine	**Feminine**	
Definite NP	***de*** jongen		***het*** kind
	chłopie***c***	dziewczyn***a***	dzieck***o***
	the boy	*the girl*	*the child*
Definite NP with adjective	***de*** goed***e*** jongen		***het*** goed***e*** kind
	dobr***y*** chłopie***c***	dobr***a*** dziewczyn***a***	dobr***e*** dzieck***o***
	the good boy	*the good girl*	*the good child*
Indefinite NP with adjective	een goed***e*** jongen		een goed kind
	dobr***y*** chłopie***c***	dobr***a*** dziewczyn***a***	dobr***e*** dzieck***o***
	a good boy	*a good girl*	*a good child*

Hypotheses

We have shown that native Dutch speakers show a clear P600 in response to both finiteness and gender violations (Loerts, Stowe, & Schmid, Under Review). The presence of a, delayed and/or reduced, P600 in L2 learners would reveal that they initiate processes of repair/re-analysis. If post-puberty learners cannot establish native-like representations and/or processing in the L2, the ERP data should reveal qualitatively different processes as reflected by the absence of any effects in response to morphological violations. If, however, proficiency is the most important factor distinguishing native and L2 morphological and syntactic processing, we would expect to find native-like P600 effects only in higher proficiency L2 learners and no P600 or possibly an N400 in low proficiency L2 learners reflecting initial stages of syntactic processing (Osterhout et al., 2004). Linguistic similarity might also be a decisive factor in L2 processing (Sabourin & Stowe, 2008), which would result in differences between the processing of tensed auxiliaries plus past participles (different in the two languages) and modal verbs plus infinitives (similar in the two languages) and possibly also differences in the processing of gender-marking on determiners (unique to the L2) and gender-marking on adjectives (similar in the two languages). Differences in the processing of finiteness and gender violations, on the other hand, cannot entirely be attributed to L1 transfer in the present study, because both structures are present, and to a certain degree also similar in L1 Polish. These differences are thus more likely to reflect differences in the input frequency and reliability of the gender-marking cues. If input frequency is indeed crucial, differences in the acquisition and processing of more frequent and salient common gender-marking as opposed to neuter gender-marking would also be hypothesized as well as a positive influence of length of residence.

7.2. Methods

7.2.1. Participants

Participants in the present study were 28 native Dutch speaking adults (16 female, aged 34-60, mean age=48) and 26 late advanced learners of Dutch with L1 Polish (20 female, aged 24-52, mean age=34). The late L2 learners had all acquired L2 Dutch after the age of 20 (range: 20-34) and the average length of exposure (LoR) to the L2 was 9 years (range 2-27). Participants were recruited from a non-university population by means of advertisements in the local newspaper, the online Polish newspaper ('Polonia'), and social media. All participants were healthy, right-handed (assessed according to Oldfield, 1971), had normal hearing and none suffered from neurological problems and/or language disorders. They were paid for their participation.

L2 proficiency of the Polish-Dutch learners was assessed through a combination of measures that have previously successfully been used to assess proficiency, as outlined below.

1. A C-Test to assess general proficiency, in which participants were asked to fill in the gaps in two texts of about 60 to 70 words each. The first sentence of each text was complete to provide context and the rest of the text contained approximately 20 word ending gaps. Each answer was rated using the scoring system proposed by Keijzer (2007) and Schmid (2005), which takes into account the correctness of the lexical items used, agreement errors, and spelling errors;
2. A self-assessed proficiency measure was based on the scores on a 5-point Likert scale on several items reflecting Dutch listening, reading, speaking, and writing;
3. A film retelling task was used to assess oral proficiency in the L2 (see Perdue, 1993). Participants watched an excerpt of about 10 minutes from the silent 1936 Charlie Chaplin film *Modern Times* and were asked to re-tell the story they had been watching. The semi-spontaneous speech data obtained was rated by three native Dutch linguists on fluency, pronunciation, intonation, syntax, and lexicon on a scale from 1 (very basic) to 6 (native-like) to establish a holistic (oral) proficiency score (interrater reliability across domains was excellent, with all *r*s > .836). Furthermore, all NP constructions that had to be preceded either by a gender-marked determiner or by a gender-marked adjective were counted and classified as being either correct or incorrect. Due to the asymmetry in gender classes and the potential reflection of this in the acquisition of common and neuter gender (see, e.g. Unsworth & Hulk, 2010), a gender production accuracy score was provided for common and neuter nouns separately.
4. The accuracy scores on the acceptability judgement task, as assessed during the ERP experiment, were also taken into account in the general proficiency score. During the ERP experiment (see below), participants listened to correct sentences and sentences containing violations of finiteness or violations of gender concord. After each sentence, participants were asked to judge the overall acceptability of the sentence

they had heard. Again, due to possible asymmetries in the comprehension of finiteness and gender on the one hand and common and neuter gender on the other hand, separate accuracy scores were obtained for these three different constructions.

Based on the overall scores in percentages of the above-mentioned measures, participants were divided into two groups of equal size, which were then classified as high or low proficiency L2 learners. In addition, after completion of the ERP experiment, participants' explicit knowledge of the Dutch gender system was assessed by asking them to circle the article (i.e. *'de'* or *'het'*) belonging to the nouns that had been presented to them during the experiment. Each noun occurred three times in this gender assignment task to avoid performance at chance level. These scores were not used in the overall proficiency score, but served to select only those ERP epochs for analyses that corresponded to nouns of which the L2 learners knew the gender.

Level of education, age at time of testing, and age of acquisition did not differ significantly between high and low proficiency learners (also see Table 7-3). Wilcoxon rank-sum tests with Bonferroni correction showed that both the overall proficiency scores as well as the separate proficiency measures (cloze test scores, self-rating scores, scores on the free speech data from the film retelling task, an acceptability judgement task, and a gender assignment task) for the high and low proficiency learners were statistically independent (all ps < 0.05). In addition, high proficiency learners tended to have lived in the Netherlands longer as compared to low proficiency learners ($p < 0.05$).

The proficiency overview in Table 7-3 also clearly reveals differences between performance on common and neuter nouns. All groups, including native speakers, show more difficulty with neuter gender and the use of correct neuter gender-marking in free speech as well as the ability to assign gender to neuter nouns appears to be especially problematic. The scores on the gender assignment task also reveal, however, that most of the high proficiency L2 learners are able to assign the correct gender to neuter nouns while some appear to be unable to assign gender to common nouns. This discrepancy between the acquisition and use of common and neuter gender was confirmed by the lack of correlations between proficiency scores on common and neuter nouns in free speech and in the offline gender assignment task. Scores on the acceptability judgement task, however, did correlate positively with each other (Spearman's $r = 0.63$, $p < 0.01$). As can be seen in Table 7-3, there is substantial overlap between the scores on the several tasks in lower and higher proficiency L2 learners. Detailed inspection of the data revealed that some participants tend to score extremely high on the C-Test, while scoring rather low on gender assignment and production. Other participants, on the other hand, scored within native ranges on tasks related to gender while their C-Test and holistic proficiency score was rather low. The division of low and high proficiency learners was therefore based on a combination of these scores to minimize the effects of variance on the different tasks. The continuous overall proficiency scores of the L2 learners were used in correlational analyses to assess their individual impact on the amplitude of the potentially present P600 effect and overcome potential problems caused by averaging over low and high proficiency L2 learners respectively.

Table 7-3: Scores on several proficiency measures that were used to calculate an overall proficiency score. The table provides the scores and ranges on the different measures of proficiency that were used to assess an overall level of proficiency of the L2 learners. The overall proficiency level is printed in bold and the scores on the off-line gender assignment task were not added to the overall proficiency score, but used to analyse only those nouns of which the L2 learners knew the gender. As can be seen from the ranges of the scores in the high and low proficiency groups, there is substantial overlap between these two groups. Due to the variability of participant's scores on the several tasks, however, the use of one of these scores as a measure of proficiency would not provide reliable results and hence, a composite measure of L2 proficiency was used in the present study.

Group				**General Proficiency Measures**		**Data from free speech**			**Acceptability Judgements Task (% correct)**			**Offline Gender Assignment (% correct)**	
	AoA (years)	**LoR** (years)	**Overall Proficien-cy**	**C-Test** (%)	**Self-Rating** (%)	**Holistic Proficien-cy** (%)	**Common gender-marking** (%)	**Neuter gender-marking** (%)	**Finiteness**	**Common**	**Neuter**	**Common**	**Neuter**
Natives *(range)*	x	x	**94*** ***(78-98)***	97.04* *(90-100)*	97.57* *(84-100)*	94.03* *(91-99)*	97.74* *(85-100)*	93.68* *(79-100)*	96.47* *(81-100)*	91.28* *(67-100)*	89.41* *(52-100)*	x	x
HP learners *(range)*	23.8 *(20-29)*	12† *(3-27)*	**75*†** ***(65-88)***	91.97*† *(81-97)*	87.38*† *(75-100)*	70.51*† *(60-89)*	84.90*† *(63-100)*	62.42*† *(29-100)*	79.76*† *(62-95)*	61.85*† *(47-76)*	61.68*† *(49-82)*	84.90† *(23-100)*	66.39† *(41-85)*
LP learners *(range)*	26.3 *(20-34)*	6† *(2-15)*	**56*†** ***(49-63)***	66.71*† *(25-94)*	67.31*† *(40-84)*	51.88*† *(36-86)*	71.02*† *(36-92)*	34.18*† *(0-80)*	61.54*† *(38-86)*	48.03*† *(36-67)*	50.57*† *(35–63)*	64.55† *(23-94)*	38.97† *(9–74)*

* Scores differed significantly between natives and L2 learners
† Scores differed significantly between high and low proficiency L2 learners

ERP data of two of the subjects from the native control group were removed from the analyses due to too much contamination of artefacts (blinks and/or heavy alpha activity), resulting in two populations of equal size (*n*=26).

7.2.2. Stimuli

The same stimuli as in Loerts et al. (Under Review) were used containing a total of 440 experimental sentences presented auditorily. Sentences were digitally recorded in a soundproof booth with a sampling rate of 44100 Hz by a female native speaker of Dutch.

Finiteness

80 Finiteness sentences (40 grammatical and 40 ungrammatical) were adapted from Sabourin and Stowe (2008). Sentence endings were added to the original sentences to avoid potential sentence-final wrap up effects.

These sentences had previously successfully elicited a P600 response in both native speakers and late L2 learners from languages that are similar to Dutch. The present study included these sentences to test whether these constructions would also elicit a P600 in L2 learners with a typologically different native language. Some of the sentences contained an auxiliary followed by a past participle, and the ungrammatical counterpart was created by replacing the past participle with an infinitive form of the verb (example sentence 7-3). This construction differs slightly from Polish, where the auxiliary does not precede the past participle but is rather (in some constructions) added to the past participle as a suffix (Migdalski, 2006). The other finiteness construction used was similar in Dutch and Polish and contained modal auxiliaries followed by an infinitive form and the ungrammatical variants of these sentences contained a past participle instead of a finite form of the verb (example sentence 7-4).

*7-3. Ze **heeft** alleen haar beste vriendin **uitgenodigd/*uitnodigen** voor haar verjaardag.*
She **has** only her best friend **invited/*invite** for her birthday.
*7-4. Ik **wil** dit jaar naar Spanje **gaan/*gegaan** op vakantie.*
I **want** this year to Spain **to go/*gone** on holiday.

Markers for later ERP analysis in the finiteness condition were placed on the exact time-point at which the ungrammaticality becomes apparent. In some cases (e.g. finite verbs), this was the onset of the verb, but in other cases this point was only present after a prefix of the verb (e.g. after *'uit-'* in example sentence 7-3).

Grammatical Gender

To assure knowledge of the critical nouns in the grammatical gender condition, 45 common and 45 neuter high frequency nouns were selected from the 5000 words occurring most frequently in the Spoken Dutch Corpus (Corpus Gesproken Nederlands;

CGN) as compiled by The Centre for Language and Speech Technology (Dutch HLT Agency or TST-centrale).

These nouns were used to create 180 grammatical sentences in which gender agreement either occurred on the determiner (definite NPs) or on the adjective (indefinite NPs).[11] Additionally, definite determiners were either followed by the noun immediately (DN: determiner + noun), or by an intervening adjective (DaN: determiner + adjective + noun) to check for potential effects of locality. Table 7-4 provides an overview of the three constructions used and their ungrammatical counterparts. These ungrammatical counterparts were created by violating gender-marking on the determiner or the adjective. In the definite constructions (DN and DaN), where gender is only marked on the determiner, ungrammatical variants were created by substituting the definite articles (i.e. replacing the common determiner *'de'* with *'het'* and replacing *'het'* with *'de'* in the neuter conditions). In the indefinite constructions (IaN), where gender is only marked on the adjective, the incorrect form of the modifying adjective was used to create an ungrammatical counterpart (i.e. adding the common suffix *–e* in neuter conditions or using bare adjectives in the common conditions).

Because the gender mismatch between the preceding determiner/adjective and the target noun becomes apparent after hearing the onset of the critical noun, markers were placed at this position in the sentence for later ERP analysis.

Table 7-4: Example sentences of the different constructions containing gender violations. Examples are divided into NP-construction and gender congruency. The critical NPs are printed in bold and English equivalents of the sentences are presented in italic below every sentence. Note that each condition equally contained neuter and common nouns as target words and only one example is given in the above table.

NP	Gender	Example Sentence
DN	Congruent	Vera plant rode rozen in **de tuin** van haar ouders. *(Vera is planting red roses in the$_{COM}$ garden$_{COM}$ of her parents.)*
	Incongruent	Vera plant rode rozen in ***het tuin** van haar ouders. *(Vera is planting red roses in *the$_{NEU}$ garden$_{COM}$ of her parents.)*
DaN	Congruent	Bianca leest (…) **het nieuwe boek** van haar favoriete schrijfster. *(Bianca is reading the$_{NEU}$ new book$_{NEU}$ of her favourite writer.)*
	Incongruent	Bianca leest (…) ***de nieuwe boek** van haar favoriete schrijfster. *(Bianca is reading the$_{COM}$ new book$_{NEU}$ of her favourite writer.)*
IaN	Congruent	Sjoerd schrijft **een lange brief** aan zijn verloofde. *(Sjoerd writes a long$_{COM}$ letter$_{COM}$ to his fiancée.)*
	Incongruent	Sjoerd schrijft **een lang brief** aan zijn verloofde. *(Sjoerd writes a long$_{NEU}$ letter$_{COM}$ to his fiancée.)*

[11] Each noun was used twice: once with highly constraining preceding context and once with a neutral context. These sentences were used to assess the impact of semantic expectancy in native speakers (see Loerts et al., Under Review). To avoid abundance and small numbers of items by splitting the data into all possible variables, semantic expectancy is not discussed in the present paper and the focus lies solely on morphological processing.

Following a Latin Square design, different lists were created to assure each participant was presented only one version of each item in a certain condition. In addition, 125 plausible filler sentences were added and randomized with the experimental sentences (45 gender and 40 finiteness sentences per list). Each list was divided into a practice session followed by 4 runs that could each be followed by a short break. The conditions were spread even across lists and experimental sentences were randomized in such a way that the same conditions never occurred twice in a row.

7.2.3. Procedure

Participants were seated at a distance of about 60 – 70 cm from a computer screen in an isolated soundproof dimly-lit room. Before each session, the sound of the speakers was adjusted to a volume of between 73dB and 77dB SPL. During the experiment, continuous EEG signal was recorded while participants listened to correct sentences and sentences containing morphological violations. Subjects were instructed to minimize eye blinks and body movements when listening to the sentences. In addition, they were asked to look at a central fixation cross to minimize eye movements while listening and instructed to blink only when a string of asterisks appeared on the screen (****). After each sentence the phrase *goed?* ('correct?') appeared on the screen and participants had to judge the acceptability of the sentence they had heard by pressing one of two buttons using their right hand.

7.2.4. Recording and analysis

Accuracy and reaction times were collected using E-prime software (Schneider, Eschman, & Zuccolott, 2002). The continuous EEG was recorded with 64 tin electrodes integrated in a plastic cap (Electro-Cap International). The positions of the electrodes were based on the extended 10-20 system (also see Figure 7-1). Blinks and eye movements were monitored by means of 4 electrodes that were placed on the temples and above and below the left eye. Two electrodes were placed for reference on the mastoids and a ground electrode was placed on the sternum. Scalp electrode impedances were reduced to below 10 KΩ, except that the temporal and in some cases frontal electrodes (Fp1 and Fp2: also see Figure 7-1) could sometimes only be reduced to below 20 KΩ. The raw EEG signal was sampled at 500 Hz.

The EEG data time-locked to the target words were analysed using Brain Vision Analyzer 2.0. Data were filtered using a low-pass and high-pass filter of 50 Hz and 0.25 Hz respectively and divided into segments from 500 ms before to 1500 ms after the critical words. For each participant, the data was corrected for blinks using the Gratton and Coles (1989) method after which trials containing artifacts were rejected. A baseline period was set from 200 ms to 0 ms before the onset of the critical words to normalize the data. For statistical analyses, averaged ERPs were calculated per subject per condition and per time window of 50 ms.

To assure reasonable knowledge of the noun as well as its gender in the late L2 learners, only those trials were analysed that contained nouns to which the L2 learner had correctly assigned the gender at least twice (as based on the offline gender assignment task, see 'participants'). Due to the relatively small amount of correct responses on the acceptability judgement task for some less proficient L2 learners as well as a skewed distribution of the number of accurate responses among conditions, trials were included regardless of the accuracy of these judgements.

Assuming that the L2 speakers show a similar pattern, we expected a P600 in response to both the gender and the finiteness violations, based on the results from our previous study (Loerts et al., Under Review). To be able to explore the onsets and durations of the possible ERP effects in the late L2 learners, however, statistical analyses were carried out on 50 ms latency windows from stimulus onset until 1500 ms after the onset of the critical word.

A total of 6 regions of interest (ROIs) containing 6 electrodes each were selected based on anterior/posterior and hemisphere resulting in a left-anterior ROI (FP1, AF3, AF7, F3, F5, F7), a right-anterior ROI (FP2, AF4, AF8, F4, F6, F8), a left-central ROI (FC3, C1, C3, C5, CP3, CP5), a right-central ROI (FC4, C2, C4, C6, CP4, CP6), a left-posterior ROI (P3, P5, P7, PO3, PO7, O1), and a right-posterior ROI (P4, P6, P8, PO4, PO8, O2). These ROIs are depicted in Figure 7-1.

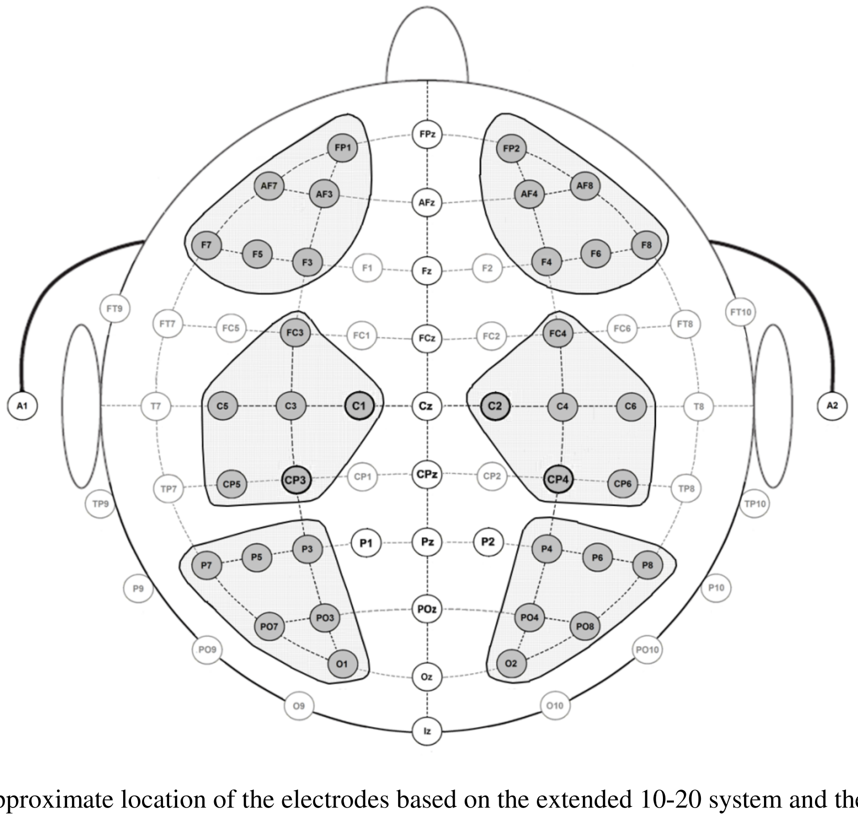


Figure 7-1: Approximate location of the electrodes based on the extended 10-20 system and the layout of the 6 regions of interest used for analyses.

Accuracy scores, reaction times, and ERP epochs in response to correct sentences and their ungrammatical equivalents were analysed to investigate morphological processing using analyses of variance. ANOVAs with repeated measures were performed, including two levels of grammaticality (grammatical vs. ungrammatical), two levels of type (gender vs. finiteness), two levels of hemisphere (left vs. right), three levels of anterior-posterior (anterior vs. central vs. posterior), and either L1 (Dutch vs. Polish) or Proficiency (high vs. low) as between-subject factor. For the analyses of finiteness sentences, the two-level factor of construction (finite vs. past participle) was added to explore potential differences. Additionally, type of noun phrase (definite vs. indefinite; determiner-marking versus adjective-marking) and type of gender (common vs. neuter) were analysed separately in combination with grammaticality of the gender sentences to see whether this caused any within- or between-subjects effects. Additional separate analyses were performed in step-down fashion when significant interactions were found, collapsing over factors that did not cause interactions. Reported p-values reflect the application of the Greenhouse-Geisser correction for repeated measures with more than one degree of freedom (Greenhouse & Geisser, 1959).

To further investigate the potential impact of L2 proficiency on the amplitude of the effects, correlation analyses were performed. On the basis of the between-groups analyses, the difference in amplitude was calculated for the effect of grammaticality for both the gender and the finiteness sentences (ungrammatical - grammatical) per Region of Interest. Correlation coefficients were then calculated between the average difference amplitude and the continuous L2 proficiency score.

7.3. Results

7.3.1. Behavioural data

Only known items, i.e. items for which the L2 learners had assigned the correct determiner at least twice in the gender assignment task, were analysed. In a separate analysis comparing known versus unknown items, however, neither accuracy rates nor reaction times appeared to be affected by the presence or absence of conscious knowledge of the item's gender.

Accuracy

Dutch native speakers were quite accurate with an overall mean score of 94% (95% confidence intervals = [90,98]), but the L2 learners as a group performed only slightly above chance level with an overall mean score of 66% (95% CI = [63,70]). Both groups were more accurate in judging grammatical structures (Natives: $F(1,27) = 13.543$, $p = 0.001$; L2 learners: $F(1,24) = 76.317$, $p < 0.001$), but late L2 learners had significantly more difficulty in detecting ungrammatical structures than native Dutch speakers ($F(1,52) = 53.274$, $p < 0.001$). An additional type x grammaticality x group interaction ($F(1,52) = 19.236$, $p < 0.001$) revealed that late L2 learners made most errors when judging gender mismatches. Low proficiency learners had slightly more difficulty detecting the errors in the ungrammatical sentences than high proficiency learners ($F(1,24) = 3.687$, $p = 0.067$).

Accuracy rates were not significantly affected by the type of finiteness violation (infinitive vs. past participle), but a phrase x grammaticality x group interaction for the gender sentences ($F(2,104) = 5.054$, $p < 0.01$) showed that natives performed significantly better when judging structures containing the incorrect determiner as opposed to the less salient ungrammatical structures containing an incorrectly inflected adjective ($F(1,27) = 13.133$, $p = 0.001$). L2 learners did not show this effect of phrase (F(2,48) = 1.096, $p = 0.341$).

Interestingly, both native speakers and L2 learners are significantly less accurate in judging sentences containing neuter nouns than they are with common nouns (Natives: $F(1,27) = 13.264$, $p = 0.001$; L2 learners: $F(1,24) = 149.811$, $p < 0.001$). This gender effect was significantly larger in the late L2 learners ($F(1,52) = 98.539$, $p < 0.001$) who had a mean accuracy of 81% on common and only 29.5% on neuter nouns as compared to an accuracy of 92.5% for common and 86.5% for neuter nouns in the native control group.

Reaction times

Subjects were overall faster to judge ungrammatical structures ($F(1,52) = 4.609$, $p < 0.05$). Reaction times were not significantly affected by the type of finiteness violation (infinitive vs. past participle), but the faster responses for ungrammatical items were especially present in determiner-noun mismatches and less for structures containing an adjective-noun mismatch ($F(2,104) = 5.057$, $p < 0.01$). A group x grammaticality x phrase interaction ($F(2,104) = 3.771$, $p < 0.05$) revealed that this latter effect was bigger in native Dutch speakers. Overall, subjects were slightly faster when judging neuter nouns as compared to common nouns ($F(1,52) = 4.805$, $p < 0.05$), which was not significantly affected by group or proficiency (Fs < 2.063, ps < 0.157).

7.3.2. Electrophysiological data

The grand average waveforms suggest that both finiteness violations and gender violations elicit a long-lasting positivity in native speakers (see Figure 7-2) that is preceded by a central-posterior negativity for finiteness violations. Late L2 learners, regardless of their proficiency level, do seem to show a similar positivity in response to violations of finiteness (Figure 7-3; p.159), but the ERP responses to gender violations suggest a positivity only for high proficiency learners (Figure 7-4; p.161).

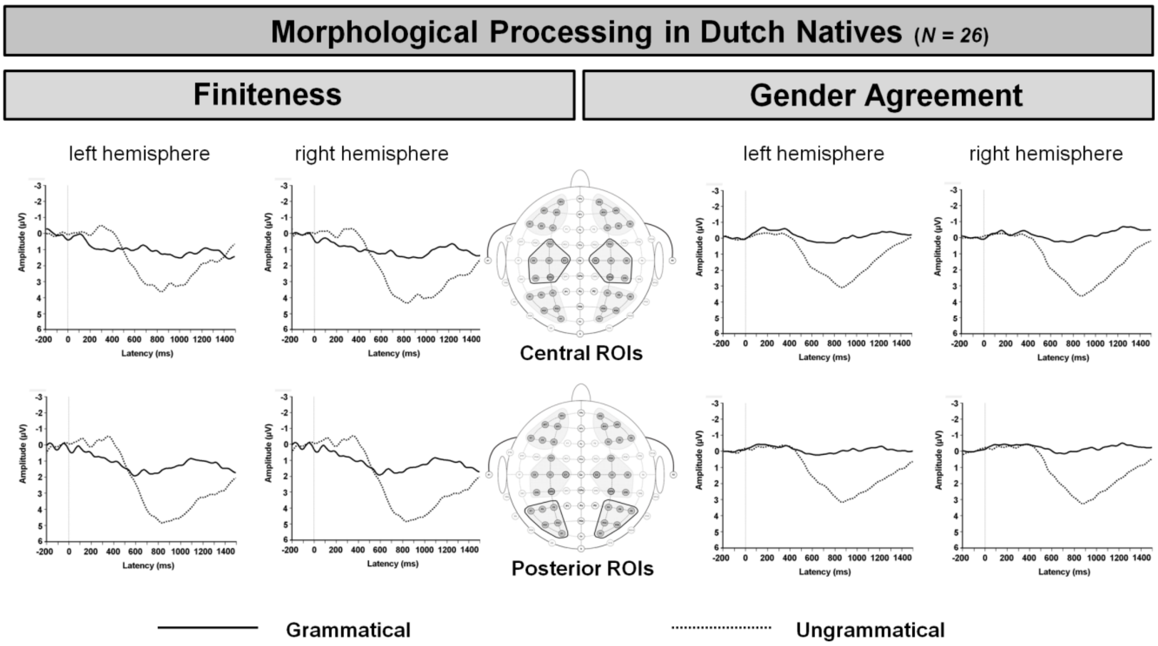


Figure 7-2: ERPs from the native Dutch speakers in response to the finiteness violations (left column) and the sentences containing a gender mismatch (right column).

Finiteness Violations

Grand-averages of the finiteness violations in Dutch native speakers are depicted in the left panel of Figure 7-2. These grand averages clearly show that the positivity in response to finiteness violations is preceded by a negativity for ungrammatical items in central and posterior regions.

When comparing brain activation in response to sentences containing gender agreement violations and sentences concerning violations of finiteness (type), a grammaticality by L1 (Dutch vs. Polish) interaction was found from 200 to 450 ms (Fs(2,49) > 3.267; ps < 0.05), which was additionally affected by the factor anterior/posterior from 300 ms to 450 ms (Fs(4,98) > 2.993 ; ps < 0.05). Post-hoc ANOVAs for the factors region and L1 revealed an interaction of type by grammaticality from 250 to 400 ms only for the native Dutch speakers (Fs(1,25) > 4.174; ps < 0.05), which was caused by a significant negativity for ungrammatical finiteness items only (Fs(1,25) > 6.087; ps < 0.05). This negativity was bilaterally present in both central (Fs(1,25) = 5.337; ps < 0.05) and posterior regions (Fs(1,25) = 5.507; ps < 0.05), as can also be seen in the grand averages of Figure 7-2. The timing, distribution, and polarity resemble the N400 effect commonly found for semantic violations or violations reflecting semantic integration difficulties. L2 learners, regardless of their proficiency level, showed no effects of grammaticality in these time windows (Fs(1,12) < 1.541; ps > 0.238; Fs < 1).

As can be seen in Figure 7-2, the N400-like effect in response to finiteness violations is followed by a large positive deflection in the native Dutch speakers. This effect of grammaticality reached significance in central ROIs from 650 to 1150 ms (Fs(1,25) > 5.019; ps < 0.05), where the effect shifted to a slightly right-lateralized effect over central electrodes from 1150 until 1500 ms post stimulus (Fs(1,25) > 4.564; ps < 0.05). In posterior regions, the P600-like effect was bilaterally present from 700 until 1350 ms (Fs(1,25) > 5.223; ps < 0.05).

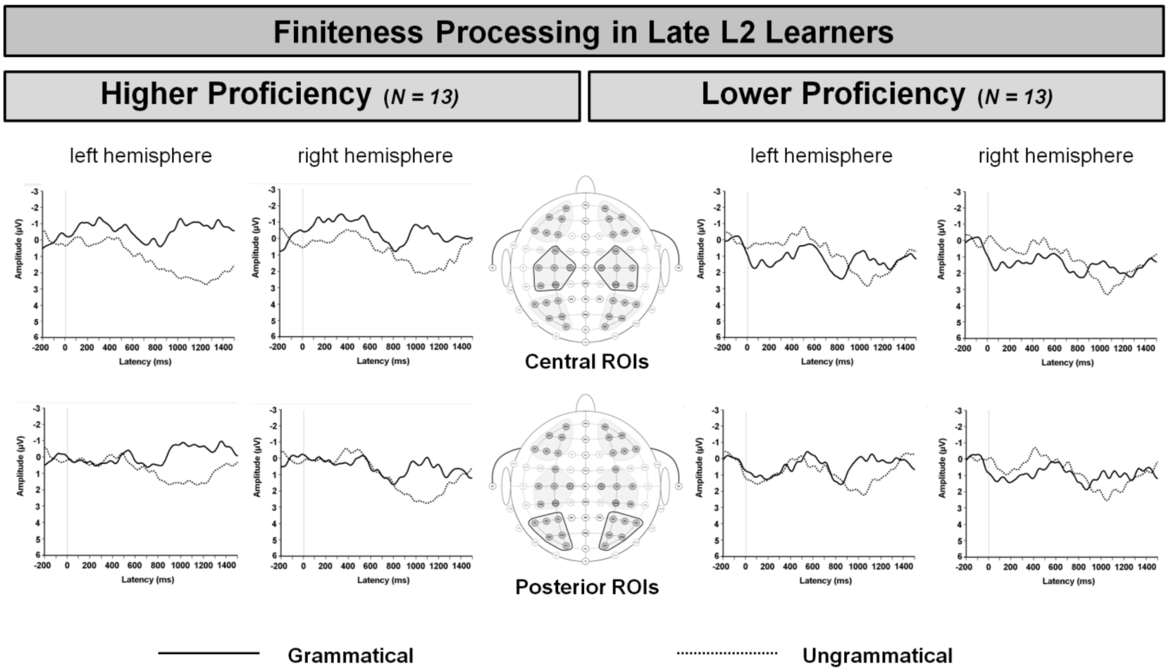


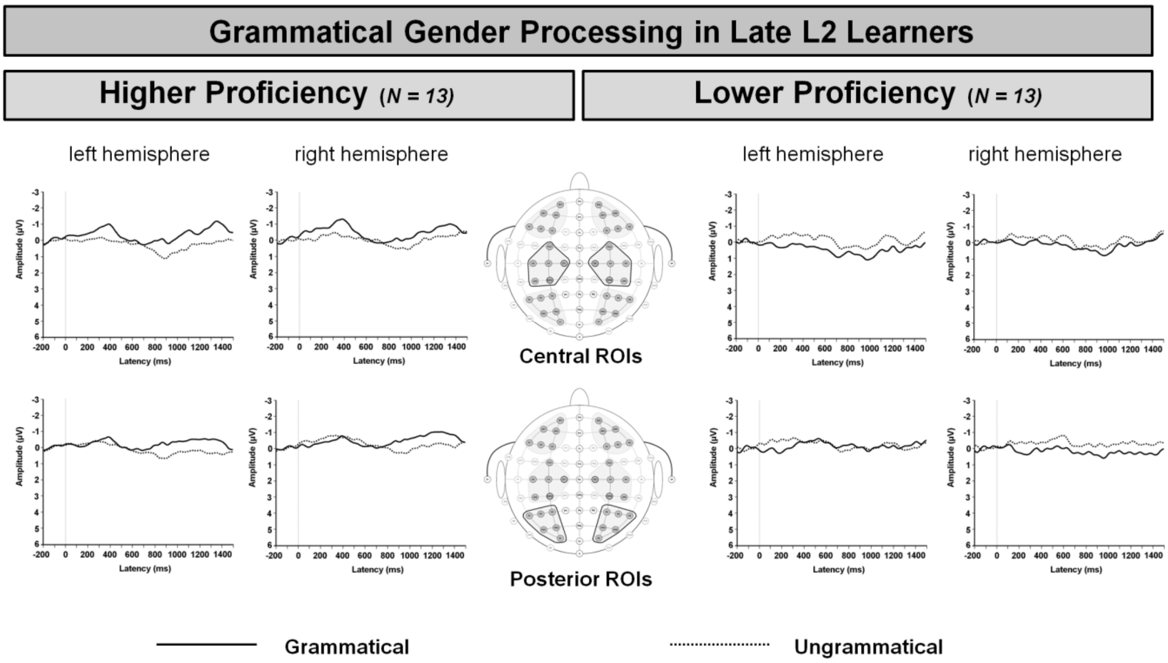
Figure 7-3: ERPs from the higher proficiency (left column) and lower proficiency (right column) L2 learners in response to violations of finiteness.

As can be seen in Figure 7-3, both high proficient and low proficient L2 learners appear to show a positive deflection for ungrammatical finiteness items in central and posterior areas that resembles the native P600 effect. A significant effect of grammaticality, however, was initially only present in native speakers, as also revealed by a significant interaction of L1 x grammaticality from 650 to 900 (Fs(1,50) > 4.557; ps < 0.05). Post-hoc ANOVAs on the factors proficiency confirmed the presence of a significant positivity for ungrammatical finiteness items in high proficient learners in posterior regions from 850 to 1100 ms post onset (Fs(1,12) > 6.906; ps < 0.05) and in central ROIs from 900 to 1200 ms (Fs(1,12) > 5.483; ps < .05). As suggested by the presence of a trend towards a grammaticality x proficiency interaction within the group of L2 learners ($F(1,24) = 3.448$, $p = 0.076$), the positivity for low proficient L2 learners only reached significance in posterior regions around 900 ms post stimulus (F(1,12) = 14.820, $p < 0.01$) and was only marginally present from 950 to 1000 ms (F(1,12) = 3.720, $p = 0.078$). While the positivity for ungrammatical finiteness items in low proficiency learners reached significance in both central an posterior areas from 1000 to 1050 ms (Central: F(1,12) = 7.243, $p < .05$; Posterior: F(1,12) = 28.667, $p < 0.001$), it was again only marginally significant from 1050 to 1150 ms (Central: Fs(1,12) = 4.243, ps < .062; Posterior: Fs(1,12) = 3.792; ps < 0.075). The P600-like effect in high proficiency learners, on the other hand, remained significant in central ROIs until 1400 ms post onset (Fs(1,12) > 7.441; ps < .05). No effects of construction (finite vs. past participle) were found in either native speakers or L2 learners.

Morphological Processing: Gender Agreement
From 500 to 550 ms, there was a trend towards a significant difference between natives and L2 learners and the effect of grammaticality ($F(1,50) = 3.476$, $p = 0.068$). Post-hoc analyses by group and the factor anterior/posterior showed that the gender mismatches elicited a significant P600 in native speakers across the entire scalp, as can also be seen in Figure 7-2 (Frontal: ($F(1,25) = 10.391$, $p < 0.01$; Central: $F(1,25) = 12.838$, $p < 0.001$; Posterior: $F(1,25) = 6.229$, $p < 0.05$). L2 learners, however, showed no effects of grammaticality (F<1) in this time-window.

The overall P600 in response to gender mismatches reached significance from 550 to 1300 ms post stimulus ($Fs(1,50) > 7.246$, ps < 0.01) and was significantly affected by L1 throughout these time-windows (Fs(1,50) > 5.234, ps < 0.01). This interaction was mainly caused by a large P600 that was present in the native control group from 600 to 1350 ms (Fs(1,25) > 7.305, ps < 0.01), which was significantly more pronounced in central and posterior regions from 650 ms onwards (F(2,50) = 5.724, p < 0.05), but bilaterally present across the entire scalp until 700 ms (Frontal: Fs(1,25) > 5.509, ps < 0.05; Central: Fs(1,25) > 30.039, ps < 0.001; Posterior: Fs(1,25) > 20.380, ps < 0.001). From 700 ms onwards, the native gender P600 remained significant in central regions until 1400 ms post onset (Fs(1,25) > 6.129, ps < 0.05) and in posterior regions until the end of the ERP epoch around 1500 ms post onset (Fs(1,25) > 4.269, ps < 0.05). In central regions, the P600 was slightly more pronounced in the right hemisphere from 750 to 950 ms (Fs(1,25) > 4.079, ps < 0.054) and from 1000 to 1300 ms (Fs(1,25) = 5.688, ps < 0.05), comparable to the P600 in response to finiteness violations.

An anterior/posterior x grammaticality x L1 interaction from 650 to 1350 ms post stimulus (Fs(2,100) > 4.902, ps < 0.05) revealed differences in the presence and/or distributions of the grammaticality effects in natives and L2 learners. This interaction was initially caused by the absence of a grammaticality effect within late L2 learners (Fs < 1). From 850 to 1000 ms post stimulus, however, an interaction between grammaticality and proficiency was found within the L2 learners that was mostly present in posterior regions (Fs(1,24) > 4.885, ps < 0.05). Post-hoc analyses for the two proficiency groups revealed a significant positivity for ungrammatical items for high proficient L2 learners only (Fs(1,12) > 5.743; ps < 0.05). This P600-like effect for high proficiency learners, which is also visible in Figure 7-4, remained significant from 850 to 1100 ms in posterior regions (Fs(1,12) > 5.596; ps < 0.05) and marginally so in central regions from 950 to 1100 ms (Fs(1,12) > 4.090; ps < 0.066). This positivity was only a trend from 1100 to 1200 ms (Fs(1,12) > 3.756; ps < 0.076), but reliably present again in central and posterior regions from 1250 to 1400 ms (Fs(1,12) > 6.562; ps < 0.05; Posterior: (Fs(1,12) > 6.168; ps < 0.05). No significant effects of grammaticality were found in the low proficient L2 learners (Fs(1,12) < 2.061; ps > 0.177). Type of



noun phrase (definite vs. indefinite; determiner-marking versus adjective-marking) did not significantly impact the effects of grammaticality in natives or in L2 learners.

Figure 7-4: ERPs from the higher proficiency (left column) and lower proficiency (right column) L2 learners in response to gender mismatches.

Interestingly, when the gender of the target noun was entered as a factor into the analyses, a trend towards a grammaticality x proficiency x gender interaction was found within the group of L2 learners from 900 – 950 ($F(1,24) = 3.644$, $p = 0.068$), which reached significance from 950 – 1000 ($F(1,24) = 5.700$, $p < 0.05$). Post-hoc ANOVAs for the factor proficiency revealed a grammaticality x gender interaction within high proficient L2 learners from 950 to 1400 ms post stimulus (Fs (1,12) > 4.793, $p < 0.05$), which was absent in the low proficient L2 learners (Fs<1). When analysing common and neuter nouns separately, high proficiency learners showed a significant overall P600 for common nouns (Fs(1,12) > 8.880, ps < 0.05), but no grammaticality effects were found in response to neuter nouns (Fs (1,12) < 1.266, ps > 0.283). This common P600 effect was unaffected by type of noun phrase, suggesting similar effects for violations between determiner-noun (with or without an intervening adjective) and adjective-noun constructions. Low proficient learners of Dutch showed no effects of grammaticality regardless of the gender of the target noun (Fs (1,12) < 1.349, ps > 0.268).

Morphological Processing: Finiteness versus Gender

The most striking difference between native processing of finiteness and grammatical gender is the presence of a negativity in the finiteness condition. Additionally, the re-analysis process in response to gender violations started earlier (around 500 ms post stimulus) as compared to the finiteness P600 that reached overall significance around 650 ms. Within native speakers, the gender P600 was initially present all over the scalp, while the finiteness P600 was mainly present in frontal and central regions from 550 to 600 ms ($F(2,50) = 6.680$, $p < 0.01$). Similarly, a trend towards a type x grammaticality interaction ($F(1,25) = 3.823$, $p = 0.062$) suggested a slightly larger P600 in response to gender mismatches in posterior regions, probably related to the fact that the P600 in response to finiteness violations was preceded by an N400 in these regions. After 600 ms post stimulus, the overall P600 for natives (Fs(1,25) > 13.678, ps< 0.001) did not differ per type anymore (Fs<1), except for a trend towards a slightly larger gender P600 from 900 to 950 ms ($F(1,25) = 3.280$, $p = 0.082$).

The overall positivity for high proficiency L2 learners was significant in central and posterior regions from 850 until around 1250 ms post stimulus (Central: Fs (1,12) > 6.592, ps < 0.05; Posterior: Fs (1,12) > 6.834, ps < 0.05) and was only significantly bigger for finiteness violations from 1000 to 1050 ms (Central: $F(1,12) = 13.560$, $p < 0.01$; Posterior: $F(1,12) = 8.090$, $p < 0.05$) and marginally so from 1050 to 1100 ms (Central: $F(1,12) = 4.305$, $p = 0.060$; Posterior: $F(1,12) = 3.899$, $p = 0.072$). The absence of a type x grammaticality interaction after 1100 ms suggested comparable P600 effects for finiteness and gender in this group of high proficient L2 learners. Low proficient learners initially showed no effects of grammaticality. In the time windows from 900 to 1200 ms post onset, however, low proficient L2 learners showed a marginal type x grammaticality interaction (Fs (1,12) = 3.964, ps < .070), which was caused by a positivity for ungrammatical finiteness items only (Central: Fs(1,12) = 7.243, ps < .05; Posterior regions: Fs(1,12) = 28.667, ps < 0.001).

Correlational Analyses of the effects

The grand average waveforms in Figure 7-3 and Figure 7-4 suggest differences in the P600 between lower and higher proficiency learners of Dutch, as was also confirmed by statistical analyses. To further examine potential individual differences, correlation analyses were performed on the mean P600 amplitudes from the time-window in which all groups showed a reliable grammaticality effect, i.e. from 900 to 1150 ms.

Within the native control group, no significant correlations were found between any of the personal characteristics or proficiency scores and the amplitude of the P600 in response to ungrammatical finiteness items. The amplitude of the native gender P600 in the left posterior ROI, however, correlated positively with the overall proficiency score (Common P600: $r_s(26) = 0.42$, $p < 0.05$; Neuter P600: $r_s(26) = 0.37$, $p = 0.0618$). Separate analyses of the different scores showed that this was mainly driven by a positive correlation between the amplitude of the common gender P600 and participant's scores on the acceptability judgement task for common nouns (Left Central: $r_s(26) = 0.40$, $p < 0.05$; Left Posterior: $r_s(26) = 0.41$, $p < 0.05$).

Within the group of late L2 learners, a significant negative correlation between their age of acquisition (AoA) and the amplitude of the finiteness P600 in the left central ROI ($r_s(26) = -0.43$, $p < 0.05$) revealed larger P600 effects for those participants who started acquiring Dutch at an earlier age (also see Figure 7-5). No correlation was found for length of residence (LoR), but the finiteness P600 amplitude in L2 learners did correlate positively with the percentage of correctly produced neuter gender-marking in the film-retelling task (Left Central: $r_s(26) = -0.42$, $p < 0.05$). The importance of the ability to use neuter gender was also suggested by a trend between the finiteness P600 amplitude and the percentage of correctly assigned determiners in the offline gender assignment task (Left Posterior: $r_s(26) = -0.35$, $p = 0.084$).

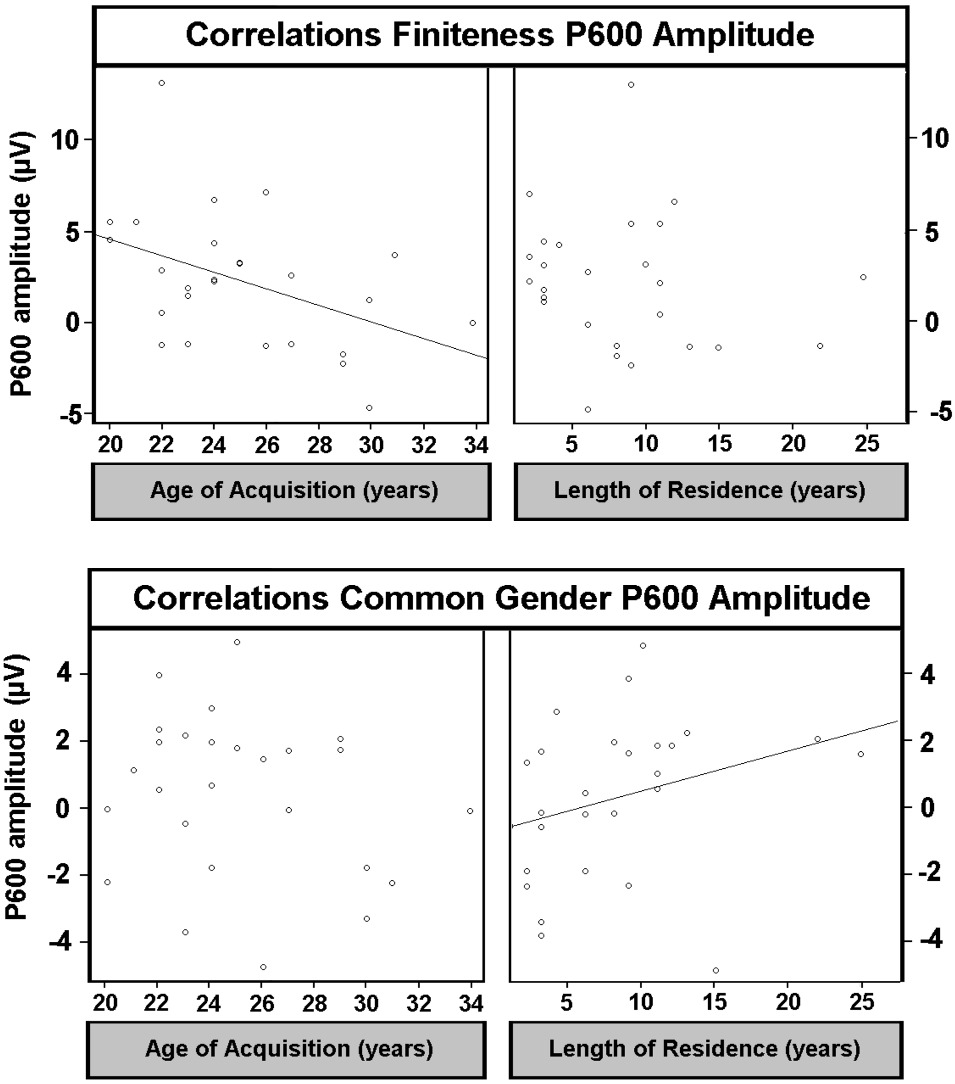


Figure 7-5: The correlational structures between Age of Acquisition (left panel) and Length of Residence (right panel) and the amplitude of the P600 in L2 learners in response to finiteness violations (upper panel) and common gender violations (lower panel) in the left-posterior ROIs.

As expected on the basis of the ERP analyses, no correlations were found with the difference in amplitude between grammatical and ungrammatical neuter nouns. Interestingly, as opposed to the amplitude of the finiteness P600 in L2 learners, the amplitude of the P600 in response to common nouns correlated positively with LoR (Left Central: $r_s(26) = 0.42$, $p < 0.05$; Right Central: $r_s(26) = 0.40$, $p < 0.05$), but not with AoA, suggesting that increased exposure to the Dutch, rather than an earlier age of onset, increases the amplitude of the repair or re-analysis process in L2 learners, as also depicted in Figure 7-5. The amplitude of the common gender P600 in both central and posterior ROIs additionally correlated positively with the overall proficiency scores of the L2 learners (all $r_s(26) = 0.40$, ps < 0.05) as well as with most of the proficiency measures separately (all $r_s(26) > 0.39$, ps < 0.05) except for the acceptability judgement task scores and the offline gender assignment task scores on neuter nouns.

7.4. Discussion

The present study recorded event-related potentials in response to spoken Dutch sentences in native and late L2 learners of Dutch with different levels of proficiency and different lengths of residence in the Netherlands (LoR). Importantly, age of acquisition (AoA) did not differ between lower and higher proficiency L2 learners. The auditory stimuli contained violations of finiteness and grammatical gender violations to investigate the potential influence of the frequency and reliability of morphological structures in the input. It was hypothesized that, since words from a language with a non-salient and asymmetrical gender system, such as the Dutch gender system, largely have to be acquired word-by-word (Unsworth, 2008), the acquisition of gender should require more input than other morphological structures, such as finiteness. Hence, the acquisition and processing of Dutch gender should be affected by the input quantity and length of exposure (Gathercole & Thomas, 2005).

Morphological processing: Finiteness

Within native speakers, violations of finiteness elicited a biphasic pattern including the hypothesized P600, which was preceded by a less expected N400. Sabourin (2003) had previously used these finiteness violations in sentence-final position in a visual experiment and did report a negativity. One possible explanation of such a negativity constituted a sentence final wrap-up effect. The present study shows, however, that this negativity remains present when the violations occur in sentence-medial position and thus cannot be a wrap-up effect.

When considering the finiteness violations in the present study, e.g. '*Ze heeft alleen haar beste vriendin* ***uitgenodigd/*uitnodigen*** *voor haar verjaardag*' (She has only **invited/*invite** her best friend for her birthday), the auxiliary '*heeft*' (has) can only be followed by a past participle form of a verb (note that the same holds for the sentences in which modal auxiliaries had to be followed by the infinitive). When hearing such a sentence, a subject is thus waiting to hear a past participle and when suddenly confronted with a non-finite form of the verb, this constitutes a violation of expectancy. An effect of expectancy might have an effect on semantic-related components, such as the N400, because the absence of a past participle hinders thematic roles assignment. This was also hypothesized by Hinojosa, Moreno, Casado, Muñoz and Pozo (2005), who found an N400 in response to a sentence like *'El luchador ganador* ***el*** *combate'* (The winning fighter **the** combat) in which participants were confronted with an article when they expected to encounter a verb. Similar N400 effects have been found when nouns in a prepositional phrase were replaced by a verb (Gunter & Friederici, 1999), and biphasic N400-P600 responses have been reported in response to violations in the number of arguments of transitive verbs (Friederici & Frisch, 2000). These syntactic N400 effects reflect expectancies that are based on the verb's selectional restrictions (e.g. Friederici, Pfeifer, & Hahne, 1993). The question arises why other types of prepositions elicit an early frontal negativity instead of an N400. As has also been noted by Hinojosa et al. (2005), N400s with a more anterior distribution could be related to the rapid presentation of stimuli. We know that increased word presentation rate causes the N400 to display a rather frontal distribution (Kutas, 1987). Further research should compare these N400-like effects to find out what kinds of expectancy results causes what kinds of negativities and whether these are of a more syntactic (LAN) or thematic (N400) nature.

Late L2 learners did not show an N400 in response to violations of finiteness and this might well be related to the lack of auxiliary verbs in Polish (Migdalski, 2006). If the mother tongue provides the basis for L2 grammar, at least in initial stages of L2 acquisition, then a form of the verb '*hebben'* (to have) or *'zijn'* (to be) may not lead to the syntactic expectation of a past participle. According to this explanation, however, positive transfer should occur for expectations based on modal verbs, which are expected to be followed by infinitives in both the L1 and the L2. It should be noted that the distribution of violations of expected past participles and infinitives was unbalanced (26 past participle and 14 infinitive constructions). The potential presence of a similar negativity in response to infinitive constructions might thus have been too small due to a smaller number of items.

Alternatively, the absence of a negativity in L2 learners might also be due to working memory problems, as has been suggested by Clahsen and Felser's Shallow Structure Hypothesis (2006a). According to the SSH, L2 learners have severe difficulty in processing non-local violations, such as these violations of finiteness, because they are unable to retain information intervening the two verbal elements in working memory. Their lack of sufficient processing resources then results in less 'in depth' processing. It is difficult to disentangle this issue on the basis of the present data alone and future research should look into the issue of similarity between languages and the influences of working memory.

The native finiteness N400 is followed by a central-posterior P600 from 650 to around 1350 ms post stimulus which is slightly more pronounced in the right hemisphere after 1150 ms. While L2 learners did not show any N400-like effects related to the verb's selectional restrictions, high proficiency L2 learners did show a delayed and reduced P600 from 850 to 1250 in central and posterior regions and low proficiency L2 learners showed a delayed and reduced P600 from 900 to 1150 ms post stimulus that was mainly present in posterior areas. Within group analyses revealed positive correlations between the amplitude of this P600 effect and proficiency scores related to the knowledge and use of neuter gender-marking. This result suggests that L2 proficiency specifically related to a difficult and abstract feature such as neuter gender serves as a valuable predictor for native-like morphological processes of repair and/or re-analysis. P600 amplitude additionally correlated negatively with age of acquisition (AoA), but not with length of residence (LoR), suggesting that earlier AoAs tend to increase native-like processes of repair and/or re-analysis, even in learners that have acquired the L2 between 20 and 34 years of age. The observed positive correlation between AoA and P600 amplitude argues against the presence of a sensitive period and suggests a linear function of age that fits within the predictions of the Competition Model for L2 learning (e.g. MacWhinney, 2008), in which age-related effects are explained in terms of L1 entrenchment that obviously impacts older learners more severely than relatively younger learners.

These results suggest that both high and low proficiency learners are able to re-analyse violations of finiteness in a qualitatively similar way to natives, which argues against the claim made by Sabourin and Stowe (2008) that languages need to be similar to allow direct transfer. There are instead quantitative differences between natives and L2 learners and between high and low proficiency learners, as visible in a delayed and reduced P600 effect possibly reflecting the need for more time, especially in low proficiency L2 learners, to initiate the re-analysis process and try to integrate the incongruent verb (Rossi, Gugler, Friederici, & Hahne, 2006).

Morphological processing: Gender agreement

The gender P600 for natives started around 500 ms and continued until the end of the ERP epoch and was present across the entire scalp. This P600, apart from its earlier onset, did not differ from the finiteness P600 in native speakers, suggesting that both types of violations are regarded as non-semantic phenomena that elicit similar processes of repair and/or re-analysis.

A delayed and reduced P600 from 850 to around 1400 ms was found only for high proficiency learners and only for violations on common nouns. The absence of a gender P600 for low proficiency learners, who also showed a delayed and reduced finiteness P600, suggests that finiteness is acquired somewhat more easily than grammatical gender. This is also evident from the low accuracy scores on gender as compared to finiteness sentences in both L2 proficiency groups and the fact that, within high proficiency learners, the P600 for finiteness tended to be slightly bigger than the P600 for common gender, possibly reflecting a more in depth re-analysis. The common gender P600 in high proficiency learners correlated positively with all proficiency measures, except for the score on the acceptability judgement task and the offline gender assignment task on neuter nouns. This result is in contrast with the fact that proficiency related to neuter gender served as a valuable predictor for native-like morphological processes of repair and/or re-analyses as reflected by the finiteness P600. This result, however, is in line with the observed discrepancy in proficiency between common and neuter gender agreement in L2 learners and suggests that proficiency on neuter nouns can only predict native-like processing that is unrelated to the processing of (common) gender agreement. Although this seems counterintuitive, the individual data also reveal that while some of these highly proficient learners perform at ceiling with common nouns and score below 30% accurate on neuter nouns, others are excellent in assigning neuter gender while they score at chance on common nouns. The acquisition and processing of common and neuter gender thus likely depends on different strategies and common gender proficiency is more likely to reflect greater overall proficiency in the L2 than neuter gender proficiency.

The presence of a P600 in response to common gender violations in highly proficient Polish learners of Dutch contradicts the claim by Sabourin and Stowe (2008) that native-like processes of repair and/or re-analysis depend on the degree of similarity between the L1 and L2 gender system. As Clahsen and Felser (2006a) have pointed out, the presence of a P600 in German-Dutch learners, but not in L2 learners from Romance languages, might well have reflected the relatively higher proficiency of the German L2 learners as revealed by the scores on the acceptability judgement task. The presence of a P600 in response to common gender only in high proficiency and not in low proficiency learners from a language with a gender system that is not similar to the Dutch gender system shows that L2 proficiency, possibly in combination with L1 background, plays a crucial role in predicting native-like processing of grammatical gender.

The fact that a delayed and reduced P600 in high proficiency L2 learners is only found in response to violations on common nouns and not on neuter nouns, even though the participants knew the gender of these nouns offline, fits within the predictions of the Competition Model. The cues for common gender in Dutch are not only more available since common nouns are more frequent, but the cues are generally also more reliable while the availability of cues for neuter gender is limited. Hence, the default common gender is acquired prior to neuter gender. Moreover, while the amplitude of the finiteness P600 was affected by AoA, the amplitude of the common gender P600 correlated positively with LoR. This result showed that longer residences in the Netherlands, and hence larger amounts of linguistic input related to grammatical gender, can be related to larger, more native-like re-analysis processes in response to violations related to common gender. Similar results have been reported by Foucart & Frenck-Mestre (2012), who only found a P600 for gender agreement violations on post-posed adjectives, which is the canonical structure in French. Violations on pre-posed adjectives, a less frequent position in French, elicited an N400 possibly reflecting early stages of syntactic processing. The absence of an N400 or a P600 in response to gender violations in the low proficiency L2 learners in the present study thus suggests that these grammatical and ungrammatical structures cannot be distinguished by the learner (Steinhauer et al., 2009), which also corresponds to the low accuracy rates on the acceptability judgement task.

Furthermore, the common gender P600 was unaffected by type of noun phrase, suggesting similar effects for violations between determiner-noun (with or without an intervening adjective) and adjective-noun. This, in combination with the variability in the scores on the gender assignment task suggests no increased difficulty for L2 learners with less salient adjective-noun gender-marking (i.e. the presence or absence of the schwa-ending) as compared to determiner-noun agreement (i.e. the presence of *'de'* vs. *'het'*). It should be noted, however, that Polish lacks determiners and, from a competition perspective, determiners are unique to the L2 and do not suffer competition from the L1, while adjectival inflection is relatively similar in the L1 and the L2. In other words, even though adjective-noun agreement might generally be more difficult for L2 learners, determiner-noun constructions might be relatively less difficult for Polish-Dutch learners due to a lack of competition from the L1 which in turn results in the absence of a difference between the two constructions.

In sum, the present study shows that proficiency plays an important role in the online processes in both native and L2 learners with higher proficiency leading to higher amplitude P600 effects. Unfortunately, due to the lack of any early effects, no conclusions can be drawn on the basis of the present data regarding early automatic versus later and more controlled processes in natives versus L2 learners. The presence of similar late components in the natives and high proficiency L2 learners, however, does again show that highly proficient L2 learners can reach native-like levels even if they learned their L2 late in life (Rossi et al, 2006: Foucart & Frenck-Mestre, 2012). The presence of a native-like repair or re-analysis process in response to morphological violations was dependent on proficiency as well as the frequency, availability and reliability of morphological cues in the input. Finiteness violations, a construction that is frequently and reliably present in the input, elicited a repair or re-analysis process in both natives and L2 learners. The acquisition and processing of a non-transparent and asymmetrical gender system, however, is highly dependent on proficiency and input quantity as suggested by the presence of a delayed and reduced P600 only in high proficiency L2 learners and only in response to violations on the default common gender. These results suggest that more available and reliable structures are acquired earlier, but, given enough time and exposure, L2 learners can process gender in a similar way as compared to natives.

Chapter 8

GENERAL DISCUSSION

On what the eyes and brains reveal while native and second language learners listen to Dutch and its gender

Learning a new language, and especially learning a new gender system, has often been shown to be one of the major stumbling blocks in the acquisition of a second language, especially when the L2 is acquired after puberty. The main question addressed in this dissertation is whether late L2 learners from a typologically different language are able to acquire and process the Dutch language, and in particular the covert Dutch gender system, in a similar way as native speakers. More specifically, the aim was to investigate whether the differences between natives and L2 learners of Dutch concern fundamental or partial differences in the representations and/or the processing of the gendered language.

The question of whether post-puberty learners are able to acquire and process a grammatical gender system other than their own has previously mostly been investigated by examining the acquisition of Romance languages. It is not inconceivable that the gender systems that have predominantly been the focus of these investigations are acquired somewhat differently than the Dutch gender system. This might either be due to the availability of morphophonological cues that are present on both the agreement targets and the noun itself (e.g. Spanish and Italian) or due to the relatively higher availability, saliency and consistency of agreement cues in the input (e.g. French, but also German). Moreover, most studies to date have looked at the L2 acquisition of these gender systems by L2 learners from a typologically similar language. This is especially problematic since these learners might be able to use their L1 to positively transfer genders and gender agreement rules that are similar in the two languages. The present dissertation therefore examined the acquisition and processing of a covert and asymmetrical gender system (i.e. Dutch) by learners from a typologically different language background with a gender system (i.e. Polish).

Another issue about L2 learning concerns the way in which learning takes place. First language learners typically learn by listening and interacting with other speakers, rather than via written materials using adult analytical skills. Most studies to date have compared native and L2 acquisition and processing by investigating highly educated (college) participants, who are particularly likely to have made use of their analytical skills. The late Polish-Dutch learners studied in the present dissertation have mostly acquired the L2 in an informal setting, so the sample of participants mostly concerned non-university students with various ages and educational backgrounds. Although these criteria made the group of learners more heterogeneous than is the case in studies that draw their participants from a university student pool, this more diverse population might constitute a more realistic portrayal of L2 learners and therefore allow generalization of the results to (L2) speakers outside of the university campus.

A third issue with previously reported studies is that most studies have made use of written stimuli. The processing of written material depends heavily on the amount of reading experience with the L2, a skill that (L2) learners do not always possess in equal measure. The late L2 learners in the present study have primarily acquired the L2 through listening, and written comprehension might therefore underestimate their actual abilities. Therefore, auditory stimuli were used to study online processing in both the eye tracking and the event-related potentials experiment.

The studies presented in this dissertation furthermore make an important methodological contribution with regard to the extensive range of data used to assess the level of proficiency of the L2 learners. Especially since the late L2 learners tested in the present dissertation have mostly learned L2 Dutch through listening and using the language in everyday life, we are most likely investigating the results of implicit learning strategies where the acquired gender system is entirely covert. It was thus crucial to include a sophisticated and multifaceted measure of proficiency to be able to test beyond the surface level of proficiency. Research up to date has mostly used a single measure to determine the level of proficiency, most often a C-Test or some form of a placement task. It is questionable, however, to what degree such formal tasks actually mirror proficiency in general. Some learners are very adept at retrieving words from their lexicon and are good in speech production while they do not comprehend spoken language equally well. Others understand almost everything they hear in the L2, but are less capable of telling a story using fluent speech and correct pronunciation. This potential asymmetry between different proficiency measures was indeed obvious in the lack of consistent correlations between the different scores used to assess proficiency level in the studies presented here (see also *Chapter 3*). For this reason, a composite measure of proficiency level can have added value when assessing the effect of L2 proficiency on whether language processing is native-like.

Before being able to answer the question of whether L2 comprehenders of Dutch can process gender in a native-like way, however, it was necessary to investigate how native Dutch speakers use and process Dutch and its grammatical gender system. This seemingly elementary question is of the utmost importance, especially when considering the rather curious manifestations of gender in the Dutch language. Before summarizing and discussing the results of the various experiments conducted in the present dissertation, the next section will first outline the challenges of the Dutch gender system for the learner (either L1 or L2), after which the most important theories with respect to the acquisition of a second language and an L2 gender and their corresponding predictions will be summarized in §8.2.

8.1. The challenges of the Dutch gender system

The Dutch gender system is rather special and not entirely comparable to other Indo-European gender systems. The Dutch two-way gender system is expressed by definite articles (‘de_{COM}’ versus ‘het_{NEU}’), attributive adjectives (with the presence or absence of the suffix *–e* on adjectives modifying common versus neuter nouns respectively), and a subset of pronouns (e.g. ‘die_{COM}’ versus ‘dat_{NEU}’). As opposed to some Romance languages (e.g. Spanish and Italian), Dutch nouns generally do not display any morphophonological cues that can be used to derive the underlying gender of the noun. Knowledge of the gender category, however, is a prerequisite for being able to apply gender agreement rules. Due to the covertness of Dutch gender, the language learner is forced to acquire the system based on the agreement targets, which are the only clear indication of the presence of a gender system and of the gender category of the noun. In addition to the fact that common and neuter nouns themselves do not carry any differentiating cues with respect to their gender, the two genders do not occur equally often in the input. Common gender, as the name already suggests, has evolved from a collapse of the original masculine and feminine genders and hence comprises around 75% of all Dutch nouns, while only 25% of Dutch nouns have neuter gender.

Considering the fact that the language learner mainly has to rely on the agreement system to acquire Dutch gender, it is additionally problematic that the agreement cues are not always saliently present in the input. The definite determiners are phonetically differentiating and clearly noticeable (‘de_{COM}’ versus ‘het_{NEU}’), but the schwa-ending that is either present (common) or absent (neuter) on the attributive adjective is less noticeable, especially in spontaneous speech. In addition to these factors – asymmetry, skewedness and opaqueness – which make the acquisition of Dutch gender difficult, there are further idiosyncrasies of the system rendering it yet further confusing to the learner. The common definite determiner, the common schwa-ending on adjectives, and the common pronouns are identical to the determiners, adjectives, and pronouns used to refer to all plural nouns. On the other hand, all Dutch nouns, regardless of their gender, receive neuter gender-marking when used as diminutives. The agreement cues that are present in the input are thus not reliable for the language learner, because neuter nouns can receive common gender-marking when used in the plural and common nouns can receive neuter gender-marking when used as diminutives. Increasing this asymmetry of cues is that gender-marking on determiners is clear in definite, but not in indefinite noun phrases where the indefinite determiner 'een' precedes both common and neuter nouns. On the other hand, gender-marking on the adjective is only distinctive when an indefinite noun is modified, while in definite NPs, the common schwa-ending is used on the adjective regardless of the gender of the nouns they modify. Gender-marking in Dutch is thus either present on the determiner (definite NPs) or on the adjective (indefinite NPs) or absent entirely in indefinite NPs lacking an adjective (see also Roodenburg & Hulk (2008) for a detailed description of the puzzles on grammatical gender in Dutch).

Considering the lack of predictability, systematicity, reliability and saliency of gender and gender cues in Dutch, it is not surprising that the Dutch gender system is acquired relatively late even by native speakers (i.e. around the age of 6) as compared to the acquisition of gender by native speakers of German, French, and Spanish, who appear to master their gender system around the age of 3 (Van der Velde, 2004; Szagun et al., 2007; Lew-Williams & Fernald, 2007). The lack of predictability, systematicity, reliability and saliency of gender cues provides an interesting starting point for the investigation of both native and second language processing of grammatical gender in Dutch. If native speakers acquire their gender system with relative difficulty, is it still processed with the same speed and ease as other grammatical phenomena in the language? And are late L2 learners of Dutch from a typologically different language able to acquire and process the Dutch language and its gender system in a way similar to native speakers? This latter question is especially interesting with respect to the Dutch gender system since the late L2 learners tested in the present dissertation have mostly learned L2 Dutch through listening and using it in everyday life. Hence, we are most likely investigating the results of implicit learning strategies where the gender system is entirely covert.

8.2. The challenges of L2 gender from a theoretical perspective

The questions of whether late L2 learners can ever attain a native-like level and whether the processing of a second language is only partially or potentially fundamentally different from the processing of a native language can be viewed from various theoretical frameworks which in their turn, have specific predictions. The purpose of the studies presented in the current dissertation was not to prove or disprove one of these frameworks as such. Instead, we aimed at evaluating specific predictions of these different theoretical frameworks with respect to the processing of grammatical gender in Dutch by native and late L2 learners based on the results of various online experiments. More specifically, the aim was to investigate to what extent L2 learners from a typologically different background can attain native-like levels of processing. Previous research by Sabourin (2003) has suggested that L2 learners not only need to have an L1 gender system to be able to attain native-like proficiency in L2 gender, but the gender systems of the L1 and L2 additionally need to be similar. In Sabourin's study, similarity between gender systems was, however, potentially confounded with proficiency level: German-Dutch L2 learners (whose L1 gender system is similar to the Dutch gender system) outperformed Romance-Dutch L2 learners (whose L1 gender system differs from the Dutch system), but they also appeared to have an overall higher level of proficiency. The present studies allowed testing potential competition effects in the acquisition of an L2 gender system by investigating L2 learners from a typologically different language background, but with an L1 gender system, similar in some ways to the L1 Romance group studies by Sabourin. The composite proficiency measure used additionally allowed for a detailed examination of the effects of proficiency on the acquisition and processing of an L2 gender system and the potential influence of the native language on these processes.

The only theoretical framework which makes specific predictions both with respect to the acquisition of the covert and asymmetrical gender system and the influence of L1-L2 similarity is the Competition Model (Bates and MacWhinney, 1987). This model not only allocates an important role to input frequency, but also to the reliability and saliency of structures and cues in the input. The Competition Model assumes that acquisition occurs on the basis of syntactic, semantic and morphological cues in the input. As discussed above, in the case of the Dutch gender system, acquisition mainly has to rely on morphological cues that are present on the agreement targets. The cues on agreement targets are available and frequent, but they are not reliable in that all nouns can receive common gender-marking in the plural and all nouns receive neuter gender-marking when used as diminutives. Additionally, gender-marking on the determiner is only present in definite NPs and attributive adjectives are only marked for gender in indefinite NPs. The acquisition of the Dutch gender system should, according to the Competition Model, additionally be affected by a difference in saliency between definite determiner-noun constructions and indefinite adjective-noun constructions. Gender-marking on definite determiners is more salient than the schwa-ending on attributive adjectives and hence definite determiner-noun constructions are likely to be acquired prior to indefinite adjective-noun constructions. The Competition Model predicts that, in many ways, L2 acquisition should mirror L1 acquisition, since it relies on the same cues in the input.

However, the Competition Model also allocates an important role to the native language of the L2 learner (MacWhinney, 2008). Structures that are different in the two languages are more challenging to acquire and use. This difficulty is explained mainly as a result of competition between the native and second language. Features that are unique or similar in the two languages are relatively easy to acquire; there is argued to be little role for positive transfer, so that similar constructions do not fare better than ones which do not arise in L1. Specifically related to the present study, determiners are unique for the Polish-Dutch late L2 learners tested in the present studies while adjectival inflection is different between the two languages. Consequently, based on the predictions made by the Competition Model, one would expect to find that determiners are acquired before the adjectival inflection system, because the latter system is negatively affected by competition from the native language. Related to the issue of competition and/or transfer, the eye tracking studies presented here investigated both processing of items that have the same gender in Dutch and Polish and items that do not share gender in the two languages. If Polish-Dutch late bilinguals rely on their L1 while processing the L2 and can use the L1 for positive transfer, the same gendered items should be easier to process. This might be particularly true when these concern definite determiner-noun constructions, since the cues for this structure are more salient and more reliable, than indefinite determiner-adjective-noun constructions. The acquisition of L2 Dutch gender by learners with L1 Polish thus allows us to test the predictions of the Competition Model with respect to the acquisition of L2 structures that are unique, different, or similar in the two languages.

A similarly important role for frequency has been advocated by researchers who view language learning as an emergent process, such as emergentists and proponents of usage-based accounts of language acquisition. According to Tomasello (2003), for example, there is a socio-cognitive need to understand others and this positively affects language learning. As opposed to other gendered languages, however, the use of gender-marking in Dutch is not a prerequisite for making yourself understandable nor is it necessary to know gender in order to understand others. This is mainly related to the fact that anaphoric reference, which in most gendered languages depends on gender for disambiguation, is very rarely used in Dutch. This is because the common gender does not actually have an anaphor, but rather is referred to by the historical feminine and masculine forms. These are obvious for living creatures, but not for other items referred to with common gender, like spoon (‘de_{COM} $lepel_{COM}$’) and apple (‘de_{COM} $appel_{COM}$’), because it requires the additional knowledge of the original masculine and feminine genders, a distinction that native Dutch speakers do not use consistently anymore (Audring, 2009). Nevertheless, the importance of frequency, as advocated by usage-based accounts, would predict that the more L2 exposure there has been, the more input the language learner has received and the better he or she can put the pieces of the gender system puzzle together. Similar to the predictions made by the Competition Model, a proponent of the usage-based account would expect learners to acquire common gender, which is most frequent, before neuter gender. Children have indeed been shown to overgeneralize the default common determiner *‘de’* until a relatively late age and the observation that even very early or simultaneous Turkish-Dutch bilinguals fail to (fully) attain the Dutch gender system (Seton, 2009) might result from a frequency effect. Early simultaneous bilinguals acquire two languages from birth and hence their input time is divided between the two languages. The relatively lower frequency of the input of the Dutch language, when compared to native children who grow up in a monolingual situation, might explain the Turkish early bilinguals' difficulties with the Dutch gender system, and specifically their difficulties with the less frequent neuter gender. If frequency and input are crucial in acquiring a language, as predicted by both usage-based accounts and the Competition Model, we would expect late L2 learners to acquire the more frequent common gender before neuter gender.

Finally, neurocognitive models suggest a processing view on the differences or similarities between L1 and L2 speakers. The Shallow Structure Hypothesis (SSH: Clahsen and Felser, 2006a), for example, suggests that late L2 learners over rely on semantic information and are unable to process syntactic structures in an automatic and native-like fashion. The shallow and less deep syntactic processing of the L2 would result in slower processing due to a lack of automaticity. Interestingly, the authors do not believe that the specifics of the L1 hamper L2 acquisition and claim that "whereas L1 influence of phonological and lexical properties is well attested, non-local dependencies do not seem to be susceptible to transfer effects in L2 processing. The absence of transfer effects in this domain could be due to a mapping incompatibility between learners' L1 and L2 representations, e.g. if the latter are 'shallow' in the sense described earlier." (Clahsen & Felser, 2006b: 566). Assuming grammatical gender would be regarded as a non-local phenomenon, the SSH would predict no effects of L1 transfer since the late L2 learners would not have an L2 gender system to transfer the L1 grammar to.

The hypothesis that there are no syntactic L1 transfer effects clearly conflicts with the claims for effects of the L1 suggested by the Competition Model, and also the initial influence of the L1 in L2 acquisition as proposed by Full-Transfer Full-Access (FTFA) models (e.g. Bruhn de Garavito & White, 2000). The FTFA model predicts that beginning learners, who are still in the initial stages of acquiring their L2, do not have enough knowledge of the new language to be able to use its gender system. Instead, they tend to transfer gender knowledge from their mother tongue, which may help or hinder them in dealing with the new language. Only when they have reached a more advanced proficiency level in the L2 are they able to use and process the L2 gender system without (correctly or incorrectly) transferring gender information from their L1. This latter model would thus predict the presence of L1 transfer to depend on proficiency level of the late L2 learners, with L1 transfer of gender categories only occurring in lower proficient L2 learners of Dutch. The current dissertation examined both low and high proficient learners to disentangle the issue of whether the L1 gender system does not influence L2 acquisition (SSH), whether the L1 gender system only affects L2 gender acquisition and processing for less proficient L2 speakers (FTFA), or whether there is an overall difference in the acquisition of L2 features that are either unique, different, or similar in the L1 and L2 (Competition Model).

In addition to the SSH, another neurocognitive model focussing on the online processing of language is the Declarative/Procedural Model (DP: Ullman, 2004). The

DP model would hypothesize that, at least in the initial stages of L2 acquisition, L2 learners mostly rely on their declarative memory system, which is specialized in associative learning. The acquisition and processing of grammatical rules, such as the Dutch gender system, however, involves more implicit learning and that is something the procedural system is specialized in. Due to the L2 learners' (initial) overreliance on the declarative system, they might acquire the rules of the gender system by association rather than by computing sequences. This might result in the use of lexical processing strategies while listening to constructions containing L2 grammatical gender. Several models of sentence processing have additionally focussed on the timing of these different processes. Some claim syntactic information is processed prior to any other information; while others suggest semantic and syntactic information interact continuously during each stage of processing. These competing views can be investigated by using online measures that are very sensitive to the timing of the different processes involved, such as eye tracking and event-related brain potentials.

In sum, the FTFA would predict an interaction of L1 transfer and level of proficiency, with beginning learners transferring gender from their mother tongue and advanced learners potentially having acquired the L2 system up to a native-like level. According to both the emergentists and proponents of the Competition Model, we would expect to find a frequency and/or availability effect, with common gender being acquired prior to neuter gender. Such an effect is likely to be related to proficiency level, with lower proficiency (or: less proficient) learners only having acquired common and high proficiency (or: more proficient) learners potentially having acquired both genders. The Competition Model would additionally predict determiner agreement to be acquired before adjectival agreement, since determiners are more saliently present in the input and since they are unique to the L2 while the adjectival system might suffer from competition from the L1. The SSH would predict slower processing in the late L2 learners due to a lack of automaticity. The DP model suggests differences between low and high proficient learners in that the former group might use lexical strategies to process both semantic and syntactic information.

In order to make these predictions more concrete, it is necessary to know how native speakers behave. The next section will outline how native Dutch speakers process gender and relate this to the predictions outlined above. The data from the L2 learners will be discussed in detail in §8.4.

8.3. Native processing of Dutch gender

Listener's eyes are strongly guided by information in the speech stream (e.g. Tanenhaus et al., 1995). The question examined in *Chapter 4* was whether grammatical gender information in Dutch, despite its non-transparent and asymmetrical system, can also be used by native speakers to facilitate comprehension. In other words, after hearing the common determiner *'de'*, will native speakers stop fixating on a picture containing a house, because the word house is neuter in Dutch (‘het$_{NEU}$ huis$_{NEU}$’) and hence cannot be preceded by a common determiner? When native Dutch participants were asked to, for example, click on *‘de rode appel’* or *‘een rode appel’* (‘the$_{COM}$ red apple’ or ‘a red$_{COM}$ apple’), they looked relatively more towards the competitor picture when this depicted an item which shared common gender with the target, than when it had neuter gender. No such effect was found when the target was of the less frequent neuter noun class. This difference between the two types of gender is possibly due to the unreliability of the cue provided by the determiners, including the fact that neuter gender-marking, such as the neuter definite determiner *‘het’* and bare adjectives, can precede all nouns when used as diminutives. In other words, when asked to click on *‘het rode’* or *‘een rood’* (‘the$_{NEU}$ red’ or ‘a red$_{NEU}$’), the phrase can still be followed by the noun *‘appeltje’*, which is the diminutive form of the common noun *‘appel’* (‘apple$_{COM}$’). This first experiment thus showed that native Dutch speakers can use common, but not neuter gender-marking to facilitate comprehension.

Similar to at least common gender-marking, sentence context is used to predict and anticipate upcoming words. In *Chapter 5*, the interaction between gender agreement and semantic expectancy as measured by cloze probability was investigated in order to determine whether gender processing can be influenced by the predictability of the nouns used. The results showed that nouns are more easily retrieved from the mental lexicon when they can easily be predicted on the basis of the preceding context, as reflected by a smaller N400. Interestingly, the violations of gender agreement embedded in sentences, such as *‘*het tuin’* (‘*the$_{NEU}$ garden$_{COM}$’) or *‘*de huis’* (‘*the$_{COM}$ house$_{NEU}$’), was responded to more quickly when the context was highly constraining, as reflected by an earlier P600. Semantic and syntactic information thus interact in later stages of processing and gender concord violations on nouns can only be repaired or re-analysed once they have successfully been retrieved from the mental lexicon, the ease of which depends on the preceding context.

The P600 response to gender violations did not differ from the P600 in response to more salient morphological violations of finiteness for native Dutch speakers, even though the evidence for agreement in this construction is frequent, salient and consistent, suggesting that both types of violations are treated as non-semantic phenomena that elicit comparable processes of repair and/or re-analysis (see *Chapter 7*).

Neuter is not common

The relative difficulty with neuter gendered nouns even for native speakers was visible across experiments. The proficiency measures outlined in *Chapter 3* additionally revealed that these participants produce neuter gender-marking errors in their speech.

The eye tracking experiment clearly revealed that Dutch natives only partially make use of gender for referent anticipation. Common gender-marking allowed participants to exclude visual competitors, while neuter gender-marking induced a cohort effect or even an opposite pattern when compared to the exclusion effect for common gender-marking. In other words, while a same-gender competitor slows down the process of locating a common target, a same-gender competitor appeared to speed up the process of locating a neuter target. The observed preference for the target, when the gender information in the speech stream could not disambiguate between neuter target and neuter competitor, is very difficult to explain, but was strengthened by the observation that the observed colour effect (i.e. later responses when a competitor shared colour with the target) decreased when target and competitor additionally shared gender (also see §4.4. for a more detailed discussion). In line with previous research (Cubelli et al., 2011), we hypothesized this effect might reflect the fact that nouns from the same gender category activate each other, which in turn facilitates processing and speeds up reaction times. It is not surprising that this facilitation effect is especially visible in the eye movement data for nouns from the neuter gender category. Neuter nouns make up only 25% of all Dutch nouns and due to the smaller number of nouns there are to activate, activation takes place at a faster pace and hence speeds up the recognition process. Future research might further investigate this issue, potentially also in other languages that have gender categories that differ in size (e.g. Polish).

Despite the asymmetry between common and neuter gender, however, native Dutch speakers showed the same ERP response to violations of more frequent common and violations of less frequent neuter gender and they did not show statistically different responses to violations on definite determiners and less salient attributive adjectives. The results of the ERP experiment additionally revealed that the brain response to gender violations was all but inferior to the response to more salient violations of finiteness (see *Chapter 7*). These combined findings show that despite the unreliability of the cues in the input (with common gender-marking being identical to the gender-marking used to modify plural nouns and neuter gender-marking being used for all nouns when used as diminutives), native Dutch speakers show a robust effect in response to gender violations.

Pre-lexical activation of gender

Many studies on gender activation and representation have supported a post-lexical approach to gender processing (see *Chapter 4* for more information). For example, behavioural studies have shown that natives are faster to respond to nouns that are preceded by congruent gender-marking. Similarly, eye tracking studies have reported inhibitory gender effects in response to nouns with phonological overlap, but not when the onset was unique to the target noun (Dahan et al., 2000). These results suggest that gender information is only activated after the lexical item has been retrieved from the mental lexicon. The results of the present eye tracking study, on the other hand, clearly reveal a gender effect for common nouns before the native participants had heard the unique phonemic onset of the target noun. These results thus suggest that gender information is, at least partially, activated automatically before hearing the noun in the speech stream among L1 speakers.

These results, in combination with the facilitation effect of same-gender nouns outlined above, imply that gender information of the objects on the screen is made available before participants have encountered the gender-carrying noun itself in the speech stream and these combined results argue in favour of a pre-activation account of gender processing.

The results with respect to native processing clearly reveal that gender is used to facilitate comprehension. Despite the skewed distribution and the unreliability of the gender agreement cues in the input, native speakers of Dutch have a very stable representation of gender, as revealed by a robust effect in response to both common and neuter gender violations. There is, however, an asymmetry between common and neuter nouns, with only the more frequent and disambiguating common gender-marking being used to anticipate upcoming referents in the speech stream. The observed asymmetry between common and neuter nouns is, however, only present in anticipating upcoming referents and not when trying to repair or re-analyse after hearing gender violations on common and neuter nouns. Dutch appears to be a very special case in this respect to the processing of gender and it may therefore be especially difficult for L2 learners to acquire. The next section will discuss the results related to these late L2 learners of Dutch and the extent to which they are sensitive to L2 gender-marking as well as the extent to which they are able to use it.

8.4. Late L2 processing of Dutch gender

Language comprehension in Dutch natives is facilitated by both semantic and morphological context. The question addressed in *Chapter 6* was whether late L2 learners of Dutch can use morphological information in the L2, i.e. gender-marking on determiners or adjectives, to facilitate noun selection as native speakers do. Alternatively, they may tend to rely on gender information as instantiated in their mother tongue and this may affect gender processing in the L2. The results in *Chapter 6* indicate clearly that unlike native speakers, late L2 learners were not able to inhibit neuter nouns after hearing the common definite determiner *'de'* or an inflected common colour adjective such as *'rode'* ('red$_{COM}$'). Looks towards the target picture were also not affected by the presence of a competitor that shared gender with the target in Polish. This was the case regardless of the L2 learners' proficiency level, demonstrating that gender from the L1 was neither correctly nor incorrectly transferred to the L2. However, late Polish-Dutch learners were slower in looking towards the target picture when it had the minority neuter gender in Polish, regardless of the gender of the target picture in Dutch. This result is difficult to explain, but suggests that while L1 gender is not transferred, it *is* activated while listening to speech in the second language (this finding is discussed in more detail below).

Late L2 learners are thus not able to rapidly use gender-marking in the second language in a facilitatory fashion. This does not necessarily mean, however, that they are unaware of or unable to process L2 gender after the noun is presented. The fourth and final question, which was addressed in *Chapter 7*, was whether late L2 learners can process morphological structures in a native-like manner and aimed at investigating the influence of L2 proficiency and the availability and reliability of morphological structures in the input. The results suggested that, at lower proficiency levels, late L2 learners are not able to process gender violations in a native-like manner. They are, however, capable of processing morphological structures that are reliably present in the input (i.e. finiteness constructions), despite their relatively lower proficiency. The acquisition and processing of a non-transparent and asymmetrical gender system, on the other hand, appears to depend more crucially on proficiency and amount of exposure. Only high proficient learners of Dutch showed sensitivity to violations of gender agreement and, interestingly, only for the default common gender. In other words, high proficient L2 learners of Dutch engaged in processes of re-analysis or repair after hearing a violation such as '**het tuin*' ('*the$_{NEU}$ garden$_{COM}$'), but not after hearing a violation such as '**de huis*' ('*the$_{COM}$ house$_{NEU}$').

Common is more frequent, more reliable, and easier to acquire

As mentioned in the previous section, the relative ease of acquiring and processing common as opposed to neuter gender was visible across experiments. Even native Dutch speakers produce neuter, but not common, gender-marking errors in their speech (see *Chapter 3*).

The eye tracking experiments outlined in *Chapter 6* additionally showed that although late L2 learners of Dutch are not able to use common or neuter gender to facilitate language comprehension, they did show a relative ease in fixating on the target when it was a common as opposed to a neuter gendered noun. In other words, late L2 learners showed difficulty in acquiring and processing the Dutch gender system in general, but they had more difficulty in using and processing the less available and reliable cues for neuter gender in particular. The slowed processing of neuter as compared to common suggests that these late learners are sensitive to the difference between the two genders, but they are not able to use it pre-lexically to anticipate upcoming referents.

The proficient L2 speakers' greater awareness of common gender is underlined by the ERP results. Even though the L2 learners knew the gender of the neuter nouns in the ERP experiment, they did not attempt to repair or re-analyse gender violations on these neuter nouns (as in '**de huis*'; '*the$_{COM}$ house$_{NEU}$'). The finding of a delayed and reduced P600 in high proficiency L2 learners in response to violations on common nouns only, such as '**het tuin*' ('*the$_{NEU}$ garden$_{COM}$'), corresponds to the fact that the cues for common gender in Dutch are more available and more reliably present in the input. The importance of input frequency in this respect is additionally strengthened by the finding that the positive effect of amount of L2 exposure mainly affected the use and processing of common nouns (more details on this effect are discussed below). Knowledge and processing of neuter gender, on the other hand, appeared to remain problematic even after years of exposure.

It is interesting to note that some late L2 learners appear to perform better in producing and comprehending common and others are better at producing and comprehending neuter. Hawkins (2001) has previously suggested that L2 learners establish a default and one might suggest that the learners have all taken a different strategy with some of them using common and others neuter as the default gender. This, however, seems highly unlikely when considering that some of these learners have chosen to base their default on an input frequency of about 25%. Taken together, these results suggest that, even though not all speakers use the same strategies in acquiring gender, they are overall better at processing the more frequently and reliably occurring common gender. The asymmetry in the input may make it more difficult for (late) L2 learners to acquire, in addition to cue validity, reliability and saliency. Furthermore, it is very well possible that this asymmetry manifests itself in the input from Dutch natives in ways which have not yet been explored. Either way, the results of the studies presented here again reveal the wide variety of acquisition processes and eventual outcomes in second language acquisition and future research might try and explore these individual differences in further detail.

No transfer, but co-activation

One major question in SLA research is whether having a gender system in the mother tongue is a prerequisite for being able to fully acquire an L2 gender system or whether anyone can in principle eventually become native-like in an L2 (the Failed Functional Features Hypothesis (FFFH) versus the Full-Transfer Full-Access models (FTFA), see §1.2.1.). Since the present dissertation only tested a group of late L2 learners from a gendered language (i.e. Polish), it is not possible to answer questions pertaining to whether or not the L1 functional feature make-up can help or prevent learners from becoming fully native-like. It is, however, interesting to examine whether, as predicted for example by the FTFA, late L2 learners activate and use lexico-syntactic knowledge such as gender from their native language, especially in earlier stages of L2 acquisition. As opposed to the FTFA, the Shallow Structure Hypothesis (SSH) would predict no influence of the L1 grammatical gender system due to the unavailability of an L2 grammar system to transfer the L1 grammar to.

The eye tracking experiment (*Chapter 6*) showed that, at least for non-cognates and for languages that differ typologically, the L1 gender system is not transferred to the second language. L1 gender is, however, activated, as suggested by slower responses of the late L2 learners in response to nouns that had the minority neuter gender in Polish. This effect is rather difficult to explain, but might be related to the fact that the neuter gender class is the smallest gender category in Polish and additionally has a more restricted phonological density as compared to both masculine and feminine nouns in Polish (Dąbrowska, 2004). The true origin of this minority class co-activation effect remains speculative, but future research might look into the nature of smaller gender classes and a specific experiment might address the question of whether Polish neuter nouns would cause this exact same difficulty in native processing of the Polish language.

The activation of a neuter gender class from the mother tongue while listening to the L2 adds evidence to the pre-lexical activation account of gender information discussed in §8.3. While advanced late learners from a typologically unrelated language do not appear to transfer gender from their mother tongue to the L2, at least with respect to non-cognates, they do appear to activate (the minority neuter) gender from their mother tongue even before hearing the noun in the speech stream. The fact that gender information from the L1 is activated for non-cognate nouns additionally suggests that gender information does not necessarily require the output of the lexical processing system for its operation. Together with the finding that native speakers use common gender-marking even before the (onset of the) noun has been heard, the results of the present dissertation strongly suggest that L1 (and possibly L2) gender information from the lexicon is activated before the noun has been retrieved from the mental lexicon and provide evidence for a more automatic, pre-activation account on gender processing.

The absence of an L1 transfer effect suggests that the late L2 learners tested in the present studies appear to be able to dissociate the Dutch items from their Polish equivalents. From the FTFA point of view, a reason for the absence of L1 transfer effects might be that the majority of subjects tested were rather highly proficient, as was also suggested by the proficiency measures outlined in Chapters 2 and 3. The results, however, also suggest that the Polish-Dutch L2 learners are not proficient enough to use gender in the L2 to facilitate language comprehension. Another explanation for the absence of L1 transfer might be related to the typological differences between Polish and Dutch. As outlined in Chapter 1, Polish-Dutch bilinguals do have a gender system in the native language, but this gender system is not at all compatible with the Dutch one and Dutch gender largely has to be acquired on an item-by-item basis. The role of L1 transfer and/or interference might thus be more important for languages that are similar. Effects such as those reported by Sabourin & Stowe (2008) might thus reflect a proficiency effect instead of an effect of the mother tongue, suggesting that similarity between the two languages is not necessary except that it facilitates overall acquisition (also see below for more details). Future research should further investigate the issue of proficiency and L1-L2 similarity by examining L2 gender acquisition in proficiency-matched speakers from a similar language background that either have or do not have a gender system in their mother tongue.

The Competition Model would additionally hypothesize an effect of saliency and/or competition from the L1 which would have resulted in a difference in acquiring and processing the determiner system as opposed to the adjectival system. Determiners are unique to the L2 and more saliently present in the input, which would make them easier to acquire than the adjectival system, which is different in the two languages and less saliently present in the input. There was no effect of structure among highly proficient speakers, suggesting that they have acquired both the unique and salient determiners as well as the less salient adjectival system.

Enough time for gender

While L1 transfer might only play a role in L2 acquisition and processing for languages that are typologically similar, the similarities between languages and their gender systems does not appear to be a prerequisite for native-like processing of gender violations in Dutch, as previously suggested by Sabourin (2003). Instead, at least for Polish-Dutch L2 learners, native-like processing of grammatical gender in the L2 appears to be a matter of time and a matter of proficiency.

Regardless of their proficiency in Dutch, the late L2 learners tested in the experiments reported here showed a repair or re-analysis process after hearing phrases such as *'*ze heeft uitnodigen'* ('*she has invite') where an infinitive is heard when this should be a past participle. Violations of gender agreement, on the other hand, were only repaired or re-analysed by highly proficient L2 learners and only when these violations affected the default common gender. Despite the equality of both the finiteness and the gender constructions with respect to L1 and L2 similarities, L2 learners showed severe difficulty with processing gender as opposed to finiteness. In other words, they are proficient enough to process finiteness structures that are different from the structures used in their mother tongue, suggesting that they are able to process difficult and competing finiteness violations in a more or less native-like fashion. Probably due to the less available and reliable cues in the input, grammatical gender thus takes a relatively long time to acquire. The importance of input frequency was additionally confirmed by the fact that the sensitivity in response to gender violations was dependent on the amount of time the participant had spent in the Netherlands. Amount of exposure not only positively influenced the processing of gender violations, but the learners' ability to correctly assign gender to nouns and to use the correct gender-marking in production also increased as more time had been spent in the Netherlands (also see §3.2 and §3.5).

These combined results strongly suggest that, at least in a non-transparent and asymmetrical system, amount of exposure is one of the key factors of success in both using and processing an L2 gender system. As predicted by the Competition Model, structures that are more frequently and reliably present in the input are acquired earlier, while it takes a longer time and more practice before late L2 learners are able to process the covert and asymmetrical Dutch gender system.

8.5. Summary & concluding remarks

The main question addressed in this dissertation was whether late L2 learners from a typologically different language are able to acquire and process the Dutch language, and in particular the covert Dutch gender system, in a similar way as native speakers. Across experiments, the overall findings first of all underline that the gender system in Dutch is different than it is in other gendered languages. Dutch native speakers only partially tend to use gender for referent anticipation. The results of the eye tracking experiment show that common definite determiners and common inflected adjectives in indefinite constructions help to eliminate competitors, while neuter prenominal gender-marking results in a cohort effect. This suggests that the distributional asymmetry has a correlate in the underlying linguistic structure for native speakers and neuter gender-marking appears to serve as a non-disambiguating and hence non-facilitatory cue. In all probability, neuter gender cannot be used to disambiguate upcoming nouns in the speech stream because every noun can be preceded by neuter gender-marking when used as diminutives. This asymmetry is not present in the ERP data, where natives do not appear to differentiate between violations on common and neuter nouns and neither do they reveal differences in the processing of highly salient definite determiner-noun constructions and less salient indefinite determiner-adjective-noun constructions. There may be an asymmetry present in the underlying structure, but violations of both common and neuter gender-marking are processed the same by native speakers of Dutch as well as violations on definite determiners and attributive adjectives. Similarly, the speeded up repair or re-analysis process for words that are easily predictable on the basis of the preceding sentence context was unaffected by the gender of the noun. These results, together with the finding that native Dutch speakers process gender violations similar to other morphological violations, suggest that gender agreement is a very robust phenomenon despite the unreliability of the cues in the input.

Where Dutch natives appear to have a perfectly reliable response (P600) despite the skewed distribution, the asymmetrical input appears to cause problems for the late L2 learners. They were not able to pre-activate L2 gender to facilitate language comprehension and, in line with both usage-based accounts and the Competition Model, a frequency and/or availability effect was found, with common gender being acquired prior to neuter gender. Late L2 learners of Dutch were only able to process violations of the frequent common gender, but not of the less frequent neuter gender category. The lack of consistent input due to the asymmetrical Dutch gender system and its abstract rules thus appears to cause severe difficulty in the acquisition of the Dutch gender system and especially the acquisition of the infrequent and unreliable neuter gender. No evidence was found in favour of the Competition Model's hypothesis, however, that determiners are acquired before the adjectival system, since determiners are unique to the L2 and the adjectival system might suffer from competition from the L1.

Interestingly, opposing the predictions made by the FTFA and the Competition Model, no effects of L1 transfer or competition were found, regardless of the proficiency level of the L2 learners. If the late L2 learners did make use of their L1 gender system in initial stages of L2 acquisition, they now all appear to have gone past that stage. Hence, we cannot provide evidence in favour or against the predictions made by the FTFA and the DP model concerning the use of the L1 gender system and the use of the declarative system in lower proficiency speakers respectively. Both the lack of L1 transfer effects and competition effects might have resulted from the general higher level of proficiency of the L2 learners under investigation. This interpretation would also explain the absence of a lexical processing strategy in response to L2 grammatical structures in initial stages of L2 acquisition.

Although most of the late L2 learners appeared to be highly proficient in their L2, or at least beyond the beginning stage of L2 acquisition, differences in proficiency level significantly impacted the late L2 learner's brain activation. Sensitivity towards gender violations was only present in highly proficient learners and only for violations on common nouns. The sensitivity towards these common violations was delayed and reduced when compared to native processing, which might reflect a lack of automaticity as has been predicted by the SSH. The lack of automaticity would also explain why higher proficiency learners can recognize violations of common gender-marking, as reflected by the P600, but appear to be unable to use the knowledge that they have of gender to anticipate the referent. In other words, they recognize errors in lexical clusters, but they do not internalize gender to make use of it for disambiguation. Related to this issue is the finding that highly proficient speakers who perform fairly well on assigning neuter gender do not have a P600 in response to violations with regard to these neuter nouns. Even though they appear to be able to acquire neuter gender assignment (which is lexical-semantic information); they appear to be unable to recognize corresponding agreement rules (syntactic information). This discrepancy likewise suggests a form of shallow or less automatic processing of gender in the L2.

The ability of highly proficient L2 learners from a typologically different language background to process gender violations on common nouns shows that similarity between the two languages is not a prerequisite for successful L2 gender acquisition, but most likely facilitates overall acquisition of an L2. Late Polish-Dutch L2 learners are able to process frequent and salient morphological violations similar as native speakers do. The processing of morphological violations is influenced both by proficiency and cue availability and reliability, with only highly proficient L2 learners being able to process violations of common, but not neuter, gender-marking. High proficient L2 learners can overall tell, in the frequent common gender, when it is incorrectly used, but they are unable to do so in the infrequent case. Taken together, the effects of input frequency and reliability of the morphological cues in the input, as outlined by the Competition Model, appear to be the best way to account for the lagged and skewed acquisition of this L2 system.

In sum, the present dissertation shows that native speakers use all information present in the input to facilitate language comprehension, including morphological information such as gender. Additionally, native speakers of Dutch process gender violations in a similar way as other morphological structures, such as violations of finiteness. Even in such a confounded gender system, native Dutch speakers thus appear to be able to show robust responses to all violations, regardless of frequency and locality. Before encountering the noun, however, the more frequent common determiner only helps to exclude less frequent neuter nouns, but not the other way round. These results show that grammatical learning in the L1 is different between different languages. Nevertheless, native Dutch speakers, despite all the pitfalls, appear to develop a good system.

Late L2 learners from a typologically different language are not able to use gender information in the same facilitatory way as native speakers do, but high proficient L2 learners are able to show a similar repair process in response to violations on common gender as native speakers do. If late L2 learners have spent enough time in the country of their L2 and if they are proficient enough, they are able to recognize gender errors on common nouns after the fact, as suggested by the P600 that was only present in response to violations on common nouns. They are, however, not able to use it to facilitate comprehension, suggesting that there is no pre-activation of L2 gender information. Native-like processing thus appears to depend highly on the proficiency level, the availability and reliability of the structures in the input, and the amount of time participants have had to listen and use their second language.

Despite the residual response to grammaticality in Dutch natives, gender in Dutch is not used for the only purpose that it is generally used for, which is anaphoric reference. The fact that gender is never used to identify a referent in Dutch, and hence never used to disambiguate referents, might be the reason why L2 learners have so much difficulty acquiring the system. In addition, Dutch is in a process of change, which makes it even more difficult to acquire, but at the same time interesting to follow. Considering the results and the previous speculations of why it is so difficult to acquire for L2 learners and even early bilinguals, there is still one unresolved issue that future researchers should try to disentangle: how come natives do get it? The answer to this question might be related to the on-going change and the potentially increasingly simplified Dutch gender system. If the changes over time can explain the difficulties in learning, one might expect to find differences in the use and processing of Dutch gender between younger and older speakers of the language. Dutch gender might, after all, be a grammatical system that may have outlived its usefulness.

REFERENCES

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: an ERP study. *Language and Cognitive Processes, 18*, 405-430.

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time couse of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Aristotle. (1407, Translated with an Introduction and Notes by H.C. Lawson-Tancred in 1991.). *The Art of Rhetoric.* London: Penguin Classics.

Audring, J. (2009). *Reinventing Pronoun Gender.* Utrecht: LOT.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Baayen, R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*, 12-28.

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

Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience, 17*, 137-153.

Barr, D. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, Special Issue: Emerging Data Analysis, 59,* 457-474.

Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*, 27-30.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of Language Learning* (pp. 157-194). Hillsdale, NJ: Lawrence Erlbaum.

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception and Psychophysics, 57*, 847-862.

Blom, E., Polišenskà, D., & Weerman, F. (2008). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research, 24*, 297-331.

Bloomfield, L. (1933). *Language.* New York: Henry Holt and Co.

Bobb, S., & Mani, N. (2012). Categorizing with Gender: Does Implicit Grammatical Gender Affect Lexical Access in Young Children? *Talk presented at International Conference on Infant Studies (ICIS XVIII).* Minneapolis, Minnesota, June 7-9.

Bongaerts, T. (1999). Ultimate attainment in foreign language pronunciation: The case of very advanced late language learners. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 133-159). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Brouwer, S., Unsworth, S., & Mak, P. (2010). Processing grammatical gender in Dutch. *Talk presented at Generative Approaches to Language Acquisition North America (GALANA 4).* Toronto, Canada, Sep 3.

Bruhn de Garavito, J., & White, L. (2000). L2 Acquisition of Spanish DPs: the Status of Grammatical Features. In C. Howell, S. Fish, & T. Keith-Lucas (Eds.), *BUCLD 24: Proceedings from the 24th Annual Boston University Conference on Language Development* (pp. 164-175). Somerville, MA: Cascadilla Press.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning, 39*, 535–594.

Chomsky, N. (1995). *The minimalist program.* Cambridge, Massachusetts: MIT Press.

Clahsen, H. (2008). Behavioral Methods for Investigating Morphological and Syntactic Processing in Children. In I. Sekerina, E. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 1-27). Amsterdam: John Benjamins.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10,* 564-570.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

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

Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research, 24*, 267-295.

Cornips, L., van der Hoek, M., & Verwer, R. (2006). The acquisition of grammatical gender in bilingual child acquisition of Dutch (by older Moroccan and Turkish children): the definite determiner, attributive adjective and relative pronoun. In J. van de Weijer & B. Los (Eds.), *Linguistics in the Netherlands 2006* (pp. 40-51). Amsterdam: John Benjamins.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1283-1296.

Coulson, S., King, J., & Kutas, M. (1998). Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes, 13*, 21-58.

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The Effect Of Grammatical Gender on Object Categorisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, pp. 449-460.

Dąbrowska, E. (2004). Rules or schema? Evidence from Polish. *Language and Cognitive Processes, 19,*, 225–271.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, p. 317–367.

Dahan, D., Swingley, D., Tanenhaus, M., & Magnuson, J. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42*, 465-480.

Davidson, D., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience, 21*, 433-446.

De Houwer, A. (2005). Early bilingual acquisition: Focus on morphosyntax and the separate development hypothesis. In J. Kroll, & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 30-48). New York, NY, US: Oxford University Press.

De Houwer, A., & Gillis, S. (1998). Dutch child language: an overview. In S. Gillis, & A. De Houwer (Eds.), *The Acquisition of Dutch* (pp. 1-100). Amsterdam: John Benjamins.

Der, G., & Deary, I. (2006). Age and Sex Differences in Reaction Time in Adulthood: Results From the United Kingdom Health and Lifestyle Survey. *Psychology and Ageing, 21*, pp. 62-73.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*, 200-224.

Deutsch, W., & Wijnen, F. (1985). The Article's Noun and the Noun's Article: Explorations into the Representation and Access of Linguistic Gender in Dutch. *Linguistics, 23*, 793-810 .

Dijkstra, T., & Van Heuven, W. (2002). The architecture of the bilingual word recognition system. *Bilingualism: Language and Cognition, 5,* , 175–197.

Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*, 355–372.

Duchowski, A. (2003). *Eye Tracking Methodology: Theory and Practice.* London, UK: SpringerVerlag.

Dussias, P. (2010). Uses of eyetracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30*, 149 -166.

Fagerland, M., & Sandvik, L. (2009). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine, 28*, 1487-1497.

Field, A. (2009). *Discovering statistics using SPSS: (and sex, drugs and rock 'n roll).* London: Sage Publications Ltd.

Foucart, A. (2008). *Grammatical Gender Processing in French as a first and second language.* PhD Dissertation, University of Edinburgh and University of Provence, France.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory & Language, 66*, 226–248.

Franceschina, F. (2005). *Fossilized second language grammars: the acquisition of grammatical gender.* . Amsterdam: John Benjamins.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78-84.

Friederici, A., & Frisch, S. (2000). Verb-argument structure processing: the role of verb-specific and argument-specific information. *Journal of Memory and Language, 43* , 476–507.

Friederici, A., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 2*, 183-192.

Frisch, S., Kotz, S., von Cramon, D., & Friederici, A. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology, 14*, 336-340.

Gathercole, V., & Thomas, E. (2005). Minority language survival: Input factors influencing the acquisition of welsh. In J. Cohen, K. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th international symposium on bilingualism* (pp. 852-874). Somerville, MA: Cascadilla Press.

Gouvea, A., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25* , 149-188.

Gratton, G., & Coles, M. H. (1989). Generalization and evaluation of eye-movement correction procedures. *Journal of Psychophysiology, 3*, 14-16.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. . *Perception & Psychophysics, 28*, pp. 267–283.

Grüter, T., Lew-Williams, C., & Fernald, A. (2011). Grammatical gender in L2: Where is the problem? In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development,* (pp. 246-258.). Somerville, MA: Cascadilla Press.

Gunter, T., & Friederici, A. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*, 126–137.

Gunter, T., Friederici, A., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556-568.

Gunter, T., Stowe, L., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*, 660-676.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. (1997). *Algemene Nederlandse Spraakkunst (ANS).* Martinus Nijhoff: Groningen.

Hagoort, P. (2003a). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hagoort, P. (2003b). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage, 20*, S18–S29.

Hagoort, P., & Brown, C. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*, 1531-1549.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439-483.

Hagoort, P., Wassenaar, M., & Brown, C. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research, 16*, 38-50.

Hahne, A., & Friederici, A. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*, 339-356.

Hanulíková, A., van Alphen, P., van Goch, M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of Cognitive Neuroscience, 24(4)*, 878-887.

Hawkings, R. (2001). *Second Language Syntax: A Generative Introduction.* Malden, MA: Blackwell.

Hawkins, R., & Chan, Y.-H. C. (1997). The partial availability of universal grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research, 13*, 187-226.

Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner-noun gender concord in French and Spanish. In P. Prévost, & J. Paradis (Eds.), *The Acquisition of French in Different Contexts* (pp. 175–205). Amsterdam: John Benjamins.

Hinojosa, J., Moreno, E., Casado, P., Muñoz, F., & Pozo, M. (2005). Syntactic expectancy: An event-related potentials study. *Neuroscience Letters, 378*, 34-39.

Hopp, H. (2007). *Ultimate attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.* PhD Dissertation, University of Groningen, the Netherlands.

Howell, D. (2009). *Statistical methods for psychology.* Australia: Wadsworth, Cengage Learning .

Huettig, F. R., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171.

Hulk, A., & Cornips, L. (2006a). Neuter gender and interface vulnerability in child L2/2L1Dutch. In Unsworth, S. (Ed.), *Paths of Development in L1 and L2 acquisition: In honor of Bonnie D. Schwartz.* (pp. 107–134). Amsterdam: John Benjamins.

Hulk, A., & Cornips, L. (2006b). Between 2L1 - and child L2 acquisition: an experimental study of bilingual Dutch. In C. Lleó (Ed.), *Interfaces in Multilingualism: Acquisition and representation. (Hamburg Studies on Multilingualism)* (pp. 115-138.). Amsterdam/Philadelphia: John Benjamins.

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass, 1*, 571–591.

Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada.* . PhD Dissertation, Vrije Universiteit Amsterdam.

Koniuszaniec, G., & Błaszkowska, H. (2003). Sexism in Polish. In M. Hellinger, & H. Bußmann (Eds.), *Gender Across Languages* (pp. 259–286). Amsterdam: John Benjamins.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. *Electroencephalography and Clinical Neurophysiology, 40,* , 406-411.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*, 539–550.

Kutas, M., Lindamood, T., & Hillyard, S. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum, & J. Requin (Eds.), *Preparatory States and Processes* (pp. 217-237). Hillsdale, NJ: Erlbaum.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*, 218-233.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler, & M. Gernsbacher, *Handbook of Psycholinguistics* (pp. 659-724). New York: Elsevier.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language, 59*, 312-330.

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

Levi, D., Klein, S., & Aitsebaomo, A. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*, 963-977.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical genderin spoken word recognition. *Psychological Science, 18,* 193-198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*, 447-464.

Loerts, H. (2009). *Native and second language processing of grammatical gender: the role of the L1 and age of acquisition.* University of Groningen: Unpublished Manuscript.

Loerts, H., Stowe, L., & Schmid, M. (Under Review). Predictability speeds up the re-analysis process: an ERP investigation of gender agreement and cloze probability. Submitted to: *Journal of Neurolinguistics*.

Loerts, H., Wieling, M., & Schmid, M. (Under Review). Neuter is not Common in Dutch: Eye Movements reveal Asymmetrical Gender Processing. Submitted to*: Journal of Psycholinguistic Research.*

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*, 1-36.

Luck, S. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: The MIT Press.

MacWhinney, B. (2008). A Unified Model. In P. Robinson, & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 341-371). Mahwah, NJ: Lawrence Erlbaum Associates.

Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133-156.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*, 97–115.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In S. Unsworth, & E. Blom (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 139-162). Amsterdam: John Benjamins.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.
Martin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*, 899-917.
Matthews, P. (1997). *The concise Oxford Dictionary of Linguistics.* Oxford: Oxford University Press.
McClelland, J., & Elman, J. (1986). The Trace model of speech perception. *Cognitive Psychology, 18*, 1–86.
Migdalski, K. (2006). *The Syntax of Compound Tenses in Slavic.* Utrecht University, the Netherlands: LOT Dissertation Series.
Mills, A. (1986). *The acquisition of gender. A Study of English and German.* Berlin: Springer Verlag.
Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: an ERP study on agreement violations. *Brain Research, 1228*, 161-176.
Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*, 503-553.
Muller, H., King, J., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research, 5*, 193-203.
Müller, N. (1990). Developing two Gender Assignment Systems simultaneously. In J. Meisel (Ed.), *Two first languages. Early grammatical development in Bilingual Children.* Dordrecht: Foris.
Münte, T. F., Matzke, T., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience, 9*, 318-329.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.
Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.
Osberger, W., & Rohaly, A. (2001). Automatic Detection of Regions of Interest in Complex Video Sequences. In *Human Vision and Electronic Imaging VI*, 4299, SPIE - The International Society for Optical Engineering, Bellingham, USA, 361-372.
Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.
Osterhout, L., & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739-773.
Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition, 25*, 273-285.
Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras, & C. J. Clifton (Eds.), *The online study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 271–308). New York - Hove: Psychology Press.
Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism.* Amsterdam: John Benjamins.
Paris, G., Weber, A., & Crocker, M. (2006). German morphosyntactic gender and lexical access. *Poster presented at 12th Annual Conference on Architectures and*

Mechanisms for Language Processing (AMLaP 2006). Nijmegen, The Netherlands.
Perdue, C. (1993). *Adult Language Acquisition Cross-Linguistic Perspectives.* Cambridge: Cambridge University Press.
Perlak, D., & Jarema, G. (2003). The recognition of gender-marked nouns and verbs in Polish-speaking aphasic patients. *Cortex, 39*, 383-403.
Raaijmakers, J., Schrijnemakers, J., & 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.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.
Renkema, J. (1997). *Woordenlijst van de Nederlandse taal (Het groene Boekje), (Word List of Dutch (The Green Booklet)). Collected on behalf of the Belgian and Dutch governments.* The Hague: SDU Uitgevers.
Richardson, D., & Spivey, M. (2004). Eye-tracking: Characteristics and methods. In G. Wnek, & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering.* (pp. 568-572). Marcel Dekker.
Ritchie, W. C., & Bhatia, T. K. (2009). *The new handbook of second language acquisition.* Bingley, UK: Emerald Group Publishing Limited.
Roodenburg, A., & Hulk, A. (2008). Puzzles on grammatical gender. *Lingue e Linguaggio*, 67-92.
Rossi, E., Gugler, M., Friederici, A., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030-2048.
Sabourin, L. (2003). *Grammatical Gender and Second language Processing: An ERP study.* PhD Dissertation, University of Groningen, The Netherlands.
Sabourin, L., & de Haan, G. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22(1)*, 1-29.
Sabourin, L., & Stowe, L. (2008). Second language processing: when are first and second languages processed similarly? *Second Language Research, 24*, 397-430.
Schmid, M. S. (2005). *The Language Attrition Test Battery: A Research Manual.* University of Groningen: MS.
Schmid, M. S. (2011). *Language Attrition.* Cambridge: Cambridge University Press.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburg: Psychology Software Tools Inc.
Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research, 12*, 40-72.
Sekerina, I. A., Brooks, P. J., & Kempe, V. (2006). Gender Transparency Facilitates Noun Selection in Russian. In J. a. Levine, *The Proceedings of the 14th Annual Workshop on Formal Approaches to Slavic Linguistics.* The Princeton Meeting: Michigan Slavic Publications.
Seton, B. J. (2009). *"De appel valt niet ver van het boom": on native and non-native processing of grammatical gender violations: an ERP study.* University of Groningen: Unpublished Master's Thesis.
Sharbrough, F., Chatrian, G.-E., Lesser, R., Lüders, H., Nuwer, M., & Picton, T. (1991). American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature. *Journal of Clinical Neurophysiology, 8*, 200-202.

Slabakova, R. (2006). Is there a Critical Period for the Acquisition of Semantics. *Second Language Research, 22* , 302-338.

Smoczyńska, M. (1986). *The acquisition of Polish.* In D.I. Slobin (Ed.), The crosslinguistic study of language acquisition. Vol.1: The data. Hillsdale, NJ: Lawrence Erlbaum.

Snodgrass, J., & Vanderwart, M. (1980). A Standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174-215.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112*, 1-36.

Steinhauer, K., & Connolly, J. (2008). Event-related potentials in the study of language. In B. Stemmer, & H. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 91-104). New York: Elsevier.

Steinhauer, K., White, E., & Drury, J. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*, 13–41.

Szagun, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language, 34*, 445-471.

Szagun, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language, 34*, 445-471.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173–204.

Tomasello, M. (2003). *Constructing a Language. A Usage Based Theory of Language Acquisition.* London, UK: Harvard University Press.

Ullman, M. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature reviews - Neuroscience, 2*, 717-26.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92(1-2)*, 231-70.

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second language research, 24*, pp. 365-396.

Unsworth, S., & Hulk, A. C. (2010). L1 acquisition of neuter gender in Dutch: Production and judgement. *Proceedings of Generative Approaches to Language Acquisition 2009*, (pp. 50-51).

Van Berkum, J. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production.* Doctoral dissertation, Max Planck Institute for psycholinguistics. Nijmegen, Netherlands: Nijmegen University press.

Van Berkum, J., Brown, C., & Hagoort, P. (1999). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research, 28*, 555-571.

Van Berkum, J., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading

times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467.
Van den Brink, D., Brown, C., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 364-372.
Van der Velde, M. (2004). L'acquisition des determinants en L1: une etude comparative entre le Francais et le Néerlandais. *Acquisition et interaction en langue etrangere, 21*, 9-46.
van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research, 26*, 43-74.
Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and lexicalist grammar. *Cognition, 75*, 105-143.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1-25.
Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. In K. Forbus, D. Gentner, & T. Tegier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.
White, L., Valenzuela, E., Kozlowska-MacGregor, M., & Leung, Y. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*, 105-133.
White, L., Valenzuela, E., Macgregor, E., Leung, I., & Ben Ayed, H. (2001). The status of abstract features in interlanguage grammars: gender and number in L2 Spanish. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 792-802). Somerville, MA: Cascadilla Press.
Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: evidence from eye movement measures of natural paragraph reading. *Psychonomic bulletin & review, 19*, 73-80.
Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1272-1288.
Yamada, Y., & Neville, H. (2004). Modulation of syntax-related ERP effects by difference in tasks and variations in syntactic violation types. *Cognitive Neuroscience Society, 11*, 141.
Zubin, D., & Köpcke, K. M. (1984). Affect classification in the German gender system. *Lingua 63*, 41-96.

APPENDICES

A-1. Recruitment & Selection Questionnaire

Instructions (Dutch): Om te bepalen of u aan het onderzoek 'De verwerking van zinnen, onderzocht met Eye tracking & EEG' deel kunt nemen, verzoeken wij u onderstaande vragen in te vullen. Mocht er iets niet duidelijk zijn dan kunt u de vraag open laten. De onderzoeker neemt in dat geval contact met u op. De formulieren zullen vertrouwelijk worden behandeld. Indien u op basis van dit formulier wordt uitgesloten van deelname aan het onderzoek, dan wordt dit formulier vernietigd. Het invullen van de vragenlijst duurt slechts een paar minuten.

Instructions (English translation)**:** *To determine whether you can participate in the study "The processing of sentences, studied with Eye Tracking & EEG", we would like to ask you to complete the following questions. If something is not clear, you can leave the question unanswered and the researcher will then contact you. These forms are treated confidentially. In case you are excluded from participation based on this question form, the form will be destroyed. Completing the questionnaire will only take a few minutes.*

Personal Information (Original version in Dutch)

Name:…………………………………………………………………………………...
Address:………………………………………………………………………………...
Postal Code:……………………………………………………………………………
City:…………………………………………………………………………………….
Phone Number:…………………………………………………………………………
Email address:…………………………………………………………………………..
Sex: ☐ Male ☐ Female
Date of birth:……………………………………………………………………………
Place of birth:…………………………………………………………………………...
What is the highest level of education you have completed?
☐ Secondary School ☐ Secondary Vocational Education
☐ Higher Professional Education ☐ University

What is your current profession?...

Questions concerning (language-related) deficits

Do you have trouble reading, writing, or speaking?
☐ Yes ☐ No
If so, please specify these problems:……………………….……………………………..
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
……………………………………………………………………………………….

Do you have trouble hearing or listening?
☐ Yes ☐ No

If so, please specify these problems:…………………….……………………………..
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………….

Have you ever been treated by a psychiatrist?
☐ Yes ☐ No
If so, please specify the reason:………………………….……………………………..
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………….

Have you ever been treated by a neurologist?
☐ Yes ☐ No
If so, please specify the reason:………………………….……………………………..
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………….

Are you colour-blind?
☐ Yes ☐ No
Are you right handed?
☐ Yes ☐ No

Questions specifically related to participation in the EEG experiment

Have you ever suffered from a seizure?
☐ Yes ☐ No

Have you ever suffered from a severe head injury?
☐ Yes ☐ No

If so, please specify when:…………………………….……………………………..
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………….

Do you take any medicines?
☐ Yes ☐ No
If so, please specify the reason and the dosage:.………..……………………………..
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
…………………………………………………………………………………….

Do you take any medicines?

☐ Yes ☐ No

If so, please specify the reason and the dosage:..………….………………………………..
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………….

Do you drink alcohol?

☐ Yes ☐ No

If so, how often (number of units per week)?

Do you use drugs?

☐ Yes ☐ No

If so, how often (number of units per week)?

To check the extent to which you are right handed, please answer the following questions:

Which hand do you write with?	Left / No preference / Right
Which hand do you draw with?	Left / No preference / Right
Which hand do you use to throw a ball at a specific object?	Left / No preference / Right
In which hand do you hold a racket e.g. when playing tennis or squash?	Left / No preference / Right
Which hand do you use to brush your teeth?	Left / No preference / Right
In which hand do you hold a knife when you have to cut something (without a fork)?	Left / No preference / Right
In which hand do you hold a hammer when hitting a nail into the wall?	Left / No preference / Right
If you light a match, in which hand do you hold the match?	Left / No preference / Right
In which hand do you hold an eraser when you have to erase something on paper?	Left / No preference / Right
Which hand do you use to brush your teeth?	Left / No preference / Right
Which hand do you use to share cards during a card game?	Left / No preference / Right
Which hand do you use to cut something with scissors?	Left / No preference / Right

A-2. Sociolinguistic Questionnaire (Original version in Dutch)

Questions concerning sociolinguistic background (all participants)

Have you ever lived other countries?

☐ Yes ☐ No

If so, please specify when, where, and how long:……….…………………………………..
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

What language(s) did you acquire before starting school?

☐ Dutch
☐ Dutch and (an)other language(s)
☐ (an) other language(s):………………………………………………

What languages did you acquire at school or at work? And how proficient are you in those languages? Please give a number for each of the four skills:
1 = very bad, 2 = bad, 3 = average, 4 = good, 5 = very good
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….

What languages did you acquire outside work/school? And how proficient are you in those languages? Please give a number for each of the four skills:
1 = very bad, 2 = bad, 3 = average, 4 = good, 5 = very good
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….
Language:…………...………Lezen:…….Schrijven:…….Spreken:…….Luisteren:…….

Questions concerning sociolinguistic background (only L2 participants)

How often do you visit Poland?

☐ Less than once every 10 years / never
☐ Once every 5-10 years
☐ Once every 3-5 years
☐ Once every 2 years
☐ Once or twice a year

How long are your visits to Poland?

☐ Shorter than 2 weeks
☐ 2 to 4 weeks
☐ 4 to 6 weeks
☐ 6 weeks to 3 months
☐ Longer than 3 months

How well do you speak Dutch?

- ☐ Very bad
- ☐ Bad
- ☐ Reasonable
- ☐ Good
- ☐ Very good

How well do you speak Polish?

- ☐ Very bad
- ☐ Bad
- ☐ Reasonable
- ☐ Good
- ☐ Very good

How often do you speak Polish?

- ☐ Hardly ever
- ☐ A few times a year
- ☐ Monthly
- ☐ Weekly
- ☐ Daily

What percentage (estimate) of your friends is Polish?..%
What percentage (estimate) of your friends is Dutch?...%

Which language is easier to speak?

- ☐ Dutch
- ☐ Polish
- ☐ Equally easy

Do you live together with a partner?

☐ Yes ☐ No

If so, what is your partner's mother tongue?........……………………………………..
And what language do you use with your partner?..........………………………………..

Do you have any children?

☐ Yes ☐ No

If so, what language do you use with your children?........………………………………..

What percentage of time (estimate) do you speak Dutch with relatives?........................%

What percentage of time (estimate) do you speak Dutch with friends?...........................%

What percentage of time (estimate) do you speak Dutch with colleagues?.....................%

What percentage of time (estimate) do you speak Dutch when you´re not at home?... %

What percentage of media (radio/television) is in Dutch?..%

What percentage of reading (books/newspapers) is in Dutch?.......................................%

A-3. C-test

Instructions (Dutch):

Vul het woorddeel in

Hieronder staan 2 korte Nederlandstalige teksten. In de teksten zijn gaten gevallen. Het zijn geen hele woorden die zijn weggelaten, maar delen van woorden. Het is de bedoeling dat u uit het zinsverband probeert af te leiden welk woorddeel op de puntjes zou kunnen staan. De eerste zin is steeds helemaal intact gelaten om u een beetje op weg te helpen. U hebt maximaal 5 minuten de tijd per tekst.

Instructions (English translation):

Fill in the gap

Below are 2 small Dutch texts. Each text contains gaps. Parts of some words have been left out, but it is never the case that a whole word has been deleted. Please try and derive from the context which parts of the words could be filled in. The first sentence is always left intact to get you started. You have up to 5 minutes per text.

Tekst 1 / *Text 1*

Ik houd van Nederland en niet zo'n beetje ook. Waarom ik van het land houd is niet alleen omdat velen van wie ik houd hier leven, nee, het is me________ dan d________. De groo________ reden v________ mijn lie________ voor het land ko________ voort u_______ het feit dat al_________ zo geor______________ en syste______________ is. Er i_____ een systeem e____ het werkt. Je kan, ni_________ zonder twijfel, maar to________ met dic________ ogen er________ uitgaan d________ het recht zege________.

Tekst 2 / *Text 2*

Openlijke narcisten zijn mensen met een opgeblazen gevoel over zichzelf. Ze ei________ vaak ande________ aandacht o_____ en ko________ charmant ov_____, ond________ het feit d_____ ze wei________ besef he________ van de beho________ van anderen. Verb________ narcisten zijn weli________ net z_____ hevig met zichzelf be________ en ev________ arrogant a_____ openlijke narcisten, ma_____ ze do_____ dit o_____ een subti________ manier.

Answers Tekst 1 / *Text 1*

Ik houd van Nederland en niet zo'n beetje ook. Waarom ik van het land houd is niet alleen omdat velen van wie ik houd hier leven, nee, het is meer dan dat. De grootste reden voor mijn liefde voor het land komt voort uit het feit dat alles zo georganiseerd en systematisch is. Er is een systeem en het werkt. Je kan, niet zonder twijfel, maar toch met dichte ogen ervan uitgaan dat het recht zegeviert.

I love the Netherlands and not a little either. Why I love the country is not only because many people whom I love are living here, not, it's more than that. The biggest reason for my love of the country stems from the fact that everything is so organized and systematic. There is a system and it works. You can, not without doubt, but still with closed eyes, assume that the law prevails.

Answers Tekst 2 / *Text 2*
Openlijke narcisten zijn mensen met een opgeblazen gevoel over zichzelf. Ze eisen vaak andermans aandacht op en komen charmant over, ondanks het feit dat ze weinig besef hebben van de behoeften van anderen. Verborgen narcisten zijn weliswaar net zo hevig met zichzelf bezig en even arrogant als openlijke narcisten, maar ze doen dit op een subtielere manier.

Overt narcissists are people with an inflated sense of self. They often demand people's attention and come across as charming, despite the fact that they have little awareness of needs of others. Hidden narcissists are indeed just as heavily self-absorbed and arrogant as over narcissists, but they do so in a subtle way.

A-4. Gender Assignment Task

Instructions (Dutch):

Kies het juiste lidwoord

Voor u ligt een lijst met allerlei woorden. Geef voor elk woord aan of er ‘de’ of ‘het’ voor zou moeten staan door het juiste lidwoord te omcirkelen of het foute lidwoord door te strepen. Sommige woorden staan er meerdere keren in. Probeer zo snel mogelijk een keuze te maken.

Instructions (English translation):

Choose the correct article

In front of you lies a long list of words. For each word, choose the article that should precede this word by circling the correct one of by striking through the incorrect one. Some words are in the list more than once. Please try and choose the article as fast as possible.

de of het?		***de of het?***	
de / het	deur	de / het	bank
de / het	bed	de / het	horloge
de / het	agent	de / het	boer
de / het	nummer	de / het	huis
de / het	eiland	de / het	kamer
de / het	ster	de / het	brood
de / het	kind	de / het	fles
de / het	anker	de / het	meer
de / het	verhaal	de / het	gesprek
de / het	klok	de / het	gras
de / het	antwoord	de / het	bureau
de / het	verschil	de / het	schaap
de / het	fiets	de / het	muzikant
de / het	film	de / het	dag
de / het	appel	de / het	gang
de / het	bedrijf	de / het	steen
de / het	patiënt	de / het	dorp
de / het	hart	de / het	konijn
de / het	pet	de / het	hond
de / het	auto	de / het	glas
de / het	hek	de / het	tuin
de / het	blad	de / het	hoofd
de / het	geld	de / het	jongen
de / het	geluid	de / het	plan
de / het	bloem	de / het	publiek
de / het	woord	de / het	voorbeeld
de / het	boek	de / het	wolk
de / het	knoop	de / het	sok
de / het	feest	de / het	kerk

de of het?

de / het	man
de / het	pak
de / het	slot
de / het	varken
de / het	wet
de / het	zon
de / het	programma
de / het	moment
de / het	potlood
de / het	tafel
de / het	uur
de / het	vraag
de / het	wijn
de / het	vest
de / het	stoel
de / het	recept
de / het	oog
de / het	les
de / het	laars
de / het	mes
de / het	paard
de / het	papier
de / het	wiel
de / het	minister
de / het	vat
de / het	raam
de / het	krant
de / het	tijd
de / het	vakantie
de / het	vrouw
de / het	pan
de / het	sigaar
de / het	taart
de / het	weg
de / het	trap
de / het	vogel
de / het	kat
de / het	weekend
de / het	lamp
de / het	slang
de / het	kast
de / het	nieuws
de / het	moeder
de / het	schoen
de / het	kanon
de / het	stem
de / het	licht

de of het?

de / het	trui
de / het	vliegtuig
de / het	overhemd
de / het	meneer
de / het	land
de / het	probleem
de / het	ziekenhuis
de / het	kaars
de / het	hulp
de / het	lied
de / het	straat
de / het	piano
de / het	kikker
de / het	dokter
de / het	bus
de / het	gras
de / het	woord
de / het	brief
de / het	gesprek
de / het	schaap
de / het	doos
de / het	gezicht
de / het	strijkijzer
de / het	wereld
de / het	taal
de / het	plan
de / het	touw
de / het	sok
de / het	slang
de / het	ster
de / het	agent
de / het	deur
de / het	anker
de / het	nummer
de / het	kind
de / het	hek
de / het	verhaal
de / het	boek
de / het	antwoord
de / het	steen
de / het	dorp
de / het	publiek
de / het	licht
de / het	slot
de / het	tafel
de / het	trap
de / het	varken

de of het?

de / het	stad
de / het	wolk
de / het	appel
de / het	weekend
de / het	knoop
de / het	patiënt
de / het	hulp
de / het	les
de / het	pak
de / het	raam
de / het	weg
de / het	vakantie
de / het	vraag
de / het	taart
de / het	vogel
de / het	stad
de / het	papier
de / het	vat
de / het	pan
de / het	auto
de / het	laars
de / het	nieuws
de / het	schoen
de / het	uur
de / het	vrouw
de / het	stem
de / het	vliegtuig
de / het	wet
de / het	sigaar
de / het	probleem
de / het	meneer
de / het	overhemd
de / het	paard
de / het	recept
de / het	bank
de / het	verschil
de / het	bed
de / het	kamer
de / het	buurt
de / het	dag
de / het	kikker
de / het	hoofd
de / het	kaars
de / het	ziekenhuis
de / het	kat
de / het	moment
de / het	krant

de of het?

de / het	vest
de / het	wiel
de / het	wijn
de / het	minister
de / het	potlood
de / het	stoel
de / het	tijd
de / het	land
de / het	mes
de / het	kast
de / het	moeder
de / het	voorbeeld
de / het	jongen
de / het	tuin
de / het	gezicht
de / het	meer
de / het	klok
de / het	ster
de / het	film
de / het	bedrijf
de / het	hart
de / het	blad
de / het	geluid
de / het	boom
de / het	gang
de / het	lied
de / het	zon
de / het	lamp
de / het	man
de / het	oog
de / het	kerk
de / het	stad
de / het	strijkijzer
de / het	hond
de / het	brief
de / het	deur
de / het	slot
de / het	agent
de / het	glas
de / het	piano
de / het	kanon
de / het	slang
de / het	sok
de / het	krant
de / het	man
de / het	potlood
de / het	wet

de of het?		***de of het?***	
de / het	programma	de / het	vogel
de / het	trui	de / het	stoel
de / het	varken	de / het	paard
de / het	trui	de / het	vraag
de / het	recept	de / het	wijn
de / het	pak	de / het	taart
de / het	moment	de / het	mes
de / het	land	de / het	lamp
de / het	les	de / het	vat
de / het	minister	de / het	tijd
de / het	pan	de / het	sigaar
de / het	raam	de / het	papier
de / het	schoen	de / het	oog
de / het	vest	de / het	vrouw
de / het	trap	de / het	weekend
de / het	vakantie	de / het	kerk
de / het	bureau	de / het	kat
de / het	eiland	de / het	touw
de / het	pet	de / het	taal
de / het	doos	de / het	fles
de / het	konijn	de / het	boer
de / het	wereld	de / het	woord
de / het	touw	de / het	verhaal
de / het	laars	de / het	kind
de / het	meneer	de / het	antwoord
de / het	overhemd	de / het	bed
de / het	wiel	de / het	hek
de / het	vliegtuig	de / het	geluid
de / het	uur	de / het	boek
de / het	stem	de / het	appel
de / het	probleem	de / het	horloge
de / het	licht	de / het	huis
de / het	muzikant	de / het	bus
de / het	brood	de / het	ziekenhuis
de / het	anker	de / het	moeder
de / het	verschil	de / het	kanon
de / het	fiets	de / het	wolk
de / het	patiënt	de / het	voorbeeld
de / het	bloem	de / het	publiek
de / het	buurt	de / het	jongen
de / het	dokter	de / het	feest
de / het	nieuws	de / het	pet
de / het	programma	de / het	klok
de / het	tafel	de / het	auto
de / het	weg	de / het	fiets
de / het	zon	de / het	hart
de / het	kast	de / het	blad

de of het?

de / het	nummer
de / het	geld
de / het	straat
de / het	lied
de / het	hulp
de / het	plan
de / het	straat
de / het	taal
de / het	wereld
de / het	tuin
de / het	piano
de / het	strijkijzer
de / het	kikker
de / het	glas
de / het	doos
de / het	schaap
de / het	gras
de / het	boom
de / het	film
de / het	bank
de / het	huis
de / het	knoop
de / het	eiland
de / het	bedrijf
de / het	geld
de / het	feest
de / het	hond
de / het	horloge
de / het	dag
de / het	muzikant
de / het	bus
de / het	boer
de / het	gesprek
de / het	gezicht
de / het	meer
de / het	kamer
de / het	brief
de / het	boom

de of het?

de / het	bloem
de / het	hoofd
de / het	kaars
de / het	brood
de / het	fles
de / het	bureau
de / het	buurt
de / het	konijn
de / het	dorp
de / het	steen
de / het	gang
de / het	dokter

B-1. Materials used in the eye tracking experiment

The list below provides an overview of all the target-competitor pairs used in the eye tracking experiment in the four different conditions. Each item was used as a target, as a competitor, and twice as a distractor. Distractors were randomized and never matched the target in gender or in colour. Hence, they are not mentioned in the table below.

The first column of the table provides the NPs used in the instructions in the definite block and indefinite block respectively. Target numbers 1 to 8 and 25 to 40 concern nouns that are common in Dutch and items 9 to 24 and 41 to 48 concern neuter gendered nouns. Additionally, target items 1 to 24 concerns nouns that overlap in gender between Dutch and Polish and targets 25 to 48 are nouns that do not share gender in the two language. The Polish equivalent noun and its gender are also mentioned as well as the competitor's gender in Polish (last column).

Conditions:
DG-DC: Different Gender – Different Colour
DG-SC: Different Gender – Same Colour
SG-DC: Same Gender – Different Colour
SG-SC: Same Gender- Same Colour

	Target Dutch *Polish translation*	**Condition**	**Competitor**	**Gender Competitor in Polish**
1	de rode appel$_{COM}$ /	DG-DC	het/een bruin(e) varken$_{NEU}$	feminine
	een rode appel$_{COM}$	DG-SC	het/een ro(o)d(e) overhemd$_{NEU}$	feminine
		SG-DC	de/een gele vogel$_{COM}$	masculine
	jabłko$_{NEU}$ *(‘apple’)*	SG-SC	de/een rode bloem$_{COM}$	masculine
2	de blauwe doos$_{COM}$ /	DG-DC	het/een ro(o)d(e) hart$_{NEU}$	neuter
	een blauwe doos$_{COM}$	DG-SC	het/een blauw(e) raam$_{NEU}$	neuter
		SG-DC	de/een rode taart$_{COM}$	neuter
	pudełko$_{NEU}$ *(‘box’)*	SG-SC	de/een blauwe sok$_{COM}$	feminine
3	de rode taart$_{COM}$	DG-DC	het/een blauw(e) bed$_{NEU}$	neuter
	een rode taart$_{COM}$	DG-SC	het/een ro(o)d(e) bureau$_{NEU}$	neuter
		SG-DC	de/een bruine sigaar$_{COM}$	neuter
	ciasto$_{NEU}$ *(‘cake’)*	SG-SC	de/een rode kaars$_{COM}$	feminine
4	de gele stoel$_{COM}^{**}$	DG-DC	het/een blauw(e) horloge$_{NEU}$	masculine
	een gele stoel$_{COM}^{**}$	DG-SC	het/een ge(e)l(e) anker$_{NEU}$	feminine
		SG-DC	de/een blauwe wolk$_{COM}$	femine
	krzesło$_{NEU}$ *(‘chair’)*	SG-SC	de/een gele vogel$_{COM}$	masculine
5	de bruine sigaar$_{COM}^{*}$	DG-DC	het/een ge(e)l(e) hek$_{NEU}$	neuter
	een bruine sigaar$_{COM}^{*}$	DG-SC	het/een bruin(e) varken$_{NEU}$	feminine
		SG-DC	de/een gele zon$_{COM}$	neuter
	cygaro$_{NEU}$ *(‘cigar’)*	SG-SC	de/een bruine laars$_{COM}$	masculine
6	de blauwe piano$_{COM}^{*}$	DG-DC	het/een ge(e)l(e) strijkijzer$_{NEU}$	neuter
	een blauwe piano$_{COM}^{*}$	DG-SC	het/een blauw(e) bed$_{NEU}$	neuter
		SG-DC	de/een ge(e)l(e) ster$_{COM}$	feminine
	fortepian$_{NEU}$ *(‘piano’)*	SG-SC	de/een blauwe trui$_{COM}$	masculine

7	de gele zon$_{COM}$	DG-DC	het/een blauw(e) raam$_{NEU}$	neuter
	een gele zon$_{COM}$	DG-SC	het/een ge(e)l(e) slot$_{NEU}$	masculine
		SG-DC	de/een rode kaars$_{COM}$	feminine
	słońce$_{NEU}$ (‘sun’)	SG-SC	de/een gele vogel$_{COM}$	masculine
8	de groene boom$_{COM}$	DG-DC	het/een blauw(e) boek$_{NEU}$	feminine
	een groene boom$_{COM}$	DG-SC	het/een groen(e) oog$_{NEU}$	neuter
		SG-DC	de/een bruine schoen$_{COM}$	feminine
	drzewo$_{NEU}$ (‘tree’)	SG-SC	de/een groene kikker$_{COM}$	feminine
9	het gele anker$_{NEU}$	DG-DC	de/een rode taart$_{COM}$	neuter
	een geel anker$_{NEU}$	DG-SC	de/een gele stoel$_{COM}$	neuter
		SG-DC	het/een ro(o)d(e) vliegtuig$_{NEU}$	masculine
	kotwica$_{FEM}$ (‘anchor’)	SG-SC	het/een ge(e)l(e) wiel$_{NEU}$	neuter
10	het bruine vat$_{NEU}$	DG-DC	de/een gele ster$_{COM}$	feminine
	een bruin vat$_{NEU}$	DG-SC	de/een bruine schoen$_{COM}$	feminine
		SG-DC	het/een ge(e)l(e) strijkijzer$_{NEU}$	neuter
	baryłka$_{FEM}$ (‘barrel’)	SG-SC	het/een bruin(e) konijn$_{NEU}$	masculine
11	het blauwe boek$_{NEU}$[**]	DG-DC	de/een groene slang$_{COM}$	masculine
	een blauw boek$_{NEU}$[**]	DG-SC	de/een blauwe piano$_{COM}$	neuter
		SG-DC	het/een ge(e)l(e) anker$_{NEU}$	feminine
	książka$_{FEM}$ (‘book’)	SG-SC	het/een blauw(e) raam$_{NEU}$	neuter
12	het groene kanon$_{NEU}$[**]	DG-DC	de/een blauwe sok$_{COM}$	feminine
	een groen kanon$_{NEU}$[**]	DG-SC	de/een groene kikker$_{COM}$	feminine
		SG-DC	het/een ro(o)d(e) overhemd$_{NEU}$	feminine
	armata$_{FEM}$ (‘canon’)	SG-SC	het/een groen(e) potlood$_{NEU}$	masculine
13	het groene glas$_{NEU}$	DG-DC	de/een blauwe piano$_{COM}$	neuter
	een groen glas$_{NEU}$	DG-SC	de/een groene boom$_{COM}$	neuter
		SG-DC	het/een ro(o)d(e) bureau$_{NEU}$	neuter
	szklanka$_{FEM}$ (‘glass’)	SG-SC	het/een groen(e) oog$_{NEU}$	neuter
14	het bruine varken$_{NEU}$	DG-DC	de/een groene kikker$_{COM}$	feminine
	een bruin varken$_{NEU}$	DG-SC	de/een bruine pet$_{COM}$	feminine
		SG-DC	het/een groen(e) glas$_{NEU}$	feminine
	świnia$_{FEM}$ (‘pig’)	SG-SC	het/een bruin(e) brood$_{NEU}$	masculine
15	het rode overhemd$_{NEU}$	DG-DC	de/een bruine sigaar$_{COM}$	neuter
	een rood overhemd$_{NEU}$	DG-SC	de/een rode kaars$_{COM}$	feminine
		SG-DC	het/een bruin(e) brood$_{NEU}$	masculine
	koszula$_{FEM}$ (‘shirt’)	SG-SC	het/een ro(o)d(e) huis$_{NEU}$	masculine
16	het blauwe vest$_{NEU}$	DG-DC	de/een groene lamp$_{COM}$	feminine
	een blauw vest$_{NEU}$	DG-SC	de/een blauwe doos$_{COM}$	neuter
		SG-DC	het/een ro(o)d(e) huis$_{NEU}$	masculine
	kamizelka$_{FEM}$ (‘vest’)	SG-SC	het/een blauw(e) boek$_{NEU}$	feminine
17	het bruine brood$_{NEU}$	DG-DC	de/een blauwe trui$_{COM}$	masculine
	een bruin brood$_{NEU}$	DG-SC	de/een bruine sigaar$_{COM}$	neuter
		SG-DC	het/een ge(e)l(e) wiel$_{NEU}$	neuter
	chleb$_{MASC}$ (‘bread’)	SG-SC	het/een bruin(e) konijn$_{NEU}$	masculine
18	het groene potlood$_{NEU}$	DG-DC	de/een bruine sigaar$_{COM}$	neuter
	een groen potlood$_{NEU}$	DG-SC	de/een groene slang$_{COM}$	masculine

		SG-DC	het/een blauw(e) bed$_{NEU}$	neuter
	ołówek$_{MASC}$ (‘pencil’)	SG-SC	het/een groen(e) kanon$_{NEU}$	feminine
19	het rode huis$_{NEU}$	DG-DC	de/een gele stoel$_{COM}$	neuter
	een rood huis$_{NEU}$	DG-SC	de/een rode knoop$_{COM}$	masculine
		SG-DC	het/een groen(e) kanon$_{NEU}$	feminine
	dom$_{MASC}$ (‘house’)	SG-SC	het/een ro(o)d(e) bureau$_{NEU}$	neuter
20	het groene blad$_{NEU}$	DG-DC	de/een bruine schoen$_{COM}$	feminine
	een groen blad$_{NEU}$	DG-SC	de/een groene auto$_{COM}$	masculine
		SG-DC	het/een bruin(e) varken$_{NEU}$	feminine
	liść$_{MASC}$ (‘leaf’)	SG-SC	het/een groen(e) glas$_{NEU}$	feminine
21	het gele slot$_{NEU}$	DG-DC	de/een rode knoop$_{COM}$	masculine
	een geel slot$_{NEU}$	DG-SC	de/een gele zon$_{COM}$	neuter
		SG-DC	het/een bruin(e) varken$_{NEU}$	feminine
	zamek$_{MASC}$ (‘lock’)	SG-SC	het/een ge(e)l(e) hek$_{NEU}$	neuter
22	het bruine konijn$_{NEU}$	DG-DC	de/een groene boom$_{COM}$	neuter
	een bruin konijn$_{NEU}$	DG-SC	de/een bruine laars$_{COM}$	masculine
		SG-DC	het/een groen(e) oog$_{NEU}$	neuter
	krolik$_{MASC}$ (‘rabbit’)	SG-SC	het/een bruin(e) vat$_{NEU}$	feminine
23	het blauwe horloge$_{NEU}$	DG-DC	de/een gele bus$_{COM}$	masculine
	een blauw horloge$_{NEU}$	DG-SC	de/een blauwe sok$_{COM}$	feminine
		SG-DC	het/een ge(e)l(e) slot$_{NEU}$	masculine
	zegarek$_{MASC}$ (‘watch’)	SG-SC	het/een blauw(e) vest$_{NEU}$	feminine
24	het rode vliegtuig$_{NEU}$	DG-DC	de/een bruine laars$_{COM}$	masculine
	een rood vliegtuig$_{NEU}$	DG-SC	de/een rode taart$_{COM}$	neuter
		SG-DC	het/een blauw(e) bed$_{NEU}$	neuter
	samolot$_{MASC}$ (‘airplane’)	SG-SC	het/een ro(o)d(e) overhemd$_{NEU}$	feminine
25	de rode kaars$_{COM}$	DG-DC	het/een groen(e) oog$_{NEU}$	neuter
	een rode kaars$_{COM}$	DG-SC	het/een ro(o)d(e) hart$_{NEU}$	neuter
		SG-DC	de/een blauwe piano$_{COM}$	neuter
	świeca$_{FEM}$ (‘candle’)	SG-SC	de/een rode taart$_{COM}$	neuter
26	de bruine pet$_{COM}$	DG-DC	het/een groen(e) blad$_{NEU}$	masculine
	een bruine pet$_{COM}$	DG-SC	het/een bruin(e) vat$_{NEU}$	feminine
		SG-DC	de/een groene slang$_{COM}$	masculine
	przykrywka$_{FEM}$ (‘cap’)	SG-SC	de/een bruine schoen$_{COM}$	feminine
27	de blauwe wolk$_{COM}$	DG-DC	het/een ro(o)d(e) overhemd$_{NEU}$	feminine
	een blauwe wolk$_{COM}$	DG-SC	het/een blauw(e) boek$_{NEU}$	feminine
		SG-DC	de/een groene boom$_{COM}$	neuter
	chmura$_{FEM}$ (‘cloud’)	SG-SC	de/een blauwe piano$_{COM}$	neuter
28	de groene kikker$_{COM}$	DG-DC	het/een ge(e)l(e) slot$_{NEU}$	masculine
	een groene kikker$_{COM}$	DG-SC	het/een groen(e) glas$_{NEU}$	feminine
		SG-DC	de/een bruine pet$_{COM}$	feminine
	żaba$_{FEM}$ (‘frog’)	SG-SC	de/een groene lamp$_{COM}$	feminine
29	de groene lamp$_{COM}$*	DG-DC	het/een ro(o)d(e) bureau$_{NEU}$	neuter
	een groene lamp$_{COM}$*	DG-SC	het/een groen(e) potlood$_{NEU}$	masculine
		SG-DC	de/een rode bloem$_{COM}$	masculine
	lampa$_{FEM}$ (‘lamp’)	SG-SC	de/een groene slang$_{COM}$	masculine

30	de bruine schoen$_{COM}$	DG-DC	het/een ge(e)l(e) anker$_{NEU}$	feminine
	een bruine schoen$_{COM}$	DG-SC	het/een bruin(e) konijn$_{NEU}$	masculine
		SG-DC	de/een groene kikker$_{COM}$	feminine
	butelka$_{FEM}$ ('shoe')	SG-SC	de/een bruine pet$_{COM}$	feminine
31	de blauwe sok$_{COM}$**	DG-DC	het/een groen(e) potlood$_{NEU}$	masculine
	een blauwe sok$_{COM}$**	DG-SC	het/een blauw(e) vest$_{NEU}$	feminine
		SG-DC	de/een groene auto$_{COM}$	masculine
	skarpetka$_{FEM}$ ('sock')	SG-SC	de/een blauwe doos$_{COM}$	neuter
32	de gele ster$_{COM}$**	DG-DC	het/een bruin(e) brood$_{NEU}$	masculine
	een gele ster$_{COM}$**	DG-SC	het/een ge(e)l(e) hek$_{NEU}$	neuter
		SG-DC	de/een blauwe doos$_{COM}$	neuter
	gwiazda$_{FEM}$ ('star')	SG-SC	de/een gele bus$_{COM}$	masculine
33	de gele vogel$_{COM}$	DG-DC	het/een ro(o)d(e) vliegtuig$_{NEU}$	masculine
	een gele vogel$_{COM}$	DG-SC	het/een ge(e)l(e) wiel$_{NEU}$	neuter
		SG-DC	de/een bruine schoen$_{COM}$	feminine
	ptak$_{MASC}$ ('bird')	SG-SC	de/een gele zon$_{COM}$	neuter
34	de rode bloem$_{COM}$	DG-DC	het/een groen(e) glas$_{NEU}$	feminine
	een rode bloem$_{COM}$	DG-SC	het/een ro(o)d(e) huis$_{NEU}$	masculine
		SG-DC	de/een gele bus$_{COM}$	masculine
	kwiat$_{MASC}$ ('flower')	SG-SC	de/een rode knoop$_{COM}$	masculine
35	de bruine laars$_{COM}$	DG-DC	het/een ro(o)d(e) huis$_{NEU}$	masculine
	een bruine laars$_{COM}$	DG-SC	het/een bruin(e) brood$_{NEU}$	masculine
		SG-DC	de/een rode appel$_{COM}$	neuter
	kozak$_{MASC}$ ('boot')	SG-SC	de/een bruine pet$_{COM}$	feminine
36	de gele bus$_{COM}$*	DG-DC	het/een bruin(e) konijn$_{NEU}$	masculine
	een gele bus$_{COM}$*	DG-SC	het/een ge(e)l(e) strijkijzer$_{NEU}$	neuter
		SG-DC	de/een blauwe trui$_{COM}$	masculine
	autobus$_{MASC}$ ('bus')	SG-SC	de/een gele stoel$_{COM}$	neuter
37	de rode knoop$_{COM}$	DG-DC	het/een ge(e)l(e) wiel$_{NEU}$	neuter
	een rode knoop$_{COM}$	DG-SC	het/een ro(o)d(e) vliegtuig$_{NEU}$	masculine
		SG-DC	de/een groene lamp$_{COM}$	feminine
	guzik$_{MASC}$ ('button')	SG-SC	de/een rode appel$_{COM}$	neuter
38	de groene auto$_{COM}$	DG-DC	het/een bruin(e) vat$_{NEU}$	feminine
	een groene auto$_{COM}$	DG-SC	het/een groen(e) blad$_{NEU}$	masculine
		SG-DC	de/een rode knoop$_{COM}$	masculine
	samochód$_{MASC}$ ('car')	SG-SC	de/een groene boom$_{COM}$	neuter
39	de groene slang$_{COM}$	DG-DC	het/een blauw(e) vest$_{NEU}$	feminine
	een groene slang$_{COM}$	DG-SC	het/een groen(e) kanon$_{NEU}$	feminine
		SG-DC	de/een blauwe sok$_{COM}$	feminine
	wąż$_{MASC}$ ('snake')	SG-SC	de/een groene auto$_{COM}$	masculine
40	de blauwe trui$_{COM}$	DG-DC	het/een groen(e) kanon$_{NEU}$	feminine
	een blauwe trui$_{COM}$	DG-SC	het/een blauw(e) horloge$_{NEU}$	masculine
		SG-DC	de/een gele stoel$_{COM}$	neuter
	sweter$_{MASC}$ ('sweater')	SG-SC	de/een blauwe wolk$_{COM}$	feminine
41	het blauwe bed$_{NEU}$	DG-DC	de/een gele zon$_{COM}$	neuter
	een blauw bed$_{NEU}$	DG-SC	het/een blauw(e) trui$_{COM}$	masculine

		SG-DC	het/een groen(e) blad$_{NEU}$	masculine
	łóżko$_{NEU}$ (‘bed’)	SG-SC	het/een blauw(e) raam$_{NEU}$	neuter
42	het rode bureau$_{NEU}$*	DG-DC	de/een gele vogel$_{COM}$	masculine
	een rood bureau$_{NEU}$*	DG-SC	de/een rode bloem$_{COM}$	masculine
		SG-DC	het/een ge(e)l(e) hek$_{NEU}$	neuter
	biurko$_{NEU}$ (‘desk’)	SG-SC	het/een ro(o)d(e) hart$_{NEU}$	neuter
43	het groene oog$_{NEU}$*	DG-DC	de/een rode appel$_{COM}$	neuter
	een groen oog$_{NEU}$*	DG-SC	de/een groene lamp$_{COM}$	feminine
		SG-DC	het/een blauw(e) vest$_{NEU}$	feminine
	oko$_{NEU}$ (‘eye’)	SG-SC	het/een groen(e) blad$_{NEU}$	masculine
44	het gele hek$_{NEU}$	DG-DC	de/een blauwe wolk$_{COM}$	feminine
	een geel hek$_{NEU}$	DG-SC	de/een gele ster$_{COM}$	feminine
		SG-DC	het/een blauw(e) horloge$_{NEU}$	masculine
	ogrodzenie$_{NEU}$ (‘fence’)	SG-SC	het/een ge(e)l(e) strijkijzer$_{NEU}$	neuter
45	het rode hart$_{NEU}$	DG-DC	de/een blauwe doos$_{COM}$	neuter
	een rood hart$_{NEU}$	DG-SC	de/een rode appel$_{COM}$	neuter
		SG-DC	het/een blauw(e) boek$_{NEU}$	feminine
	serce$_{NEU}$ (‘heart’)	SG-SC	het/een ro(o)d(e) vliegtuig$_{NEU}$	masculine
46	het gele strijkijzer$_{NEU}$	DG-DC	de/een groene auto$_{COM}$	masculine
	een geel strijkijzer$_{NEU}$	DG-SC	de/een gele ster$_{COM}$	feminine
		SG-DC	het/een bruin(e) vat$_{NEU}$	feminine
	żelazko$_{NEU}$ (‘iron’)	SG-SC	het/een ge(e)l(e) slot$_{NEU}$	masculine
47	het gele wiel$_{NEU}$	DG-DC	de/een rode kaars$_{COM}$	feminine
	een geel wiel$_{NEU}$	DG-SC	de/een gele bus$_{COM}$	masculine
		SG-DC	het/een ro(o)d(e) hart$_{NEU}$	neuter
	koło$_{NEU}$ (‘wheel’)	SG-SC	het/een ge(e)l(e) anker$_{NEU}$	feminine
48	het blauwe raam$_{NEU}$	DG-DC	de/een rode bloem$_{COM}$	masculine
	een blauw raam$_{NEU}$	DG-SC	de/een blauwe wolk$_{COM}$	feminine
		SG-DC	het/een groen(e) potlood$_{NEU}$	masculine
	okno$_{NEU}$ (‘window’)	SG-SC	het/een blauw(e) horloge$_{NEU}$	masculine

* These words were initially translated differently and turned out to overlap in phonemic onset to a certain degree. Adding this factor in the model outlined in Chapter 6 revealed no significant impact. Moreover, deleting these items did not change any of the significant results.

** These words turned out to resemble another Polish word. For example, the pronunciation of the Dutch word *boek* (‘book’) resembles the Polish word for God, namely *‘Bóg’*. Similarly, the Dutch word for star (‘ster’) can refer to a rudder or wheel in Polish. No significant effects were found, however, when this variable was added in the model presented in Chapter 6.

B-2. Creating and eye tracking experiment using the E-Prime extensions for Tobii Software

The hand-out below was used during an invited talk entitled ‘Creating an Eye tracking Experiment using E-Prime Extensions for Tobii’ presented at the *Tobii Eye tracking Workshop* in Manchester on June 22nd, 2010. The overview represents the steps used to create the eye tracking experiment presented in Chapters 4 and 6.

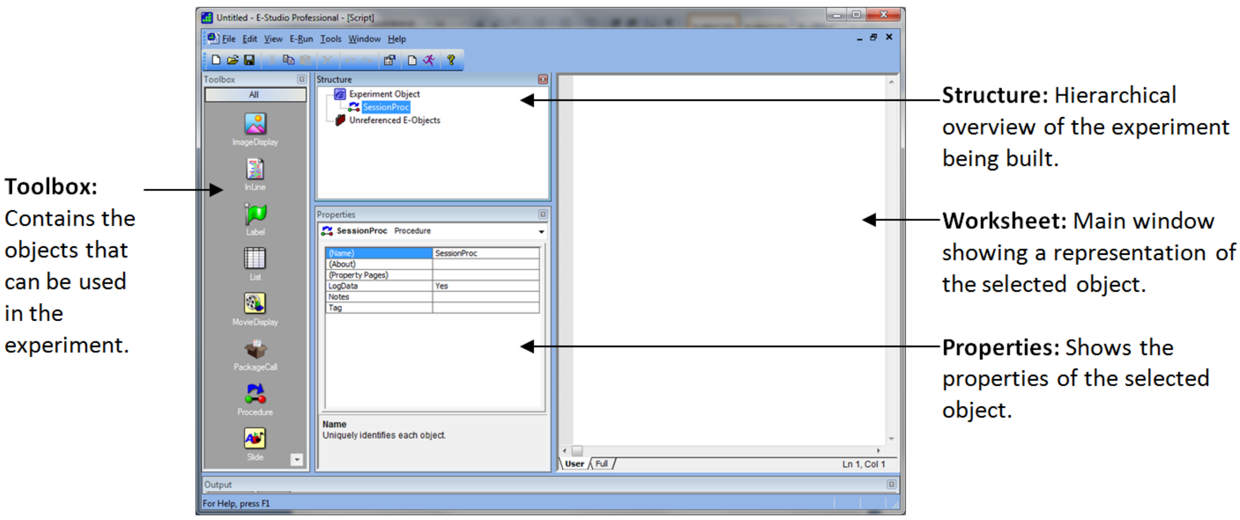
This handout will explain all practical steps in creating an eye tracking experiment using the E-Prime extensions for Tobii software. More information can be found in the E-Prime documentation and the E-Prime Extensions for Tobii manual.

1. Short Introduction to E-Prime

E-prime components:

- ***E-Studio*** → Graphical Design Environment
- E-Run → Real-Time Experiment Generator
- E-DataAid → Spreadsheet Application for E-prime Data Files
- E-Merge → Data Merging Unit
- E-Recovery → Converts txt files into ‘normal’ EDAT files
- ***E-basic*** → Full Scripting Language

E-Studio is used to design the experiment:



The basic hierarchical structure in E-prime:

- **E-Objects**: The separate events during the experiment (a picture, a prime, a fixation cross, a delay…) are chosen from the **Toolbox** menu.
- **TrialList:** A trial consists of an organized set of frames in which the stimuli are presented. The order and specifics of the trial are defined in the **TrialProc** (Trial Procedure)

- **BlockList:** A block defines a group of trials and the **BlockProc** (Block Procedure) defines the structure for one subtest.
- **SessionProc:** Session Procedure. Outlines the structure for the complete experiment. Items that apply to the whole set of tests (such as a general instruction) can be placed here.

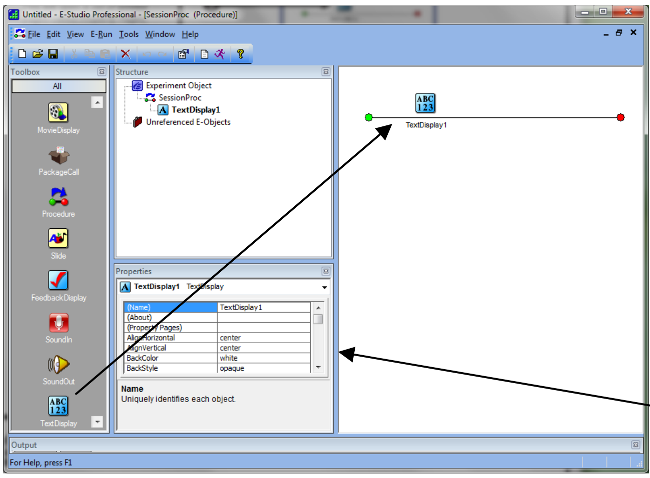
Some common E-objects in the Toolbox:

- **Text display**: a slide containing text
- **Image display**: a slide containing an image displayed to the subject
- **Slide**: Text, images, and/or sounds at the same time
- **Feedback display**: presents feedback to the subject about his/her performance
- **Wait**: inserts a delay between two other objects
- **Procedure**: organizes the frames and tools
- **List**: dictates how and how many times the procedures will be repeated
- **Inline**: if the experiment has particular features not included in E-Studio, InLine objects can be used to program these features using E-Basic code
- **Sound Out**: implements a sound within the experiment
- **Label**: Label tool is used to mark a point in the time line (useful when one wants to jump to a certain point in the experiment)
- **Package Call**: to connect E-Prime experiment to another device

E-objects are simply dragged onto the procedural timeline, after which their properties are set.

2. Building an E-prime experiment: the Visual World Paradigm

First, we will add a text display at the beginning of the *SessionProc* containing instructions for the participant.



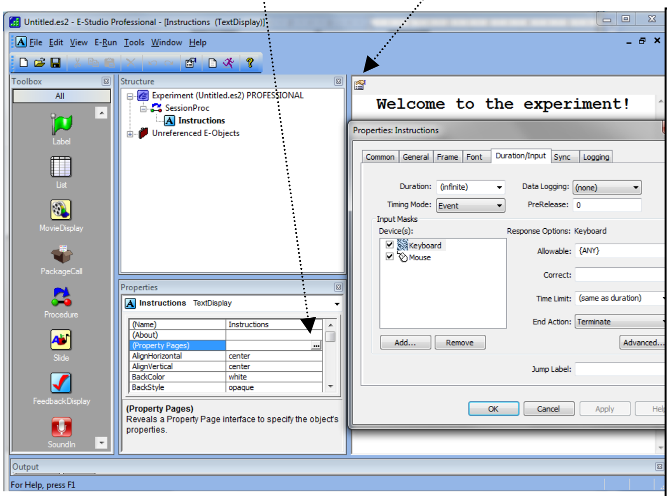
Double-click the *SessionProc* to display its procedural timeline.

Simply drag and drop a *TextDisplay* from the toolbox to the timeline.

Double-click the *TextDisplay* to open it in the workspace and type in the instructions for the experiment.

Note: Always change the name of the E-object to refer to its function.

Now, we need to tell E-Prime to display these instructions until the subject presses a key. To open the Property Pages, either click on the icon in the worksheet or select the Property Pages in the Properties Window .



Select the *Duration/Input* tab.

Duration can be set to anything in ms, but for these instructions we want the duration to depend on the subject pressing a button. Therefore, *Duration* reads 'infinite'.

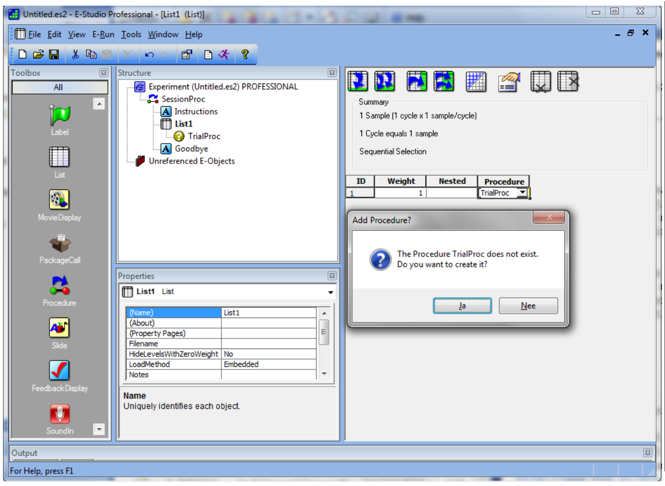
If there are no devices selected as input masks, click *add...* and select keyboard, mouse, or both. Allowable responses are in this case {ANY}, but you might want to type {SPACE} for the spacebar.

Time Limit is the same as *Duration* and *End Action* should read 'Terminate'. Now, the instructions display will stay on the screen until the subject presses any key using the mouse or the keyboard.

Note: If you want to add a different input mask, e.g. a Serial Response Box, you will first have to add this as a device for the complete experiment.

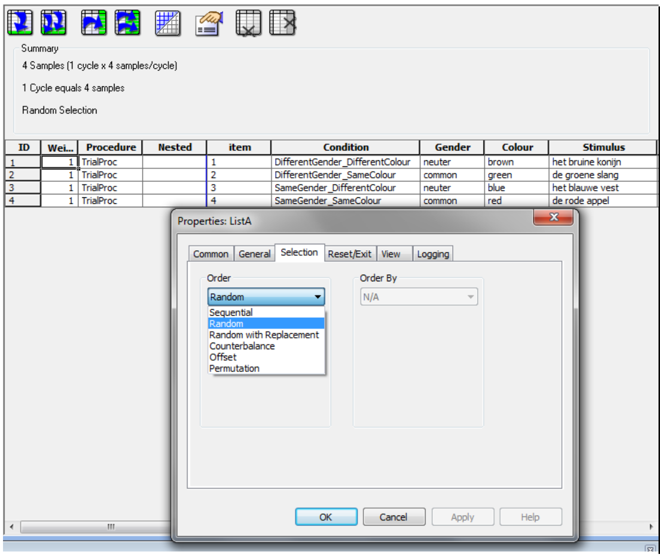
Using the above procedures, we will also add a 'Goodbye' *TextDisplay* to the end of the *SessionProc*.

You will now have to create a list with all the variables. Most experiments will also have a *BlockList* containing several lists (e.g. one with images and one with movies), but we will now proceed with a *TrialList*. This *TrialList* controls all the items in the trial sequence.



After dragging a list icon from the toolbox into the procedural timeline (in between the instructions and goodbye displays), type 'TrialProc' in the 3rd column of this list.

E-prime will notice that the *TrialProc* does not exist; click 'yes' to create it.



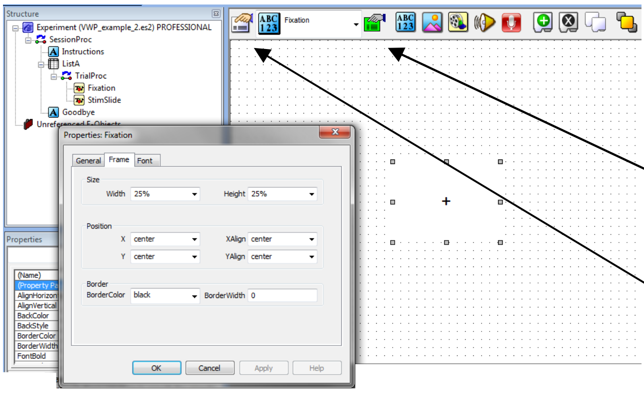
Each row in the list represents one trial and trials are added using the buttons. The 'weight' represents how many times the specific trial is run, which is once for the present experiment. The 'Nested' column allows one list to reference other lists, but we won't use that here.

The columns, or *attributes*, represent our variables. You can add columns by using the buttons.

Select the Properties of the *TrialList* and click on the *Selection* tab. Here, you can specify whether you want your trials to be presented sequentially or in random order.

Next, double click on *TrialProc* in the Structure view to open its procedural timeline. We will now specify a Procedure that will apply to all trials in the *TrialList*. In this case, we want a fixation cross to appear before the slide containing our four pictures. Normally, a simple *TextDisplay* would suffice for our fixation, but we will choose a *Slide* for both the Fixation and the Stimulus Slide, because we eventually want them to interact with the Tobii Eye Tracker.

Drag and drop two Slides, , onto the procedural timeline. Name the first one 'Fixation' and the second one 'StimSlide' (or any name you prefer).

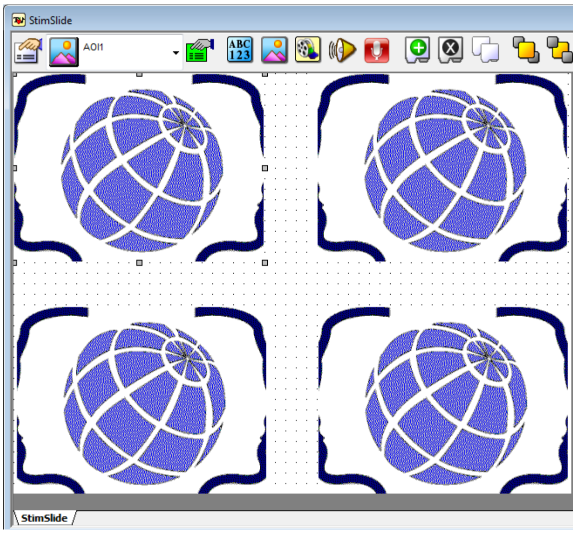


Open the Fixation slide and add a *TextDisplay* . Select the *TextDisplay* and type a '+'.

It is very important to place the fixation in the center of the screen, so select the sub-objects property pages , check the *Frame* tab and make sure that all drop-down menus under position read 'center'.

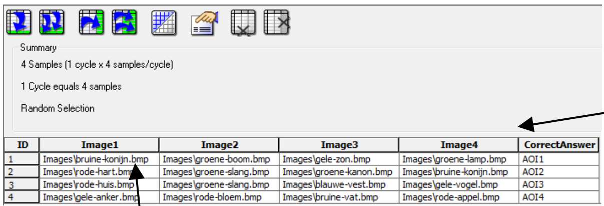
For now, we will show the fixation for 2 seconds, so select the Properties of the whole slide and go to the *Input/Duration* tab and enter 2000 for Duration.

In this VWP experiment, we will create a slide containing 4 images. We will specify the images as AOIs as in the picture on below.

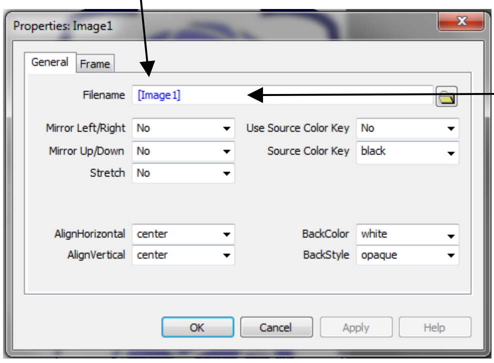


Open the 'StimSlide' and add four images, one in each corner of the slide. Just as it is important to have the fixation equidistant from the pictures, i.e. the center, it is also important to leave a space in between the pictures you will eventually use as Areas of Interest (AOIs).

Note: The Eyetracker will eventually provide E-prime with x and y coordinates of a gaze. In turn, E-prime will see whether these coordinates belong to one of our images. Consequently, you need to make sure that the coordinates to not overlap. This will be discussed in more detail below.



Then, for each picture, we have to specify which file, i.e. which picture, to display. Your *TrialList* will thus have four *attributes* (columns), each one corresponding to one picture.

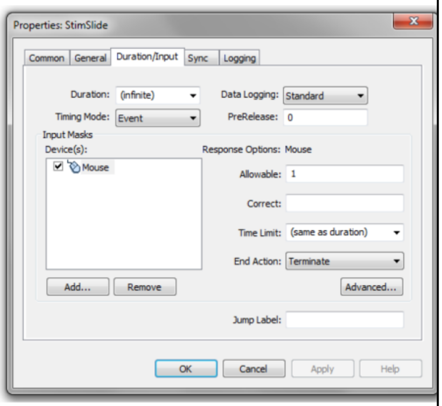


Select the sub-objects property pages of Image1.

Type [Image1] in the Filename box.

Note: By adding the square brackets, you're referring to an attribute in the TrialList. The text should turn blue.

In this VWP experiment, subjects will be asked to click on one of four pictures on the screen using the mouse.



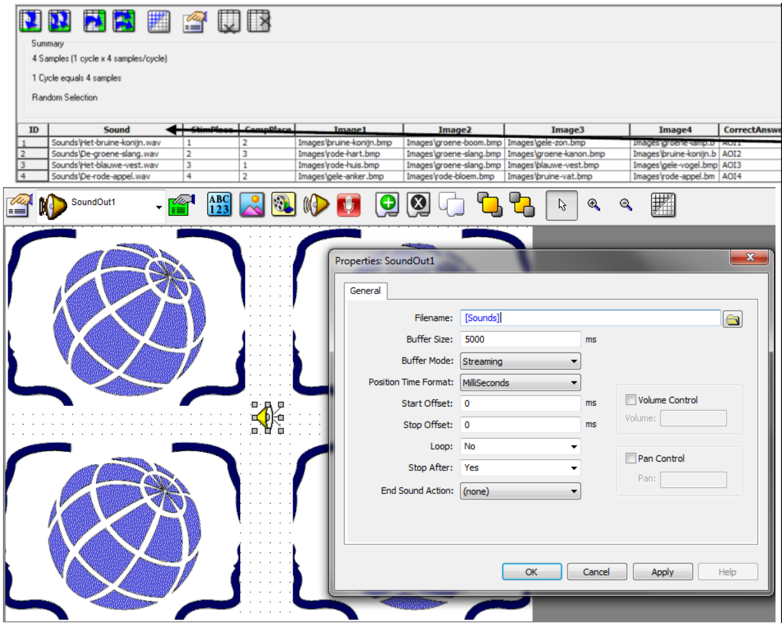
Select the Property Pages of the entire StimSlide and go to the *Duration/Input* tab. Similar to our specifications for the instructions *TextDisplay, Duration* reads 'infinite' because we want the duration to depend on the subject pressing a button.

If there are no devices selected as input masks, click *add…* and select the mouse. The allowable response is 1 (i.e. the left mouse button).

Time Limit is the same as *Duration* and *End Action* should read 'Terminate'. Now, our Stimulus Slide will be displayed until the subject clicks on one of the pictures using the mouse.

Note: Make sure the mouse cursor is shown on the screen by double-clicking the Experiment. Go to the Devices tab, highlight the mouse and click edit…. Check whether the drop-down menu for 'Show Cursor' read 'Yes'.

Note: In mot experiments, you would not want the cursor to be shown during the whole trial, e.g. not during the Fixation Slide (a moving cursor would be distracting and probably also annoying for the participant). These issues can only be solved by using InLine Scripts: one preceding the Fixation Slide to not show the cursor and one preceding the Stimulus Slide to show the cursor and place it in the center of the screen.



The only thing we need to add now are the auditory instructions, e.g. 'Click on the blue piano'. In the *TrialList*, we should thus have an attribute containing our wav-files. We'll name the attribute 'Sounds'.

Add a *SoundOut* to the StimSlide and open its Property Pages.

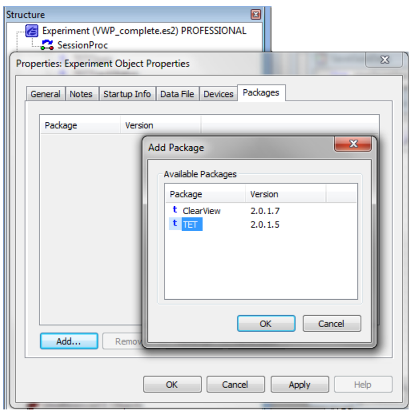
Specify the Filename that should be loaded for each trial by typing the reference to the attribute [Sounds].

3. Adding PackageCalls

In order to create an eye tracking experiment using E-prime extensions for Tobii, two methods of programming are integrated: PackageCalls and InLine Scripts.

- **PackageCalls:** pre-written scripts to automatically perform various functions,
- **InLine Scripts:** user-defined scripts using E-basic scripting language.

Before being able to add PackageCalls, you have to add the TET package file to the experiment. This Package contains the routines that are used to communicate with the TET software.



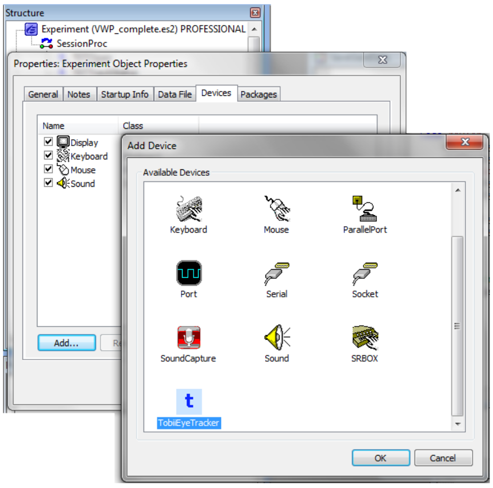
Open the Experiment Object Properties by double clicking on the Experiment Object.

Click on the *Packages* tab. This tab contains all available packages.

Click *add..* and then select TET.

Note: If you would also like to record using ClearView you can add this package here as well.

In order to connect to the eye tracker, we must also add this as a device.



We are still in the Experiment Object Properties.

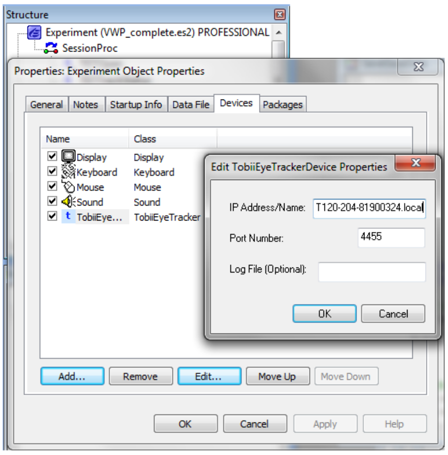
Click on the *Devices* tab. This tab allows you to add all kinds of devices, such as mouse, keyboard, serial response box, and so on.....

Click *add..* and then select *TobiiEyeTracker.*

Note: do not change the order of (selected) devices: it has a purpose! The Display Device must open before TobiiEyeTracker can, so never move TET to the top!

Make sure all devices that you need are selected before pressing *ok*.

In order for E-Prime to be able to communicate with the TET server, we must add its I.P. address. So, first check the I.P. address of the computer that houses the TET server, i.e. the computer containing E-studio.



We are still in the *Devices* tab within the Experiment Object Properties.

Highlight the *TobiiEyeTracker* and click the *Edit…* button. Enter the I.P. address.

The Port Number is provided automatically with default number 4455.

Note: If you're using the T-Series of Tobii (T60/T120), you will need to add the extension '.local' to the I.P. address.

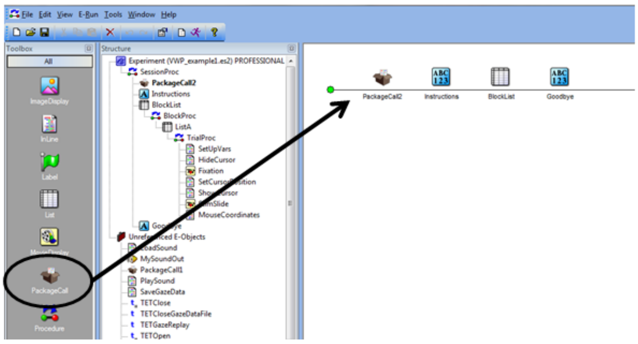
Required packages in every experiment:

1. TETOpen: used to signal the Eye Tracker to begin collecting eye movement data
2. TETTrackStatus: used to display a track status window after calibration and prior to the beginning of the experiment
3. TETOpenGazeDataFile: used to open the tab delimited GazeData file that will collect all eye movements (.gazedata)
4. TETStartTracking: used to start tracking/recording eye movements
5. TETStopTracking: used to stop tracking/recording eye movements
6. TETCloseGazeDataFile: used to end data collection
7. TETClose: used to close the TET Server

Note: Always rename the TET packages to reflect the Routine!

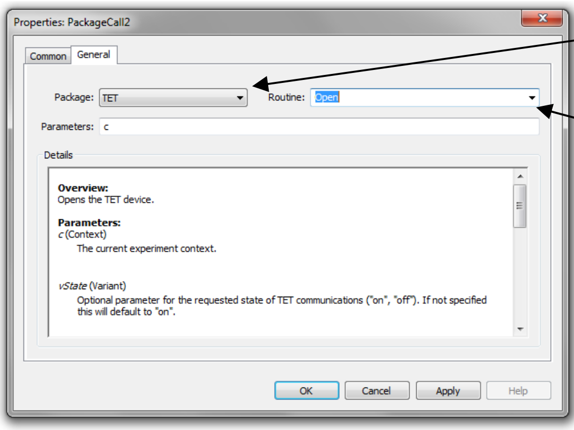
Placement of PackageCalls

First, we will add the TET_Open and TET_Close Routines to begin and stop collecting eye movement data. We want to collect and record eye gaze data for those portions that will be analysed eventually. For this short experiment, we'll start recording at the beginning of the experiment. One may choose to start collecting eye movements later (at the trial level, for example).



PackageCalls are added, like any other E-Object, from the toolbox, by dragging it and dropping it into the procedural timeline.

After double-clicking the PackageCall, you should first rename it to refer to its routine, in this case TETOpen.



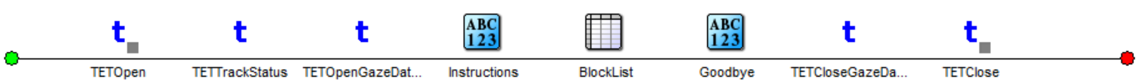
First select the Package TET from the Package dropdown list.

Then select the correct Routine, in this case Open, from the Routine dropdown list.

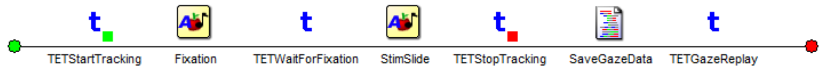
The default parameters are 'c', but you can add any parameter mentioned in the Details of the PackageCall.

This procedure is repeated for the other required PackageCalls: simply drag and drop them in the appropriate spot in the timeline, rename them, and select TET Package and the appropriate routine. See the E-Prime Extensions for Tobii manual for more information on each of the PackageCalls.

The timeline of the *SessionProc* will now look something like this:



Within the *TrialProc*, we will add two packages to start and stop tracking. The *TETStartTracking* will be placed at the beginning of the Trial, because (in most experiments) we want to start gathering eye movement data here. The procedure for adding these is exactly the same as explained above. For these required PackageCalls it often suffices to use the default parameter 'c'.

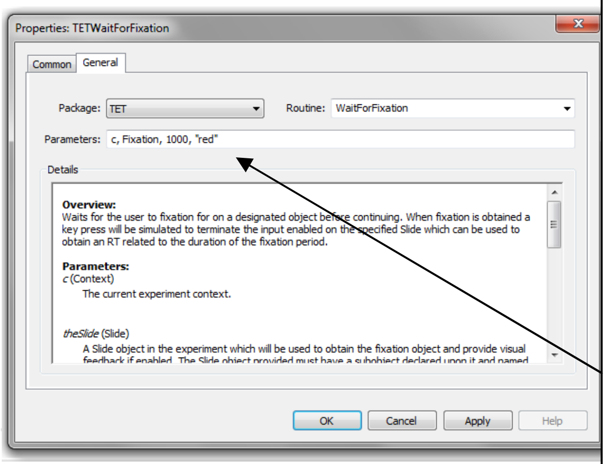


Optional packages

As can be seen from the Trial timeline above, we will use some additional Routines: *TETWaitForFixation* and *TETGazeReplay*.

TETWaitForFixation

It is very helpful, but also very important to have an initial centralized fixation point where the subject is instructed to look at the beginning of each trial. We add *TETWaitForFixation* to make sure people have fixated the crosshair and the trial will only start once they have looked at the fixation for a certain amount of time. Drag and drop another *PackageCall* into the *TrialProc* and place it directly after the slide containing the fixation. Again, select TET from the Package list and select the correct routine, in this case *WaitForFixation*, from the routine list.



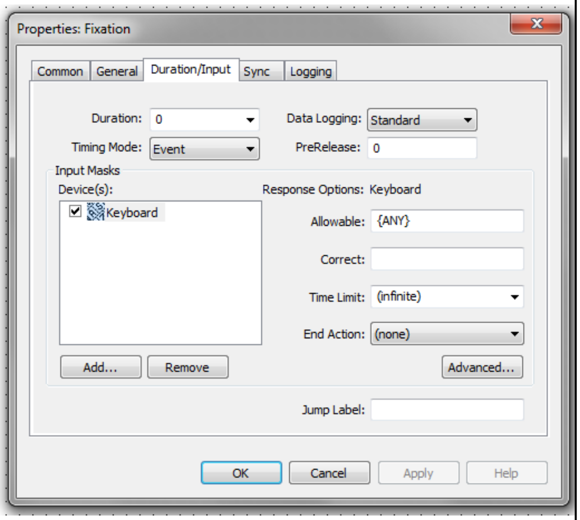
In this case, we need more parameters than just the default 'c'. We add:

- Fixation; refers to the slide containing the fixation
- 1000: refers to the time (ms) the subject has to fixate on the fixation before starting the trial
- "red" (optional): this parameter is used to create a coloured square around the fixation, if fixation to the object takes place.

The Parameters box reads: c, Fixation, 1000, "red"

Note: If you want the coloured square around the Fixation, make sure the fixation sub-object has a border. Select the Fixation sub-object Property Pages, open the Frame tab and check whether BorderWidth > 0.

We will have to change the properties of the slide to work in conjunction with the *TETWaitForFixation* PackageCall. Select the Fixation Slide (not the object) and open its Property Pages. Then select the *Duration/Input* tab.



The Duration should be '0', to ensure that the object is displayed on the screen, but to ensure the *PackageCall* is executed immediately after.

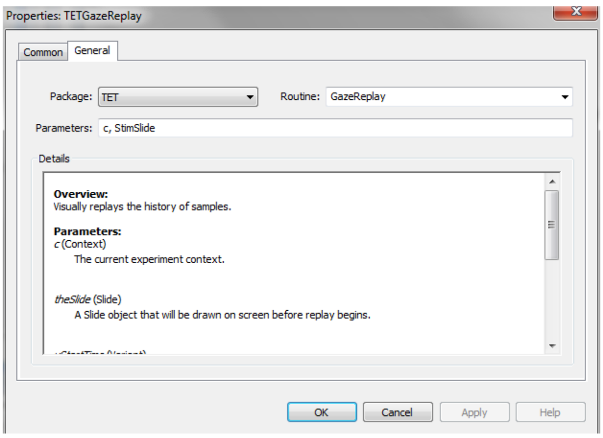
Data Logging should be 'Standard', *Timing Mode* should read 'Event' and the *PreRelease* '0'.

The *PackageCall* will enter a loop that will only exit once a response has been made, by actively accessing the GazeData information. We will add the Keyboard as *Device* in the *Input Masks*.

In the response options window, make sure that *Allowable* reads {ANY}. The *Time Limit* will be {infinite} and *End Action* reads {none}.

We now made sure that once a fixation on the object has been made for 1000 ms (we specified that in the *TETWaitForFixation* PackageCall), a key press will be simulated and the experiment will move on.

Another optional Routine is the *TETGazeReplay*, which is used to visually replay the eye movements during the experiment.



Add a new *PackageCall* to follow the *TETStopTracking* PackageCall (at the end of the *TrialProc*) and choose the Package TET and the Routine *GazeReplay*.

We have to specify which Slide object has to be drawn on the screen before replay begins. In this case, we want to see a *GazeReplay* on our *StimSlide.*

In addition, we have to add a start time stamp for the first gaze point to be included in the replay, which will be the *StimSlide.OnsetTime.* If not specified, we will see the gaze from the onset of the Fixation.

The Parameters box reads: c, StimSlide, StimSlide.OnsetTime.

Others PackageCalls:

- TET_Connect: establishes a connection with the TET over the network
- TET_Disconnect: disconnects with the TET over the network
- TET_ClearHistory: clears the history of accumulated eye tracking samples
- TET_WriteGazeData file: writes a line of data

4. Adding InLine Scripts to create an appropriate Output file

We have already asked the TET Server to open a tab delimited GazeData file that will collect all eye movements (.gazedata). We still need to specify what details we want to have in this

GazeData file. Because PackageCalls contain pre-written routines, we need to make a user-defined InLine Script in which we can specify the information we want to have in our output. The Output file will contain all the information of our specified variables per eye gaze observation with each sample resulting into one row in the output.

The scripting language used in E-prime is called E-Basic and there are a couple of ground rules.

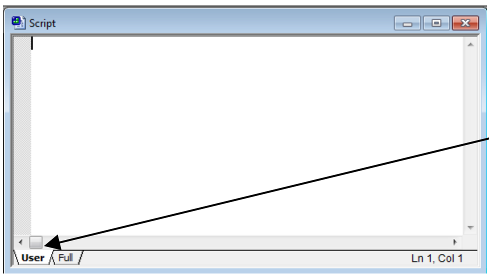
- Comments/notes in the script are set off from the code by a single quotation mark ' to and are displayed in green.
- Keywords or symbols recognized by E-Basic as part of the language are displayed in blue.
- User-entered script concerning statements is displayed in black.
- Strings of data values (referencing attributes in the TrialList) are displayed in red.

User-written script is generally placed in one of three places in an experiment:

- **User Script Window:** Next to Full Script (which is generated by E-Studio). Used to declare global variables. Also, it should be used to declare functions and subroutines (sub..end sub), which will interfere with the automatically generated script if put outside the Usertab and in Inline.
- **InLine Objects at any given time:** to insert segments of user-written script within an experiment. This is the most common placement of script.
- **InLine Objects at the beginning of the experiment:** to initialize variables: For example, global variables declared on the User tab must be initialized prior to their use, and it is common to do so at the beginning of the experiment.

Note: variables or attributes declared within an E-Prime experiment are limited to the scope of the Procedure in which they are defined! Some need to be declared at the local Trial Level, others at the Session Level.

We will use the UserScript provided by E-prime extensions for Tobii, which can be found in "...My Documents\My Experiments\Tobii\Tutorials\TET\FixedPositionAOI\Userscript.txt



Open the Userscript.txt, select all and copy it.

Open the script view (Alt+5) and select the User tab of the script.

Paste the copied text into the User Script Window.

The code provided in theUserScript.txt can be copied to any experiment. You will need to add and alter the variables you want to include in the output. For most experiments, you would want to add many, if not all, variables specified in your TrialList.

ID	Nested	item	Condition	Gender	Colour	Stimulus	Competitor	Sound	StimPlace	CompPlace	Image1	Image2	Image3	Image4	CorrectAnswer
1		1	DifferentGender_DifferentColour	neuter	brown	het bruine konijn	de groene boom	Sounds\Het-bruine-konijn.wav	1	2	Images\bruine-konijn.bmp	Images\groene-boom.bmp	Images\gele-zon.bmp	Images\groene-lamp.bmp	AOI1
2		2	DifferentGender_SameColour	common	green	de groene slang	het groene kanon	Sounds\De-groene-slang.wav	2	3	Images\rode-hart.bmp	Images\groene-slang.bmp	Images\groene-kanon.bm	Images\bruine-konijn.bmp	AOI2
3		3	SameGender_DifferentColour	neuter	blue	het blauwe vest	het rode huis	Sounds\Het-blauwe-vest.wav	3	1	Images\rode-huis.bmp	Images\groene-slang.bmp	Images\blauwe-vest.bmp	Images\gele-vogel.bmp	AOI3
4		4	SameGender_SameColour	common	red	de rode appel	de rode bloem	Sounds\De-rode-appel.wav	4	2	Images\gele-anker.bmp	Images\rode-bloem.bmp	Images\bruine-vat.bmp	Images\rode-appel.bmp	AOI4

In addition, you also want to include response (RT) and accuracy (ACC) measures from E-prime. They should be added in three places within the UserScript.

The first place to enter your variables is at the beginning of the User Script Window. The Type Statement is used to declare the UserEyeGazeData data types.
Some common data types:

- String: Sequences of characters
- Boolean: True (1) or False (0) value
- Integer: Whole number ranging from –32767 to 32767
- Long: Whole number ranging from –2,147,483,648 to 2,147,483,647

Here, we declare the data types of our 4 AOIs. We will name and specify the referents of these Areas of Interest later on.
We want to include some response and accuracy measures in the output, as calculated by E-prime.
These measures all refer to our variables in the TrialList.
The variable 'UserDefined_1' (name can be changed) will contain information on the slide that is being presented.

```
' For each new piece of user defined data that you add you must perform the following ste
'   1. Add a new member to the end of the UserEyeGazeData data structure defined below.
'   2. Edit the script in the UserWriteGazeDataFile subroutine to append a tab character
' new column label to the section that writes out the column labels.
'   3. Edit the script in the UserWriteGazeDataFile subroutine to append a tab character
' the current value of the new variable.
'
'   Type UserEyeGazeData data structure is used to keep track of information specific to
' data per eye gaze observation.  You may use it to help you to track and log any additio
' experiment related data that you need to associate with the gaze data.
Type UserEyeGazeData
    TrialId As String
    StateName As String
    AOI1 As String
    AOI2 As String
    AOI3 As String
    AOI4 As String
    AOI As String
    AOIStimulus As String
    CRESP As String
    RESP As String
    ACC As String
    RT As Long
    Condition As String
    Gender As String
    Colour As String
    item As String
    Stimulus As String
    StimPlace As String
    Competitor As String
    CompPlace As String
    UserDefined_1 As String
    ' 1. Declare additional User Defined values here as needed
End Type
```

The Sub Statement here is used to enter a subroutine, which is composed of a series of commands combined into a unit. This unit may be run by a call to the subroutine from within an InLine object.

This is the second place to add our variables. It is located in the middle of the User Script, below 'Append additional user defined data'.

Here, you will need to add an ebTab line for each of the new data values you created in the previous step. This part of the script will make sure that, for each of the variables we specify, a column will be made and names accordingly in the .gazedata output file.

Note: You can add anything you'd like, but it is advisable to keep the order consistent across these three steps.

```
' Write a line of tab delimited data to the user defined gaze data file.
Sub UserWriteGazeDataFile( c As Context, theGazeData As TobiiEyeTrackerResponseData, th

    Dim strOut As String

    ' Need to have a valid data structure.
    Debug.Assert Not theGazeData Is Nothing

    ' If this is the first call to the file since it was opened then write
    ' the column headers
    If TET_GetGazeDataFileLineNumber() = 0 Then

        ' Common prefix
        strOut = "Subject" & ebTab & "Session" & ebTab & "ID" & ebTab & "TETTime"

        ' Append standard data from TET
        strOut = strOut & _
            ebTab & "RTTime" & ebTab & "CursorX" & ebTab & "CursorY" & ebTab & "Timesta
            ebTab & "XGazePosLeftEye"  & ebTab & "YGazePosLeftEye"  & ebTab & "XCameraP
            ebTab & "XGazePosRightEye" & ebTab & "YGazePosRightEye" & ebTab & "XCameraP

        ' Append additional user defined data
        strOut = strOut &_
            ebTab & "TrialId" & _
            ebTab & "AOI1" & _
            ebTab & "AOI2" & _
            ebTab & "AOI3" & _
            ebTab & "AOI4" & _
            ebTab & "AOI" & _
            ebTab & "AOIStimulus" & _
            ebTab & "CRESP" & _
            ebTab & "RESP" & _
            ebTab & "ACC" & _
            ebTab & "RT" & _
            ebTab & "Condition" & _
            ebTab & "Gender" & _
            ebTab & "Colour" & _
            ebTab & "item" & _
            ebTab & "Stimulus" & _
            ebTab & "StimPlace" & _
            ebTab & "Competitor" & _
            ebTab & "CompPlace" & _
            ebTab & "UserDefined_1"
            ' 2. Declare additional column labels for User Defined columns here as need

        ' Write the line to the file
        TET_WriteGazeDataFile c, strOut
    End If
```

Finally, we have to add an ebTab line in the UserEyeGazeData group for each of the new data value created before. Locate the 'Append additional user defined data', which is located at the end of the User Script.

This part of the script will write the observed values of our specified variables to the output.

Note: the user defined columns are empty if there's no valid observation!

```
    ' Write the new line of data values to the file

    ' Common prefix
    strOut = c.GetAttrib( "Subject" ) & ebTab & c.GetAttrib( "Session" ) & ebTab & TET_GetGazeDataFileLineNumber() & ebTab & Format$( TET_ConvertToMilliseconds( theGazeData.TimestampSec, theGazeDat

    ' Append standard data from TET
    strOut = strOut & _
        ebTab & theGazeData.RTTime & ebTab & theGazeData.CursorX & ebTab & theGazeData.CursorY & ebTab & theGazeData.TimestampSec & ebTab & theGazeData.TimestampMicrosec & _
        ebTab & theGazeData.XGazePosLeftEye  & ebTab & theGazeData.YGazePosLeftEye  & ebTab & theGazeData.XCameraPosLeftEye  & ebTab & theGazeData.YCameraPosLeftEye &  ebTab & theGazeData.DiameterP
        ebTab & theGazeData.XGazePosRightEye & ebTab & theGazeData.YGazePosRightEye & ebTab & theGazeData.XCameraPosRightEye & ebTab & theGazeData.YCameraPosRightEye & ebTab & theGazeData.DiameterP

    ' Append additional user defined data
    ' NOTE: User defined columns may be empty if there was not a valid observation
    strOut = strOut & _
        ebTab & theUserEyeGazeData.TrialId & _
        ebTab & theUserEyeGazeData.AOI1 & _
        ebTab & theUserEyeGazeData.AOI2 & _
        ebTab & theUserEyeGazeData.AOI3 & _
        ebTab & theUserEyeGazeData.AOI4 & _
        ebTab & theUserEyeGazeData.AOI & _
        ebTab & theUserEyeGazeData.AOIStimulus & _
        ebTab & theUserEyeGazeData.CRESP & _
        ebTab & theUserEyeGazeData.RESP & _
        ebTab & theUserEyeGazeData.ACC & _
        ebTab & theUserEyeGazeData.RT & _
        ebTab & theUserEyeGazeData.Condition & _
        ebTab & theUserEyeGazeData.Gender & _
        ebTab & theUserEyeGazeData.Colour & _
        ebTab & theUserEyeGazeData.item & _
        ebTab & theUserEyeGazeData.Stimulus & _
        ebTab & theUserEyeGazeData.StimPlace & _
        ebTab & theUserEyeGazeData.Competitor & _
        ebTab & theUserEyeGazeData.CompPlace & _
        ebTab & theUserEyeGazeData.UserDefined_1
        ' 3. Declare additional User Defined values here as needed

    ' Write the line to the file
    TET_WriteGazeDataFile c, strOut
End Sub
```

We have now declared all the variables in the UserScript. The only thing left to do is to make sure the data is written to a tab delimited GazeData file. We will have to create an InLine script and you can copy and paste the code that's provided by E-prime extensions for Tobii, which can be found in "…My Documents\My Experiments\Tobii\Tutorials\TET\FixedPositionAOI\SaveGazeData.txt

Drag a new InLine object from the toolbox and drop it after the TETStopTracking PackageCall. Rename the InLine object to read SaveGazeData. After copying the code from SaveGazeData.txt into the InLine script, rename it to SaveGazeData. Now, we need to add our user defined variables.

```
' Get access to the critical stimulus object and the slide state t
Dim theSlide As Slide
Dim theState As SlideState
Set theSlide = StimSlide
Set theState = theSlide.States(theSlide.ActiveState)
```

First, the variables are set allowing access to the critical slide that contains the Areas of Interest (AOIs) that we will specify in the StimSlide later.

Note: When you're using several inline scripts, you might choose to set all the variables at once in an InLine at the beginning of the SessionProc.

```
Dim theGazeData As TobiiEyeTrackerResponseData
Dim theUserEyeGazeData As UserEyeGazeData
Dim nMaxHistoryCount As Long
Dim n As Long

' Set defaults for the user defined data
theUserEyeGazeData.TrialId = c.GetAttrib( c.GetAttrib("Running") & ".Sample" ) ' obtain
theUserEyeGazeData.AOI1 = c.GetAttrib( "Image1" )
theUserEyeGazeData.AOI2 = c.GetAttrib( "Image2" )
theUserEyeGazeData.AOI3 = c.GetAttrib( "Image3" )
theUserEyeGazeData.AOI4 = c.GetAttrib( "Image4" )
theUserEyeGazeData.AOI = ""
theUserEyeGazeData.AOIStimulus = ""
theUserEyeGazeData.CRESP = StimSlide.CRESP
theUserEyeGazeData.RESP = StimSlide.RESP
theUserEyeGazeData.ACC = StimSlide.ACC
theUserEyeGazeData.RT = StimSlide.RT
theUserEyeGazeData.Condition = c.GetAttrib( "Condition" )
theUserEyeGazeData.Gender = c.GetAttrib( "Gender" )
theUserEyeGazeData.Colour = c.GetAttrib( "Colour" )
theUserEyeGazeData.item = c.GetAttrib( "item" )
theUserEyeGazeData.Stimulus = c.GetAttrib( "Stimulus" )
theUserEyeGazeData.StimPlace = c.GetAttrib( "StimPlace" )
theUserEyeGazeData.Competitor = c.GetAttrib( "Competitor" )
theUserEyeGazeData.CompPlace = c.GetAttrib( "CompPlace" )
theUserEyeGazeData.UserDefined_1 = ""
```

Now, we need to assign values to the columns that will be output into the GazeData file.

Here, we specify that the four columns named AOI1, AOI2, AOI3, and AOI4 will record the displayed images specified in the attributes for each recorded eye gaze sample.

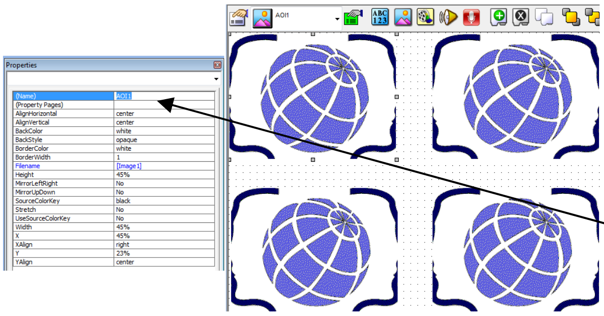
Empty quotation marks are used to indicate an empty string value and is used if you do not want to specify a default value. In this case, AOI will record the AOI the subject was looking at at that time and AOIStimulus will record the corresponding bmp image.

```
' See how many samples are currently saved in the history
nMaxHistoryCount = TobiiEyeTracker.History.Count

' Write out all the accumulated gaze data
For n = 1 To nMaxHistoryCount

    ' Get the next gaze data point
    Set theGazeData = CTobiiEyeTrackerResponseData( TobiiEyeTracker.History( n )
    If Not theGazeData Is Nothing Then

        ' Determine which E-Prime object was running when this sample was taken
        If theGazeData.RTTime >= Fixation.OnsetTime Then
            theUserEyeGazeData.UserDefined_1 = "Fixation"
        End If
        If theGazeData.RTTime >= StimSlide.OnsetTime Then
            theUserEyeGazeData.UserDefined_1 = "StimSlide"
        End If
```

The second part of the SaveGazeData InLine controls how the GazeData is associated with the objects presented on the screen. Here, we specify what should be recorded for the empty string value 'Userdefined_1'.

In the output, a column called Userdefined_1 will show whether the gazedata was recorded during presentation of the Fixation or the StimSlide.

Now, one important thing needs to be done. We need to specify our Areas of Interest (AOIs), which are the areas on the screen which represent critical objects (in our case the images). We specified AOI columns in the UserScript and we need to populate them with data.

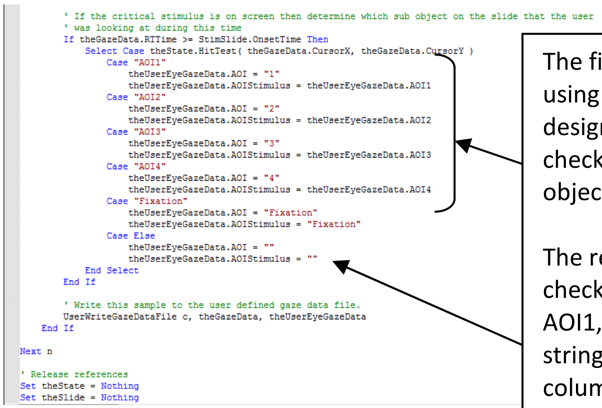


We will rename the sub-objects as:

AOI1 AOI2

AOI3 AOI4

To rename the image, select the image and type in the new name in the Properties Window.



```
' If the critical stimulus is on screen then determine which sub object on the slide that the user
' was looking at during this time
If theGazeData.RTTime >= StimSlide.OnsetTime Then
    Select Case theState.HitTest( theGazeData.CursorX, theGazeData.CursorY )
        Case "AOI1"
            theUserEyeGazeData.AOI = "1"
            theUserEyeGazeData.AOIStimulus = theUserEyeGazeData.AOI1
        Case "AOI2"
            theUserEyeGazeData.AOI = "2"
            theUserEyeGazeData.AOIStimulus = theUserEyeGazeData.AOI2
        Case "AOI3"
            theUserEyeGazeData.AOI = "3"
            theUserEyeGazeData.AOIStimulus = theUserEyeGazeData.AOI3
        Case "AOI4"
            theUserEyeGazeData.AOI = "4"
            theUserEyeGazeData.AOIStimulus = theUserEyeGazeData.AOI4
        Case "Fixation"
            theUserEyeGazeData.AOI = "Fixation"
            theUserEyeGazeData.AOIStimulus = "Fixation"
        Case Else
            theUserEyeGazeData.AOI = ""
            theUserEyeGazeData.AOIStimulus = ""
    End Select
End If

' Write this sample to the user defined gaze data file.
UserWriteGazeDataFile c, theGazeData, theUserEyeGazeData
End If

Next n

' Release references
Set theState = Nothing
Set theSlide = Nothing
```

The final part of the SaveGazeData Inline declares the AOIs using the HitTest method in E-prime, which was originally designed to track mouse coordinates. This method allows to check whether a specific gaze was within one of the critical objects

The recorded x and y coordinates are calculated and E-prime checks whether these coordinates were within the regions of AOI1, AOI2, AOI3, or AOI4. The previously defined empty string, AOI, is now getting a value of 1 to 4 and the output column for AOI Stimulus will specify the specific image that was looked at.

If the gaze was not within one of the objects, the output will return an empty string.

Note: Generate and Run your experiment after each change to the InLine Script. Otherwise, it is hard to figure out where the scripting went wrong.

C-1. Materials used in the ERP experiment

List of all experimental sentences used in the ERP experiment. Gender agreement conditions are divided into DN (definite determiner-noun), DaN (definite determiner-adjective-noun), and IaN (indefinite determiner-adjective-noun). In addition, each sentence contains a high and low cloze variant.

Finiteness sentences either consist of a violation on the past participle or on the infinitive.

Gender agreement condition: definite determiner-noun (DN)

Common Targets

Item	Cloze	Sentence
1	high	Jakob gaat iedere zondag bidden in **de/*het kerk** in zijn woonplaats.
	low	Jakob bakt verse koekjes voor **de/*het kerk** in zijn woonplaats.
2	high	Vera plant rode rozen in **de/*het tuin** van haar ouders.
	low	Vera breit een sjaal in **de/*het tuin** van haar ouders.
3	high	Het muisje werd 's avonds opgegeten door **de/*het kat** van de buren.
	low	Het meisje werd 's avonds aangevallen door **de/*het kat** van de buren.
4	high	Nadat hij de koeien had gemolken ging **de/*het boer** aan het werk op het land.
	low	Nadat hij een gezond ontbijt had gehad ging **de/*het boer** aan het werk op het land.
5	high	De gezonde jongen gaat nooit met de lift, maar neemt altijd **de/*het trap** naar de tweede verdieping.
	low	De gezonde jongen gaat vaak onderuit, maar neemt altijd **de/*het trap** naar de tweede verdieping.
6	high	Marianne wandelt dagelijks met **de/*het hond** door het park.
	low	Marianne winkelt dagelijks met **de/*het hond** aan de riem.
7	high	De arts besloot het slechte nieuws persoonlijk aan **de/*het patiënt** te vertellen.
	low	De man besloot het slechte nieuws persoonlijk aan **de/*het patiënt** te vertellen.
8	high	De docent vraagt om stilte aan het begin van **de/*het les** over de Middeleeuwen.
	low	Annette schilt een appel aan het begin van **de/*het les** over de Middeleeuwen.
9	high	Onze nieuwe woning bevindt zich op de hoek van **de/*het straat** vlak bij het bos.
	low	Maurits en Willem hebben om acht uur afgesproken in **de/*het straat** bij het bos.
10	high	De toerist bezoekt veel monumenten tijdens **de/*het vakantie** op Ameland.
	low	De student leert nieuwe talen tijdens **de/*het vakantie** op Ameland.
11	high	Silvia gaf de huilende baby aan **de/*het moeder** en zette koffie.
	low	Silvia bracht verse koffie naar **de/*het moeder** van de tweeling.
12	high	Kees heeft alweer een lekke band en leent vandaag **de/*het fiets** van zijn zus.
	low	Kees gaat vandaag vroeg naar zijn werk en neemt **de/*het fiets** van zijn zus mee.
13	high	Pieter ligt al de hele dag op **de/*het bank** televisie te kijken.
	low	Pieter heeft per ongeluk zijn sleutels op **de/*het bank** laten liggen.
14	high	Marieke heeft geen idee hoe laat het is en kijkt naar **de/*het klok** in de keuken.
	low	Marieke heeft speciaal voor haar moeder de klok in **de/*het klok** goed gezet.
15	high	Richard wil in tachtig dagen een reis om **de/*het wereld** maken.
	low	Richard wil na zijn studie **de/*het wereld** helpen te verbeteren.

Neuter Targets

Item	Cloze	Sentence
1	high	Laurens slaapt aan de linkerkant van **het/*de bed**, omdat die kant lekkerder ligt.

	low	Laurens zit verveeld tv te kijken op **het/*de bed** in zijn slaapkamer.
2	high	Het werd pikkedonker in de keuken toen plotseling **het/*de licht** uitviel.
	low	Het werd gevaarlijk in de keuken toen plotseling **het/*de licht** uitviel.
3	high	Marijn kreeg een ontzettend moeilijke vraag van de leraar, maar hij wist **het/*de antwoord** wel meteen.
	low	Marijn kreeg het ontzettend warm en benauwd, want hij kon **het/*de antwoord** op de vraag niet geven.
4	high	Jeroen Pauw is bekend als presentator van **het/*de programma** Pauw & Witteman.
	low	Jeroen heeft een ontzettende hekel aan **het/*de programma** Pauw & Witteman.
5	high	De interviewer stelde zeer interessante vragen tijdens **het/*de gesprek** met de minister-president.
	low	De oliemagnaat stond uit verveling lange tijd te gapen tijdens **het/*de gesprek** met zijn werknemers.
6	high	De herder scheert ieder jaar de wol van **het/*de schaap** voor de verkoop.
	low	De handelaar controleert de kwaliteit van **het/*de schaap** voor de verkoop.
7	high	De bakker in ons dorp bakte **het/*de brood** volgens speciaal recept.
	low	Mijn buurvrouw maakte **het/*de brood** volgens speciaal recept.
8	high	Na het ongeluk werd de vrouw per ambulance naar **het/*de ziekenhuis** gebracht.
	low	Na de wedstrijd werd de vrouw met een taxi naar **het/*de ziekenhuis** gebracht.
9	high	De journalist kijkt iedere ochtend naar **het/*de nieuws** op televisie.
	low	De pianist is altijd erg geïnteresseerd in **het nieuws** op televisie.
10	high	Peter zette de bos bloemen naast de computer op **het/*de bureau** van de zakenman.
	low	Peter vond de speelgoedauto naast de snoeppot op **het bureau** van de zakenman.
11	high	Steven is benoemd tot de nieuwe directeur van **het/*de bedrijf** van zijn vader.
	low	Steven bezoekt iedere vrijdagmiddag **het/*de bedrijf** van zijn vader.
12	high	Johan betaalde de motor met **het/*de geld** dat hij voor zijn verjaardag kreeg.
	low	Johan was erg blij met **het/*de geld** dat hij voor zijn verjaardag kreeg.
13	high	In de winter schaatst Henk dagelijks op **het/*de meer** achter zijn huis.
	low	In de winter zit Henk graag bij **het/*de meer** achter zijn huis.
14	high	De schipper legt een stevige knoop in **het/*de touw** van de mast en vertrekt dan richting Ameland.
	low	De schipper ziet plotseling **het/*de touw** van de mast losgaan en rent er direct naartoe.
15	high	Japan staat bekend als **het/*de land** van de moderne technologie.
	low	Jacob was niet bekend met **het/*de land** van de moderne technologie.

Gender agreement condition: definite determiner-adjective-noun (DaN)

Common Targets

Item	Cloze	Sentence
1	high	Na de verkiezingen werd Wouter Bos benoemd tot **de/*het nieuwe minister** van Financiën.
	low	Iedereen vraagt zich af wie zich binnenkort **de/*het nieuwe minister** van Financiën mag noemen.
2	high	Peter is geboren in een klein plaatsje en verdwaalt vaak in **de/*het grote stad** waar hij nu woont.
	low	Peter is een fanatieke hordeloper en traint vaak in **de/*het grote stad** waar hij nu woont.
3	high	De student betaalde een te hoge huur voor **de/*het kleine kamer** boven het café.
	low	De buurman zag een prachtige plant in **de/*het kleine kamer** boven het café.
4	high	Bij de ingang van het restaurant hield de portier **de/*het grote deur** open voor de klanten.
	low	Een foto van de familie stond op het kastje vlak naast **de/*het grote deur** van de slaapkamer.

5	high	De mevrouw zette al het eten op **de/*het grote tafel** in het restaurant.
	low	De mevrouw bracht de dozen naar **de/*het grote tafel** in het restaurant.
6	high	Het toilet bevindt zich aan het einde van **de/*het lange gang** van het kasteel.
	low	Sabrina legde de boodschappen midden in **de/*het lange gang** van het kasteel.
7	high	Sandra was te laat voor de vergadering, omdat ze in haar agenda **de/*het foute tijd** had opgeschreven.
	low	Sandra begreep niets van de Franse brigadier en had daardoor **de/*het foute tijd** in haar agenda gezet.
8	high	Alle Kamerleden gingen akkoord met **de/*het nieuwe wet** tegen kindermishandeling.
	low	Alle leerlingen waren enthousiast over **de/*het nieuwe wet** tegen kindermishandeling.
9	high	Carolien werd geopereerd door **de/*het knappe dokter** van orthopedie.
	low	In de deuropening stond **de/*het knappe dokter** van orthopedie.
10	high	Nadat Tom zijn biertje opgedronken had, gooide hij **de/*het lege fles** zomaar in de prullenbak.
	low	Nadat Tom de barbecue had aangestoken, gooide hij **de/*het nieuwe wet** benzine in de prullenbak.
11	high	Diana bedankte haar kraamverzorgster voor **de/*het goede hulp** die ze kreeg tijdens haar zwangerschap.
	low	Diana vertelde haar collega's over **de/*het goede hulp** die ze kreeg tijdens haar zwangerschap.
12	high	Vera gaat vanavond naar een optreden van **de/* het goede muzikant** André Rieu.
	low	Vera hoorde laatst dat ze familie is van **de/* het goede muzikant** André Rieu.
13	high	Marlies haar artikel kan in januari worden gelezen in **de/*het nieuwe krant** van Amsterdam.
	low	Marlies doet een onderzoek naar de betrouwbaarheid van **de/*het nieuwe krant** van Amsterdam.
14	high	Het verkeer raast elke ochtend over **de/*het lange weg** die de mensen naar Rotterdam brengt.
	low	De advocaat kijkt verbaasd naar **de/*het lange weg** die langs zijn huis gebouwd wordt.
15	high	Peter zet het prachtige servies in **de/*het oude kast** van zijn opa.
	low	Peter zoekt naar belangrijke papieren in **de/*het oude kast** van zijn opa.

Neuter Targets

Item	Cloze	Sentence
1	high	Tijdens het vouwen van de envelop sneed het meisje zich aan **het/*de scherpe papier** en begon te huilen.
	low	Tijdens het spelen sneed het meisje zich per ongeluk aan **het/*de scherpe papier** van een envelop.
2	high	Kim krijgt een onbekende man aan de telefoon en vraagt zich af of ze **het/*de juiste nummer** heeft gedraaid.
	low	Kim heeft haar zaken vandaag niet op orde en vraagt zich af of ze **het/*de juiste nummer** heeft gedraaid.
3	high	Door mensen met respect te behandelen, wil Marga haar kinderen **het/*de goede voorbeeld** geven.
	low	Door iedere avond te gaan hardlopen, wil Marga proberen **het/*de goede voorbeeld** te geven.
4	high	Na het optreden kreeg de band een wild applaus van **het/*de leuke publiek** in de zaal.
	low	Na de zwemles kreeg het jongetje de zenuwen van **het/*de leuke publiek** in de zaal.
5	high	Speciaal voor de trouwerij trok Jeroen **het/*de nette pak** van zijn broer aan.
	low	Het duurde uren voordat Jeroen **het/*de nette pak** van zijn broer had aangetrokken.
6	high	Tijdens haar eerste rijles op de manege reed Leonie vol trots op **het/*de mooie paard** dat ze zelf had uitgekozen.

	low	Tijdens haar eerste vakantie was Leonie dolgelukkig met **het/*de mooie paard** dat ze zelf had uitgekozen.
7	high	Het professionele koor zong een nieuwe versie van **het/*de prachtige lied** over de liefde.
	low	Het onzekere meisje gaf haar familie een voorproefje van **het/*de prachtige lied** over de liefde.
8	high	Om het vlees in kleine stukjes te snijden gebruikte Marco **het/*de nieuwe mes** van de slager.
	low	Om zijn vervelende vrienden te pesten gebruikte Marco **het/*de nieuwe mes** van de slager.
9	high	Op 31 december eten we oliebollen tijdens **het/*de laatste uur** van het jaar.
	low	Karin en Greet kijken al weken lang uit naar **het/*de laatste uur** van het jaar.
10	high	Tijdens de finale van de spellingwedstrijd spelt Tom **het/*de lastige woord** foutloos.
	low	Tijdens zijn vakantie in Gelderland leert Tom **het/*de lastige woord** 'aubergine' te spellen.
11	high	Toen hij zijn ouders kwijt was geraakt schreeuwde **het/*de kleine kind** totdat iemand hem optilde.
	low	De automonteur liep graag langs water met **het/*de kleine kind** van zijn broer.
12	high	De koeien renden het weiland in en aten **het/*de hoge gras** langs de sloot.
	low	De jongens speelden met autootjes in **het/*de hoge gras** langs de sloot.
13	high	Mark wil erg graag zijn vriendin ten huwelijk vragen, maar wacht op **het/*de juiste moment** om dit te doen.
	low	Mark wil zijn nieuwe recept uitproberen, maar heeft nog niet **het/*de juiste moment** gevonden om dit te doen.
14	high	Bianca leest vol spanning **het/*de nieuwe boek** van haar favoriete schrijfster.
	low	Bianca zoekt al weken naar **het/*de nieuwe boek** van haar favoriete schrijfster.
15	high	De lerares vertelde de kinderen **het/*de mooie verhaal** van roodkapje.
	low	De lerares was niet bekend met **het/*de mooie verhaal** van roodkapje.

Gender agreement condition: indefinite determiner-adjective-noun (IaN)

Common Targets

Item	Cloze	Sentence
1	high	Sasha sliep ontzettend slecht nadat ze in de bioscoop **een enge/*eng film** had gezien.
	low	Sasha moest lachen van de zenuwen nadat ze bij haar vriend **een enge/*eng film** had gezien.
2	high	Sjoerd schrijft **een lange/*lang brief** aan zijn verloofde.
	low	Sjoerd krijgt **een lange/*lang brief** van zijn verloofde.
3	high	Tijdens een wandeling door Den Haag verdwaalde Sonja en kwam terecht in **een nare/*naar buurt** aan de rand van de stad.
	low	Vanuit het bosrijke gebied vertrokken de wandelaars in de richting van **een nare/*naar buurt** aan de rand van de stad.
4	high	Tijdens het optreden bleek de nieuwe zangeres **een mooie/*mooi stem** te hebben.
	low	Tijdens de vlucht bleek de nieuwe piloot **een mooie/*mooi stem** te hebben.
5	high	De Fransman eet altijd stokbrood en drinkt daarbij **een goede/*goed wijn** uit Frankrijk.
	low	De Groninger eet altijd boerenkool en neemt daarbij **een goede/*goed wijn** uit Frankrijk.
6	high	De wandelaar struikelde over **een grote/*groot steen** in het bos.
	low	De ambtenaar kocht **een grote/*groot steen** voor in zijn voortuin.
7	high	De kok besloot de aardappelen in **een diepe/*diep pan** te koken.
	low	De vrouw besloot de brandnetels in **een grote/*groot steen** te koken.
8	high	Gelukkig is er morgen **een nieuwe/*nieuw dag** om door te werken.
	low	Pieter verheugt zich op **een nieuwe/*nieuw dag** om door te werken.
9	high	Dennis werd na te hard rijden bekeurd door **een strenge/*streng agent** met een boze blik.

	low	Dennis werd na het feestje gevolgd door **een strenge/*streng agent** met een boze blik.
10	high	Jorien spreekt alleen maar Nederlands en heeft besloten **een nieuwe/*nieuw taal** te gaan leren.
	low	Jorien verveelt zich nogal snel en heeft daarom besloten **een nieuwe/*nieuw taal** te gaan leren.
11	high	De Aboriginal zag gisteren voor het eerst **een witte/*wit man** in zijn gebied.
	low	De schrijfster beschreef laatst voor het eerst **een witte/*wit man** in haar roman.
12	high	Tijdens het debat beantwoordt de politica **een lastige/*lastig vraag** over het milieu.
	low	Tijdens het feest krijgt de organisator **een lastige/*lastig vraag** over het milieu.
13	high	In het bejaardentehuis wandelt **een oude/*oud vrouw** door de tuin.
	low	Op het politiebureau ziet Pim **een oude/*oud vrouw** koffie drinken.
14	high	Myra heeft een nieuwe klasgenoot en ze vindt hem **een leuke/*leuk jongen** die ook nog knap is.
	low	Myra zit net op de middelbare school en ze heeft daar **een leuke/*leuk jongen** van haar leeftijd ontmoet.
15	high	Met zijn bolhoed en nette pak doet Jesper zich voor als **een oude/*oud meneer** met veel geld.
	low	Aan de overkant van de straat zag Jesper **een oude/*oud meneer** met een koffer lopen.

Neuter Targets

Item	Cloze	Sentence
1	high	Na de vechtpartij liep Joris nog weken rond met **een blauw/*blauwe oog** en een pijnlijke vinger.
	low	Na het schoolbal zag Joris zijn vader met **een blauw/*blauwe oog** en een pijnlijke vinger.
2	high	De rijke zakenman neemt zijn minnares mee naar de Ardennen voor **een lang/*lange weekend** in een luxe hotel.
	low	De minnares is compleet verbaasd als ze **een lang/*lange weekend** in een luxe hotel mag doorbrengen.
3	high	Toen de populaire jongen Liesbeth mee uitvroeg kreeg ze van schaamte **een rood/*rode hoofd** en begon te stotteren.
	low	Toen de jongen naar buiten liep zag hij Liesbeth bij de stalling staan met **een rood/*rode hoofd** van schaamte.
4	high	Voor de inzamelactie bedacht Dorothea **een goed/*goede plan** om snel geld in te zamelen.
	low	Het verslag bestond onder andere uit **een goed/*goede plan** om snel geld in te zamelen.
5	high	Toen de actrice alle make-up had verwijderd zag ik dat ze toch **een mooi/*mooie gezicht** had.
	low	Toen de columnist goed keek vond hij dat de actrice **een mooi/*mooie gezicht** had.
6	high	Sandra is geboren en getogen in **een klein/*kleine dorp** in Limburg.
	low	Sandra gaat graag met haar moeder naar **een klein/*kleine dorp** in Limburg.
7	high	Voor haar achttiende verjaardag organiseerde Lieke **een groot/*grote feest** op het strand.
	low	Vlak voordat de vrouw een ongeluk had gekregen was ze bij **een groot/*grote feest** op het strand.
8	high	Tijdens de speurtocht hoorde Pim plotseling **een vreemd/*vreemde geluid** achter zich.
	low	Tijdens de speurtocht kwam er plotseling **een vreemd/*vreemde geluid** uit de bosjes.
9	high	Haar fiets vergrendelde Natasja met **een sterk/*sterke slot** gemaakt van staal.
	low	Naast het keukenblok vond Natasja **een sterk/*sterke slot** gemaakt van staal.
10	high	Met een jacht voer het jonge stel naar **een prachtig/*prachtige eiland** in de Zuidzee.
	low	Na het huwelijk gingen de ouders naar **een prachtig/*prachtige eiland** om verder te feesten.
11	high	Voor de speciale groentesaus gebruikte Arjan **een oud/*oude recept** dat hij van zijn oma kreeg.

	low	Speciaal voor oudejaarsavond haalde Arjan **een oud/*oude recept** van zijn oma van zolder.
12	high	De broers lijken veel op elkaar, maar als je goed kijkt zie je **een groot/*grote verschil** tussen beide.
	low	De broers staren uren naar de spiegel en zien dan **een groot/*grote verschil** tussen beide.
13	high	Het gelukkige en pasgetrouwde stel ging samenwonen in **een oud/*oude huis** aan het water.
	low	Aan de overkant van het meer zagen de puberale jongens **een oud/*oude huis** aan het water.
14	high	Karel heeft zo'n ongelofelijke dorst dat hij in één keer **een groot/*grote glas** water opdrinkt.
	low	Karel rent de keuken in en ziet daar **een groot/*grote glas** water op een dienblad staan.
15	high	Volgens de minister vormt overgewicht bij kinderen **een groot/*grote probleem** in de samenleving.
	low	Volgens de voetballer zal zijn jaloerse vrouw **een groot/*grote probleem** vormen dit voetbalseizoen.

Finiteness agreement condition

Past Participle

Item	Sentence
1	Ze heeft alleen haar beste vriendin **uitgenodigd/*uitnodigen** voor haar verjaardag.
2	Het rookverbod is vorig jaar **ingegaan/*ingaan** en geldt in heel Nederland.
3	De directeur heeft de situatie aan de werknemers **uitgelegd/*uitleggen** tijdens de lunchpauze.
4	Pim is al twee keer te laat op zijn werk **gekomen/*komen** door een staking bij de spoorwegen.
5	Overal in de stad worden goedkope huizen **gebouwd/*bouwen** voor studenten en jongeren.
6	Die man heeft gisteren een groot ongeluk **veroorzaakt/*veroorzaken** vlakbij het treinstation.
7	De bevolking heeft twee jaar geleden een nieuwe regering **gekozen/*kiezen** en dat moet bijna weer gebeuren.
8	Ik heb hem nog nooit **geloofd/*geloven** toen hij zijn sterke verhalen vertelde.
9	Ik had mijn auto verkeerd **geparkeerd/*parkeren** en vond een fikse bekeuring onder de ruitenwisser.
10	Die man heeft zijn hele leven hard **gewerkt/*werken** en gaat volgende maand met pensioen.
11	De directeur heeft alle regels al **bepaald/*bepalen** zonder met zijn medewerkers te overleggen.
12	Ik ben vreselijk boos op hem **geworden/*worden**, omdat hij had gelogen over zijn minnares..
13	Hij heeft per ongeluk de thee naast het kopje **geschonken/*schenken**, maar hij ruimt het netjes weer op met een doekje.
14	Ik heb die roman al gelezen/*lezen toen ik nog maar 12 jaar was.
15	De trein werd gisteren meteen **gerepareerd/*repareren** na de heftige botsing met een goederenwagon.
16	Hij heeft het in Zuid-Spanje veel te warm **gehad/*hebben** en is blij dat hij weer in Nederland is.
17	Hij had zich plotseling **omgedraaid/*omdraaien** toen hij zijn naam hoorde.
18	De slager heeft zich in zijn vinger **gesneden/*snijden** met een scherp mes.
19	Jouw les is om 11 uur **begonnen/*beginnen** in de zaal hiernaast.
20	Ik heb in Utrecht **gewoond/*wonen** totdat ik 18 jaar was.
21	Ze hebben van de directeur een cadeautje **gekregen/*krijgen** voor de kerstdagen.
22	We hebben de hele middag op hem **gewacht/*wachten** bij het gemeentehuis.
23	Hij heeft uitsluitend over zijn verleden **gepraat/*praten** tijdens het interview.
24	Ik ben vroeger een keer in de sloot **gevallen/*vallen** met de fiets.
25	Sofie heeft gisteren vijf kilometer **gelopen/*lopen** door de bossen in Drenthe.
26	Ik heb in maart een huis **gekocht/*kopen** aan de Amsterdamse gracht.

Infinitive

Item	Sentence
1	Hij beloofde vóór acht uur te **komen/*gekomen**, maar om negen uur was hij er nog niet.
2	Je moet te lang op de bus **wachten/*gewacht** in Groningen.
3	Wij willen dit jaar met de bus naar Spanje **gaan/*gegaan** om ons jubileum te vieren.
4	Joop moet op zaterdag zijn boodschappen **doen/*gedaan** voor de hele week.
5	Zij kan deze puzzel **oplossen/*opgelost** binnen vijf minuten.
6	Ik kan dit jaar helaas niet op vakantie **gaan/*gegaan**, omdat ik daar geen geld voor heb.
7	Niemand kan zich de winter in Alaska **voorstellen/*voorgesteld**, want het is daar ongelooflijk koud.
8	Ik zag die mensen in het restaurant **dansen/*gedanst** op salsamuziek.
9	Jongeren kunnen zich maar moeilijk voor politiek **interesseren/*geïnteresseerd** tegenwoordig.
10	De mensen van het bureau willen hem advies **geven/*gegeven** over de verkoop van zijn huis.
11	Hij gaat een cursus **volgen/*gevolgd** aan de Universiteit van Utrecht.
12	Hij probeert me altijd aan het lachen te **maken/*gemaakt** door grapjes te vertellen.
13	We vonden het leuk om naar die voorstelling te **kijken/*gekeken**, maar het was jammer dat er geen pauze was.
14	Ik kan haar soms moeilijk **begrijpen/*begrepen**, omdat ze zo onduidelijk praat.

Nederlandse

SAMENVATTING

Als kind hebben we moeiteloos onze moedertaal (T1) geleerd en vaak kunnen we veel van de onderliggende regels van de taal niet eens opnoemen. Het leren van een nieuwe taal op latere leeftijd is echter vaak een moeilijk en langzaam proces, waarbij we juist veel regeltjes uit ons hoofd moeten leren om de taal te kunnen begrijpen en te gebruiken. Hoewel veel late tweede taalleerders uiteindelijk redelijk vloeiend worden in hun tweede taal (T2), zal een late T2 leerder vrijwel nooit verward worden met een moedertaalspreker. We herkennen de late T2 leerder, oftewel de leerder die pas na de puberteit met de T2 in aanraking is gekomen, vooral aan het accent. T2 sprekers staan echter ook bekend om het maken van specifieke, vaak grammaticale, fouten die moedertaalsprekers niet snel zullen maken. Eén taalkundig aspect waar T2 leerders moeite mee blijven hebben, zelfs na jarenlange ervaring met de T2, is grammaticaal geslacht of *genus*, een classificatiesysteem dat in vele, maar niet in alle talen voorkomt. T2 leerders van een taal met een genussysteem worden geconfronteerd met de taak van het toekennen van een geslacht aan een zelfstandig naamwoord dat vaak geen enkel geslachtskenmerk heeft. Zo moet een leerder van het Nederlands zien te begrijpen dat het onzijdige lidwoord *'het'* wordt gebruikt wanneer je naar een mes verwijst, terwijl een lepel moet worden voorafgegaan door het commune lidwoord *'de'*. Zelfs als de leerder al bekend is met een geslachtensysteem vanuit de moedertaal, heeft hij of zij hier in eerste instantie vaak weinig aan: diezelfde lepel is mannelijk in het Duits (*'der Löffel'*), terwijl deze het vrouwelijke geslacht toegewezen krijgt in Spanje (*'la cuchara'*).

In ***hoofdstuk 1*** van dit proefschrift wordt het begrip *genus* of geslacht, en specifiek de bijzondere wijze waarop het zich in de Nederlandse taal manifesteert, uiteengezet. In talen met een genussysteem worden naamwoorden ingedeeld in verschillende categorieën. Die indeling lijkt in het Nederlands, in tegenstelling tot de genussystemen in talen als het Spaans en Italiaans, zelden gebaseerd op fonologische en/of morfologische aspecten van het woord. Met andere woorden, de klanken en vormen van Nederlandse zelfstandige naamwoorden geven vaak geen enkele informatie over het geslacht van het naamwoord: we spreken bijvoorbeeld van *'**het** huis'*, maar van *'**de** kluis'*. Het Nederlandse genussysteem is niet alleen speciaal vanwege de ondoorzichtige regels, maar ook door een asymmetrie in zowel de verdeling van categorieën als de toepassing van de grammaticale regels. De Nederlandse taal kent *de*-woorden (commuun genus) en *het*-woorden (onzijdig genus), waarbij commuun geslacht door de samensmelting van de oorspronkelijke mannelijke en vrouwelijke geslachten de grootste groep vormt (75% van alle zelfstandige naamwoorden). Het geslacht van woorden moet overeenkomen met bepaalde lidwoorden (*'de_{COM}'* en *'het_{NEU}'*), verschillende voornaamwoorden (*'die_{COM}'* en *'dat_{NEU}'*) en, in het geval van onbepaalde naamwoordgroepen, met bijvoeglijke naamwoorden (*'mooi**e**$_{COM}$'* en *'$mooi_{NEU}$'*).[12] Deze regels hoeven echter niet consistent te worden toegepast. Zo hoeft men geslachtsmarkering op bijvoeglijke naamwoorden alleen te gebruiken in onbepaalde naamwoordgroepen (*een **mooie** auto'* en *'een **mooi** paard'*), terwijl adjectieven in bepaalde naamwoordgroepen altijd op *'–e'* eindigen, de uitgang die normaal voor commune zelfstandige naamwoorden staat (*'de **mooie** auto'* en *'het **mooie** paard'*). Daarbij worden de commune vormen van het lidwoord (*'de'*), het bijvoeglijke naamwoord (*'–e'*), en het voornaamwoord (*'die'*) ook gebruikt voor alle zelfstandige naamwoorden in het meervoud, ongeacht hun onderliggende genus. Daarentegen worden alle woorden onzijdig wanneer zij als verkleinwoord gebruikt worden, zoals in *'het autootje'*. De leerder van het Nederlands moet het geslachtensysteem dus zien te verwerven op basis van asymmetrische en onduidelijke regels, waarvan bovendien het commune geslacht veel vaker en duidelijker aanwezig is. Desondanks zijn moedertaalsprekers doorgaans in staat om, zonder na te denken, de juiste geslachten en hun bijbehorende vormen te gebruiken. Zo spreken Nederlandse moedertaalsprekers zonder enige moeite, en vaak zelfs zonder zich bewust te zijn van de onderliggende geslachten, over *'**de** kluis'*, *'**het** huis'*, *'een **mooi** paard'* en *'een **mooie** auto'*. Het is dan ook uitermate interessant om te achterhalen in hoeverre Nederlandse moedertaalsprekers genus gebruiken en of ze dit hetzelfde verwerken als andere grammaticale constructies in de taal. Daarnaast is het interessant om te onderzoeken of een dergelijk systeem op latere leeftijd aangeleerd kan worden en of deze late leerders genus in hun T2 op eenzelfde wijze gebruiken en verwerken als moedertaalsprekers.

Om antwoord op deze vragen te krijgen werden 28 volwassen Nederlandse moedertaalsprekers en 27 Poolse Nederlanders die pas op latere leeftijd (na hun 20ste levensjaar) Nederlands hebben geleerd met elkaar vergeleken. In ***hoofdstuk 2*** wordt uitgebreid uitgelegd op welke wijze en waarom deze twee groepen met elkaar zijn vergeleken. Een van de redenen voor het kiezen van Poolse Nederlanders is dat de

[12] De afkortingen COM en NEU verwijzen hier naar de Engelse benaming voor commuun (*'common'*) en onzijdig (*'neuter'*) geslacht.

Poolse taal een complex en uitgebreid geslachtensysteem kent. Echter, zowel de Poolse taal zelf als het genussysteem lijken in zeer weinig opzichten op het Nederlands en het Nederlandse genussysteem. De onderzochte T2 leerders zijn dus bekend met de diverse categorieën en de mogelijke consequenties van deze categorieën op het gebruik van taal, maar kunnen vrijwel nooit woorden direct vertalen vanuit de moedertaal naar de T2 (zoals dit vaak wel mogelijk is met bijvoorbeeld Duits en Nederlands). De resultaten van diverse metingen gerelateerd aan de vaardigheid van de groep T2 leerders in het Nederlands worden uiteengezet in ***hoofdstuk 3***. Hieruit bleek dat de groep T2 leerders gemiddeld al 9 jaar (wisselend van 2 tot 27 jaar) in Nederland woonde. Diverse testen van zowel hun productie als het begrip van de Nederlandse taal wezen uit dat de groep uitermate vaardig was in de T2. De algemene T2 score correleerde echter zelden met hun specifieke vaardigheden in het begrijpen en gebruiken van zowel commuun als onzijdig geslacht. Alleen zeer vaardige T2 leerders bleken in staat om voor een aanzienlijk deel van de gebruikte naamwoorden het correcte lidwoord te selecteren, maar slechts enkele van hen waren ook daadwerkelijk in staat om deze kennis tijdens een verhalende taak te gebruiken. Deze resultaten stemmen overeen met diverse taalkundige theorieën, waarin gesteld wordt dat het leren van een abstract grammaticaal systeem, zoals het Nederlandse genussysteem, op latere leeftijd niet meer volledig mogelijk is. Dit zou kunnen komen doordat de hersenen minder 'rekbaar' zijn geworden ofwel doordat we last hebben van een 'cognitieve overbelasting' aan processen. Bewijs voor of tegen dergelijke aannames komt tot op heden vooral van zogenoemde *offline* experimenten waar men kijkt naar een eindproduct, zoals de hierboven genoemde vaardigheden om talige informatie achteraf te begrijpen of te gebruiken. Het zou echter best zo kunnen zijn dat, in de tijd voordat T2 leerders hun verbale of non-verbale reactie gegeven hebben, de onderliggende processen van taalverwerking hetzelfde zijn als in moedertaalsprekers. Om te onderzoeken of deze meer onbewuste verwerking van gesproken taal, en dan voornamelijk de verwerking van geslacht, ook verschilt tussen moedertaalsprekers en T2 sprekers zijn twee verschillende experimenten uitgevoerd. In het eerste experiment werden oogbewegingen gevolgd om te kijken hoe en wanneer mensen talige informatie gebruiken. Tijdens het tweede experiment werd hersenactiviteit gemeten tijdens het luisteren naar correcte en incorrecte zinnen. Alvorens in te gaan op de resultaten van de T2 leerders, volgt eerst een samenvatting van de resultaten betreffende het gebruik en de verwerking van geslacht door Nederlandse moedertaalsprekers.

Het eerste experiment, waarin oogbewegingen werden gevolgd, was gebaseerd op de bevinding dat mensen ontzettend snel gebruik maken van klanken en talige informatie. Men blijkt daarbij sterk de neiging te hebben om naar objecten te kijken die overeenkomen met de informatie die zij horen. In ***hoofdstuk 4*** worden de resultaten gegeven van het onderzoek dat betrekking had op de vraag of en hoe Nederlandse moedertaalsprekers informatie betreffende het geslacht van woorden kunnen gebruiken om hun taalbegrip te versnellen. Om dit te onderzoeken werden oogbewegingen gemeten terwijl de proefpersonen gesproken instructies volgden om op één van de vier weergegeven plaatjes op een computerscherm te klikken (bijvoorbeeld: 'Klik op de_{COM} rode $appel_{COM}$'). De vier plaatjes op het scherm bestonden uit (1) het doelplaatje, (2) een plaatje dat hetzelfde geslacht en/of dezelfde kleur had als het doelplaatje en (3) twee plaatjes die nimmer hetzelfde geslacht of dezelfde kleur hadden als het doelplaatje. Nederlandse moedertaalsprekers hadden moeite met het lokaliseren van het doelplaatje wanneer er nog een plaatje met dezelfde kleur aanwezig was. Ze bleken echter weer sneller te zijn wanneer het doelplaatje het commune geslacht had en het concurrerende plaatje onzijdig was. Met andere woorden, na het horen van *'de rode'* of *'een rode'* keek men niet meer naar het plaatje met een rood huis, aangezien *'huis'* onmogelijk het volgende naamwoord kon zijn. Onzijdige lidwoorden en adjectieven ontlokten niet hetzelfde effect. Dit stemt overeen met het feit dat elk plaatje op het scherm voorafgegaan kan worden door *'het rode'* en/of *'een rood'* wanneer het als verkleinwoord wordt gebruikt.

Net als commune geslachtsmarkering wordt context vaak gebruikt om de volgende woorden in de zin te voorspellen. In ***hoofdstuk 5***, het tweede experiment, wordt uitgelegd hoe de hersenactiviteit, die gerelateerd kan worden aan een bepaalde (taalkundige) stimulus (een zogenaamde 'event-related potential' of ERP), verandert wanneer men naar Nederlandse zinnen en in het bijzonder naar informatie over geslacht luistert. Wanneer moedertaalsprekers een grammaticale fout horen, proberen zij dit vaak zelf te repareren. Dit reparatieproces wordt doorgaans geassocieerd met een langdurige toename aan positieve activiteit die piekt rond 600 milliseconden na de geconstateerde (grammaticale) fout. Om te bekijken of de verwerking van geslachtsinformatie beïnvloed wordt door context en of schendingen van het asymmetrische genussysteem in het Nederlands ook het bekende reparatieproces zouden ontlokken, luisterden de proefpersonen naar zinnen die ofwel correct ofwel incorrect waren en waar het zelfstandige naamwoord ofwel voorspelbaar ofwel niet voorspelbaar was. De fouten bestonden uit (1) voor de hand liggende schendingen van werkwoorden zoals in *'*zij heeft haar **uitnodigen**'* en (2) schendingen van geslacht tussen het bepaald lidwoord of het bijvoeglijk naamwoord en het daaropvolgende zelfstandige naamwoord zoals in *'*de bakker bakt **de brood**'*. Beide schendingen ontlokten een vergelijkbare P600 in de moedertaalsprekers, waaruit blijkt dat ook schendingen van geslacht in een abstract genussysteem als duidelijke grammaticale fout worden beschouwd. Ook bleek dat de genusfouten sneller door moedertaalsprekers werden gerepareerd wanneer het zelfstandige naamwoord voorspelbaar was. Met andere woorden, mensen repareerden ***'*de brood'*** sneller in '*de bakker bakte ***de brood***' dan in '*de buurvrouw maakte ***de brood***'. Dit resultaat laat zien dat context het onbewuste herstel van grammaticale fouten kan versnellen.

Eerder onderzoek heeft laten zien dat tweede taalleerders, zeker in de beginstadia van het leerproces, soms vertrouwen op de kennis vanuit hun moedertaal. In sommige gevallen zullen de twee talen vrijwel volledig overeenkomen en zal de overdracht van kennis en regels van de moedertaal naar de T2 geen fouten veroorzaken. In andere gevallen kan het echter ook tot fouten leiden en genussystemen kunnen hier, mede door de diverse verschillen tussen deze systemen, erg gevoelig voor zijn. De vraag of late T2 leerders in staat zijn om hun nieuwe genussysteem te gebruiken of dat zij (ofwel correct ofwel foutief) gebruik maken van hun vertrouwde T1 genussysteem is uitgebreid onderzocht in ***hoofdstuk 6***. De resultaten laten zien dat T2 leerders, in tegenstelling tot moedertaalsprekers, niet in staat zijn om Nederlandse commune lidwoorden en bijvoeglijke naamwoorden te gebruiken om de activatie van onzijdige woorden te onderdrukken. Ook laten de data zien dat, in tegenstelling tot wat enkele theorieën binnen tweede taalverwerving suggereren, late T2 leerders geen (ofwel correct ofwel foutief) gebruik maken van geslachtsinformatie uit hun moedertaal tijdens het luisteren naar zinnen in hun tweede taal. Dit zou wel degelijk te maken kunnen hebben met de diverse verschillen tussen de Poolse en Nederlandse taal. De overdracht van informatie uit de moedertaal naar de tweede taal zou beperkt kunnen zijn tot talen die tot op zekere hoogte veel op elkaar lijken (zoals het Duits en het Nederlands).

Dat T2 leerders geslachtsmarkering niet kunnen gebruiken om het aantal potentiële kandidaten in de woordenlijst te verminderen, betekent uiteraard niet dat ze het verschil tussen commuun en onzijdig geslacht niet kennen. Daarom wordt in ***hoofdstuk 7*** uitgebreid ingegaan op de hersenactiviteit van T2 leerders terwijl zij luisterden naar correcte en incorrecte zinnen in het Nederlands. De duidelijke en voor de hand liggende schendingen van werkwoorden (*'*zij heeft haar* ***uitnodigen****')* ontlokten een vertraagde, maar een met moedertaalsprekers vergelijkbare P600 in T2 leerders. Een vergelijkbare P600 in reactie op schendingen van genus (*'****het tuin****'*) werd in deze groep echter alleen gevonden voor zeer hoog bekwame T2 leerders die al langere tijd in Nederland woonden en alleen bij fouten die betrekking hadden op *de*-woorden. Schendingen van onzijdig geslacht, zoals *'*de brood'*, werden niet door de T2 leerders als fout verwerkt en gerepareerd. Deze resultaten suggereren dat het abstracte Nederlandse genussysteem wel degelijk geleerd kan worden, maar in verhouding tot voor de hand liggende grammaticale structuren duurt het langer voordat late T2 leerders het systeem onder de knie hebben. Bovendien blijkt men alleen het vaker voorkomende commune geslacht te kunnen verwerven en fouten hiervan op eenzelfde wijze als moedertaalsprekers te kunnen repareren. Fouten van het minder voorkomende onzijdige geslacht blijken, zelfs voor zeer vaardige T2 sprekers van het Nederlands, niet als grammaticale fout te worden herkend en gerepareerd.

Samengevat laten de uitkomsten van de experimenten zien dat geslacht, net als context, door moedertaalsprekers gebruikt kan worden om taalverwerking te vergemakkelijken. Dit geldt echter niet voor onzijdige geslachtsmarkering, zoals *'het'* of *'een rood'*, aangezien deze markering aan alle naamwoorden vooraf gaat wanneer zij als verkleinwoord worden gebruikt. Hoewel het Nederlandse geslachtssysteem abstract is en vaak woord voor woord geleerd moet worden, beschouwen moedertaalsprekers geslacht wel degelijk als een duidelijk grammaticaal systeem, zoals bleek uit eenzelfde effect voor werkwoord- en geslachtsschendingen. Waar moedertaalsprekers sterk reageren op fouten betreffende zowel *de*-woorden als *het*-woorden, blijkt de asymmetrie van het Nederlandse genussysteem in tweede taalverwerving wel degelijk voor problemen te zorgen. Naarmate de T2 leerders langer blootgesteld worden aan de Nederlandse taal en meer vaardigheid ontwikkelen, blijken zij uiteindelijk wel in staat om schendingen van commuun geslacht op een vergelijkbare manier als moedertaalsprekers te verwerken. Fouten met betrekking tot de minder voorkomende *het*-woorden blijken door hen dan echter nog steeds niet als fout te worden herkend. Hoewel ze fouten van *de*-woorden achteraf wel als grammaticale fout konden verwerken, lieten zowel hun productie als hun oogbewegingen zien dat ze deze kennis niet van te voren kunnen activeren en op een adequate manier kunnen toepassen of gebruiken om hun taalbegrip te vergemakkelijken. Ook blijkt uit de experimenten in dit proefschrift dat T2 leerders met een typologisch andere taalachtergrond geslachtsinformatie vanuit hun moedertaal niet overdragen naar de tweede taal. Deze resultaten, en dan vooral de moeilijkheid van de verwerving van het minder frequente onzijdige geslacht, suggereren dat de hoeveelheid van informatie over grammaticale structuren en de hoeveelheid oefening die T2 leerders hebben gehad van cruciaal belang zijn voor de verwerving ervan. In ***hoofdstuk 8*** wordt gespeculeerd over de toekomst van het Nederlandse geslachtensysteem. Door de asymmetrie en de abstracte regels blijft het een raadsel hoe Nederlandse moedertaalsprekers dit systeem zo goed onder de knie krijgen en zo fors reageren op zowel fouten van de frequente *de*-woorden als de minder frequente *het*-woorden. Recente studies suggereren dat het Nederlandse geslachtensysteem constant aan het veranderen is en in de loop der tijd simplistischer wordt. De toekomst zal uitwijzen of het Nederlandse geslachtensysteem zich staande zal kunnen houden of dat het zijn beste tijd heeft gehad.

Acknowledgements

DANKWOORD

Het boekje is af. Na vier lange, soms lastige, maar ook leuke en interessante jaren, is er een einde gekomen aan mijn tijd als promovendus. Een aantal keren heb ik gedacht dat het boekje er niet zou komen, maar zie hier! Het boekje is echt af, geprint en al. Vlak voordat het manuscript bij de drukker afgeleverd diende te worden, kwam ineens dat verschrikkelijke besef: 'ik moet ook nog een dankwoord schrijven'. Lange tijd heb ik gezocht naar redenen om hier onderuit te komen, maar hoe langer ik erover na ging denken, des te meer ik mij realiseerde dat ik best veel te danken heb aan een grote groep mensen. Mijn familie en vrienden, mijn collega's in het Harmoniegebouw, mijn mede-promovendi aan de Rode Weeshuisstraat, en de diverse mensen die mij hebben geholpen op het NeuroImaging Center. Als ik dan toch een dankwoord moet schrijven, zou ik het liefste iedereen persoonlijk willen bedanken, maar ik realiseer mij maar al te goed dat de kans groot is dat ik dan mensen ga vergeten die ik nu juist niet had willen vergeten. Daarom, speciaal voor jullie, diegenen die hier niet met naam en toenaam genoemd staan, maar wel heel belangrijk zijn geweest: bedankt! Voor een selecte groep wil ik dan best een uitzondering maken, mijn gêne opzij zetten, en hen uitgebreid bedanken.

Allereerst wil ik mijn promotor Monika Schmid bedanken voor de interessante brainstormsessies in het begin, het vele (en snelle) leeswerk van alle versies van (delen van) dit proefschrift en de constructieve feedback daarop in de latere jaren, en op het einde vooral de hulp bij de laatste, lastige loodjes van de voltooiing van dit proefschrift. Vooral haar volledige vertrouwen in mij, vanaf het prille begin, heeft mij enorm goed gedaan. Graag bedank ik ook mijn copromotor Laurie Stowe, die ik na mijn scriptie gelukkig mee kon nemen als begeleider tijdens mijn promotieonderzoek, voor de hulp bij de EEG experimenten en de uitgebreide feedback die ik altijd kreeg op mijn geschreven werk. Uiteraard bedank ik Laurie ook voor de heerlijke, en soms verrassende, maaltijden en de gezellige avonden die we bij haar thuis hebben doorgebracht.

I would also like to thank the members of the reading committee: Giuli Dussias, Cheryl Frenck-Mestre, and Petra Hendriks. I am grateful to them for taking the time to read and evaluate my work.

De opleiding die ik voor mijn hele PhD-avontuur heb genoten is onmisbaar geweest. Ik had het geluk dat ik na mijn Bachelor door kon gaan met een vrij nieuw Masterprogramma: Toegepaste Taalwetenschappen. Een programma bedacht en samengesteld door zeer speciale mensen die mij toendertijd alles over tweede taalverwerving hebben geleerd: Kees de Bot, Wander Lowie en Marjolijn Verspoor. Jullie behoren zonder enige twijfel tot de beste docenten die er op dit moment op de universiteit rondlopen. Door jullie goede opleiding, vertrouwen, en misschien ook wel een beetje durf, mocht ik in september aanschuiven bij de Master TTW. En deze keer als trotse en zeer vereerde collega. Het zou best zo kunnen zijn dat wij elkaar in de toekomst nog vaker tegen zullen komen. Ik kijk daar in elk geval erg naar uit!

Dan wil ik, vanuit de grond van mijn hart, de mensen bedanken die mee hebben gedaan aan dit onderzoek als proefpersoon. Het is zeker niet niks om twee keer af te spreken met een onbekende die je dan vervolgens heel serieus vraagt om een half uur lang op plaatjes op een scherm te klikken. Als klap op de vuurpijl krijg je dan, tijdens de tweede afspraak, een badmuts op waarna er (met een stompe naald weliswaar) allemaal gel in je haar komt te zitten om vervolgens weer 3 kwartier afgezonderd in een cabine naar zinnen te luisteren. Jullie deden allemaal met plezier mee en wilden me op alle mogelijke manieren wel helpen met het onderzoek. Vooral de tweetalige mensen verdienen een bedankje voor de manier waarop zij, stuk voor stuk, actief op zoek gingen naar mensen die mee zouden kunnen doen met het onderzoek. Ik heb vaak gedacht dat ik niet genoeg mensen zou kunnen vinden, maar het heeft me ontzettend veel goed gedaan dat sommige van jullie niet schroomden om 3 uur in de trein naar Groningen te komen onder het motto ‘dan maken we er een leuk uitje van’. Dit boekwerk had niet bestaan zonder jullie hulp en ik zal jullie daar eeuwig dankbaar voor zijn. Daarbij wil ik ook de diverse sociale media, het webportaal Polonia, en stichting Humanitas Groningen bedanken voor hun hulp bij het vinden van geschikte proefpersonen voor mijn onderzoek.

Dan zou ik nu een ellenlange opsomming kunnen geven van al mijn collega’s en alles wat zij voor mij hebben gedaan, maar zoals ik al eerder liet blijken behoed ik mij liever voor vergeten namen en noem een selecte, maar specifieke groep. Joyce (a.k.a. ‘the voice’) Haisma moet dan de eerste zijn. Zonder er ooit iets voor terug te vragen heeft zij, zonder te klagen, honderden zinnen ingesproken, waarvan elke zin ook nog eens meerdere malen achter elkaar. Dan wil ik ten tweede John Hoeks, Diana Dimitrova, en Ryan Taylor bedanken. Diana was in Groningen één van de eerste die een auditief EEG experiment in E-prime op ging zetten en zonder de hulp van haar en haar collega’s was het mij waarschijnlijk nooit gelukt om in die korte tijd mijn experiment in elkaar te zetten. *Many thanks also to Agnieszka Czachór for helping me translate my materials into Polish!*

Wat technische zaken betreft heb ik een hoop te danken aan Peter Albronda. Niet alleen wilde hij me altijd wel helpen met van alles en nog wat, ook heb ik altijd genoten van de leuke gesprekken tijdens het testen van proefpersonen op het NiC. Ook de mensen binnen het NWO Vici project van Petra Hendriks wil ik bedanken. In de eerste plaats dat ik gebruik mocht maken van hun Eye Lab en in de tweede plaats voor de hulp, van vooral Gisi Cannizzaro en Sanne Kuijper, bij het opzetten, roosteren, en aanleren van het gebruik van alle voorzieningen in het Eye Lab.

Toen de experimenten eenmaal liepen werd ik gelukkig vaak bijgestaan door diverse mensen, maar Bregtje Seton is wel een hele speciale. Wat hebben wij veel samen getest en gezellig gekletst over diverse onderwerpen, maar ik kan mij zeker ook nog de momenten herinneren dat wij gebaald hebben van alles wat mis ging, of gewoon omdat we moesten klagen over alles wat ons op dat moment dwars zat (en dat was soms best wel veel). Ik heb altijd enorm genoten van onze samenwerking en eigenlijk heb ik je het laatste jaar wel een beetje gemist!

Ingewikkelde experimenten vragen doorgaans om ingewikkelde analyses en wat heb ik de haren uit mijn hoofd getrokken toen ik de oogbewegingen moest gaan analyseren. Gelukkig heb ik met statistici kunnen overleggen, zoals Lammertjan Dam en Mik van Es, waardoor ik ontzettend veel geleerd heb over diverse analyse-methoden. Uiteindelijk kwam ik in contact met Martijn Wieling, en wat was ik daar blij mee! Ik heb niet alleen ontzettend veel van hem geleerd, maar ook heb ik veel plezier beleefd aan de (soms lange) dagen die we besteed hebben aan het, steeds opnieuw, analyseren van mijn dataset.

Some people are fortunate enough to be able to meet the right people in the right place at the right time. After my initial difficulties with several programmes, I got hooked to the E-Prime Extensions for Tobii software to present my experiment and record eye movement data. Fortunately, this knowledge led me to give a presentation at the Tobii eye tracker workshop in Manchester. I want to thank Tobii and the organizers for inviting me and I am thankful for the opportunities I got after this event. Especially my trip to Bangor and the ESRC Centre for Research on Bilingualism was very valuable to me. Many thanks to the group of Virginia Mueller Gathercole, and especially Hans Stadthagen and Rocio Perez Tattam for the nice discussions and talks about eye tracking designs and analyses!

I also received invaluable advice from the people at the Tanenhaus lab at the University of Rochester during my visit in November 2011. In particular, Mike Tanenhaus, who helped me to understand my eye tracking data and the analyses of it far better. Special thanks I would also like to express to Anne Pier Salverda, Florian Jaeger, Meredith Brown, and Kristen Gorman, for looking into my experiment and the data analysis in order to get a better understanding of the (sometimes weird) effects that I had found.

De onderzoeksschool van BCN (school of Behavioural and Cognitive Neurosciences) heeft een belangrijke rol gespeeld in de ontwikkeling van zowel mijn algemene als specifieke kennis van onderzoek, cognitie, en neurowetenschappen. Voornamelijk Diana Koopmans, Janine Wieringa, en Hedwig van Oosten wil ik bedanken voor de organisatie van diverse workshops, lezingen, en symposia, maar ook voor hun gezelligheid tijdens de koffiepauzes van deze activiteiten of gewoon op de gang. Ook wil ik mijn collega's van de BCN nieuwsbrief bedanken, waaronder Evelyn Kuiper-Drenth, Nynke Penninga, Deniz Baskent, en Léon Faber. In mijn tijd als redactrice bij de BCN heb ik veel geleerd over allerlei disciplines en ik ben ervan overtuigd dat dit een positieve uitwerking heeft gehad op mijn gehele PhD.

De portiers van het Harmoniegebouw en de medewerkers van de ICT afdeling en de Audiovisuele Dienst hebben mij een aantal keren uit de brand geholpen en ik wil hen dan ook bedanken voor het helpen met het doorsturen van proefpersonen en het oplossen van computer gerelateerde problemen, iets wat voor mij inmiddels zeker geen onbekend terrein meer is.

Binnen het CLCG wil ik ook iedereen bedanken, maar Wyke van der Meer wil ik hier graag met naam en toenaam inzetten. Wij promovendi aan de RuG mogen in onze handjes knijpen dat we Wyke altijd hadden om onze problemen, klein of groot, professioneel of persoonlijk, mee te bespreken en oplossingen te zoeken.

Naast de mensen die ik hierboven genoemd heb zijn er natuurlijk veel meer collega's die mij de afgelopen jaren hebben geholpen, met wie ik gezellige koffiepauzes heb doorgebracht en/of de nodige borrels en stapavonden mee heb beleefd, waaronder: Ruggero, Ildikó, Veerle, Tim, Karin, Nienke, Emi, Paolo, Sanne, Susanne, Jacolien, Nienke, Kars, Tim, Petra, Farah, Tal, Laura, Rimke, Jan, Rika, Radek, en Anja. In het bijzonder wil ik hier Katharina, Gulsen, Tedi, Ankelien en Sieuwke noemen, maar ook de mensen die ik lang voor mijn promotie al kende, voornamelijk Bernou en Josine. *Many thanks also to those whom I had lovely chats and discussions with at conferences in and outside of the Netherlands: Laura, Eleonora, Alice, Holger, and Susan.*

En dan is er geen ontkomen aan om ook de meest belangrijke mensen in mijn leven te bedanken. Aan de ene kant is dit natuurlijk een uitgelezen kans om eindelijk eens te zeggen hoe blij ik ben dat ik jullie in mijn leven heb. Aan de andere kant wil ik dit graag kort houden, omdat dit nieuw terrein is waar ik mij niet echt in thuis voel en ik ervan uitga dat jullie allemaal weten hoe ik over jullie denk en dat ik jullie voor geen goud zou willen missen. Pappa, mamma, Marijke, Wouter, Hessel en natuurlijk Folkert. Een paar dagen geleden stond er nog een heel lief en sentimenteel verhaal op papier over en voor jullie, maar uiteindelijk vond ik deze woorden te persoonlijk om met de rest van de (taalkundige) wereld te delen. Dat vertel ik jullie later wel. Daarom in dit dankwoord voor jullie een toast op de toekomst: op ons, en alles wat er nog komen gaat!

Groningen dissertations in linguistics

GRODIL

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.
12. Ton van der Wouden (1994). *Negative Contexts*.
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar*.
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar*.
15. Maarten de Wind (1995). *Inversion in French*.
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.

28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. H. Wee (1999). *Definite Focus.*
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority substitutions.*
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic* Structures.
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology.*
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*

59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships.*
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning.*
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*

88. Jelena Prokić (2010). *Families and Resemblances*.
89. Radek Šimík (2011). *Modal existential wh-constructions*.
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*
93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Cagri Coltekin (2011).*Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*
99. Herman Heringa (2012). *Appositional constructions.*
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia*.
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*.
104. Gisi Cannizzaro (2012). *Early word order and animacy*.
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model.*
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*
107. Veerle Baaijen (2012). *The development of understanding through writing.*
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*
109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance dependencies.*
110. Hanneke Loerts (2012). *Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender.*

GRODIL
Center for Language and Cognition Groningen (CLCG)
P.O. Box 716
9700 AS Groningen
The Netherlands