Doctoral theses at NTNU, 2024:104

Katrine Falcon Søby

My mistake!

Production and perception of grammar anomalies

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Department of Language and Literature
University of Copenhagen
Department of Nordic Studies and Literature

NTNU
Norwegian University of
Science and Technology

UNIVERSITY OF
COPENHAGEN

Katrine Falcon Søby

My mistake!

Production and perception of grammar anomalies

Thesis for the Degree of Philosophiae Doctor

Trondheim, April 2024

Norwegian University of Science and Technology
Department of Language and Literature
University of Copenhagen
Department of Nordic Studies and Literature

NTNU
Norwegian University of Science and Technology

UNIVERSITY OF COPENHAGEN

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Department of Language and Literature
University of Copenhagen
Department of Nordic Studies and Literature

ISBN 978-82-326-7802-0 (printed ver.)
ISBN 978-82-326-7801-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:104

Printed by NTNU Grafisk senter

Institutions:	University of Copenhagen The Norwegian University of Science & Technology
Departments:	Department of Nordic Studies and Literature, UCPH The Department of Language and Literature, NTNU
Author:	Katrine Falcon Søby
Title and subtitle:	My mistake! Production and perception of grammar anomalies
Supervisor:	Line Burholt Kristensen
Co-supervisor:	Mila Vulchanova
Submitted:	August 27th 2023

UNIVERSITY OF COPENHAGEN

NTNU
Norwegian University of Science and Technology

Summary

This thesis examines the production as well as the perception of *naturally occurring grammar anomalies*, focusing on verb-third (V3) word order in written Danish (and Norwegian). In these verb-second (V2) languages, the norm prescribes that the finite verb must occur as the second constituent in the sentence. Sentences with V3 word order do not confine to this norm as they have the finite verb in the third position, e.g. **Nu han spiser* ‘Now he eats’ for *Nu spiser han*. V3 is rare in L1 Danish, but is common in Danish as a second language (DSA). The thesis consists of three research articles: the first examines learners’ production of V2 and V3, while the second and third examine L1 users’ attention to V3 (and other anomalies), as well as their online processing of V3. In contrast to previous research on DSA, which has mainly been qualitative, this thesis is based on quantitative studies. As something new, it examines which factors especially have influence on V3 production, including crosslinguistic influence. As for perception, the thesis contributes knowledge about L1 users’ processing of anomalies produced by learners, and about allocation of attention to different types of grammar anomalies.

Article 1 builds on a cross-sectional corpus study of texts from 217 learners of Danish (A2-B1 level). The study examined effects of the learners’ language background (V2 or non-V2), their CEFR level, and the complexity of the first three sentence constituents on correct V2 production. The study showed that V2 is not difficult for all learners, as learners with another V2 language as their L1 had significantly higher shares of V2. For non-V2 learners, the share of V2 significantly increased with CEFR level. Finally, V2 is not equally difficult in all contexts, as increasing complexity of the first constituent and the subject was negatively correlated with the share of V2 (both significantly).

Article 2 is based on an error detection study, in which 211 Danish high school students read texts containing different types of errors: syntactic errors (V3), morphological agreement errors (verb inflections, gender mismatches in NPs) and orthographic errors. The study showed that different types of errors do not attract the same amount of attention, and that V3 is a prominent anomaly. In general, it was found that error rates in an L1 corpus could be used in the hypotheses to predict detection rates for different subtypes of grammar errors. However, frequency was not the only possible explanation. Finally, the study also showed that high scores in a grammar quiz and high levels of self-reported annoyance with errors affected error detection positively.

Article 3 builds on an eye-tracking study of L1 Norwegian users’ online processing of V3 anomalies. Participants read sentences with sentence-initial adverbials, followed by either V2 or V3 word order. The study showed that L1 users reacted immediately to V3 word order, as indicated by increased fixation durations and more regressions out on the misplaced subject, and subsequently on the verb. Participants recovered quickly, already on the following word.

This thesis can inform future eye-tracking and language processing models on allocation of attention to, and processing of, different types of grammar anomalies. It both presents a tentative model of factors in anomaly detection and a preliminary model of how perception patterns can be affected by what is common in L1 production (inspired by prediction theory). This may lead to more robust models which can accommodate naturally occurring, non-standard variation. From a didactic perspective, knowing which anomalies attract more attention, or require additional processing costs, can both help language instructors and school teachers prioritize grammatical focus areas in the classroom.

Danish summary

Denne afhandling omhandler produktion og perception af *naturligt forekommende grammatikafvigelser* med fokus på V3-ordstilling i skriftligt dansk (og norsk). I V2-sprog som disse foreskriver normen, at det finitte verbal skal stå på andenpladsen i deklarative sætninger. Sætninger med V3-ordstilling følger ikke denne norm, da de har det finitte verbal på tredjepladsen, fx **Nu han spiser* for *Nu spiser han*. V3 forekommer sjældent hos L1-brugere af dansk, men findes hyppigere i dansk som andetsprog (DSA). Afhandlingen er baseret på tre forskningsartikler. Den første omhandler learneres produktion af V2 og V3, mens de to andre omhandler L1-brugeres opmærksomhed på V3 (og andre typer afvigelser) samt deres processering af V3. I modsætning til tidligere forskning i DSA, som primært har været kvalitativ, bygger denne afhandling på kvantitative studier og undersøger som noget nyt, hvilke faktorer der især har indflydelse på produktion af V3, inkl. tværsproglig indflydelse. I forhold til perception bidrager den med viden om L1-brugeres processering af afvigelser, som er produceret af learnere, og om opmærksomhedstildeling ift. forskellige typer af grammatikafvigelser.

Artikel 1 bygger på et tværsnitsstudie af tekster fra 217 learnere med DSA (A2-B1). Studiet undersøgte effekterne af learnernes sproglige baggrund (V2 over for ikke-V2), deres CEFR-niveau og kompleksiteten af de første tre sætningsled målt på korrekt V2-produktion. Undersøgelsen viste, at V2 ikke er vanskeligt for alle learnere, eftersom learnere med et andet V2-sprog som L1 havde signifikant højere andele af V2. For ikke-V2-learnerne steg andelen af V2 signifikant med CEFR-niveau. Endelig er V2 ikke lige vanskeligt i alle kontekster. Andelen af V2 er lavere for komplekse subjekter og komplekse led i forfeltet (begge signifikant).

Artikel 2 er baseret på et fejlfindingsstudie, hvor 211 danske gymnasieelever læste tekster, der indeholdt forskellige typer fejl: syntaktiske fejl (V3), morfologiske fejl (verbalbøjninger; genusafvigelser i nominalsyntagmer) og ortografiske fejl. Undersøgelsen viste, at forskellige typer af fejl ikke tiltrækker sig lige meget opmærksomhed, og at V3 er en fremtrædende afvigelse. Generelt viste studiet, at fejlrater i et L1-korpus kunne bruges i hypoteserne til at forudsige fejlfindingsraterne for forskellige undertyper af grammatikfejl. Dog udgjorde fejlenes frekvens ikke den eneste mulige forklaring på mønstrene. Endelig viste undersøgelsen, at høje scorer i en grammatikquiz – og høje scorer, når eleverne angav deres irritation over sprogfejl i et spørgeskema – begge påvirkede fejlfindingsraterne positivt.

Artikel 3 bygger på et eyetrackingstudie af L1-taleres online processering af V3-afvigelser i norsk. Deltagerne læste sætninger med sætningsinitiale adverbialer fulgt af enten V2- eller V3-ordstilling. Studiet viste, at deltagerne reagerede hurtigt på V3-ordstilling, målt i form af længere fikseringer på subjektet og verbalet samt flere regressioner væk fra subjektet og verbalet. Effekterne forsvandt dog hurtigt igen, allerede på det efterfølgende ord.

Afhandlingen kan informere fremtidige eyetracking- og sprogprocesseringsmodeller om opmærksomhedstildeling og processering ift. forskellige typer grammatikafvigelser. I afhandlingen præsenteres både en foreløbig model over faktorer, der kan påvirke opmærksomhedstildelingen, og en foreløbig model for, hvordan perceptionsmønstre kan påvirkes af, hvad der er udbredt i L1-produktion (inspireret af *prediction theory*). Dette kan forhåbentlig føre til mere robuste modeller, som kan rumme naturligt forekommende sproglig variation, herunder grammatikafvigelser. Fra et didaktisk perspektiv kan viden om, hvilke afvigelser der især tiltrækker opmærksomhed eller kræver øgede processeringsomkostninger, være til nytte for undervisere i dansk og DSA ift. at prioritere grammatiske fokusområder i undervisningen.

Acknowledgements

First and foremost, I would like to thank my supervisor Line Burholt Kristensen for giving me the opportunity to be part of the *Broken Grammar and Beyond* research group. I also want to thank Line for her great support during every step of the way, her patience, enthusiasm and valuable feedback.

Moreover, I wish to thank my co-supervisor Mila Vulchanova, who welcomed me warmly to The Language Acquisition and Language Processing Lab at the Norwegian University of Science and Technology in Trondheim, where I spent an interesting and challenging semester. I thank all of the employees at the Department of Language and Literature for good company during a cold winter. Special thanks go to Evelyn Milburn and Valentin Vulchanov, who co-authored one of the articles in the thesis, and to Charlotte Sant and Heming Strømholt Bremnes for helping me create stimuli in Norwegian.

I am grateful for having been part of a wonderful research environment at the University of Copenhagen with great colleagues. I thank Byurakn Ishkhanyan, co-author of one of the articles in the thesis, for always being most enthusiastic about solving any issue I encountered in R. I am also grateful to Sabine Gosselke Berthelsen for interesting discussions, great company, and for spending many hours fighting with SurveyXact. Furthermore, I wish to thank Elisabeth Engberg-Pedersen, Kasper Boye, Tanya Karoli Christensen, Jan Heegård Petersen and Mads Poulsen, for always being happy to answer questions.

I would like to thank all of my old colleagues and students at Københavns Sprogcenter (Copenhagen Language Center) for contributing to the learner corpus.

Finally, I wish to thank my family and friends, and especially my husband, Peter, and son, Niels, for their support.

Contents

List of figures

List of tables

Abbreviations

A list of abbreviations used in the extended introduction (sections 1-6) and in the extended discussion (section 8).

CEFR	The Common European Framework of Reference for Languages
CLI	Crosslinguistic influence
DEF	Definite
DP1	Danish Program 1
DP2	Danish Program 2
DP3	Danish Program 3
DSA	Danish as a second language
IMP	Imperative
INF	Infinitive
L1	First language
L2	Second language
N	Neuter gender
PRS	Present tense
PST	Past tense
U	Uter gender
V2	Verb-second (word order)
V3	Verb-third (word order)

1 Introduction

In this thesis, I examine the production and perception of *naturally occurring grammar anomalies*, defined as cases of convention-breaking grammar in naturally occurring texts (cf. section 2.2). The thesis focuses on verb-third word order (V3), primarily in written Danish, and to some extent in Norwegian. Both Danish and Norwegian are a so-called verb-second (V2) languages. V2 is rare across the world's languages, but all Germanic languages (apart from modern English) have V2 word order. V2 also occurs in a few other languages, such as Estonian, some Rhaetoromance languages/dialects, Breton, Sorbian, and Kashmiri (Holmberg 2015). Languages are characterized as verb-second "when the finite verb is obligatorily the second constituent, either specifically in main clauses or in all finite clauses" (Holmberg 2015:342). Danish, Swedish, Norwegian, German and Dutch have V2 in main clauses, while e.g. Icelandic has V2 in all finite clauses (Holmberg 2015).

Traditionally, in descriptions of Danish, the so-called main clause frame is characterized by having a first position in which any constituent (apart from the finite verb, negation and modal particles) can be placed, and a second position which is occupied by the finite verb (Jensen & Christensen 2013:2, with reference to Diderichsen 1974). Example (1a) is written by a second language learner of Danish. Here, the adverbial *I 1948* 'in 1948' is in first position, followed by the subject *Maria* in second position, and the finite verb *begyndte* 'began' is in third position – not in second as prescribed by the main clause frame (1b). In other words, the sentence in (1a) has verb-third (V3) word order.

(1) a. **I 1948* <u>*Maria begyndte*</u> *at studere medicin*
In 1948 Maria began to study medicine
'In 1948, Maria began to study medicine'

b. *I 1948 begyndte Maria at studere medicin*
In 1948 began Maria to study medicine
'In 1948, Maria began to study medicine'

The anomalous sentence with V3 word order in (1a) does not express a different propositional content than the grammatical sentence with verb-second word order in (1b), but V3 word order is generally not in accordance with norms of written standard Danish, in which V3 is also rarely found (cf. section 3.3, Article 2). In contrast, V3 is relatively common in second language learner production (e.g. L2 Danish: Holmen 1994, Lund 1997, Søby & Kristensen 2019. L2 Norwegian: Brautaset 1996, Hagen 1992, Johansen 2008. L2 Swedish: Bolander 1989, Hyltenstam 1978. L2 German: Håkansson et al. 2002).

In the thesis, the notion *first language* (L1) is used to refer to the language that someone learns to speak first (Cambridge Dictionary 2023), or languages in cases of simultaneous bilinguals. In Article 2 of the thesis, L1 users of Danish are operationalized as those who had acquired Danish before the age of 6, in line with Hyltenstam and Abrahamsson (2003). In Article 3, L1 users of Norwegian are operationalized as participants who were monolingual until starting school. *Second languages* (L2s) are defined as "referring to an additional language learned after the L1, which in many cases is not chronologically the second in the series of languages a given learner has learned or attempted to learn." (Jarvis 2017:13). In Article 2 and Article 3, the term *native speaker* is sometimes used, however, in the rest of the thesis, the more neutral terms *L1 user* or *L1 speaker* are used instead.

V3 is not only found in L2 production. L1 users speaking in a multi-ethnic urban vernacular use V3 with the same meaning as an equivalent sentence with V2, but as part of a different stylistic practice (cf. Quist 2008), e.g. in Denmark (Quist 2008), Norway (Hårstad & Opsahl 2013), Sweden (Kotsinas 2000), and Germany (Freywald et al. 2015). Since word order norms in this vernacular are different from standard Danish, V3 is not clearly ungrammatical (convention-breaking) in this context.

This thesis examines L2 Danish learners' written production of V2 and V3 (Article 1) in a corpus study. Furthermore, L1 Danish speakers' attention to V3, to other types of grammar anomalies, and to common orthographic anomalies are examined using an error detection paradigm (Article 2). Finally, L1 Norwegian users' online processing of V3 is examined using eye-tracking (Article 3). Throughout the thesis, the terms *V2 anomaly* and *V3 anomaly* are used interchangeably.

The other types of grammar anomalies that are examined in the study in Article 2 are different naturally occurring morphological anomalies (cf. Article 2 for a more thorough description):

- Confusion of present tense and infinitive, e.g. **skal har* 'shall have.PRS[1]' for *skal have* 'shall have.INF' (here produced by an L2 user), and **angre* 'repent.INF' for *angrer* 'repent.PRS' (as e.g. produced by an L1 user).
- Missing gender agreement between indefinite articles and nouns, e.g. **en budget* 'ART.U budget.N' for *et budget* 'ART.N budget.N', as well as between adjectives and nouns, e.g. **en rigtig godt uge* 'a really good.N week.U' for *en rigtig god uge* (both produced by L2 learners).

These anomalies were chosen as they (together with V3) represent a broad range of anomaly types, and they are all attested in natural L1 and/or L2 production, but with different frequencies (cf. section 3.3, Article 2) or production patterns. For example, L1 users primarily confuse infinitive and present tense when the two verb forms are homophone, as in the case of *køre* 'drive.INF'and *kører* 'drive.PRS' which are both pronounced [ˈkʰøːɐ]. Confusions are less common in L1 texts when the two verb forms are heterophone, e.g. *rejse* 'travel.INF' [ˈʁɑjsə] vs. *rejser* 'travel.PRS' [ˈʁɑjˀsɐ] (Article 2). This does not seem to be the case in L2 texts (L1 English), where both homophone and heterophone verb pairs are confused (Hansen et al. 2019).

Grammar anomalies are a natural part of texts seen in everyday life, e.g. in online news articles (Rathje & Kjærgaard 2009), and in newspapers (Brandt 1996, Kristensen et al. 2007). They may appear for many different reasons. Writers may not be aware of, or master, grammatical conventions, or they may be distracted, or forget to make the necessary changes when editing the sentence or text elsewhere. Both L1 and L2 users may also be influenced by conventions of other languages they speak, i.e. crosslinguistic influence (Jarvis 2017). Although grammar anomalies are common in everyday texts, attention to, and processing of, grammar anomalies have not been a focal point within the current major models of eye movement control in reading (Reichle 2003, Reichle et al. 2009, Engbert et al. 2005). The E-Z Reader model (Reichle et al. 2009) does in fact make assumptions about eye movements in response to severe syntactic violations, but not other types of anomalies. Often, grammar anomalies are treated as a homogenous group (with the cover term *syntactic violations/anomalies*) (Article 2), although grammar anomalies involve many different subtypes (e.g. word order anomalies, anomalous verb inflections, or missing gender or number agreement in NPs) which are not necessarily noticed to the same degree or processed in the same way. Studying the production and the perception of these naturally occurring anomalies is crucial if future models of

[1] The glosses in the thesis do not adhere to The Leipzig Glossing Rules, for increased readability. I prefer to show the Danish words in their entirety, without hyphens marking morpheme boarders.

reading, or language processing in general, are to be able to accommodate actually occurring non-standard variation.

The thesis focuses on V3 anomalies, as both the production and the perception of these anomalies are relevant to examine. V2 is often described as notoriously difficult to master for L2 learners (e.g. Bolander 1990, Hagen 1992), but previous studies of other V2 languages than Danish have indicated that it may not be equally challenging to all learners, due to crosslinguistic influence from the L1 (e.g. Bohnacker 2006, Johansen 2008). V2 may also not be equally challenging in all sentential contexts (e.g. Bolander 1989, Brautaset 1996, Hagen 1992, Hyltenstam 1978, Johansen 2008). However, most previous studies have only included a few learners or have not used inferential statistics to understand the separate contributions of the learners' language background, the learner's proficiency level (CEFR[2]) and the complexity of the sentence constituents (Article 1). Knowing more about the factors which may influence whether a learner produces V2 or V3 word order can help language instructors and learners to increase the share of V2.

As for perception, little is known about attention allocation in relation to different types of grammar anomalies, including V3 (Article 2). Knowing which anomalies attract more attention can both help language instructors and learners prioritize grammatical focus areas and inform future reading or sentence processing models. As argued in Article 3, there is generally little research on L1 users' processing of non-standard syntax, which is surprising given the prevalence of this type of variation in a world with increased global mobility. Online processing of V3 has, to my knowledge, only been examined in Swedish using EEG (Andersson et al. 2019, Yeaton 2019, Sayehli et al. 2022). Using eye-tracking to examine online processing of V3 (as in Article 3) contributes with important input to processing models for two reasons. First, V3 is a naturally occurring anomaly in both oral and written production, in contrast to randomly scrambled words, as in a previous eye-tracking study on anomalous word order (Huang & Staub 2021). Second, previous eye-tracking studies of ungrammaticality have primarily addressed morphosyntactic anomalies (Article 2, Article 3), and it cannot a priori be known whether word order anomalies elicit the same effects as anomalies involving morphological changes. Together, the perception studies in the thesis (Article 2, Article 3) can inform future processing models about allocation of attention to, and processing of, grammar anomalies. This can hopefully lead to more robust models that can accommodate input from non-proficient language users and other types of non-standard variation.

[2] The Common European Framework of Reference for Languages (Council of Europe 2001).

1.1 BGB research questions and scope of the thesis

The thesis was written as part of the research project *Broken Grammar and Beyond* (BGB), led by Line Burholt Kristensen, and financed by The Independent Research Fund Denmark. The aim of the BGB project was to investigate how naturally occurring grammar anomalies are processed and comprehended by L1 users of Danish. The first step was thus to conduct corpus studies of Danish texts, either written by L1 users or L2 learners, to identify which grammar anomalies are produced by the two groups. The L1 corpus is based on 71 essays written by Danish high school students for their final exam, which were collected by the Danish Language Council (cf. section 3.1). The L2 corpus is based on texts written as part of module tests from 217 students at a language school, collected by me (cf. section 3.2). A selection of the grammar anomalies found in the corpora were used as stimuli in psycho- and neurolinguistic experiments of anomaly processing. The three research questions of the BGB project are presented below.

RQ 1: What are the types and frequencies of grammar anomalies produced in the writing of L1 Danish speakers compared to that of L2 Danish speakers?

RQ 2: How are different types of naturalistic grammar anomalies (characteristic of L1 and L2 usage respectively) processed and comprehended by L1 language users?

RQ 3: What are the neural underpinnings of the comprehension processes from RQ2?

Within this framework, I have focused on the production of V3 anomalies in L2 Danish, and the perception of V3 anomalies by L1 users (of Danish and Norwegian):

RQ1, PRODUCTION:
How does linguistic context and the language background of the writer affect V2 production in written L2 Danish?

RQ2, PERCEPTION:
How do L1 users process V3 anomalies compared to other types of grammar anomalies?

1.2 Overview of articles and relation to published works

Below, an overview of the three articles included in the thesis is given. The research questions related to production of V2 are investigated in Article 1, based on a corpus study, while the research questions related to perception of V2 are examined in Article 2 and Article 3. Using an error detection paradigm, Article 2 investigates potential factors affecting attention to different types of anomalies (including V3 anomalies). It is examined whether anomaly frequencies in L1 Danish can be used in the hypotheses to predict L1 users' attention to anomalies, thus combining production and perception. Article 3 focuses specifically on online processing of V2 vs. V3 word order in an eye-tracking paradigm, providing insights on the time course of the effects of V3 processing. While the first two studies were conducted in a Danish context, the eye-tracking study was conducted during my research stay at The Norwegian University of Science & Technology, and thus the experiment was in Norwegian.

1. **Katrine Falcon Søby** & Line Burholt Kristensen (submitted to *Bilingualism: Language and Cognition*). V2 is not difficult to all learners in all contexts – a cross-sectional study of L2 Danish.

   This article reports the results of a cross-sectional corpus study of the production of V2 vs. V3 word order in written L2 Danish (in sentences with non-initial subjects). Studies of L2 Swedish and Norwegian, which are primarily qualitative, indicate that factors such as crosslinguistic influence and certain characteristics of the constituents involved may modulate the difficulty of producing the correct order of verb and subject (e.g. Bolander 1989, Hagen 1992, Bohnacker 2006). Using statistical models (GLMM), we examined effects of the learners' language background (V2 or non-V2), their CEFR level, and the complexity of the first three sentence constituents, on V2 production. We found that V2 is not difficult for all learners *per se* or in all contexts. Learners with another V2 language as their L1 had a significantly higher share of V2 than non-V2 learners. The share of V2 also increased significantly with CEFR level for the non-V2 learners. Finally, we found effects of sentence complexity, so that the share of V2 decreased significantly with the length of the first constituent and for subjects consisting of multiple words.

2. **Katrine Falcon Søby**, Byurakn Ishkhanyan & Line Burholt Kristensen (2023). Not all grammar errors are equally noticed: error detection of naturally occurring errors and implications for eye-tracking models of everyday texts. Published in *Frontiers in Psychology*, *14*, 1-28.

   This article examined whether some types of naturally occurring anomalies attract more attention than others during reading, measured by detection rates in an error detection study. Danish high school students (N = 211) read texts containing different types of errors: syntactic errors (V3), morphological agreement errors (verb inflections; gender mismatches in NPs) and orthographic errors. The three types of grammar errors were presented in different conditions, varying in relation to e.g. placement in the sentence, frequency and phonological similarity to the correct form. It was examined whether there was a link between the type of errors that participants did not detect, the type of errors that participants produced themselves (as measured in a subsequent grammar quiz), and the type of errors that are typical of high school students in general (based on error rates in the L1 corpus). If an error is infrequent in L1 production, it may cause a larger surprisal effect and thus be more attended to. The study showed that different types of errors do not attract the same amount of attention, and that V3 is a prominent anomaly. In general, we found that corpus error rates could be used in the hypotheses to predict detection rates for different subtypes of grammar errors. Frequency, however, was not the only possible explanation, as phonological similarity with the correct form was often entangled with frequency. Finally, the study showed that high scores in the grammar quiz and high levels of self-reported annoyance with errors affected error detection positively. We did not measure eye movements in this study, but the differences in error detection patterns point to shortcomings of existing eye-tracking models of eye movement control in reading (Reichle 2003, Engbert et al. 2005). Based on our results, we give our recommendations for current and future processing models.

3. **Katrine Falcon Søby**, Evelyn Arko Milburn, Line Burholt Kristensen, Valentin Vulchanov & Mila Vulchanova (2023). In the native speaker's eye: Online processing of anomalous learner syntax. Published in *Applied Psycholinguistics*, *44*(1), 1-28.

In the last article of the thesis, L1 Norwegian users' online processing of V3 word order was examined in an eye-tracking experiment. Although there is extensive research on learners' production of syntactic anomalies, surprisingly little is known about how these anomalies are processed by L1 users, and to what extent they may disrupt online processing. In Article 2, we found that V3 word order was a prominent anomaly, measured by detection rates, to L1 Danish readers. Danish and Norwegian are to a great extent are mutually intelligible (Vikør 2015), but compared to Danes, Norwegians are often described as being more receptive to linguistic variation (Torp 2004). Even in this context, we found that participants reacted immediately to V3 word order, indicated by increased fixation durations and more regressions out on the subject, and subsequently on the verb. Participants appeared to recover from seeing the anomaly equally fast, however. Hopefully, this knowledge can be used to create more robust sentence processing models in the future, which can accommodate various types of non-standard variation. The findings can also – together with new studies of processing of other L2 anomalies in Norwegian – help language instructors and students prioritize which aspects of grammar to focus on.

In addition to the three articles included in the thesis, I have contributed to the following articles (in Danish) during the time of the PhD scholarship:

- **Katrine Falcon Søby** & Line Burholt Kristensen (2019). Hjælp! Jeg har mistede min yndlings rød taske. Et studie af grammatikafvigelser [Help! I have lost my favorite red bag. A study of grammar anomalies]. *Ny forskning i grammatik*, *26*, 89-104.

- Line Burholt Kristensen, Jørgen Schack & **Katrine Falcon Søby** (in press). Om unge der har skulle bøje modalverber, men ikke har turde, ikke har kunne eller ikke har ville [About young people who should have inflected modal verbs, but who have not dared to, not been able to or not wanted to]. *Ny forskning i grammatik*, *30*.

In Søby & Kristensen (2019), we examined the production of grammar anomalies in a subset of our L2 corpus (texts from 28 learners with L1 English). The study was used to develop fundamental concepts (e.g. *grammar anomaly*), and part of section 2 in the thesis is based on this article.

1.3 Contributions to co-authored articles

In Article 1, I designed the study. In Article 2, I was the main contributor in creating the specific research questions and designing the study, in collaboration with my supervisor and co-author Line Burholt Kristensen. In Article 3, I came up with idea for the study and designed it in collaboration with co-authors Valentin Vulchanov and Mila Vulchanova (co-supervisor). In Article 1, I collected the data (learner texts), and in Article 2 and Article 3, I created experimental stimuli, set up the experiments and carried them out. In Article 1 and Article 3, I carried out the analyses, while my co-author Byurakn Ishkhanyan wrote the code for the analyses in Article 2, which I later adapted and used. The collection and analysis of eye-tracking data in Article 3 was supervised by co-authors Evelyn Milburn, Valentin Vulchanov and Mila Vulchanova. I wrote the first drafts for all three articles, apart from one short section, "Processing of V3 – evidence from EEG" in Article 3. Finally, I revised the manuscripts according to suggestions and comments from my co-authors. Thus, in Article 3, my supervisor and co-author Line Burholt Kristensen was involved late in the writing process.

1.4 Structure of the thesis

The thesis consists of three parts. **Part I** forms an extended introduction to the three articles and presents theoretical and methodological issues which are relevant to the articles, but which are not addressed or elaborated in the articles. **Sections 2-4** primarily concern **PRODUCTION** of anomalies, while **sections 5-6** concern **PERCEPTION** of anomalies.

In **section 2**, I present a definition of *grammar* (Boye & Harder 2012), define *grammar anomalies*, and review previous studies on anomaly production in L1 and L2 Danish. **Section 3** describes how the L2 corpus was built and annotated. It also provides an overview of the anomalies found in the L1 and L2 corpora. **Section 4** presents preliminary findings from an online survey among language instructors in Danish as a second language, which I conducted in collaboration with Sabine Gosselke Berthelsen. The survey informs about language instructors' practices and prioritizations in relation to grammatical focus areas, including V2, and their perception of students' difficulties with V2. It thus supplements the corpus study in Article 1 with an instructor perspective.

Section 5 elaborates on theoretical issues concerning anomaly perception which are not thoroughly described in Article 2 or Article 3. Both perception studies focus on the role of the frequency of the anomaly in L1 production – and thus its predictability, but the notion *prediction* is not defined in the articles. Therefore, section 5 presents such a definition (Kuperberg & Jaeger 2016) and it describes how prediction theory can be related to anomaly processing. Finally, it presents a tentative model of possible factors involved in anomaly detection. **Section 6** presents the two research paradigms used to examine anomaly perception: error detection and eye-tracking, and discusses how the paradigms can supplement each other. Also, it discusses what it may add to combine studies of production and perception when examining grammar anomalies.

Part II forms the main part of the thesis, as it contains the three research articles (**section 7**). **Part III** consists of an extended discussion (**section 8**), which summarizes the main findings of the three articles, discusses limitations, and presents possible directions for future research.

The references for the three articles can be found after each article in section 7. The references for the extended introduction and extended discussion are found in **section 9**. Supplementary materials for the three articles are found in the Appendix in **section 10**.

Part I

Theoretical and methodological background

2 Grammar and grammar anomalies

All three articles in the thesis examine production or perception of grammar anomalies, but the fundamental terms *grammar* and *grammar anomalies* are not discussed in the articles. Within linguistic theory, “there is a stricking lack of clarity about what is means to be a grammatical or lexical expression” (Boye & Harder 2012:1). Thus, this section presents a definition of grammar by Boye & Harder (2012) (section 2.1) and discuss how *grammar anomalies* can be defined (section 2.2). Section 2.3 briefly reviews previous studies on orthographic anomalies and grammar anomalies in written L1 Danish. These are relevant to the error detection study in Article 2, in which we use anomaly frequencies in L1 Danish in the hypotheses to predict which anomalies are noticed more. Section 2.4 provides an overview of previous studies on grammar anomalies in L2 Danish, which is mainly relevant for Article 1, in which V2 production in L2 Danish is examined. Previous studies have mainly been qualitative and, in relation to V2 production, they have neither focused on the role of learners’ language background nor sentential context. Section 2.4 also introduces and defines the notion *crosslinguistic influence* (CLI).

Sections 2.1, 2.2 and 2.3 are based on Søby & Kristensen (2019), in which fundamental concepts, such as *grammar anomaly*, were developed.

2.1 What is grammar?

In this thesis, I use a broad definition of grammar, the usage-based GRAM theory (Boye & Harder 2012). The central idea of the theory is that “grammar is constituted by expressions that by linguistic convention are ancillary and as such discursively secondary in relation to other linguistic expressions” (Boye & Harder 2012:1). Lexical elements are defined as having the potential to be discursively primary, i.e. “convey the primary point of an utterance” (Boye & Harder 2012:13).

According to this theory, grammar refers to grammatical morphemes and grammatical words or the combination of elements (i.e. syntax) (Søby & Kristensen 2019). These are both words that traditionally have been described as function words (e.g. conjunctions, articles and auxiliary verbs), but also some types of pronouns and some prepositions. Prepositions can be difficult to categorize as either grammatical or lexical (cf. Messerschmidt et al. 2018). The different types of grammatical elements are always discursively secondary and can never convey the primary point of an utterance. In the utterance *the cow eats*, only the lexical elements *cow* or *eat-* can convey the primary message, not the grammatical word *the* or the grammatical suffix *-s* (Søby & Kristensen 2019). The status of an element as grammatical or lexical can be decided using different addressability or focus tests (Boye & Harder 2012, Kristensen & Boye 2016). For example, the pronoun *man* 'one' is grammatical in the sentence *man vasker dem* 'one washes them' (Kristensen & Boye 2016), as it cannot be focused, e.g. by using a focus particle **kun/ikke/især man vasker dem* 'only/not/especially one washes them' or a cleft construction *det er man, der vasker dem* 'it is one who washes them'.

2.2 What are grammar anomalies?

In the introduction, I defined grammar anomalies as cases of convention-breaking grammar. In the extended introduction, in Article 1, and in Article 3, I primarily use the term *anomaly* to refer to convention-breaking language. Particularly in those cases where the norm might be changing or depend on lectal variation, I prefer the term *anomaly* (instead of *error*). Some types of syntactic anomalies are not defined as errors in grammars (the grammars may not describe this particular phenomenon), but they may still conflict with the norms of the majority of language users. When referring to studies on error analysis, error detection or proofreading, I use the notion *error* interchangeably with *anomaly* in accordance with previous studies. In the error detection study in Article 2, I also use the term *error*, as the stimuli in the study are clearly convention-breaking in the context of writing essays for school.

In the literature on second language acquisition, and especially error analysis studies, different definitions of the notion *error* has been presented over the years (cf. Søby & Kristensen 2019). According to Pawlak (2013:3), these definitions are "far from satisfactory". Lennon (1991a) doubts whether the notion can be defined, as many factors can influence what is considered an error, such as the situation, the interlocutor, style, and production pressure on the speaker. Thewissen (2015:65) describes the border between actual *errors* and *infelicities* as "a notoriously grey area for anyone who has ever got their hands dirty doing error analysis work."

As highlighted in Søby & Kristensen (2019), the most common way to define *errors* within research in second language acquisition is to compare L2 utterances to what an L1 user would produce under similar circumstances (Pawlak 2013). This is for example seen in Lennon's (1991b) often cited working definition of *error:*

> "a linguistic form, combination of forms which in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts" (Lennon 1991b:182).

As pointed out by Søby & Kristensen (2019), comparing to L1 usage is, however, problematic, as L1 users produce anomalies as well (e.g. Blom et al. 2017). Some anomaly types are common, although the norms are clear, e.g. leaving out an *-r* on verbs in present tense **hun <u>køre</u>* 'she drive.INF' for *hun kører* 'she drive.PRS', i.e. using infinitive for present tense (Article 2). Thus, I define anomalies according to norms or *langue* (Harder 2010) (in line with Søby & Kristensen 2019):

> "*Langue* is not a set of actually attested regularities, but the set of options/constraints that attested regularities have generated as the context of any future potential linguistic acts[...]" (Harder 2010:302).

The challenge of defining anomalies in relation to a norm is that, as mentioned above, norms can change or depend on lectal variation. Norms may also vary in written vs. oral production. Therefore, sometimes a choice between different norms has to be made, and it must be stated explicitly which norm anomalies are being compared to (Søby & Kristensen 2019).

When tagging anomalies in corpora, researchers do not know the writer's norms, or *langue* (Søby & Kristensen 2019). They cannot know how or why an anomaly occurred or what the writer intended to write. The researchers can only relate to how they as recipients perceive the anomaly and categorize it. Such an introspective approach has limitations. Researchers may not be aware that a norm is changing. They may also judge texts written by learners differently than texts written by L1 users, being less open towards alternative forms when the writer is an L2 user (Søby & Kristensen 2019).

When categorizing anomalies, one has to guess what the writer intended to write, i.e. explicitly set out a suggestion for a correct target form (cf. section 3.3). Lüdeling & Hirschmann (2015:140) argue that "the identification and classification of grammar errors is far from easy" and that researchers already here have to interpret the data. The authors argue that in order to address "how a given learner expression would be used by native speakers" (Lüdeling & Hirschmann 2015:140), an alternative expression must be provided. This alternative expression, i.e. assumption of which correct utterance corresponds to the anomalous utterance, is referred to as the *target form* (alternatively, *target hypothesis* or *reconstruction*).

When localizing a potential anomaly in the L1 or L2 corpus, I (in line with Søby & Kristensen 2019) have considered whether it was anomalous in relation to the norms of the Danish Language Council (which e.g. schools are obliged to follow by law (DSN 2023)), to Danish grammar books, and according to my own linguistic intuition. These three sources are not always consistent. For example, I would consider using the uninflected form *ny* 'new' instead of the inflected form *nye* 'new.DEF' as anomalous in the noun phrase *min <u>ny</u> lejlighed* 'my new apartment', which sounds archaic to me. The uninflected form is often seen in the L2 corpus, as adjective inflections are frequently omitted. However, according to the official dictionary (Jervelund et al. 2012), it is optionally to use the suffix, and therefore, *ny* is not tagged as anomalous, although it may be considered as anomalous (Søby & Kristensen 2019).

In Article 1 it is described that sentences with *måske* 'maybe' and *så* 'so/then' were generally excluded from the analysis. When the adverb *måske* 'maybe' is in in first position, it can both be succeeded by verb-subject (i.e. V2) or subject-verb word order (i.e. V3) (cf. Article 1, Beijering 2010, Boye 2005). Thus, the success rate could not be decided in this context. Sentences with *så* 'so/then' posed another challenge, as the prescribed word order could often not be decided. *Så* 'so/then' can either be used as an adverbial or as a conjunction (either introducing main or subordinate clauses). When used as a coordinating conjunction (conveying a result or a consequence), it is followed by subject-verb word order, but when it used as an adverb ('then'), it is followed by verb-subject word order (cf. Article 1).

An example without *så* 'so/then' from the L2 corpus, in which the prescribed word order was also difficult to decide, is seen in (2a). The first part of the sentence, until the comma, may be interpreted as a subordinate clause *Når jeg kommer tilbage fra turen* 'When I come back from the trip' in first position (2b). The order of word and subject should then be SV in the subordinate clause, and VS in the main clause (i.e. two word order anomalies should be tagged). However, the sentences may also be two coordinate clauses as in (2c). In this case, the first sentence should be tagged as an example of correct VS (V2) after an adverbial. The solution became to exclude such sentences (Article 1).

(2) a. *Bagefter kommer jeg fra tur,* ***vi skal*** *går til stranden sammen.*
Afterwards come I from trip we shall go to beach.DEF together

b. *Når jeg kommer tilbage fra turen,* ***kan vi*** *gå til stranden sammen.*
When I come back from trip.DEF can we go to beach.DEF together
'When I come back from the trip, we can go to the beach together'

c. *Bagefter kommer jeg tilbage fra turen.* ***Vi kan*** *gå til stranden sammen.*
Later come I back from trip.DEF. We can go to beach.DEF together
'Later I come back from the trip. We can go to the beach together'

To sum up, in this thesis I define grammar anomalies in relation to norms (in line with Søby & Kristensen 2019). Based on the GRAM theory, grammar anomalies are defined as 1) anomalies related to combinatorics and/or 2) related to use of, or spelling of, grammatical items (i.e. grammatical words or grammatical affixes) (Søby & Kristensen 2019). Most anomalies are both related to combinatorics and the item side, as seen in the example from Søby & Kristensen (2019) in (3). Here, the combination of the quantifier *hele* 'all' and the noun *nat* 'night' is anomalous, as a noun suffix denoting definite form (a grammatical item) is missing. The target form is *hele natten* 'all night.DEF'.

(3) **Vi skal danse hele nat*
We shall dance all night
'We're going to dance all night'

In section 3.3, an overview of the anomalies found in the L1 and L2 corpora is presented. The anomalies are divided into the categories: syntax, morphology, orthography, missing word, superfluous word, and choice of word. Punctuation anomalies were not tagged, although they are also related to grammar. However, in the L1 corpus, there were so many anomalous commas that tagging these would remove the focus from morphology and syntax. In the L2 corpus, anomalous commas (among other punctuation anomalies) were not tagged, because not all learners in the L2 corpus have the prerequisites for placing commas correctly in Danish, as comma rules may be introduced late in the programs.

The categories syntax and morphology are always related to grammar. For example, syntactic anomalies like V3 word (cf. example 1b) are related to combinatorics. The morphological anomalies examined in the thesis (confusion of finite and non-finite verb forms, gender agreement in NPs) are related to use of grammatical items (articles and suffixes). The anomalies within the orthography category are not related to grammar, e.g. *tet* for *tæt* 'close', where the root of a lexical element is anomalous (Søby & Kristensen 2019). Anomalies in the categories missing and superfluous words may be related to grammar as in (4) where a grammatical word, the infinitive marker *at* 'to', is superfluous (example from Søby & Kristensen 2019). The category choice of words also contains grammar anomalies, either when grammatical words are confused, as in (5a) where the conjunction *og* 'and' is used for the infinitive marker *at* 'to', or when words from different word classes are confused, as in (5b), where the verb *adressere* 'address/direct' is used for the noun *adresse* 'address' (examples from Søby & Kristensen 2019).

(4) **der er ikke så meget de kan at gøre*
there is not so much they can to do
'there is not so much they can do'

(5) a. **det er dejligt og bo i København*
it is lovely and live in Copenhagen
'it is lovely to live in Copenhagen'

b. **min adressere*
my address.INF
'my address'

2.3 PRODUCTION: Anomalies in written L1 Danish

This section briefly presents previous studies on anomalies in written L1 Danish. Although I have primarily been involved in the L2 corpus, anomaly frequencies in L1 Danish are used in the hypotheses to predict detection rates in Article 2. It is an important point in Søby & Kristensen (2019) that it is difficult to compare the findings in the L1 corpus, i.e. anomaly types and frequencies, to previous studies of anomalies in L1 Danish (e.g. Blom et al. 2017, Brink et al. 2014, Johannsen 2012), because their focus have not been on reporting grammar anomalies, and they have used more narrow definitions of grammar than us (cf. section 2.1, Boye & Harder 2012). The latter is also the case in previous studies on perception of anomalies (cf. section 5). This important point is elaborated below (as reviewed in Søby & Kristensen 2019).

As seen in Jervelund's (2007) thorough review of previous studies of anomalies in written Danish, the focus has often been on spelling proficiency. The most important distinction has been between spelling errors and non-spelling errors, while grammar anomalies are either not reported or seen as a residual category. For example, the confusion of the intransitive verb *springe* 'explode' and the transitive *sprænge* 'blow up', or the use of *lærere* 'teachers' instead of *lærerne* 'the teachers' are often considered as letter confusions, rather than related to grammar (Jervelund 2007 in Søby & Kristensen 2019).

Later studies of anomalies in written language have focused on the role of phonology, e.g. Johannsen (2012) and Brink et al. (2014) who distinguish between punctuation errors, errors not affecting the phonology of the word, and errors which do change the phonology of the word, when examining high school texts. Punctuation anomalies are in most cases grammatical, but other than that, the categorization does not distinguish between grammar anomalies and anomalies not related to grammar. Both in Johannsen's (2012) study of high school essays and in a later study of university papers (Blom et al. 2017), punctuation errors (especially commas) are the most frequent anomaly type, in fact 90 % of all anomalies are punctuation errors in Johannsen (2012) (Søby & Kristensen 2019).

Unlike previous studies, Blom et al. (2017) do not primarily focus on either spelling or phonology. The study examines a broad range of anomaly categories: punctuation, spelling, semantics, reference, layout, concord, idioms, and syntax. Using a broad definition of grammar (Boye & Harder 2012), punctuation, reference, concord and syntax could be seen as related to grammar. However, many of the other categories also contain anomalies which I (Søby & Kristensen 2019) would consider grammatical, especially the category spelling. For example, *<u>af</u> *vide* 'of know' instead of *at vide* 'to know' is considered as a confusion of consonants, and not as related to grammar, even though the infinitive marker *at* 'to' is a grammatical item. Furthermore, confusion of the present tense verb *synes* 'think' and its past tense form *syntes* 'thought' is considered a spelling mistake (but of the type inflection deviations). Lack of the letter *r* in present tense verbs or excess *r* in infinitives are also categorized as spelling mistakes. The category semantics contain missing and superfluous words, which may be grammatical as well (Søby & Kristensen 2019).

Kristensen & Søby (2022) found that 56 % of the anomalies in the L1 corpus (cf. section 3.3) are related to grammar, disregarding punctuation. Thus, a large part of anomalies in young Danes' written production is related to grammar, when using the definition by Boye and Harder (2012).

2.4 PRODUCTION: Grammar anomalies in L2 Danish

This section provides an overview of previous studies on grammar anomalies in L2 Danish, defines the term *crosslinguistic influence* (CLI), and elaborates on previous studies on V2 production in L2 Danish (Holmen 1994, Lund 1997) which are only superficially described in Article 1.

To my knowledge, there are no previous quantitative studies comparing the distribution of different types of anomalies in L2 Danish, besides Søby & Kristensen (2019). Previous studies on grammar in L2 Danish have primarily examined language acquisition in longitudinal studies of a few learners (e.g. Lund 1997, Holmen 1994), or focused on the order of acquisition of specific grammatical phenomena, such as adjective inflections and placement of negation in subordinate clauses (Glahn et al. 2001).

Søby & Kristensen (2019) found that 55 % of the anomalies produced in a subset of the L2 corpus (28 learners with L1 English, A2-B1) were related to grammar, disregarding preposition anomalies (which are not easily categorized as grammatical or lexical, cf. Messerschmidt et al. 2018), as well as punctuation. This result comes with reservations. First, the finding is based on a small corpus (5,685 words). Second, we did not distinguish between proficiency levels. Thewissen (2013) has examined all anomalies in a large corpus of written L2 English (B1-C2). She found that when learners' proficiency increases, so does overall accuracy (and the progress is seemingly strongest between B1 and B2). Thus, the share of grammar anomalies found in Søby & Kristensen (2019) may have differed on different proficiency levels, if they had been analyzed separately.

Previous research in Danish as a second language has had little focus on the role of crosslinguistic influence or transfer, although crosslinguistic influence from one's first language (L1) play an important part in L2 learning, potentially affecting all language levels, from phonology to discourse (Mitchell et al. 2013:16). In this thesis, I follow the common convention of using the terms *crosslinguistic influence* (CLI) and *transfer* interchangeably (Tenfjord et al. 2017:2), as defined by Jarvis (2017):

> “Transfer refers to the ways in which a person’s knowledge of one language can affect his or her learning, knowledge and use of another language” (Jarvis 2017:12).

As Jarvis describes, “some researchers are unhappy with the term *transfer* because they feel it implies that something is literally being moved from one place to another” (Jarvis 2017:12). Jarvis argues that, although most researchers today do not assume that e.g. structures or meanings are literally transferred from one part of the brain to another, the original meaning of the Latin *transferre* ‘carry across’ is still relevant to many cases of crosslinguistic influence, “where language learners appear to *carry over* certain patterns of language use from one language into their use of another language” (2017:2).

In contrast to research in L2 Danish, there are several studies of CLI in L2 Swedish (e.g. Andersson et al. 2019, Bohnacker 2006, Ringbom 1985, Sayehli et al. 2022) and L2 Norwegian (Gujord 2017, Janik 2017, Ragnhildstveit 2017, Szymańska 2017). For example, Andersson et al. (2019) compared L1 and L2 users’ processing of V3 in Swedish in an EEG study. They found that L1 German learners (who have an L1 with V2) displayed more “nativelike” ERP-effects than L1 English learners (who do not have an L1 with V2). In L2 Norwegian, many studies of CLI have been carried out, based on the large Norwegian Andrespråkskorpus ‘Second language corpus’ (ASK 2015), e.g. in relation to adjective inflection (Janik 2017), gender assignment (Ragnhildstveit 2017), tense-aspect morphology (Gujord 2017), and use of spatial prepositions (Szymańska 2017).

Article 1 contains a review of previous studies of V2 production which have focused on the role of CLI (L2 German: Håkansson et al. 2002, Bohnacker 2006. L2 Norwegian: Johansen 2008. L2 English (L1 German or Dutch): Rankin 2012, Westergaard 2003) and the role of the sentential context, including constituent complexity (L2 Swedish: Hyltenstam 1978, Bolander 1989. L2 Norwegian: Hagen 1992, Brautaset 1996, Johansen 2008). As mentioned in the introduction, most previous studies on V2 production have only included a few learners or have not used inferential statistics to understand the separate contributions of the learners’ language background, the learner’s proficiency level and the complexity of the sentence constituents, as in Article 1.

The findings of the two previous studies on language acquisition, including V2 production, in L2 Danish are only briefly described in Article 1. In Lund’s (1997) longitudinal study, she examined the oral and written production of six learners (L1 Dutch, English, Spanish, and Portuguese), during their first five and a half months with intensive Danish classes. In a qualitative analysis, she describes that only the two speakers of Dutch, likely due to their L1 being a V2 language, achieved “some stability” in producing V2 in declarative clauses (Lund 1997:158). Holmen’s (1990, 1994) longitudinal study of six speakers of L1 Albanian, English and Vietnamese focused on their syntactic development in oral production, during their first 3-15 months in Denmark. In the quantitative part of the analysis, Holmen (1994) found a general increase in the length and number of utterances, and in other complexity measures such as the number of verbs or NPs per utterance. In general, the learners also gradually used more morphological coding, grammatical words and, for the majority, more subordination. The learners, in general, had a high share of V2 after *det* 'it' in first position, when it was part of a unanalysed chunk such as *det ved* [V] *jeg* [S] *ikke* 'I don't know' (in line with Bolander's (1989) findings for oral L2 Swedish). Gradually, the learners produced more sentences with other constituents than the subject in first position, but the word order varied between SV (i.e. V3) and VS (i.e. V2) (Holmen 1990:159). Seemingly, the two learners with L1 English had the most success with producing V2.

3 Method: Corpus studies

The present section describes how the L1 and L2 corpora were built and annotated, which is not elaborated in either Article 1 or Article 2 using the corpora. The L1 corpus was used for the error detection study in Article 2 to examine whether anomaly frequencies in L1 Danish can be used in the hypotheses to predict L1 readers' attention to anomalies during reading. The construction of the L1 corpus is only briefly described (section 3.1), as the texts for this corpus were collected by the Danish Language Council. The main focus of this section is on the construction of the L2 corpus (section 3.2), which is used in Article 1 to examine learners' V2 production. Within the BGB project, the aim of building the corpora was to examine which grammar anomalies occur in L1 and L2 Danish. Section 3.3 describes the mark-up of both corpora, which was based on the same annotation system, and provides an overview of the anomalies found in the L1 and L2 corpora.

3.1 The L1 corpus

In 2016, The Danish Language Council collected 187 essays from the final exam in the high school subject *Danish*, from high schools located in different parts of the country. A subset of these essays were shared with the BGB research project and used for the L1 corpus, which consists of 71 essays, written by 71 participants from three different high schools across the country (one in Jutland, one on Funen and one on Western Zealand). In total, the L1 corpus consists of 127,957 words, and the mean number of words per participant is 1,802.21 words (SD = 433.83 words).

In Denmark, there are three different upper secondary education programs: STX, HTX, and HHX. While they all prepare for higher education, they have different profiles. STX is a general examination program, HTX is a technical examination program with a STEM profile (i.e. Science, Technology, Engineering, and Math), and HHX is a commercial examination program with a business profile (Ministry of Higher Education and Science 2022, Article 2). The L1 corpus contains texts from all three education programs. Twenty-three essays are from STX (the school on Funen), 22 essays are from HHX (the school on Western Zealand) and 26 are from HTX (13 from the school in Jutland, 13 from the school on Western Zealand).

A major advantage of this sampling method is the easy access to essays written by students from both STX, HTX, and HHX. In the error detection study in Article 2, students from all three programs participated as well. Although the corpus does not contain texts from the exact same students, which participated in the experiment, it still provides insights into the types of anomalies that young people across the three examination programs produce. One limitation of using texts which have already been collected is that additional information about the participants cannot be provided. Unfortunately, the Danish Language Council did not collect information on the language background of the participants. Thus, it is not known when the students started acquiring Danish, but in either case, they are highly proficient speakers, as they have (almost) completed an upper secondary education program in Danish.

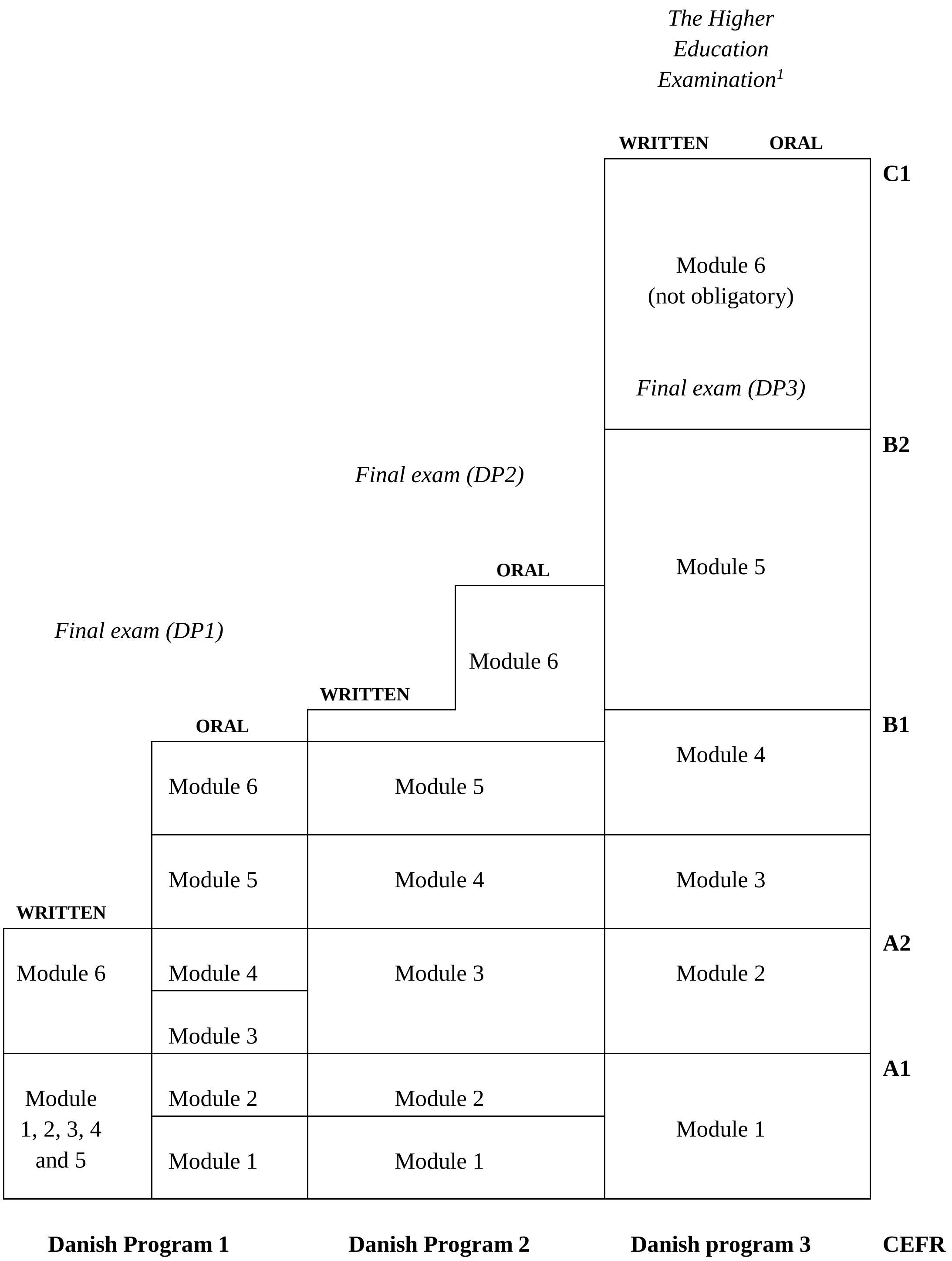
3.2 The L2 corpus

The purpose of this section is to elaborate on the construction of the L2 corpus by introducing the official Danish language programs (section 3.2.1), the data collection (section 3.2.2), and the learners and the texts (section 3.2.3). Furthermore, in section 3.2.3, the L2 corpus is compared to the Norwegian Andrespråkskorpus ‘Second language corpus’ (ASK 2015), to discuss the advantages and limitations of the Danish corpus. The ASK corpus is one of very few large-scale learner corpora with texts in a target language other than English (Tenfjord et al. 2017).

3.2.1 The official Danish language programs

In Denmark, there are three official Danish language programs called Danish Program 1, 2 and 3. When students enrol, they are assigned to one of the programs, primarily based on their educational background. Danish Program 1 (DP1) is for students who cannot read and write using the Latin alphabet. Danish Program 2 (DP2) is for students who have a short educational background and who are expected to have a relatively slow progression. Danish Program 3 (DP3) is for students with a medium or long educational background, who are expected to have a rapid progression (Ministry of Immigration and Integration 2019). All programs consist of a series of modules (see Figure 1). After each module, students must pass a module test in order to continue in the program. The tests consist of one or more of the disciplines speaking, reading, listening and writing, depending on the module. The data used in this thesis are all from the writing assignments. Module tests are not graded and are not final exams. The local teacher decides whether the student passes the test and can progress in the program.

Figure 1. The structure of the official Danish language programs and the CEFR levels (Council of Europe 2001) corresponding to the completed modules. Each program consist of six modules, although the sixth module at DP3 is not obligatory. The programs are completed after taking a final exam (both oral and written).



[1]*The Higher Education Examination*[1] (in Danish: *Studieprøven*) can e.g. be used to apply for admission to higher education programs (in Danish).

3.2.2 Data collection

The data collection took place at the language school *Københavns Sprogcenter* (*Copenhagen Language Center*), located in Vesterbro in Central Copenhagen from September 2017 to March 2018. The school was one of the largest providers of subsidised Danish courses in Copenhagen at that time. A statistic from the school paper *Verdenspressen, 17* from April 2018 showed that 3,309 students from 131 different countries were enrolled at that time. The top five of countries which students came from, was made up by the UK (196), India (184), China (161), Italy (156), and Poland (149). In order to receive subsidised (at that time free) Danish classes, you had to be 18 years old, live in Denmark and have a Danish identification number (CPR). Thus, a few students payed for their classes, e.g. if they did not (yet) have a CPR number. The students came from many different educational and cultural backgrounds. Most of them were working or studying in Denmark or had moved there because of their spouse.

The data collection was approved by *The Faculty of Humanities' Research Ethics Committee* in August 2017. The first data were collected in September 2017, and the last in March 2018, primarily in two rounds. I worked as a teacher at the school and collected module tests from my own classes. Furthermore, I announced on the school's intranet that I was collecting data for a research project, and many of my colleagues volunteered to collect data in their classes as well. Teachers who wished to help received a written instruction on how to collect data in order to secure consistency.

When handing in their written tests to the teacher, students received a document with a short summary of the project (specifying the purpose of the data collection, the aims of the research project and the anonymization and registration process). If they wished to participate in the study, they received a consent form to sign. All written information was provided in Danish or English, depending on the student's preference. The participants also filled out a questionnaire with information about their gender, age, nationality, years of living in Denmark, L1 and other languages. The tests were anonymized, including names, phone numbers, addresses etc. mentioned in the written essays.

An advantage of collecting tests from the same school was that many variables, apart from proficiency level and language background, were kept as constant as possible (e.g. test procedures, teaching materials, numbers of lessons offered etc.). We aimed at getting as many students to participate as possible, from as many different programs and modules as possible, but the data collection depended on which classes had planned tests at that time, and if they had a test, whether the teacher wanted to assist. Thus, it was not easy to balance the corpus in relation to programs and modules (and thus CEFR levels). Unfortunately, it was not possible to return to the language school later and collect more tests, as the school's contract with the municipality was not renewed due to losing a public tendering, not long after the data collection ended.

3.2.3 The L2 learners and the texts

Altogether, texts from 217 students (138 women; mean age 30.9 years, *SD* 7.2 years) were included in the corpus (37,304 words in total). Six students had participated twice, either because they had retaken a test or because they had completed a new module during the time course of the data collection. Thus, one of their tests was excluded. The test with the lowest number of participants was kept, i.e. if a learner in DP3 had both participated after module 3 and 4, the test from module 4 was kept, as there was a larger number of tests from module 3 than 4 in the corpus. Tests from one class (8 students) were also excluded, as the teacher had corrected the texts before submitting them to me.

The learners have around 52 different L1s and 65 different nationalities. Table 1 provides an overview of all learners' L1 (and DP and module). For most languages, there are only a few speakers.

In the corpus, most data are from DP3, as seen in Table 2, which also provides an overview of the participants' Danish program and module, among other things. There are no data from DP1 and only data from one module 3 class at DP2 (ten learners). At DP3, there are written tests after modules 2, 3, and 4. The written test after module 3 at DP2 is identical to the test after module 2 at DP3. All module tests are national, i.e. the same all over the country. When the data were collected, there were three different versions of each module test. In this way, students who did not pass a test would get another test for the retest. Some of the tests are currently still in use, and thus, the exact formulation and instruction in the tasks are not reported in this thesis. The writing assignments include tasks such as writing notices, invitations, or emails. On B1 level (after module 4 at DP3), the tasks are more complicated, e.g. to write a debate piece for a newspaper. The assignment usually specifies who the addressee of the text is, what the purpose is etc. At DP3, the test after module 2 consists of one task (30 minutes), the test after module 3 consists of two tasks (15 and 30 minutes), and the test after module 4 consists of two tasks (each 45 minutes).

The assignments were written by hand. During the tests, students could not use computers or smartphones to look up words. They were, however, allowed to bring a dictionary and use a verb list with inflections that was handed out at the test. The handwritten texts were digitized by the student assistants in the research group.

Table 1. Overview of learners' L1, DP and module. Slashes mark that learners have indicated to have multiple L1s. DP = Danish Program, M = module.

L1	DP2 M3	DP3 M2	DP3 M3	DP3 M4	Total N of learners
English	4	16	11	7	38
Spanish	1	9	4	2	16
German		7	3	3	13
Portuguese	1	6	3	2	12
Russian		6	3	2	11
Italian		7	1	2	10
Polish		6	2	2	10
Dutch		2	4	1	7
Arabic	1	5			6
French		4	2		6
Greek		3	2	1	6
Lithuanian		4	2		6
Romanian		2	4		6
Hungarian		2	2	1	5
Bulgarian		3	1		4
Turkish		2	1	1	4
Bahasa Indonesia		3			3
Finnish		3			3
Malayalam		3			3
Thai		2	1		3
Vietnamese	1		1	1	3
Farsi		2			2
Filipino		1		1	2
Icelandic		2			2
Chinese	1			1	2
Korean		2			2
Odia (Oriya)		1	1		2
Serbian		2			2
Tamil	1	1			2
Urdu		2			2
N = 1: Afrikaans (3.2), Albanian/Croatian (3.2), Assamese (3.2), Bangla (Bengali) (3.2), Bisaya (3.2), Cebuano (3.4), English/Italian (3.2), Gujarati (3.2), Hindi (3.2), Cantonese (3.2), Croatian (3.2), Lamnso (3.4), Lettish (3.3), Luganda (3.2), Nandi (3.2), Nepali (3.2), Sindhi (3.2), Somali (3.2), Tagalog (3.2), Telugu (3.2), German/Russian (3.2), Ukranian/Russian (3.3), Hungarian/Slovak/German (3.3), Unknown (3.2)					
Total	**10**	**127**	**51**	**29**	**217**

Table 2. Overview of Danish program, module, CEFR level and months of teaching

	Participants	CEFR	Months of teaching
Danish Program 2	**10**		
Module 3	10	A2	9 months
Danish Program 3	**207**		
Module 2	127	A2	5 months
Module 3	51	A2+	8 months
Module 4	29	B1	12 months
Total	**217**		

Table 2 also provides an overview of participants' estimated CEFR level. As seen in Figure 1, describing the Danish language programs, some modules correspond to levels described in the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001). For example, a passed module 2 test on DP3, in principle corresponds to A2 level. B1 level is in principle achieved on module 4 (DP3). The level in-between, I have named A2+ (i.e. after module 3) (cf. Article 1). These levels are only rough estimates of students' proficiency. I do not know if students passed the test or not. Testing all students' proficiency levels in other ways would have acquired more resources. The rough estimate of students' proficiency in some cases underestimate their competence. Sometimes highly proficient students with prior knowledge of Danish have to complete a low module as a formality before moving on in the program. In other cases, the rough estimate will overestimate their proficiency. Students with relatively low proficiency sometimes make it through to module 4 by taking all previous tests several times, but are not able to pass the final exam at DP3.

Finally, Table 2 provides an estimate of how long students in the different modules have been taking Danish classes. In practice, many factors make it difficult to estimate exactly how many months students at different modules have studied Danish (as elaborated below), but the estimates in the table are based on the language school's new and (stricter) policy for when students *had* to take a test (whether they were ready or not). This policy became effective from the 1st of January 2018, i.e. in the middle of the data collection period. Before this date, the teacher could decide whether students were ready to take the test, or needed extra time. The number of lessons offered during the time course of the module also varied, depending on the type of the course. The school offered day classes (3 x 4 lessons per week), evening classes (2 x 3 lessons per week) and extra intensive classes, such as online classes (4 lessons per week) at DP3. Thus, a student taking evening classes had fewer lessons than a student in a day class. To complicate matters further, the number of lessons offered were also cut down in this period – especially day classes which went from three to two days a week after module 3. It was very common that students changed classes, for instance from day to evening, if they found a job, or had changes in their schedules. This could either mean that they had to take a test earlier or later than scheduled in their old class. Also, it was common that people travelled for long periods or stopped coming to classes and then had to start the module over again. Thus, the estimates in Table 2 are first of all official guidelines from the school, reflecting a course where the student follows the program strictly and passes all tests on time.

The aim of building the L2 corpus was first and foremost to examine which grammar anomalies occur in L2 Danish, and the corpus fulfills this purpose. It provides insights into which anomalies learners on different CEFR levels (A2-B1), with many different language backgrounds and nationalities, produce. The learners have attended courses at the same school and have generally used the same standard teaching materials (e.g. at DP3: Thorborg & Riis 2010 (module 1), Slotorub & Moreira 2014 (module 2), Slotorub & Moreira 2011 (module 3), Langgaard 2011 (module 4)). Generally, these teaching materials introduce Danish word order as subject-before-verb (with exceptions) (cf. section 4.1, Article 1).

The Norwegian Andrespråkskorpus 'Second language corpus' (ASK 2015) consists of texts from more than 1700 adult learners (620,000 words). In comparison, our L2 corpus is small, but it is apparently the first attempt of building a corpus of written L2 Danish. As argued in section 2.4, there are no previous quantitative studies comparing the distribution of different types of anomalies in L2 Danish, and previous studies on grammar acquisition in L2 Danish have primarily made longitudinal studies of a few learners (e.g. Lund 1997, Holmen 1994), or used oral elicitation of specific grammatical phenomena (Glahn et al. 2001).

Like the L2 corpus, ASK consists of texts written as part of standardized tests (an intermediate test or a higher-level test of L2 Norwegian), but one difference is that all learners in ASK passed their tests and are assumed "to be at or above the proficiency level associated with a passing score for that test" (Tenfjord et al. 2017:3). This roughly corresponds to the B1 and B2 CEFR levels. In hindsight, being able to exclude texts from learners who did not pass the test would improve the estimates of the learners' CEFR levels in the L2 corpus. Furthermore, it would have been an advantage if the L2 corpus also contained texts from learners on B2 and C1 level, covering all CEFR levels represented in the official Danish language programs.

The ASK contains texts written by learners from 10 different language backgrounds, which both include non-European, Indo-European but non-Germanic, and Germanic languages. Our L2 corpus contains texts from learners with approximately 52 different L1s, but with few learners in each group. This makes it challenging to examine crosslinguistic influence from the L1.

Neither the ASK or our corpus contains texts written by the learners in their L1, about the same topic, written under similar conditions (or by L1 users who are as comparable as possible to the learners). As argued by Tenfjord et al. (2017:9) L1 texts can be useful when examining crosslinguistic influence "in order to determine the extent to which the L2 patterns really do reflect L1 tendencies."

Finally, in both corpora, the texts were written in a formal testing situation (to prompts not selected by the learners), and thus, likely reflect the learners' best attempts at producing correct text. However, this may have consequences for the naturalness and authenticity of the language (Tenfjord et al. 2017).

3.3 Markup and overview of L1 and L2 anomalies

This section briefly describes how the anomalies in the L1 and L2 corpus were tagged and presents an overview of the anomalies found in the two corpora. The purpose of the tagging was to get a quick overview of the anomalies in the L1 and L2 texts. Thus, the anomaly categories used were mainly created based on practical considerations (e.g. being easy to apply), instead of being theory-driven. In Article 1, using L2 corpus data, and in Article 2, using L1 corpus data, more detailed categorizations are made.

The anomalies were tagged using XML (Extensible Markup Language). XML is a markup language without predefined tags. Instead, researchers can define the tags and attributes (which provide extra information about the tags) themselves. Kristensen and I developed a manual for tagging anomalies in collaboration with Jørgen Schack and Philip Diderichsen from the Danish Language Council. The manual described all tags and attributes and was used for both corpora. The purpose of the anomaly categories in the tagging manual was first and foremost to create categories which were easy to apply when tagging. The categories were tested on both L1 and L2 texts and adjusted accordingly.

Anomalies were divided into six provisional categories: orthography, morphology, syntax, choice of word (lexis), missing word, and superfluous word. For every anomaly, a target form in accordance with norms was suggested (cf. section 2.2, Lüdeling & Hirschmann 2015), a subcategory was chosen, and where relevant, the word class was specified. For example, when tagging the anomaly **skal <u>har</u>* 'shall have.PRS' for *skal have* 'shall have.INF', the verb form *har* 'have.PRS' would be tagged with the category morphology, the suggested target form would be *have* 'have.INF', and the word class would be specified as "verb_present tense". The subcategory would be "confusion of finite and non-finite verb forms". This category covers all cases where a finite verb form is used instead of a non-finite verb form (infinitives or participles) and vice versa, e.g. also past tense for infinitive **Vi skal <u>spiste</u>* 'We shall ate.PST'. This category was primarily made, because these anomalies were frequent in both corpora, and much variation in the types of confusions were seen (especially in the L2 corpus). Thus, it made the tagging process faster to group them in one large category.

As Lüdeling & Hirschmann (2015) point out, there can be more than one target hypothesis for any given anomalous utterance, and even with the same target hypothesis there can be different descriptions of the anomaly. Anomaly categories e.g. depend on the grammatical model and the research question. In our case, we used a broad definition of grammar (cf. section 2.1, Boye & Harder 2012), which influenced the tagging process: Whether a given anomaly could be related to grammar was focal and thus considered first (for more details, see Søby & Kristensen 2019).

The texts in the L1 corpus were partly tagged by employees in The Danish language Council and partly by researchers in the BGB project. The texts from the 28 learners with L1 English used in the corpus study in Søby & Kristensen (2019) were tagged by Kristensen or me. The remaining texts in the L2 corpus were primarily tagged by student assistants in the BGB project. After the tagging, all texts were proofread by Kristensen, me or another student assistant. As part of the corpus study in Søby & Kristensen (2019) on grammar anomalies in a subset of the L2 corpus, we calculated our interrater reliability score (85 %) based on texts from six learners. Agreement was highest within the categories orthography and morphology and lowest for choice of words (the categories missing word and superfluous word were not used as independent categories in this study).

The anomaly types which I have focused on in this thesis (V3, confusion of infinitive and present tense, and missing gender agreement in NPs) are all relatively simple to tag and categorize. However, especially for the V3 anomalies, cases were found in which much interpretation and repair was needed, or in which the intended word order could not be decided (cf. section 2.2, Article 1). The provisional categorization of all anomalies in the corpus made it easy to extract the relevant anomalies and re-categorize them later. However, it still required a time-consuming sorting process afterwards. This may have been reduced if the provisional categories were based on more theoretically well-founded categories.

Table 3 presents the overall categories of anomalies found in the L1 and L2 corpora. In total, there are more anomalies in the L2 corpus per 1,000 running words. Danish high school students produce most orthographic anomalies (32 % of all anomalies), closely followed by morphological anomalies (28 %). Syntactic anomalies make up the smallest proportion of the categories (5 %). The L2 learners produce most morphological anomalies (35 % of all anomalies), followed by choice of word (24 %) and orthography (20 %). Superfluous words make up the smallest proportion of the categories (5 %), closely followed by syntax (7 %).

Table 3 also show frequencies for four subcategories of grammar anomalies, relevant to this thesis: V3 word order, confusion of finite and non-finite verb forms, missing gender agreement for articles, and anomalous adjective inflections. In Article 2, more details are found concerning frequencies of specific subtypes within these anomaly categories in the L1 corpus, e.g. confusion of infinitive and present tense.

Based on the provisional categories, the most frequent anomalies made by the L2 learners within the morphology category are anomalous use of definiteness (N = 440), number (N = 420), adjective inflections (N = 411), and confusion of finite and non-finite verb forms (N = 351). Other types of anomalies in relation to verb forms are also common, e.g. anomalous use of imperative (N = 102), such as imperative used in a declarative sentence, or use of another form in an imperative clause, e.g. *<u>Svare</u> mig asap* ‘Answer.INF me asap’ for *Svar mig asap* ‘Answer.IMP me asap’. Furthermore, anomalous forms of the main verb when using present perfect or past perfect (N = 73) are common, e.g. *har <u>mistede</u>* ‘have lost’, where *mistede* is a past tense form, compared to *har mistet* ‘have lost’ which is present perfect. Finally, other non-authorized verb forms or anomalous use of tense in a given context is common (N = 255).

Within the syntax category, the most frequent anomalies in the L2 corpus are placement of adverbial (N = 174), followed by V3 word order (N = 123) and anacoluthon (N = 73), i.e. syntactic inconsistency or incoherence within a sentence which results in a breakdown in communication. Finally, overuse of V2 word order is common (N = 61), e.g. after conjunctions, *når <u>har vi</u> pause* for *når vi har pause* ‘when we have a break’ (cf. Article 1).

Table 3. Types of anomalies in the L1 and L2 corpora

	L1 corpus (127,957 words; 71 participants)			**L2 corpus** (37,304 words; 217 participants; A2-B1)		
Category	**Anomalies (N)**	**Per 1,000 words**	**% of all anomalies**	**Anomalies (N)**	**Per 1,000 words**	**% of all anomalies**
Syntax	**168**	**1.31**	**4.82**	**518**	**13.89**	**6.71**
V3 word order	10	0.08	0.29	123	3.30	1.59
Morphology	**960**	**7.50**	**27.55**	**2712**	**72.70**	**35.11**
Verb inflections (confusion of finite and non-finite forms)	194	1.52	5.57	351	9.41	4.54
Articles (gender agreement)	23	0.18	0.66	111	2.98	1.44
Adjective inflection (gender, number, definiteness)	120	0.94	3.44	411	11.02	5.32
Orthography	**1099**	**8.59**	**31.54**	**1574**	**42.19**	**20.38**
Missing word	**227**	**1.77**	**6.52**	**673**	**18.04**	**8.71**
Superfluous word	**369**	**2.88**	**10.59**	**394**	**10.56**	**5.10**
Choice of word	**661**	**5.17**	**18.97**	**1854**	**49.70**	**24.00**
TOTAL	**3484**	**27.23**	**100.00**	**7725**	**207.08**	**100.00**

4 Grammar in the L2 classroom

Article 1 examines V2 and V3 in learners' written production (and mainly learners at DP3), but the study does not provide insights into didactic practices at language schools: How do language instructors prioritize grammar in classes? How do they prioritize various grammatical focus areas, such as V2? Do they correct V3 word order? This section presents preliminary (descriptive) results from an online survey among language instructors in Danish as second language focusing on the practices and attitudes of language instructors across the three Danish Programs, among others in relation to V2 in written and oral production (Gosselke Berthelsen & Søby, in prep.). The survey was made in collaboration with Sabine Gosselke Berthelsen and consisted of three parts, of which the first and third are presented in section 4.1.

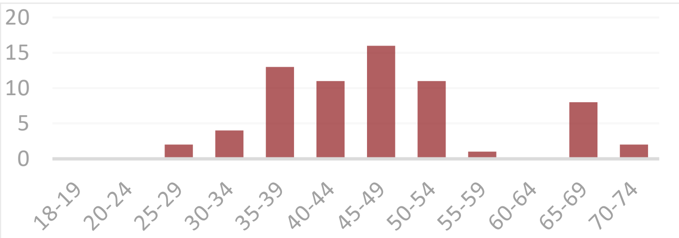
The first part examined language instructors' prioritization of pronunciation and grammar in the L2 classroom. The second and third parts examined instructors' practices and attitudes in relation to various topics within pronunciation and grammar, respectively. The second part was constructed by Gosselke Berthelsen to be used in research related to second language pronunciation and will not be elaborated here. The third part was constructed by me and is presented below.

Among other topics, the language instructors were asked about which areas of grammar their students struggle with, and so the survey informs about language instructors' perception of which anomalies are frequent in L2 production. The instructors teach at all three Danish Programmes, unlike Article 1, which is mainly based on data from DP3. Also, the instructors come from different language schools, whereas the data in Article 1 is from the same school. The data from the survey cannot be directly linked to the L2 corpus, as it comes from other instructors. In hindsight, it was also relevant to collect information directly from the instructors at Københavns Sprogcenter to supplement the learner texts. However, it was not considered at the time. Also, in retrospect, it would have been relevant to know how the teachers at Københavns Sprogcenter and the language instructors in the online survey introduce Danish word order to their students (as SV with exceptions or as V2) (cf. Article 1).

4.1 Survey among language instructors

The online survey was conducted in SurveyXact. It was primarily distributed in forums on Facebook for instructors in Danish as a second language, and through my contacts working at language schools (LinkedIn, Facebook). Gosselke Berthelsen also contacted 86 language schools from all over the country and asked if they could distribute the advertisement for the study to their instructors (link and QR code). The survey took around 10-15 minutes to complete. Participants who filled out the survey could participate in a draw for 20 chocolate gift boxes (each 350 DKK). The participants were 76 language instructors in Danish as second language with experience from the official Danish language programs, of which 10 did not complete the full survey and were excluded. Therefore, data from 66 participants were included in the analysis. Although some instructors teach in more than one of the three Danish programs, we asked them to fill out the survey based on the program they were primarily associated with. Eight instructors from DP1 participated, 28 from DP2, and 30 from DP3 (56 female, 9 male, 1 other). Thus, the distribution on Danish Programs is different than in the L2 corpus and Article 1, where 207 learners (95 %) are from DP3 and the remaining 10 from DP2. The age distribution of the instructors is seen in Figure 2. Most participants were between 35-55 years.

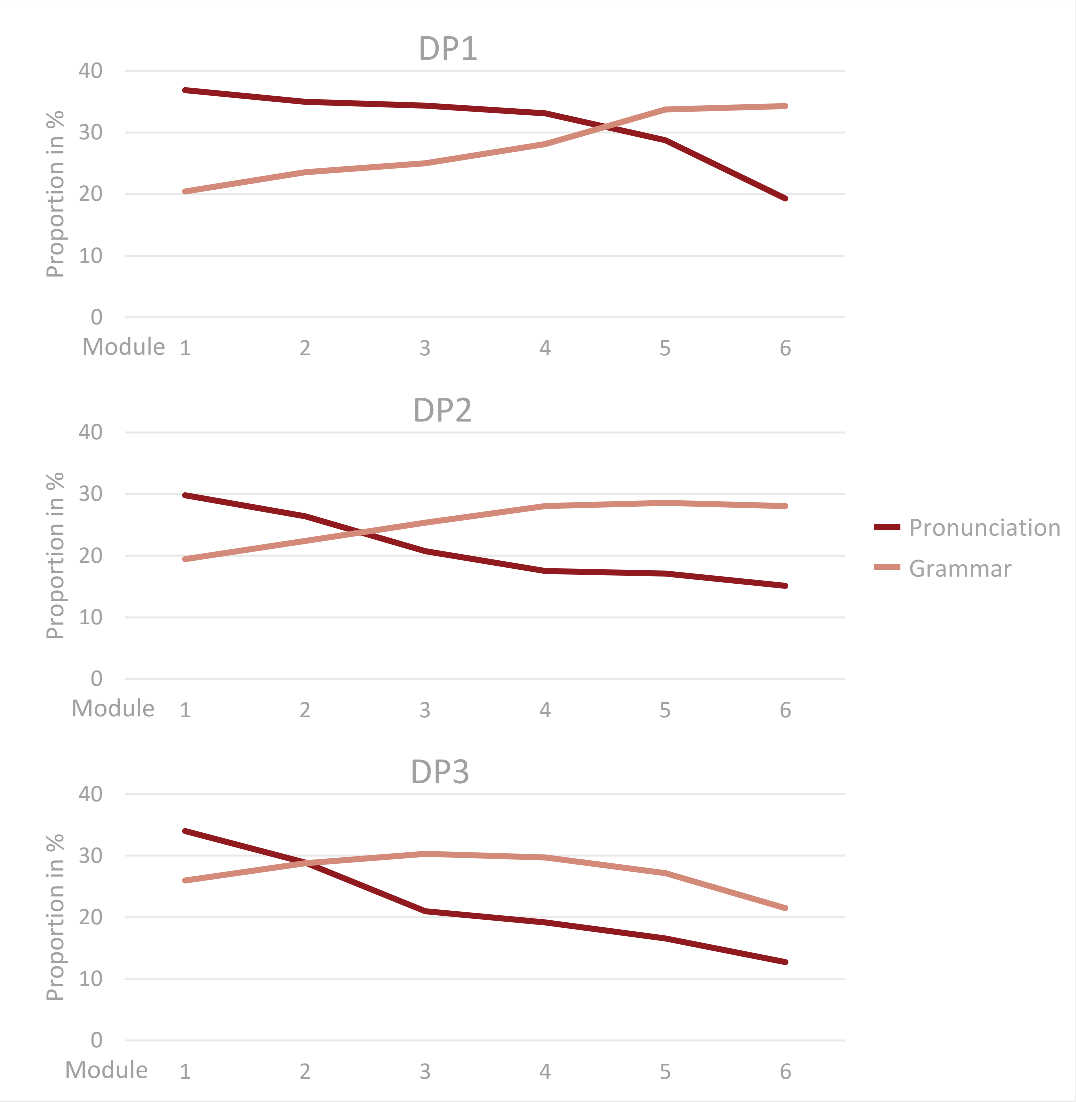
Figure 2. Language instructors' age distribution (the x-axis shows the number of instructors)



Sixty-four instructors had Danish as their L1, of which five indicated that they had an additional L1 as well. Two instructors were L2 users of Danish. The distribution of participants' age, gender and language background was similar across programs (Gosselke Berthelsen & Søby, in prep.).

We asked participants to give an estimate of the percentage of time spent on pronunciation and grammar (in the class room) in the different modules. First, they provided the percentages for pronunciation for modules 1-6. Second, they provided the percentages for grammar. We stressed that it was an abstract task and that we knew that they could only provide a rough estimate (or guess if they did not have experience with one of the modules). The results in Figure 3 show that instructors estimate to spend more time on pronunciation than grammar initially, but later, the focus on grammar increases, while the focus on pronunciation decreases, and eventually more time is spent on grammar (according to instructors). The shifting point depends on the DP. For DP3, it is around module 2, for DP2, between modules 2 and 3, and for DP1, between modules 4-5. The L2 corpus (Article 1) primarily contains data from DP3 modules 2-4. According to the survey, instructors in these modules estimate that they spend 26-30 % of classroom time on grammar.

Figure 3. Language instructors' estimated time (in %) spent on pronunciation and grammar in classes



To quantify time spent on activities which are often intertwined with each other, or other activities, is likely to feel artificial or abstract. We also did not define the concepts of pronunciation and grammar in the instructions, leaving room for individual interpretation of what constitutes teaching in the two areas.

Participants were asked how content they were with the distribution of pronunciation and grammar in classes. They answered on a slider from 1-100, where

0 was *not content at all* and 100 was *exceptionally content*. The average was 75 (range: 27-100; *SD* = 18), indicating that instructors generally were content with the distribution of pronunciation and grammar in classes. This was followed by an optional open question: "*If you could choose freely, is there anything you would prioritize differently (e.g. something you would like to spend more/less time on)?*" Thirty-four instructors answered the question. Six instructors mentioned that there is not enough time in general, and seven remarked that preparing for specific tasks in the module tests take up too much time (e.g. cloze tasks). Fourteen would like to spend more time on pronunciation and six on oral language skills. One mentions writing skills and one mentions syntax. In spite of both being primed to think about pronunciation and grammar by the scope of the questionnaire, participants predominantly mention pronunciation here, and very few mention grammar.

After having examined how much time instructors estimate they spend on grammar in general, my aim was to examine how they prioritize V2 after non-initial subjects compared to other areas of grammar. In the part of the survey focusing on grammar, participants were asked about grammar in written production and oral production separately. The first question was: "*Which areas within grammar and spelling do you think are important for your students to master in written production when they are about to finish their Danish Program?*" Below the question were eight focus areas, each with a slider from 0-100 (0 = *not important at all*, 100 = *extremely important*). They were all presented at once, so that participants could rank them relatively to each other. The focus areas were chosen because they represent a broad range of anomalies which are typically found in L2 Danish, with V3 word order being of particular interest. I also wanted to include areas which I suspected were not seen as very important (e.g. articles and spelling), to check whether participants varied their answers. They were presented with correct examples as seen below. Notice that the terms *inversion* and *reversed word order* were used instead of V2, as these are common in teaching materials.

1. Verb inflections (e.g. *at spise* 'to eat', *han spiser* 'he eats', *han spiste* 'he ate', *han har spist* 'he has eaten')
2. Correct use of articles concerning gender (fx *en bil* 'a car' (uter), *et hus* 'a house' (neuter))
3. Adjective inflections (fx *en gul bil* 'a yellow car' (uter), *to gule biler* 'two yellow cars', *et gult hus* 'a yellow house' (neuter))
4. Placement of adverbials in subordinate clauses (e.g. *Han fortæller, at han altid køber ind om onsdagen* 'He says that he always goes shopping on Wednesdays')
5. Inversion/reversed word order (V-S) in questions (e.g. *Arbejder du med IT?* 'Do you work in IT')
6. Inversion/reversed word order in declarative clauses (e.g. *I weekenden dyrker han sport* 'In the weekend, he exercises')

7. Correct use of prepositions (e.g. *om onsdagen* 'on Wednesdays', *på skolen* 'at school', *god til dansk* 'good at Danish')
8. Spelling (e.g. to write compounds in one word: *byvandring* 'city walk', remember double consonants: *butikken* 'the shop', remember letters and syllables which are not (always) pronounced: *sjældent* 'rarely', *virkelig* 'really')

The question was repeated for oral language – except from number eight (spelling). The results are seen in Figure 4 and 5 below.

Figure 4. Average of instructors' ratings regarding the importance of focus areas in WRITTEN production (0-100; 0 = *not important at all*, 100 = *extremely important*)

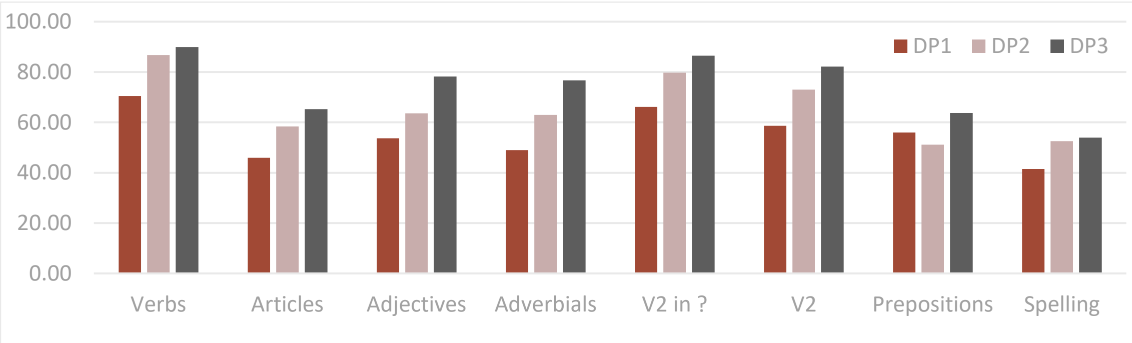
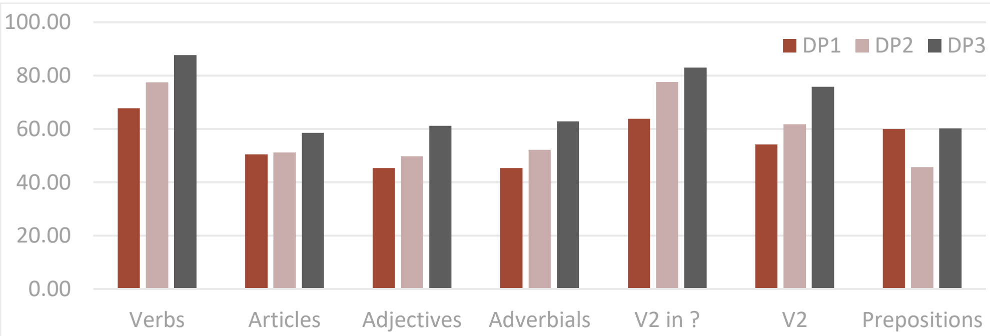


Figure 5. Average of instructors' ratings regarding the importance of focus areas in ORAL production (0-100; 0 = *not important at all*, 100 = *extremely important*)



Generally, the importance of all focus areas are rated higher by instructors in DP3 than DP2, who again rate most areas higher than instructors in DP1. This is not surprising, considering the aims of the three different programs. In written production, it seems that correct verb inflections are considered most important, followed by V2 word order in questions and declaratives (with non-initial subjects). Both anomalous verb inflections and V3 anomalies are included as stimuli in the error detection study in Article 2. Least important in written production is spelling (misspellings are also examined in Article 2). In oral production, the averages are generally lower than in written production. Verb inflections have the highest ranking at DP1 and DP3, while V2 in questions are marginally higher at DP2. As in written production, V2 in questions is rated higher than V2 in declarative clauses, which mostly is in the top three (apart from DP1 where prepositions are ranked higher). In oral production, it depends on the DP which areas have the lowest rankings: At DP1 it is adjectives and adverbials, at DP2 it is prepositions, and at DP3 it is articles. However, both articles, adjectives, placement of adverbials and prepositions are low in ranking, compared to verb inflections.

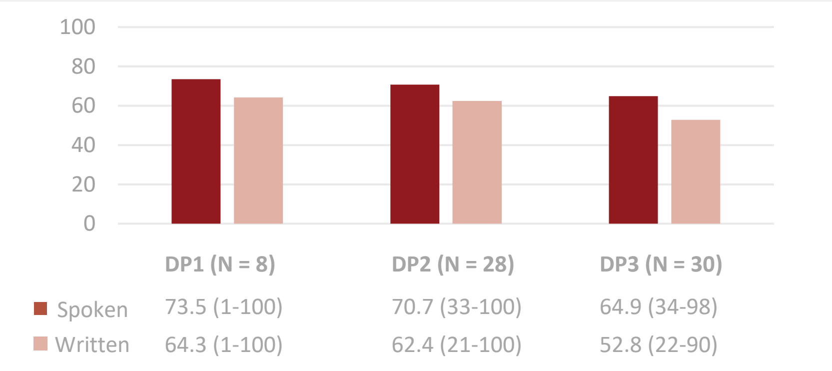
Based on the survey, it seems that DP3 instructors generally find V2 in declarative sentences important for their students to master. It may, although, be that the instructors who are willing to fill out a questionnaire about grammar (and pronunciation), are also the ones who are particularly interested in grammar and therefore finds it important.

In order to examine which areas of grammar the instructors considered as difficult, they were asked the following (obligatory) open question: *Which areas of Danish grammar do you experience that your students generally struggle with, or ask for more instructions in?* Four instructors answered "all the classic things" or "everything this survey is about". Most instructors mentioned word order or syntax (21), followed by use of prepositions (20), verb inflections (19), placement of adverbials (14), inversion (14), adjective inflection (10) and distinguishing between definite and indefinite forms (10). Only two mention the use of the indefinite articles *en/et* 'a'. In total, 34 mention word order or inversion, plus the four who mentioned everything, corresponding to 58 %. Three instructors stated that it differs, depending on the language background of the learner. Two instructors from DP1 pointed out that their students do not really ask about grammar – they want "vocabulary and speech acts". Many of the areas mentioned by the instructors are related to syntax, although morphological anomalies are more frequent in the L2 corpus than syntactic anomalies; there are 72.70 morphological anomalies per 1,000 running words vs. 13.89 syntactic anomalies (cf. section 3.3). All of the anomalies mentioned in the survey are common in the L2 corpus: e.g. use of definiteness, number, adjective inflections, and anomalous verb forms. Within the syntax category, the most frequent anomalies in the L2 corpus were placement of adverbials and V3 word order.

When comparing what instructors at DP3 prioritize and what instructors (in general) consider as difficult for students, "word order"/"inversion" and verb inflections are both mentioned. Use of prepositions is also often mentioned as something students struggle with, but this is considered less important to master. Adjective inflection is not among the most mentioned areas which students struggle with, but is almost prioritized as high as V2 in declarative clauses at DP3. Finally, use of the indefinite articles was not often mentioned as difficult for students (only by two instructors), and is also not highly prioritized.

Previous studies of V2 and V3 word order in L2 production often describe V2 after non-initial subjects as difficult for learners to acquire (e.g. Håkansson 1988, Bolander 1990, Hagen 1992). One of the aims with the survey was to examine language instructors' point of view concerning the difficulty of acquiring V2. Hence, the instructors were asked specifically about difficulties with inversion: "*How easy or difficult do you experience that it generally is for your students to use inversion of subject and verb (reversed word order) in declarative clauses such as "I weekenden dyrker han sport" 'In the weekend he exercises'?*" The answers were marked with a slider from 0 (= *very easy*) to 100 (= *very difficult*), for oral and written production, respectively. Figure 6 shows the mean values. Instructors generally find that students struggle more in spoken language than written, and teachers in DP1 rate difficulty higher than teachers in DP2, which is again higher than for DP3. The mean for DP3 is close to 50 for written production, which must be interpreted as it neither being considered particularly difficult or easy.

Figure 6. Language instructors' mean ratings of students' general difficulty with using V2 in declarative clauses with non-initial subjects on a scale from 0-100 (0 = *very easy*; 100 = *very difficult*)

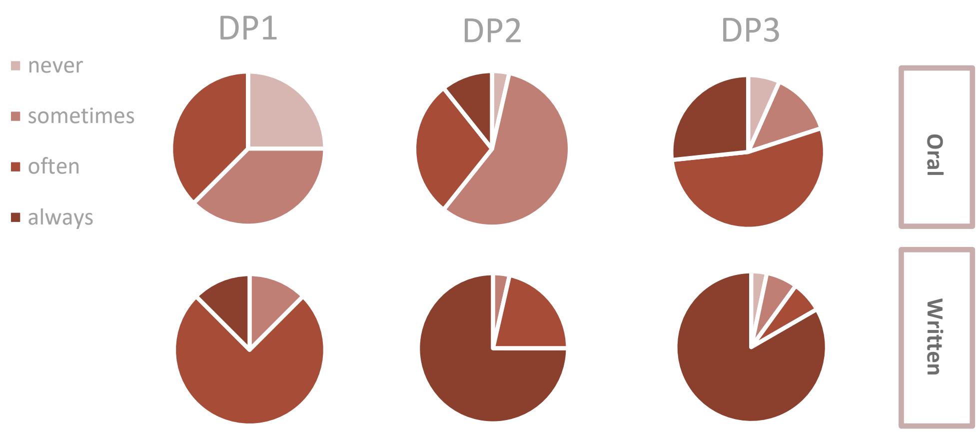


Retrospectively, knowing the results from Article 1 about the role of learners' language background, it may have been better to ask the instructors about non-V2 learners and V2 learners separately, in case the instructors thought that it depended upon the language background of the learner. This was explored in an open question afterwards: "*Do you experience that some students find it easier than others to use inversion/reversed word order in declarative clauses than others? For example, in relation to L1 or general progression?*" Here, 41 (62 %) of the instructors answered yes (15 answered no, and 10 did not know or did not answer the question). Nine answered yes without elaborating, 15 mentioned language background or gave examples of L1 groups who either found it particularly easy (especially learners with L1 German or Dutch) or difficult. Furthermore, instructors mentioned that students with a fast progression in general or long educational background (7) and students with high linguistic or grammatical awareness (6) find it easier to use V2 after non-initial subjects. Due to the way the question was formulated (yes/no), and with language background explicitly mentioned as a suggestion, it is difficult to conclude if it is a general impression among the language instructors that language background plays a role. This would have been relevant to know in relation to the applications of

Article 1, i.e. whether it is new information for the instructors that learners' language background play a role in V2 production.

On the final page of the survey, instructors were asked whether they correct students if they do not have inversion of subject and verb in declarative clauses (such as **I weekenden* <u>*han dyrker*</u> *sport*). The answers were given by choosing between four buttons (*almost never, sometimes, often, almost always*) for written and oral production, respectively. The answers are seen in Figure 7.

Figure 7. How often instructors (N) correct V3 in oral and written production



At DP2 and DP3, the majority of instructors always correct V3 in writing, in DP1 most teachers often correct it in writing. In oral production, V3 is corrected less than in written production. Still, in DP3, 27 % of instructors always correct it, or often do so (53 %). Again, these results cannot be directly linked to the L2 corpus, but they still provide insights into what seems to be a common practice at DP2 and DP3, i.e. that V3 is corrected in writing.

5 Perception of grammar anomalies

Many potential contributing factors may influence if and how a reader reacts to a grammar anomaly when reading (Article 2). Both of the perception studies in this thesis focus on the role of the frequency of the anomaly in L1 production (in Article 2 measured as *error rates*). The main hypothesis is that anomaly frequency, being tied to the predictability of the anomaly, can predict perception patterns, so that infrequent, and thus unexpected, anomalies attract more attention than frequent anomalies. This hypothesis is inspired by prediction-based approaches to sentence processing (e.g. Kamide 2008, Christiansen & Chater 2016), but the notion *prediction* is not further defined in the articles. Therefore, section 5.1 presents a definition of prediction by Kuperberg & Jaeger (2016) and it describes how prediction can be seen as related to anomaly processing.

Section 5.2 presents a tentative model of possible factors involved in anomaly detection. It is not a complete model, but it provides an overview of factors which could be relevant to examine or control for in future studies. Some of these factors are examined in Article 2 and Article 3, in addition to the role of anomaly frequency.

5.1 Prediction in language processing

According to Christiansen & Chater (2016), language processing is now-or-never, as speakers quickly lose track of the rapidly presented input in a conversation. The so-called *Now-or-Never bottleneck* is a fundamental constraint on language, and in order to deal with it, the brain has to compress and recode the linguistic input as fast as possible. Christiansen & Chater (2016) argue that an implication of the Now-or-Never bottleneck for language processing is that the use of prediction in language processing is essential. The view that prediction is a crucial concept is supported by other accounts of language processing (e.g. Kamide 2008, Kuperberg & Jaeger 2016, Levy 2008). Section 5.1.1 further elaborates on how the term prediction has been used and can be defined (Kuperberg & Jaeger 2016). Section 5.1.2 discusses how prediction theory can be used in relation to anomaly processing.

5.1.1 What is prediction?

Simply put, prediction implies that the context influences the state of the language processing system before the new input becomes available, and this facilitates the processing of the new input (Kuperberg & Jaeger 2016:4). Most models of syntactic parsing and lexico-semantic processing agree that the language processing system is predictive in this sense, i.e. that comprehenders anticipate "*some* structure or *some* semantic information" before the new bottom-up input is observed (Kuperberg & Jaeger 2016:2). An example of prediction in relation to the study in Article 2 could be that when participants read the indefinite article *en* 'a', which is uter, they predict that the upcoming word (adjective or noun) should also be uter.

Based on a review of studies examining prediction in language processing, Kuperberg and Jager (2016) conclude that "different subfields and different researchers have critically different conceptions of what it means to predict during language comprehension" (Kuperberg and Jager 2016:3). Also, the terms *expectation*, *anticipation*, and *prediction* are sometimes used with the same meaning and sometimes with slightly different meanings. The term *prediction* has become so loaded and ambiguous that some researchers are hesitant to use it, or even reject that it plays a part in language processing (Kuperberg & Jaeger 2016:3). In the following, it is described what most researchers agree upon when defining prediction, which is reflected in the definition of prediction is its minimal sense by Kuperberg & Jaeger (2016). Some of the controversies on what prediction entails are presented here as well (e.g. with respect to the scope of the prediction). In the thesis, I use the terms *prediction* and *expectation* interchangeably, as well as the verbs *predict* and *expect*.

According to Kuperberg & Jaeger (2016:5) most recent accounts view prediction as a graded and probabilistic phenomenon, meaning that some predictions are more likely than other and thus differ in strength. Such accounts are based on evidence of graded effects of context on processing, e.g. seen in garden path studies where the magnitude of the garden path effects depends on how much the context biases against the intended syntactic parse (e.g. Spivey-Knowlton et al. 1993 in Kuperberg & Jaeger 2016). Other evidence is e.g. found in ERP studies, where the magnitude of the N400, which arises in response to an incoming word, depends on the word's probability in that context (cloze probability) (e.g. DeLong et al. 2005 in Kuperberg & Jaeger 2016).

It remains debated whether prediction (especially in sentence parsing) takes place in a serial or parallel fashion, i.e. whether just one upcoming structure is predicted at any given time, or whether the parser can compute multiple syntactic parses in parallel, of which each has a certain degree of probabilistic support (Kuperberg & Jaeger 2016:5). If the new bottom-up input does not match the expected structure, the parser in a serial model reanalyzes and moves on to the next possibility. In a parallel model, the predicted parses are shifted or reweighted if the new bottom-up input is inconsistent with them. As highlighted by the authors, it is often difficult to experimentally distinguish between serial and parallel probabilistic prediction (and this is also not a theme in Article 2 or Article 3). However, Kuperberg & Jaeger (2016) argue that insights from computational probabilistic frameworks (e.g. Levy 2008) are "consistent with the idea that we can predictively compute multiple candidates in parallel, each with different strengths or degrees of belief" (Kuperberg & Jaeger 2016:6):

> "[…] the way that a rational comprehender can maximize the probability of accurately recognizing new linguistic input is to use all her stored probabilistic knowledge, in combination with the preceding context, to process this input. The reason for this is that we communicate in noisy and uncertain environments – there is always uncertainty about the bottom-up input, and neural processing itself is noisy […]" (Kuperberg & Jaeger 2016:6f).

I will return to this idea that comprehenders use previous experience with language and probabilistic knowledge when new input is being processed, in the next section on how prediction theory can be linked to anomaly processing.

Based on the view that prediction is a graded and probabilistic phenomenon, Kuperberg & Jaeger (2016) presents a definition of prediction in its minimal sense, meaning that this should be uncontroversial:

> "[I]n its minimal sense, prediction implies that, at any given time, we use high level information within our representation of context to probabilistically infer (hypothesize) upcoming information at this same higher level representation." (Kuperberg & Jaeger 2016:25).

An example of prediction in this minimal sense (i.e. on the same level of representation) is that the word *fly* can predict the words *kite* or *plane*. Semantic prediction on word level is well-researched, but prediction is not restricted to this according to Kuperberg & Jaeger (2016). A comprehender can use information in a context to facilitate processing of new input at multiple levels of representation. These levels range from syntactic, semantic to phonological, orthographic and perceptual. The perceptual level is not clearly defined in Kuperberg & Jaeger (2016), but I interpret it as the initial visual pre-linguistic processing level. Furthermore, comprehenders can draw upon multiple different types of information within their internal representation of context to facilitate processing (Kuperberg & Jaeger 2016:9f).

Moreover, Kuperberg & Jaeger (2016) conclude that:

> "[U]nder some circumstances, facilitation at lower level representations results from the use of higher level inferences to predictively pre-activate information at these lower level(s), ahead of new bottom-up information reaching these levels." Kuperberg & Jaeger (2016:25).

It is relatively uncontroversial that higher level lexical information can be used to predictively pre-activate upcoming potential phonemes. Within the field of sentence processing, it remains more controversial whether higher level information can be used to predictively pre-activate upcoming information on lower levels of representation (Kuperberg & Jaeger 2016:13). Evidence for predictive pre-activation is seen in EEG studies measuring effects of grammatical elements which are dependent on a subsequent predicted lexical element. For example, DeLong et al. (2005) found that participants were more surprised to see the article *an* than the article *a* in the sentence in (6), because they expected the word *kite*:

(6) *The day was breezy so the boy went outside to fly a/an …*

This shows that higher level predictions (about the upcoming word) affects lower levels of processing (of the phonology or orthography of the article), and that the context can predictively pre-activate semantic, but also upcoming phonological and orthographic information (Kuperberg & Jaeger 2016), or perhaps grammatical information, as the difference between *a*/*an* can be seen as related to grammar. Other evidence of predictive pre-activation has been found using the visual world paradigm where participants listen to auditive input and see images, while their eye movements are recorded. Studies show that if the linguistic context constrains towards semantic, syntactic or phonological properties of the upcoming word, participants tend to look at the image related to the predicted word, even before the target word is spoken (Kamide 2008, Kuperberg & Jaeger 2016). The eye-tracking study in Article 3 does not inform on predictive pre-activation, as the design is different. The whole sentence is presented at once, including the anomaly, and although eye movements are recorded on the words preceding the unexpected subject (i.e. the sentence-initial adverbial), no effects of ungrammaticality are expected before the target word is fixated. Thus, we cannot study predictions before the anomalous target word, only during processing of the target word and after.

Finally, Kuperberg & Jaeger (2016) emphasise that just because comprehenders are able to use information in a context to pre-activate multiple types of information, it does not mean than they do so in all situations:

> “[T]he degree and level of predictive pre-activation might be a function of the expected utility of prediction, which, in turn, may depend on comprehenders’ goals and their estimates of the relative reliability of their prior knowledge and the bottom-up input.” (Kuperberg & Jaeger 2016:1).

For the study in Article 3, this means that it may be that not all participants use prediction during the experiment, and those who do may not use it all the time. The bottom-up input, i.e. the sentences which were full of various grammatical anomalies, may also affect how usable participants find predictions based on prior knowledge in the experiment. Although they all carry out the same task, they may also differ in their goals.

5.1.2 Using prediction theory on anomaly processing

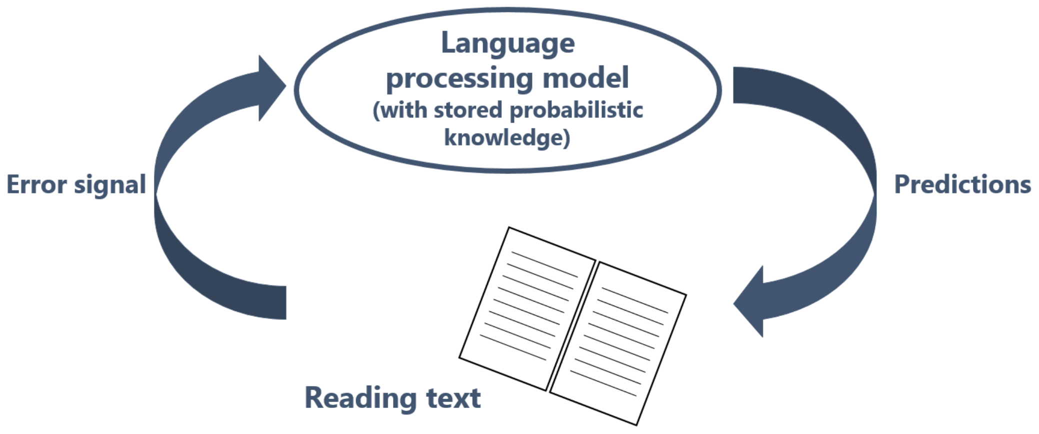
Most empirical work on predictive pre-activation has focused on lexical constraints (Kuperberg & Jaeger 2016) and effects of semantically unexpected items. This thesis focuses on grammar anomalies, and thus the theoretical foundation for the perception studies in Article 2 and Article 3 is far from comprehensive. This section presents a first step in trying to relate prediction theory to anomaly processing and to the concrete experimental stimuli in Article 2 and Article 3. It should, however, be noted that the studies in the thesis cannot directly measure prediction, as they neither measure neural effects or use an eye-tracking paradigm like the visual world paradigm. Alternatively, the increased processing times in Article 3, e.g. fixation durations, could be seen as reflecting difficulties with integration of words into the preceding context, AFTER the word has been presented, i.e. the word was not predicted by the context, but is nevertheless difficult to integrate with these words (cf. Kamide 2008 on integration interpretation). This is for instance described as integration failure in the E-Z Reader model of eye movement control in reading (Reichle et al. 2009), which is described in Article 2.

When linking processing of grammar anomalies to prediction theory, the idea is 1) that similarly to semantically unexpected items, grammar anomalies can be unexpected and thus attract attention, and 2) that based on a language user’s prior experience with language, some anomalies must be more unexpected than others to that language user.

Figure 8 is a simple illustration of how a language user’s predications can be updated after being presented to input (with or without language anomalies). The individual language user has a language processing model with stored probabilistic knowledge, which makes predictions about the upcoming input, e.g. when reading a text, as in Figure 8. If something in the text does not meet the predictions, e.g. if a word is semantically odd or appears in an anomalous position (V3), a *prediction error* occurs. Kuperberg & Jaeger (2016) use this term to describe “the difference between the comprehender's predictions at a given level of representation before and after encountering new input at that level of representation" (Kuperberg & Jaeger 2016:7).

Many different types of language anomalies (orthographic, grammatical, semantic) are assumed to potentially cause prediction errors, and these prediction errors may vary in strength. After the language user has been presented to the input, feedback about the input in the form of an *error signal* is sent back to the language processing model. The error signal informs the model on the degree to which its predictions were met, and the model is updated accordingly, i.e. the probabilities are updated. The model is always updated based on input, also in cases where no prediction errors occur. If a language user frequently sees a specific type of anomaly, e.g. infinitive for present tense **han <u>køre</u>* 'he drive.INF' for *han kører* 'he drives.PRS', we assume that his or her language processing model will be updated according to the input. It may be that the first time a language user sees an anomaly, it causes a large prediction error, but after the model has been updated (and predicts the anomaly with a certain probability), the prediction error decreases next time the anomaly is seen. If the anomaly is seen often, it may become the norm and not even cause prediction errors in the end. Unexpected input generally attracts more attention than expected input (Christensen 2021, Article 2). Thus, the more expected an anomaly is, the less attention it should attract. In other words, attention to a specific type of grammar anomaly may not only be a matter of a language user's explicit grammar awareness (i.e. whether the language user is aware of a specific grammatical rule), but also of whether the specific type of anomaly is common in the input that the language user has seen.

Figure 8. Updating of a language user's predictions based on input (with or without language anomalies)



An important question then is how individual predictions in a specific context can be measured? Studies of lexical constraints have often used the cloze probability of a word, i.e. the probability of participants using that specific word in a sentence completion task, reflecting how predictable the word is in a specific context. This measurement does not necessarily reflect what an individual predicts on a given trial, because cloze probabilities are averaged across participants (Kuperberg & Jaeger 2016). Using cloze probabilities as a measurement builds on a general assumption that language users' probabilistic predictions resemble the actual statistics of their linguistic environment (Kuperberg & Jaeger 2016).

As part of the error detection study in Article 2, participants completed a grammar quiz, in which the forced-choice grammar tasks to some extent resembled a cloze-test, but future studies with actual large-scale cloze tests of probabilities of producing various types of anomalies are needed. The scores of the grammar quiz were assumed to reflect participants' explicit grammatical awareness, and we used their total scores in the analyses. Due to ceiling effects in the quiz, we could not distinguish between probabilities for the three different subtypes of anomalies, but used the individual participants' total grammar scores as an indicator of general grammatical awareness. Instead, we used *error rates* in high school essays for the different anomaly subtypes as a measurement of frequency, and thus predictability. Error rates were calculated by dividing the number of incorrect tokens with the number of correct and incorrect tokens in the essays (for more details, see Article 2, or Søby & Kristensen (2019) on *potential occasion analysis*). This measurement was chosen over raw error frequencies, as it reflects how often a given anomaly is seen in proportion to all the cases where it could potentially have been produced. Two types of anomalies may have similar (high) raw frequencies, but if one of these anomalies is rarely seen in its correct form, the error rates for the two would differ. It is, however, difficult to say which measurement is the better, only based on speculations, and more empirical studies are needed.

Like cloze probabilities, error rates in a corpus do also not necessarily reflect the individual language user's predictions, but builds on the previously mentioned assumption that language users' probabilistic predictions resemble the actual statistics of their linguistic environment (Kuperberg & Jaeger 2016). This is seen in Article 2, where we argue that we do not assume that high school students read each other's essays, but that the anomalies that they produce in school essays are likely found in their writing in general, including informal texts directed at their peers. Additionally, we assume that the anomaly production patterns which are found in the high school texts to a large extent reflect the anomaly types found in the media and society in general. This may be a simplistic picture, as the statistics constantly change in real-life communication, as "every person we converse with will have their own unique style, accent and sets of syntactic and lexical preferences" (Kuperberg & Jaeger 2016:27). Similarly, readers are exposed to different statistical structures in the linguistic input, depending on the type of text which is being read.

To sum up, based on the idea that prediction is graded and probabilistic in nature (Kuperberg & Jaeger 2016), I assume that some types of anomalies are more expected than others, because anomalies are not equally frequent in L1 production and thus presumably differently graded in the individual models that language users use to predict upcoming language input (the studies in Article 2 and Article 3 do, however, not directly measure prediction). Article 2 examines attention to various types of grammatical and orthographic anomalies in an error detection study, and Article 3 examines online processing of V3 using eye-tracking. In the following, I present some preliminary thoughts on how the different types of anomalies in the experiments deviate from what is expected. Besides the syntactic V3 anomalies, the stimuli in Article 2 (as previously mentioned) consists of two types of morphological anomalies. One is confusion of present tense and infinitive (e.g. **han <u>køre</u>* 'he drive.INF' for *han kører* 'he drive.PRS' and **han vil <u>kører</u>* 'he will drive.PRS' for *han vil køre* 'he will drive.INF'). The other type is missing gender agreement between indefinite articles or adjectives and the noun in NPs consisting of article + adjective + noun (examples are provided below). Finally, various common orthographic anomalies were included as well.

I assume that expectations are broken on different representational levels for the different types of anomalies. For the V3 anomalies, the anomalous word (the unexpected subject) is orthographically well formed, but not syntactically (i.e. a higher representational level). For the various orthographic anomalies, expectations may already be broken at the early perceptual level, as the combination of letters is unusual, or otherwise subsequently on the orthographic level. The two types of morphological anomalies (involving verbs, articles and adjectives) can be considered well formed when considered individually, but not in their contexts. For example, the indefinite article *en* 'a' (uter gender) is a well formed word, but not in combination with the neuter noun *hus* 'house'. Thus, it is less clear which expectations are in fact broken for the morphological anomalies: expectations on the perceptual or orthographic level, because a letter is missing, or on a higher level, because the word must agree with other words.

When Danish readers see a sentence introduced with an adverbial like *kl. 14* 'at 2 o'clock', it could either be followed by a verb (e.g. *ankommer* 'arrives') or an extension of the adverbial phrase, as in *kl. 14 om onsdagen* 'at 2 o'clock on Wednesdays'. When instead a subject like *han* 'he' appears as in (7), a basic syntactic expectation is broken and there is a prediction error. This syntactic expectation could be of a more general character (what are the probabilities of a verb appearing vs. other constituents), or more specific (what are the probabilities of the specific verb *ankommer* compared to other verbs (or other words)). The error signal provides feedback to the model, which is updated, so that the probability of seeing V3 is increased. The language user likely has previous experience with V3 being rare in L1 production, and thus the probability of seeing it again is assumed only to increase a little. However, the more sentences the participant sees with V3 anomalies, the more probable they may become according to the language processing model in the specific context (the experiment), and so the prediction errors may also decrease during the experiment.

(7) **kl.* *14 han ankommer til Berlin*
o'clock 2 he arrive.PRS in Berlin
'and at 2 o'clock, he arrives in Berlin'

It is not as clear how predictions may be in relation to the verb forms in the experiment, but it seems plausible that readers predict a non-finite verb form after being presented to a modal verb, as in *han vil køre* 'he will drive'. The finite verb does not have to appear right after, as other constituents may appear in between the finite and non-finite verb, such as the negation *ikke* ('not') and the adverbial *alligevel* ('anyway') in *han vil alligevel ikke køre* 'he doesn't want to drive anyway'. When the sentence uses present tense for infinitive, as in **han vil kører* ('he will drive.PRS'), the unexpected present tense form may there cause a surprisal effect and a prediction error. For the condition infinitive for present tense, the readers may expect a finite verb after the subject, e.g. after reading a sentence with an initial *han* ('he'). Again, the subject may be extended (e.g. *han og Peter* ... 'he and Peter'), so it is not given that the finite verb is the immediate next word after *han*, but it is probable. Seeing a non-finite verb form may thus cause a prediction error.

As discussed in relation to the syntactic anomalies, it remains unanswered how specific the predictions for the verb forms are: Are the predictions updated for specific verbs, e.g. the probability of seing **vil kører* 'will drive.PRS' again, or do language users have a more general prediction of the probability of seeing *vil* 'will' + PRESENT TENSE (or "modal verb + PRESENT TENSE)?

Unlike the V3 anomalies, both types of verb anomalies are common in L1 production (especially infinitive for present tense), and so the probabilities of seeing these anomalies (according to the participants' models) are likely higher than for the V3 anomalies pre-experimentally and thus the prediction errors may be smaller in general. Once more, it is assumed that the error signal provides feedback to the model when prediction errors occur in response to the verb anomalies, after which the model is updated, and so the prediction errors in response to the verb anomalies may be decreased during the course of the experiment.

The NP anomalies in which the indefinite articles *en* or *et* do not match the upcoming adjective and noun (e.g. **en dejligt kæledyr* 'a lovely pet' or **et dejlig undulat* 'a lovely budgie') are different from the other anomalies, because it cannot be immediately decided at the point of reading the article that the use of the article is anomalous. The gender anomaly cannot be detected before the reader has read the mismatching adjective (or mismatching noun). Using either article creates an expectation to the gender of the upcoming noun (and the inflection of the adjectives in the NP, if any). A prediction error may occur when the adjective is processed, because the form of the adjective does not agree with the article. Or it may become clear to the reader at this point that *either* the article or the adjective must be anomalous, since the upcoming noun can only have one grammatical gender. Presumably, the same is the case in the conditions where the adjective does not match the article and the noun: **et dejlig kæledyr* 'a lovely pet' and **en dejligt undulat* 'a lovely budgie'). When the noun has been processed, the anomaly can be located. Perhaps, the prediction in relation to the NP anomalies is of a more general kind, i.e. that readers have a general expectation of constituents agreeing, e.g. in relation to gender. This prediction is potentially broken when the adjectives are seen, but it cannot be decided where the anomaly is, until after the noun has been processed.

Finally, the study in Article 2 included four types of orthographic anomalies, which were created based on types of misspellings which others have found to be common in L1 writing (e.g. Blom et al. 2017). These were: missing double consonants, e.g. **startskudet* for *startskuddet* 'the starting signal', split compounds, e.g. **by vandring* for *byvandring* 'city walk', missing silent letters, e.g. **siste* [ˈsisd̥ə]/[ˈsisd̥] for *sidste* [ˈsisd̥ə]/[ˈsisd̥] 'last', and reduction of syllables, e.g. **virklig* [ˈʋiɐ̯ɡ̊li] for *virkelig* [ˈʋiɐ̯ɡ̊li] 'really'. Generally, language users expect that a word is spelled in a certain way and thus has the same visual appearance, and when the word looks different, as the examples above, expectations may already be broken at the early perceptual level, or otherwise on the orthographic level (Kuperberg & Jaeger 2016). The orthographic anomalies are thus different than the other types of anomalies in the experiment, as expectations may be broken at lower representational levels. It is unclear whether language users predict types of orthographic anomalies (e.g. split compounds in general) or specific examples of split compounds, e.g. **by vandring* 'city walk'.

The speculations presented above about how prediction theory can be used in practice when examining other predictions than semantic ones, illustrate that it is complicated. Predictions are a multifaceted phenomenon which can be made on many different representational levels, as discussed when comparing the syntactic, morphological and orthographic anomalies.

Although the predictability of a grammar anomaly may well play a part for the attention allocated to that anomaly and the processing of it, it is of course not the only factor which may be relevant. Other potential factors in anomaly processing are elaborated in the next section.

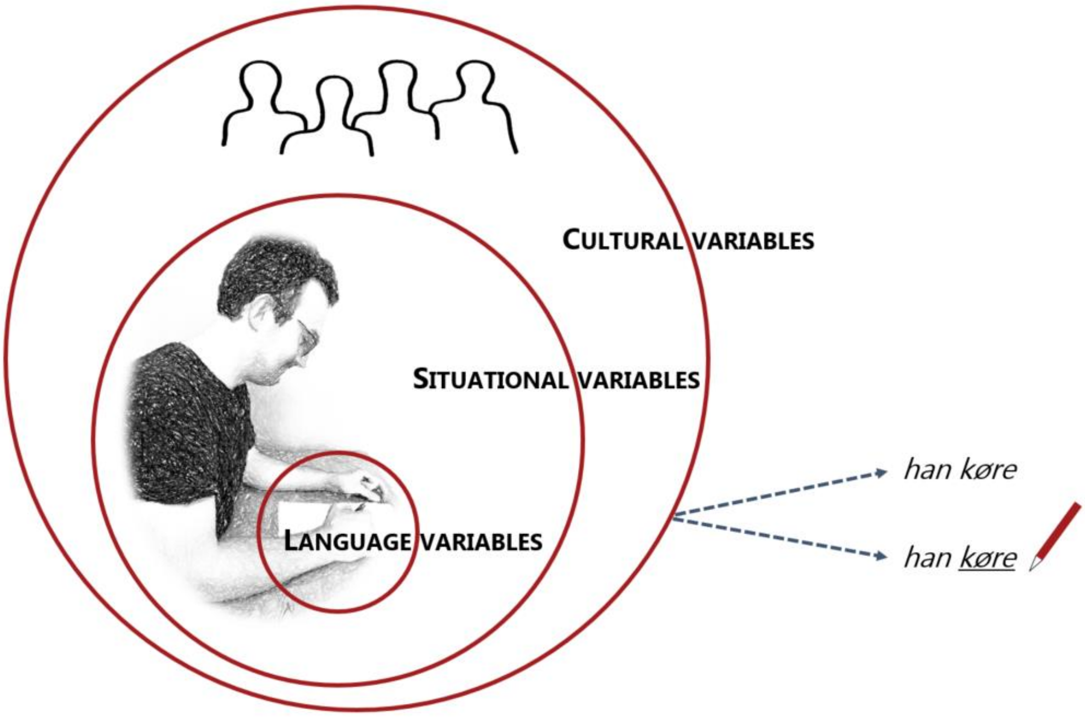
5.2 A tentative model of factors in anomaly detection

This section presents a tentative model of factors in anomaly detection. Some of these factors are examined in Article 2: Participants' explicit grammatical awareness (as measured in a grammar quiz), participants' irritation with anomalies in general, the role of anomaly frequency (and thus predictability), the anomaly's phonological similarity to the correct form (or lack of the same), and, for the V3 anomalies, the placement in the sentence (i.e. whether the V3 anomaly is presented after a short or long sentence-initial adverbial). This length manipulation is also examined in the eye-tracking study in Article 3.

The model in Figure 9 is a first attempt of providing an overview of factors potentially affecting readers' attention to anomalies, and thus the description is not exhaustive. The model is a basic illustration of a participant in an error detection study who has been asked to detect anomalies in a text, such as the verb form *køre* 'drive' in **han køre* 'he drive.INF'. The outcome of the process is that the anomaly is either detected (underlined) or not. During the process of detecting anomalies, many potential contributing factors may affect whether the reader underlines the anomaly or not. To keep the model simple and comprehensive, these contributing factors are organized in three preliminary categories: LANGUAGE VARIABLES, SITUATIONAL VARIABLES, and CULTURAL VARIABLES. The LANGUAGE VARIABLES have to do with the anomaly itself and the structure of the sentence it is part of. The SITUATIONAL VARIABLES are related to the situation, the task, and the reader, and the CULTURAL VARIABLES are related to the supra-individual level, e.g. collective language norms (Harder 2010) and discourses in society about anomalies. All three categories are elaborated below. The division of variables into different categories could have been different, but the main aim of the model is to give a comprehensive overview, not to insist on exactly this number of categories and this categorization.

The previously mentioned Figure 8 illustrated how the predications made by the individual language users' language processing models (containing stored probabilistic knowledge) can be updated based on input with or without anomalies. It does not inform on whether the language user considers a word or construction as anomalous, i.e. whether he or she would underline it in an error detection study. The potential factors involved in this decision are described in Figure 9. I assume that the language user depicted in Figure 9 has an internal language processing model similar to the one illustrated in Figure 8.

Figure 9. Tentative model of potential contributing factors in anomaly detection. The model illustrates a participant in an error detection study. During the process of detecting anomalies, many potential contributing factors might affect whether the participant underlines the anomaly or not, e.g. ****han* ***<u>køre</u>*** **'he drive.INF'. The potential contributing factors are organized in three preliminary categories: The** ***language variables*** **have to do with the anomaly itself and the structure of the sentence it is part of. The** ***situational variables*** **are related to the situation, the task, and the reader. The** ***cultural variables*** **are related to the supra-individual level, e.g. discourses in society about anomalies.**



The first group of contributing factors are the LANGUAGE VARIABLES, which include a wide range of factors related to the anomaly itself, its placement, and the structure of the sentence it is part of. This is e.g. inspired by findings from previous letter detection studies and change-blindness studies, which have examined a wide a range of factors that can influence attention during reading (e.g. Smith and Groat 1979, Sturt et al. 2004, Vinther et al. 2015, Christensen et al. 2021). In the letter detection paradigm, participants read a text while searching for (and highlighting) specific letters. The paradigm can be used to identify which words, part of words, or part of sentences readers pay particular attention to during reading (Smith & Groat 1979), e.g. by comparing how many *e*'s are detected in the beginning of words vs. in the end. The change-blindness paradigm can also be used to examine attention during reading. Here, participants see a stimulus sentence, and shortly after, they see a changed version of the sentence (e.g., a word could be substituted with another word). They are then asked whether the sentence was changed or not. Typically, participants attend less to changes in backgrounded information (Vinther et al. 2015). For example, Sturt et al. (2004) found that participants were less blind to changes in nouns that were focalized with cleft constructions compared to non-focalized nouns.

Factors connected to the anomaly itself include the frequency and thus the predictability of the anomaly (as elaborated in the previous section). Because the predictability of the anomaly may vary according to the individual language users' stored probabilistic knowledge (Figure 8) it could also be placed in the category SITUATIONAL VARIABLES.

Another factor connected to the anomaly itself is the type of anomaly, e.g. whether it is orthographic, morphological or syntactic, and which subtype within these categories it is. Previous error detection studies have found differences in attention to different anomalies, e.g. grammatical vs. orthographic anomalies (Shafto 2015, Hacker et al. 1994, Levy et al. 1992, Larigauderie et al. 2020). They have, however, come to different conclusions concerning which of the two types of anomalies are most prominent, likely due to differences in their definitions of grammar vs. orthography (Article 2). Larigauderie et al. 2020 found that typos like *toujousr* for *toujours* (which are phonologically distinct) attract more attention than grammar anomalies (e.g. gender and number agreement or misuse of past participles and infinitives), which again attract more attention than orthographic anomalies with phonological similarity to the correct form, e.g., *essentiellemment* for *essentiellement* (Larigauderie et al. 2020). The study in Article 2 also includes various orthographic anomalies with phonological similarity to the correct form, which are expected to have low detection rates. It is furthermore suggested in Article 2 that phonological similarity to the correct form could also negatively influence detection rates for grammar anomalies. For example, confusion of heterophone verb pairs such as *rejser* 'travel.PRS' [ˈʁɑjˀsɐ] and *rejse* 'travel.INF' [ˈʁɑjsə] should be easier to detect than confusion of homophone verb pairs, such as *kører* 'drive.PRS' and *køre* 'drive.INF', both pronounced [ˈkʰøːɐ] (Article 2). That the anomaly is "phonologically correct" may cause that the reading is not disturbed.

Moreover, various visual factors could also influence attention, such as segment size (long words attract more attention than short words), the shape of the word (tall letters or double letters attract more attention), and the placement of the anomaly on the line and in the sentence (elements in the start or end of a line or within a sentence tend to be more prominent than elements in the middle), as found using letter detection (Smith & Groat 1979). This, of course, comes with the reservation that it is uncertain how comparable searching for letters vs. searching for anomalies is. The placement of the anomaly within a phrase, e.g. a noun phrase, might also play a role, i.e. whether it is placed in the beginning (an article) vs. the middle (an adjective). If phrases are processed as a whole, meaning that the individual words are not all fixated or read with the same amount of thoroughness, phrase-initial anomalies should be more noticed than those in the middle.

Finally, aspects regarding information structure can increase attention to specific words, e.g. by focusing the word using focus particles or cleft constructions (Birch & Rayner 2010, McKoon et al. 1993, Sturt et al. 2004). A focalised anomaly may thus attract more attention than a non-focalised anomaly. In the error detection study in Article 2, I tried to control for this by creating four different versions of the same text, so that the anomalies in different conditions (e.g. the verb anomalies in the four conditions) were presented in the same context. I also aimed to place the anomalies in the same position on the line in the four text versions.

The second group of factors is called SITUATIONAL VARIABLES. These factors are related to the situation, the task, and the reader. Individual differences are likely to affect anomaly detection. If an individual often produces a specific type of anomaly, some grammar rule may not be fully mastered, and thus it seems likely that this person will overlook this type of anomaly in general. What constitutes a "language error" is like to vary from one person to another, and from one situation to another. The boarder between "unusual language" and "outright error" may thus be flexible according to the specific situation. This is for instance seen in studies by Konieczny et al. (1994), Hanulíková et al. (2012), and Gibson et al. (2017) who found that tolerance for various anomalies can be modulated by participants' perception of the speaker. When the speaker was perceived as an L2 user, participants' tolerance towards the anomalies increased, compared to when they were produced by an L1 user. In the EEG study by Hanulíková et al. (2012) on processing of Dutch gender agreement anomalies, this meant that the P600 effects in response to the anomalies disappeared when they were presented in a foreign accent. This is in line with Kuperberg & Jaeger (2016) who emphasise that comprehenders' predictions about the language input differ in different situations. In other words, it may not be that participants are more tolerant towards L2 users, but that they have other predictions for L2 users. In language attitude experiments, V3 has been found to be associated with immigrant status (Freywald et al. 2015). This may affect comprehenders' predictions.

Finally, the task (and the difficulty of the task) could potentially affect participants' attention to anomalies. In letter detection, less letters are found, the greater the work load is, e.g. caused by the number of simultaneous tasks (Smith & Groat 1979). Braze et al. (2002) hypothesized that anomaly detection and processing may be affected by variation in processing load prior to the anomaly. This idea was explored in Article 3 by comparing processing of V3 after short vs. long sentence-initial adverbials, i.e. a similar length manipulation as in Article 2, where the focus was on the placement in the sentence (cf. Article 3).

The third group of factors, CULTURAL VARIABLES, are related to the supra-individual level, e.g. collective language norms (Harder 2010) and discourses in society about anomalies. An individual's negative emotions, such as annoyance, towards specific anomalies (or anomalies in general) may affect attention to anomalies. These emotions could be linked to discourses in the public debate, and thus individual's irritation with anomalies is placed in this category, although it is in a grey-zone between the individual vs. supra-individual level.

Previous studies show that readers evaluate texts with different types of spelling and grammar anomalies negatively, e.g. email responses to housemate adds (Queen & Boland 2015), newspaper articles (Appelman & Schmierbach 2018), and job applications (Martin-Lacroux & Lacroux 2016). Writers of texts with anomalies also risk being judged negatively, e.g. journalists are judged as more unreliable, more incompetent, and less intelligent if they produce spelling mistakes (Blom & Ejstrup 2019a).

Some anomalies are pointed out as "typical" or "basic" in the public debate and in the prescriptive literature, whereas other anomalies are much less debated or accentuated (Blom & Ejstrup 2019b, Article 2). Due to the larger amount of attention in the media and schools, the publically debated anomalies may be more prominent to readers, e.g. the missing present tense *-r* on verbs in Danish (Blom & Ejstrup 2019b). The authors found that readers' intolerance for anomalies is modulated by the type of anomaly. They presented participants (who all found it important that journalists spelled correctly) to news with various types of spelling and grammar anomalies. Participants were more annoyed with typical and basic grammar and spelling anomalies than with atypical and complicated anomalies. The most annoying anomaly was the missing present tense *-r*. Furthermore, the authors found a correlation between participants' irritation (with a specific item) and the number of anomalies detected, so that the more anomalies participants detected in general, the more irritated they were with that item. Blom & Ejstrup (2019b) interpret this as an indication that the most competent proofreaders are more likely to be annoyed with anomalies, and the least competent proofreaders are less likely to be irritated. Finally, the participants' irritation with spelling and grammar anomalies were seemingly not related to comprehension issues. They generally did not find that the common spelling and grammar anomalies had a negative impact on text comprehension, e.g. only 3 % found that infinitive *ændre* instead of present tense *ændrer* 'change' altered the meaning of the sentence.

To sum up, when participants detect anomalies in texts, there is a wide range of potential factors which may play a part (as showed in Figure 9), and the predictability of the anomaly (as described in Figure 8) is only one of these. The tentative model in Figure 9 operates with three preliminary categories (LANGUAGE VARIABLES, SITUATIONAL VARIABLES, and CULTURAL VARIABLES). For some factors, it is debateable which category it should belong to. Also, the category SITUATIONAL VARIABLES is very broad, as it both covers individuals' inner states, and types of texts, tasks, and situations, so it could be divided into different subcategories. In order to develop the tentative model further, more studies on attention to anomalies are needed: Which factors play a part, and how are these factors related to each other and to individual differences?

As argued in both Article 2 and Article 3, understanding the factors that govern attention and reaction to different types of naturally occurring anomalies can help improve eye-tracking models of reading, so that they can accommodate everyday texts even better. Studies on attention to anomalies may also provide a better understanding of allocation of attention in language in general.

6 Method: Perception studies

This section briefly presents the two research paradigms used in Article 2 and Article 3: error detection (section 6.1) and eye-tracking (section 6.2). In section 6.3, I argue why these two paradigms were chosen and discuss strengths and weaknesses of using the paradigms. Finally, section 6.4 highlights a methodological novelty of the thesis, i.e. combining production and perception studies when examining grammar anomalies.

6.1 The error detection paradigm

In error detection or proofreading experiments, participants are explicitly asked to detect errors in text (e.g. Shafto 2015, Hacker et al. 1994, Levy et al. 1992, Larigauderie et al. 2020). In Article 2, participants were presented to a long, consecutive text on paper and asked to underline all errors they noticed when reading for comprehension. Using the error detection paradigm, it is possible to measure whether a given error is underlined by the participant or not. If the error is underlined, it is assumed that the participant has detected it and perceived it as incorrect. If the error is not underlined, it could either be because the error was not noticed or because the error was not perceived as incorrect to the reader. Alternatively, the error might not be detected (underlined), but it could still affect more sensitive measurements, such as reading time, so that the reading is slowed down. The error detection paradigm cannot provide further insights into this. Instead, eye-tracking (or perhaps self-paced reading studies) are needed.

The exact correlations between error detection and online measures, such as eye movements, is not well-explored (cf. Article 2). One study, though, which has combined error detection with eye-tracking is Huang & Staub (2021). Their stimuli sentences sometimes contained transposition errors, such as *The white was cat big*. For every sentence, the participants read, they had to judge whether it was grammatical or not. The authors did not find that undetected errors affect eye-tracking measures (such as fixation durations and regressions). Contrarily, they found that transposition errors only caused disruption in sentences which participants judged to be ungrammatical, and no disruption was found in sentences which were judged as correct by participants, although they in fact were incorrect (Huang & Staub 2021). Whether this is also the case for other types of grammar anomalies, such as morphological ones, is, to my knowledge, still unaccounted for.

As pointed out in Article 2, error detection can only provide a rough (offline) measure of attention during reading, but it can provide insights into which types of anomalies are more noticed than others, as well as insights into which factors, other than anomaly type, are likely to play a role.

6.2 The eye-tracking paradigm

This section introduces how an eye-tracker works (section 6.2.1) and what eye-trackers can measure (section 6.2.2).

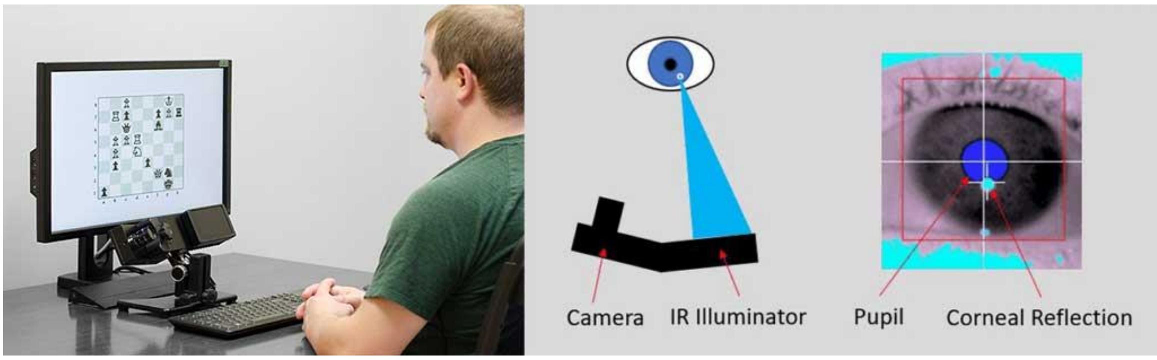
6.2.1 How does an eye-tracker work?

For the experiment in Article 3, an Eyelink 1000 eye-tracker (SR Research Ltd., Ontario, Canada) was used, similar to the eye-tracker seen in Figure 10. The experiment was carried out in the Language Acquisition and Language Processing Lab at NTNU. Below, it is briefly described how this video-based eye-tracker works. In the lab at NTNU, the participant was placed in front of a computer screen, which presented the stimulus. To minimize head movements, a chin and forehead rest was used, as seen in Figure 10. The experimenter sat in front of another screen, connected to the same computer, from which the experiment could be controlled and monitored. The actual eye-tracker was placed on the table in front of the computer screen, which presented the stimulus. The eye-tracker contains a camera which can take pictures of the participants' right eye with a sampling rate of 1,000 Hz, i.e. up to 1,000 pictures every second. The EyeLink system can calculate where on the screen the participant is looking and send this information back to the computer which controls the stimulus presentation (SR Research 2023). The most common method of estimating the point of gaze (i.e. where someone is looking on the stimulus screen) from a picture of the eye is based on tracking of the pupil and of the so-called corneal reflection (Holmqvist et al. 2011:24).

Figure 10. An eye-tracker used with a chin and forehead rest.

The eye-tracker contains an infrared illuminator, which sits next to the camera, as seen in Figure 11. The eye-tracker sends infrared light into the eye of the participant, and the light is reflected in the cornea, which covers the outside of the eye. The Eyelink software uses image-processing algorithms to identify where the pupil and the corneal reflection is (the two key locations) on each of the images sent by the camera (SR Research 2023). Like the majority of eye-tracking systems, the Eyelink 1000 exploits the fact that the relative positions of the pupil and the corneal reflection change systematically when the eye moves, so that the pupil moves faster, while the corneal reflection moves slower (Holmqvist et al. 2011:29). The eye-tracker can read the relative distance between the pupil and the corneal reflection and calculate the gaze position based on this relation. Thus, it is necessary to provide the eye-tracker with examples of how specific points on the screen correspond to specific pupil and corneal reflection relations (Holmqvist et al. 2011:29). This is done by performing a *calibration*, in which a number of points, organized in a grid, (e.g. nine as in the experiment in Article 3) are presented one at a time and fixated by the participant.

Figure 11. To the left is an image of the Eyelink 1000 Plus (a newer version of the Eyelink 1000). In the middle is an illustration of the eye-tracker seen from above: the camera and the infrared illuminator, which sends light into the eye, creating the corneal reflection. To the right is a screenshot from the experimenter's screen showing that the eye-tracker has correctly detected the pupil (in dark blue) and the corneal reflection (the light blue dot). The images are created by SR Research and used with permission (found on https://www.sr-research.com/about-eye-tracking/).



A benefit of tracking both the pupil and the corneal reflection (and not just the pupil) is that it helps the system to compensate for head movements (SR Research 2023). When the head moves, the relationship between the center of the pupil and the center of the corneal reflection remains the same, but when the eye rotates, the relationship changes (SR Research 2023).

6.2.2 What do eye-trackers measure?

> "With eye-tracking technology the eye is thought to give researchers a window into the mind" (Conklin & Pellicer-Sánchez 2016:1).

For researchers interested in language processing, eye-tracking technology can be used to monitor the eye when reading or when looking at a static image or video, perhaps combined with auditory input. Eye-trackers are primarily used to detect and measure the eye's movements and stops (Conklin & Pellicer-Sánchez 2016). The most reported event is called a *fixation* and refers to the state when the eye is still over a period of time, e.g. when the eye momentarily stops at a word during reading. They can last from "some tens of milliseconds up to several seconds" (Holmqvist et al. 2011:21f). Researchers generally consider that when a fixation is measured, attention to that position is also measured (Holmqvist et al. 2011:22). The eyes are actually not completely still during a fixation, but intra-fixational movements such as *microsaccades* (movements which bring the eye back to its original position) and *drifts* (slow movements which take the eye away from the center of fixation) are mainly studied within the field of neurology (Holmqvist et al. 2011:23).

The rapid movements from one fixation to the next, e.g. from one word to another when reading, are called *saccades*. Saccades are the fastest movement that the human body can produce, taking around 30-80 ms to complete, and during most of the saccade, we are blind (Holmqvist et al. 2011:23). Saccades back in a text (right to left in English) are often referred to as *regressions*. They can arise for several reasons; shorter regressions can be due to overshooting a target or difficulty with processing of the fixated word, longer regressions can be due to general text difficulty, ambiguity or integration difficulty (Conklin & Pellicer-Sánchez 2016).

The idea that eye-tracking can provide researchers with a "window into the mind" builds on two underlying assumptions. One is that the amount of time spent fixating a word reflects the cognitive effort which is required to process it. This means that more fixations and longer durations indicate larger processing effort. Contrarily, shorter fixations and word skipping indicate less processing effort. Within a prediction framework, processing difficulty would be reflecting a prediction error (Kuperberg & Jaeger 2016). The other assumption is that what we fixate is what is "being considered" (Conklin & Pellicer-Sánchez 2016), i.e. that attention to the fixated word is measured, as previously mentioned (Holmqvist et al. 2011:22).

Compared to self-paced reading paradigms, eye-tracking allows for more natural reading (Conklin & Pellicer-Sánchez 2016). Eye-tracking can be done without secondary tasks such as pressing a button for the next word to appear. When studying processing of grammar anomalies, it is also an advantage that the researcher can see exactly *where* in the stimulus sentence, the processing difficulties occur. This cannot be measured in a traditional reading time experiment.

In Article 2 and Article 3, I review previous eye-tracking studies of ungrammaticality and discuss numerous reservations when comparing the results of these studies. Other fundamental challenges when comparing eye-tracking studies is that terms for the numerous different measurements that exist are far from standardized (Holmqvist et al. 2011:4), and neither are data cleaning procedures (cf. Conklin & Pellicer-Sánchez 2016:464). Hessels et al. (2018) emphasize that events such as fixations and saccades are man-made entities, and not something which is objectively present in the eye-tracker signal. The algorithm of the eye-tracking system computationally defines events, e.g. fixations and saccades, in the eye-tracker signal, and there are many different event-classification algorithms available for different types of eye-trackers. Thus, depending on the type of eye-tracker being used, eye movements can appear quite differently in the eye-tracker signal (Hessels et al. 2018). This does not cause problems within the eye-tracking study in Article 3, but it causes challenges when this study is compared to other studies. If the results from Article 3 were to be reproduced, e.g. with the aim of comparing them to Danish data, the same eye-tracker and software should be used.

6.3 Combining error detection and eye-tracking

This section discusses strengths and weaknesses of using the error detection and eye-tracking paradigms when studying perception of grammar anomalies.

The aim of the study in Article 2 was to examine attention to a broad range of anomalies during reading. There were several advantages to using the error detection paradigm for this purpose. First, participants could be introduced to long, consecutive texts, simulating natural text reading (Article 2). This would not be possible using self-paced reading or EEG. Using the eye-tracking paradigm, participants can be presented to longer text passages, however, in all previous eye-tracking studies of ungrammaticality reviewed in Article 2 and Article 3 (Braze et al. 2002, Dank et al. 2015, Deutsch & Bentin 2001, Huang & Staub 2021, Lim & Christianson 2015, Ni et al. 1998, Pearlmutter et al. 1999), sentences were presented individually. If participants are presented to more than one line of text, results may be confounded, and the complexity of the analysis increases.

Second, the error detection paradigm made it possible to include a variety of grammar anomalies. Previous eye-tracking studies of ungrammaticality have included relatively few anomaly types (e.g. pragmatic vs. syntactic). Having many different types of anomalies in different conditions would likely result in a long and tiresome eye-tracking experiment (or EEG experiment).

Third, using the error detection paradigm, it is possible to get participants' feedback on where anomalies occur, in a fast way, unlike if participants have to be asked after every trial in an EEG, or eye-tracking experiment.

One limitation of using the error detection paradigm is that it may provide a less sensitive measure of attention during reading than e.g. eye-tracking, I assume. As highlighted in Article 2, the exact correlations between the two measures is not well-explored. The error detection paradigm may treat two events as the same, while they involve different eye movements. Furthermore, while the error detection experiment was fast for participants to complete, the digitization process was cumbersome.

The aim of the study in Article 3 was to focus on the V3 anomalies, and thus conduct an online processing study. Online processing of V3 has already been examined in Swedish using EEG (Andersson et al. 2019, Yeaton 2019, Sayehli et al. 2022). An advantage of using eye-tracking (vs. EEG) is that an eye-tracking experiment can be set up relative quickly and does not require excessive training. It is usually also fast to conduct (unlike EEG experiments). The cumbersome part of an eye-tracking study is the analysis of the eye movement data.

To sum up, both the error-detection and eye-tracking paradigms can be used to examine attention to anomalies during reading. In Article 2, a broad range of anomalies were examined, using a presumably rough measure of attention, and in Article 3, I focused on online processing of V3 anomalies using presumably more sensitive measures. As argued in Article 2: If differences in attention are found using the error detection paradigm, they are also likely to be found using a presumably more sensitive measure such as eye-tracking.

6.4 Combining production and perception studies

Most previous studies either focus on the production of anomalies or on the perception of anomalies. A novel take in this thesis is to both examine production and perception of grammar anomalies. The relationship between the two is especially examined in Article 2, where production data (error rates in L1 texts) are used in the hypotheses to predict perception patterns in the error detection study. Including production patterns when examining perception of anomalies is important, because production patterns can help explain some of the differences found in perception. As found in Article 2, individual differences, as measured in the grammar quiz, explain some of the variance. Furthermore, production patterns (measured as error rates in L1 production) can explain differences between different types of anomalies (e.g. syntactic vs. morphological) and some of the differences found within these types. How to operationalize and calculate production patterns is, however, not straightforward (cf. section 5.1.2), and the theoretical foundation for combining production and perception of anomalies has yet to be developed. The studies in this thesis provide an empirical foundation for developing theories by pointing to some of the similarities between production and perception patterns.

Part II

The articles

7 The articles

This section contains the three research articles and thus forms the main part of the thesis. Article 1 is an unpublished manuscript and appears in the version which is submitted to *Bilingualism: Language and Cognition*. Article 2 and Article 3 have been published and appear in the published versions.

References and appendixes for the three articles are found after each article. Supplementary materials for the published articles (Article 2 and Article 3) are found in the Appendix (section 10).

Article 1: V2 is not difficult to all learners in all contexts – a cross-sectional study of L2 Danish

Unpublished manuscript, submitted to *Bilingualism: Language and Cognition*, 1st of June 2023.

V2 is not difficult to all learners in all contexts – a cross-sectional study of L2 Danish

Katrine Falcon Søby & Line Burholt Kristensen
Department of Nordic Studies and Linguistics, University of Copenhagen

ABSTRACT

In a cross-sectional study of L2 Danish, we examined the production of correct verb-second (V2) word order. We tested the effect of 1) the learners' language background, 2) CEFR level, and 3) the complexity of the sentence constituents. The texts were written by 217 students (A2-B1, 52 different L1s). Interrogative clauses had high accuracy, but 25 % of the 491 declarative sentences with non-initial subjects had incorrect V3 word order. Our study shows that V2 is not difficult for all learners. Learners whose L1 was a V2 language had a significantly higher share of correct V2 word order and they never overused V2. For non-V2 learners, the share of correct V2 significantly increased with CEFR level. As for sentence complexity, accuracy decreased significantly with the length of the first constituent and for subjects consisting of multiple words.

1 Introduction

When learning a second language (L2), there may be crosslinguistic influence from one's first language (L1) affecting all language levels, from phonology to discourse (Mitchell, Myles & Marsden, 2013, p. 16). Previously, some areas of language use, including syntax, were believed to be impervious to crosslinguistic influence (Jarvis, 2017, p. 13), but this claim was later challenged by e.g. Bohnacker (2006). An interesting case for the study of crosslinguistic influence on syntax is verb-second (V2) word order, which is common in Germanic languages. V2 word order is often described as notoriously difficult to master for L2 learners (Bolander, 1990; Hagen, 1992), but it may not be equally challenging to all learners and in all sentential contexts (Johansen, 2008).

Here we present the first large-scale study of V2 in written L2 Danish, focusing on the order of verb and subject in sentences where the subject is not in first position. Our study is the first V2 corpus study which uses inferential statistics to understand the separate contributions of the learners' language background, the learner's CEFR level (The Common European Framework of Reference for Languages, Council of Europe, 2001) and the complexity of the sentence constituents.

2 Background

2.1 V2 word order in Danish

Across the world's languages, V2 is rare, but all Germanic languages (apart from modern English) have V2 word order. In V2 languages, "the finite verb is obligatorily the second constituent, either specifically in main clauses or in all finite clauses" (Holmberg, 2015, p. 342). The Danish sentences below have different constituents in first position. The subject and verb of the main clause are in bold. The constituent in first position is in parentheses. All except sentence (3) are grammatical and have the verb in second position. In (1), the order is subject-before-verb as in English, but in (2) and (4)-(6) the first position is occupied by a non-subject constituent. Since the verb must appear in second position, the subject occurs after the verb.

(1) ***Jeg bor*** *i Danmark nu* (subject)
I live in Denmark now
'I live in Denmark now'

(2) *Nu* ***bor jeg*** *i Danmark* (adverbial: adverb)
Now live I in Denmark
'Now, I live in Denmark'

(3) **Nu* ***jeg bor*** *i Danmark* (adverbial: adverb)
Now I live in Denmark
'Now, I live in Denmark'

(4) *Fordi jeg fik et job,* ***bor jeg*** *nu i Danmark* (adverbial: subordinate clause)
Because I got a job, live I now in Denmark
'Because I got a job, I now live in Denmark'

(5) *Ham* ***mødte jeg*** *i Danmark* (object)
Him met I in Denmark
'Him, I met in Denmark'

(6) *Det* ***er jeg*** *glad for* (prepositional object)
That am I happy about
'I'm happy about that'

Sentence (3) has ungrammatical V3 word order, because two constituents (both the adverbial and the subject) precede the verb. V3 word order is common in L2 production (L2 Norwegian: Brautaset, 1996; Hagen, 1992; Johansen, 2008. L2 Swedish: Bohnacker, 2006; Bolander, 1989. L2 Danish: Holmen, 1994; Lund, 1997; Søby & Kristensen, 2019. L2 German: Håkansson, Pienemann & Sayehli, 2002). Among second language researchers, there are different explanations of why V2 should be hard to acquire. In his influential Processability Theory, Pienemann (1998) argues that XVS word order involves a complicated movement, demanding high cognitive capacity: SVX and ungrammatical *Adv-SVX (i.e. V3) have a more basic word order which is easier to process and thus easier to produce than XVS (V2). In Pienemann's hierarchy of acquisition, V2 is acquired late – if ever. An alternative functionalistic explanation by Lund (1997, p. 162) argues that there is little communicative pressure to acquire V2 word order for declarative sentences compared to interrogative sentences, as V2 word order in declarative clauses does not have "any semantic or pragmatic function". In Danish interrogative sentences, such as (7) and (8), there is more communicative pressure, as the use of verb-before-subject indicates that the mood of the sentence is interrogative.

(7) *Bor du i Danmark?*
Live you in Denmark
'Do you live in Denmark?'

(8) *Hvor bor du?*
Where live you
'Where do you live?'

Pienemann (1998) sees V2 as difficult to all learners, but previous studies of L2 production in V2 languages suggest that crosslinguistic influence from the L1 and other L2s can reduce or increase the difficulties (L2 German: Bohnacker, 2006. L2 Norwegian: Johansen, 2008). Johansen's (2008) study of 100 learners found that L1 users of other V2 languages have an advantage when it comes to producing V2. Studies of L2 Swedish and Norwegian also indicate that production of correct V2 word order is not equally challenging in all sentential contexts (Swedish: Bolander, 1989. Norwegian: Brautaset, 1996; Hagen, 1992; Johansen, 2008). These studies call for further investigations of how the complexity of the clause may affect V2 production. A better understanding of which sentential contexts are challenging to learners can help L2 teachers of V2 languages focus their teaching.

Language teachers of L2 Danish, L2 Swedish and L2 Norwegian are faced with similar challenges, as the three languages are mutually intelligible (perhaps with "some initial difficulty" (Vikør, 2015)) and all have similar use of V2 word order. Still, there is some variation in grammaticality of V3. In Swedish, for instance, a few focalizing adverbials can be placed between the subject and verb: *Hun* [S] *bare* [A] *ville* [V] *låna min cykel* 'She just wanted to borrow my bike' (Bohnacker, 2006, p. 455), and some Norwegian dialects, most notably in northern Norway, allow V3 word order in wh-question structures (Westergaard, Vangsnes & Lohndal, 2017). Such use of V3 is ungrammatical in Danish, and the acquisition of V2 word order in L2 Danish may therefore be slightly different.

Although sentences with V3 are comprehensible despite the ungrammatical word order, the use of correct V2 word order does seem important to ensure smooth communication with L1 users. An error detection study by Søby, Ishkhanyan and Kristensen (2023) found that L1 users notice incorrect V3 word order more frequently than other types of grammar anomalies, such as confusion of verb inflections and missing gender agreement in NPs. V3 word order also disrupt L1 users' online processing, as found for Swedish by means of EEG (Andersson, Sayehli & Gullberg, 2019; Sayehli, Gullberg, Newman & Andersson, 2022; Yeaton, 2019) and for Norwegian by means of eye-tracking (Søby, Milburn, Kristensen, Vulchanov & Vulchanova, 2023).

2.2 Crosslinguistic influence and CEFR level

Crosslinguistic influence (CLI) is here used interchangeably with *transfer*, referring to "the ways in which a person's knowledge of one language can affect his or her learning, knowledge and use of another language" (Jarvis, 2017, p. 2). Pienemann, Di Biase, Kawaguchi and Håkansson (2009, p. 128) argue that learners of closely related languages do not necessarily transfer grammatical features at the initial state of learning an L2, even if the features are contained in the L1 and the L2, because transfer from the L1 is constrained by processability of the feature in the L2. This idea is supported by a study of V2 in Swedish school children's L2 German (Håkansson et al., 2002). The children produced V3 in German, despite close typological proximity between Swedish and German (both are V2 languages). Bohnacker (2006), however, argued that the use of V3 by the children in this study could be due to syntactic transfer from English (which all children have learned as an L2 prior to learning German). In order to control for influence from English, Bohnacker (2006) compared oral production data from six adult Swedish learners of German; three monolinguals learners, and three with prior knowledge of English. The learners with no knowledge of English did not produce V3 in their German, but the other group did, suggesting that learners can transfer the property of V2 from their L1, in contrast to Håkansson et al. (2002), but that "L2 knowledge of a non-V2 language (English) may obscure this V2 transfer" (Bohnacker, 2006, p. 444). Other studies have also found indications of transfer of verb-second syntax in the opposite direction, i.e. in L2 English from either L1 German and Dutch (Rankin, 2012), and from L1 Norwegian (Westergaard, 2003).

Even though CLI can occur on all language levels from phonology to discourse, the odds for encountering CLI in learner data are greatest when "the target language is related to a language the learners have already mastered" (including the L1) and when "the feature is frequent in the learners' L1" (Jarvis, 2017, p. 14). Both of these factors apply to V2 in L2 Danish, as Danish is typologically close to the Germanic V2 languages, and clauses with non-initial subjects are common in e.g. Danish, Swedish and German (Bohnacker & Rosén, 2008; Fabricius-Hansen & Solfjeld, 1994; Kristensen, 2013; Westman, 1974). Kristensen (2013) reports that 57.5 % of main clauses in a corpus of written Danish (85,000 words; newspapers, magazines etc.) have subjects in first position, 32.3 % have other constituents, including adverbials, and 1.2 % have direct objects. In written German (Fabricius-Hansen & Solfjeld, 1994) and Swedish (Westman, 1974) (newspapers and other publications), the first position has a high share of subjects (54 % for German and 64 % for Swedish), a lower share of adverbials (36.8 vs. 30.8 %) and a much lower share of objects (6.6 vs. 2.3 %).

Besides Bohnacker (2006), a few other studies of V2 in learner language have found indications of syntactic CLI. In Lund's (1997) longitudinal study of six learners of L2 Danish (L1 Dutch (2), English (2), Spanish (1), and Portuguese (1)), only the two Dutch speakers, whose L1 is a V2 language, achieved "some stability" in producing V2 in declarative clauses (Lund, 1997, p. 158). Johansen (2008) examined sentences with non-initial subjects in a standardized test for 100 learners of Norwegian (The Language Test for Adult Immigrants) from the ASK test corpus (ASK, 2015), roughly corresponding to the B1 CEFR level (Tenfjord, Jarvis & Golden, 2017, p. 3). The learners generally had a high rate of V2 in the sentences, 86.5 %, but the 20 L1 speakers of Dutch and German had even higher success rates, 98.4 % and 100 %. The other eight language groups ranged from 69.4 % to 93.6 %. These findings support the idea that having V2 in one's L1 makes it easier to produce V2 in an L2 – even though there may be syntactic CLI from other non-V2 L2s (Bohnacker, 2006).

To our knowledge, there are no quantitative studies comparing the role of CLI on V2 production on different CEFR levels. There are a few studies (Bolander, 1990; Brautaset, 1996) who report a progression in V2 production with increasing proficiency, but they do not differentiate between users with V2 and non-V2 background. In Johansen (2008), all the 100 learner texts were at B1 level. The interplay between the learner's CEFR level and language background is therefore unaccounted for.

2.3 The role of sentence complexity

Sentence-internal factors may also influence the production of V2 in learner language. Previous studies have investigated the role of the sentential context for L2 Swedish (Bolander, 1989, 1990; Hyltenstam, 1978) and L2 Norwegian (Brautaset, 1996; Hagen, 1992; Johansen, 2008), but not for L2 Danish. Comparing the results of these studies is difficult. Some studies use categories where material and syntactic function are intertwined (Bolander, 1989, 1990; Brautaset 1996). Other studies are mainly qualitative (Hagen, 1992; Johansen, 2008), and it seems that only Hyltenstam (1978) has used inferential statistics to test effects of e.g. constituent complexity.

All of the previous investigations, besides those of Bolander, are based on written production, both essays and elicited material (see Table 1). The largest studies are those of Hyltenstam (160 learners), Bolander (60 learners) and Hagen (38 learners). Johansen's (2008) qualitative analysis only includes 19 learners. The learners in the studies are on different levels, with some studies comparing the same learners at different stages (Brautaset, 1996; Hagen, 1992; Hyltenstam, 1978). As shown in Table 1, most studies do not include learners with V2 background (Bolander, 1990; Brautaset, 1996; Johansen, 2008) or they do not report the L1s (Hagen, 1992). In Hyltenstam (1978) there are, however, a few learners with L1 German. As these previous studies have not attempted to study the separate contributions of sentence-external factors, such as the learner's language background vs. the sentence-internal factors, such as sentence complexity, it is still an open question if the role of sentence complexity applies equally to all learners.

Most studies have focused on the role of the constituents in the first three sentence slots when examining whether there are favorable and unfavorable contexts for producing XVS word order. The studies include a mixture of structural (syntactic functions, material, complexity of the material, frequency) and semantic descriptions.

Table 1. Overview of previous studies of constituents in sentences with V2 vs. V3

Study	**Target language**	**Learners (N)**	**V2 learners included?**	**Task/material**	**Topics of investigation**
Hyltenstam (1978)	Swedish	160	3 learners with L1 German	Elicited written material (fill the gap) x 2 (5 weeks inbetween)	Quantitative study of pronominal vs. nominal subjects and of auxiliary verbs vs. main verbs.
Bolander (1989, 1990)	Swedish	60	No	Oral	Quantitative study of the functions/material of the first constituent (idiosyncratic categories), of the types of subjects and of the types of verbs (semantics).
Hagen (1992)	Norwegian	38	Not reported	Essays after 2, 4 and 7 months	Qualitative analysis of various characteristics of the first constituent, subject and verb with regard to complexity. No quantitative analysis.
Brautaset (1996)	Norwegian	12	No	Essays after 2, 6 and 8 months	Quantitative analysis of long adverbials (> one word) vs. short adverbials (one word) in first position, of adverbials consisting of subordinate clause vs. adverb phrases, of pronominal vs. nominal subjects and of auxiliary verbs vs. main verbs.
Johansen (2008)	Norwegian	19 (in the qualitative part)	No	Standardized test (B1)	Qualitative analysis of the frequency of the first constituent (highly frequent vs. somewhat frequent vs. unique) and of the complexity of the entire sentence (using an intricate scoring system). No numbers are reported.

As shown in Table 1, the studies differ with respect to the principles for categorization of sentential constituents, with respect to modality (written vs. oral language) and task type. Their conclusions differ, probably due to methodological differences and the different approaches to categorization. It is, for instance, not clear if nominal vs. pronominal subjects affect the share of V3. Hagen (1992) propose a general hypothesis: sentences with V3 generally include features which burden the language user's capacity to process information to a higher degree than the sentences with V2 (Hagen, 1992, p. 34). In the following, we summarize some key findings regarding the first constituent, the subject and the verb.

The role of constituents in first position

Bolander (1989, p. 76) found a high share of correct V2 after objects in first position, 82 %. However, her data is oral L2 Swedish, and many of the object-initial sentences are of the type *det tror jag* or *det tyckar jag* 'That I think' which could be learned as chunks. In her study, the most common first constituent (in sentences with non-initial subjects) are adverbs, which only have an accuracy rate of 36 %. For subordinate clauses in first position, the share of V2 is much lower (19 %). In the correct sentences, Brautaset (1996) found similar shares of adverbials consisting of subordinate clauses vs. adverbial phrases (although with different tendencies on different levels). It should be noted that Brautaset only reports the distribution in correct sentences, leaving out important information about the sentences where learners have produced V3 word order.

Hagen (1992) almost exclusively found adverbials in first position in sentences with non-initial subjects. Based on his data, he hypothesized that learners produce more V3 when the constituent in first position is heavy (measured by the number of words). Brautaset (1996) could not confirm this hypothesis, as she found a *higher* share of long adverbials (more than one word), 88.7 %, vs. short (one word) adverbials, 86.3 %, but again only in the correct V2 sentences.

The role of the subject

Previous studies have mainly investigated pronominal vs. nominal subjects, and the results are inconclusive. Bolander (1990) reports the highest shares of V2 in speech when the subjects were NPs (56 % V2) or first person pronouns (43 %), compared to e.g. second or third person pronouns (22 %). However, the shares may be different in written production. Brautaset (1996) found higher shares of pronominal subjects (89.8 %) than nominal subjects (79.6 %) in the V2 sentences, across levels. Hyltenstam (1978) did not find differences between pronominal and nominal subjects. Finally, Hagen (1992) hypothesized that a subject which is either a pronoun or long and heavy, favors V2. The last part seems contradictive to his general hypothesis on processing load, but is not further explained.

The role of the verb

Previous studies have mainly compared single verbs to complex verbs (such as a modal followed by an infinitive). Hyltenstam (1978) did not find differences, but Brautaset (1996) generally found higher shares of single verbs than complex in the V2 sentences, apart from on the highest proficiency level.

Bolander (1990) reports that V2 is often found with verbs expressing opinion or belief, as in the examples with *det tyckar jag* 'That I think'. These results may, however, be specific to oral V2 production where frequent and fixed (chunk-like) OVS sequences are more dominant than in written language (Kristensen, 2013). Hagen (1992) hypothesized, based on this findings, that frequent or short verbs favors V2.

Sentence complexity across constituents

To conclude, methodological differences in the previous qualitative and quantitative V2 corpus studies make it difficult to characterize the role of sentence complexity across the board. Still, many observations concerning the first three sentence constituents in relation to production of V2 vs. V3 evolve around complexity (sometimes intertwined with frequency, cf. Johansen (2008)) and indicate that V3 is more frequent for complex XVS sequences. This idea resonates with Skehan's

(1998) Limited Capacity Hypothesis, which argues that learners, because of limited attentional resources, constantly have to balance between focusing on the accuracy or the complexity of their output.

It is also not clear if the tendencies are the same for non-V2 learners and for learners whose L1 is also a V2 language. It is therefore relevant to systematically study the role of sentence complexity with unequivocal principles for categorization and to compare the role of sentence complexity for learners with a V2 vs. non-V2 background.

2.4 The current study

In the current cross-sectional study, we examine the production of V2 vs. V3 in texts written by students (A2-B1) enrolled in an official Danish language program. Sentences with initial subjects are not a specific challenge to learners of V2 languages, and the study therefore only covers sentences with non-initial subjects. Using a statistical model, we test whether there are effects on V2 production of 1) learners' CEFR level, 2) learners' L1, and 3) the complexity of the first constituent, the verb and the subject.

- **Hypothesis 1**. We expect to find a progression from A2 to B1, reflected in a higher share of V2. This hypothesis is based on findings from Brautaset (1996) and Bolander (1990) who report a progression in V2 production with increasing proficiency.
- **Hypothesis 2**. We expect that learners with another V2 language as their L1 (V2 learners) have a higher share of V2 than learners with an L1, not featuring V2 (non-V2 learners). This hypothesis is based on the study by Johansen (2008), who found higher shares of V2 for V2 learners than for non-V2 learners.
- **Hypothesis 3**. We expect that increasing complexity of the three first constituents will affect the share of V2 production negatively.
  - **3.1**. We expect that the complexity of the first constituent, measured roughly as the number of words, negatively affects V2 production. From previous studies, it is unclear whether heavy constituents in general affect V2 production, but Bolander (1990) found lower shares of V2 after subordinate clauses.
  - **3.2**. We expect that the complexity of the subject, operationalized roughly as one word vs. multiple words, negatively affects V2 production. Previous studies have compared pronominal vs. nominal subjects with contradictory results. This comparison is to some extent intertwined with our comparison between one vs. multiple words.
  - **3.3**. We expect that the complexity of the verb, operationalized roughly as single vs. complex, negatively affects V2 production (based on findings from Brautaset (1996)).

3 Method

3.1 Data collection

In Denmark, there are three official Danish language programs. Students are assigned to one of the programs, primarily based on their educational background. Danish Program 1 (DP1) is for students who cannot read and write using the Latin alphabet. Danish Program 2 (DP2) is for students who have a short educational background and who are expected to have a relatively slow progression. Danish Program 3 (DP3) is for students with a medium or long educational background, who are expected to have a rapid progression (MII, 2019). All programs consist of a series of modules. After each module students need to pass a module test to continue in the program. We collected the written module tests at the school *Copenhagen Language Center* in 2017 to 2018, as part of a larger research project, i.e. the data were not collected specifically for this V2 study. The written tests consisted of 1-2 writing assignments, which varied according to the level. Students were allowed to use dictionaries or verb lists. The students' handwritten texts were digitized and anonymized.

In total, texts from 217 students were collected (138 women; mean age 30.9 years, *SD* 7.2 years) (Søby, to appear; Søby & Kristensen, 2019). The participants had around 52 different L1s (cf. Appendix, Table 5). The five most dominant L1s were English (N = 38), Spanish (N = 16), German (N = 13), Portuguese (N = 12) and Russian (N = 11). Twenty-four students had another V2 language as their L1 (German (N=13), Dutch (N = 7), Icelandic (N = 2), Afrikaans (N = 1), German/Russian (N = 1)).

Figure 1.A shows which Danish Program and module the participants attended. Most participants were from Danish Program 3. The distribution between modules is skewed (most students are in module 2), but the distribution of V2 and non-V2 learners concerning DPs and modules is relatively similar.

The module tests only give rough estimates of students' proficiency and do not necessarily reflect the actual levels of the learners. We do not know if students passed their test or not. Students may also be more proficient than indicated by their current module.

4 Analysis

4.1 Markup principles and exclusion criteria

All declarative clauses with non-initial subjects (XVS/*XSV) and all interrogative sentences ((X)VS) were marked. Due to the creative and sometimes surprising nature of learner language, we based the markup on two principles. Firstly, we ignored morphological and orthographical anomalies, as well as anomalous word choice, and focused on the order of what we interpreted as a plausible verb and a plausible subject. For instance, examples (9-10) are both considered as correct V2 word order – in spite of the anomalous morphological form of the constituents. In (9), two non-finite verbs, the participles *tabt* 'lost' and *haft* 'had', are used instead of finite verbs (e.g. past tense *tabte* 'lost' and *havde* 'had'), but they are correctly placed in second position. In (10), the subject *os* 'us' is in oblique form instead of nominative *vi* 'we', but is correctly placed after the verb.

(9) [...] *firste* ***tabt*** ***jeg*** *min pengepang og begafter* ***haft*** ***jeg*** *indbrud.*
first lost.PTCP I my wallet and then had.PTCP I break-in
'first I lost my wallet, and then I had a break-in.'

(10) *I dag* ***havde os*** *en eksamen* [...].
Today had us.OBL an exam
'Today, we had an exam.'

Secondly, the punctuation is not always consistent with the content of the sentence. In these cases, the classification is based on the content of the sentence. For example, (11) was tagged as an interrogative sentence despite the lack of a question mark, because the interrogative pronoun *hvad* 'what' is used.

(11) *Så hvad* ***synes i*** *om hendes liv.*
So what think you about her life
'So what do you think about her life?'

Finally, we excluded 35 sentences from the analysis for two reasons. Sentences with the adverb *måske* 'maybe' in first position were excluded (N = 11), because the adverb both can be succeeded by verb-subject (12) and subject-verb word order (13) (Beijering, 2010; Boye, 2005). Thus, we cannot determine the success rate in this context.

(12) *Måske* ***kan i*** *se Netflix sammen* […].
Maybe can you watch Netflix together
'Maybe you can watch Netflix together.'

(13) *Måske* ***vi skal*** *også ser fjernsyn.*
Maybe we shall also watch TV
'Maybe we'll watch TV too.'

Furthermore, 24 sentences were excluded because the prescribed word order could not be decided. The majority of these cases (N = 18) included the word *så* 'so/then' which can either be used as a conjunction (introducing main or subordinate clauses), or an adverbial. When *så* is used as a coordinating conjunction, it is used to convey a result or a consequence, and is followed by subject-verb word order as in (14). In (15), *så* is used as an adverb ('then') followed by verb-subject word order (both non-authentic examples).

(14) *Jeg er sulten, så* ***jeg spiser****.*
I am hungry so I eat
'I'm hungry, so I eat.'

(15) *Først spiser jeg, så* ***drikker jeg****.*
First eat I then drink I
'First I eat, then I drink.'

In some cases, we could not determine which of the two meanings the learner intended to use, and we were therefore unable to determine if the word order was correct. Example (16) is written by an Icelandic learner, and here *så* is followed by verb-subject. The intended meaning of *så* may either be to convey a consequence (in which case *så* should have been used as a coordinating conjunction with subject-verb order) or to convey the meaning 'therefore' or 'then' (in which case *så* is an adverb in first position requiring verb-subject word order). Due to this ambiguity, example (16) was excluded from the analysis.

(16) *Jeg blive snart en gamle mand, så* ***vil jeg*** *gerne skåle nogle drikker med mine bedste*
I become soon an old man, so will I toast some drinks with my best

venner […].
friends
'I will soon be an old man, so I would love to drink some drinks with my best friends.'

In total, 491 declaratives and 158 interrogatives were included in the analysis. We included interrogatives when tagging word order, because previous studies have compared learners' use of VS in declaratives and interrogatives (e.g. Lund, 1997).

4.2 Statistical models

Data were analyzed using generalized linear mixed models for binomial data in RStudio (R Core Team, 2021, ver. 2022.02.1), including the lme4 package (Bates, Maechler, Bolker & Walker, 2015, ver. 1.1.27.1). P-values were obtained using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017, ver. 3.1.3) and this formula:

V3_or_V2 ~ first_constituent_complexity + subject_complexity + verb_complexity + CEFR + L1_V2 + (1|participant)

The dependent variable was whether the declarative sentence was incorrect (V3) or not (V2). The fixed effects were complexity of the first constituent (number of words in first position), subject complexity (one word vs. multiple words), whether the verb was complex (no vs. yes), CEFR level (1-3), and L1 (non-V2 vs. V2). For subject and verb complexity we expected that the main challenge for learners would be to process more than one word, but for first constituents (which vary in length and usually exceed one word) we expected that the length (and not just the categorical difference between one vs. multiple words) could have an effect. The model also included random intercepts for participant. Comparisons were coded using sum contrasts (Schad, Vasishth, Hohenstein & Kliegl, 2020), so that non-V2 learner was coded as -0.5 and V2 learner was coded as 0.5. For subject complexity, one word was coded as -0.5, and multiple words as 0.5. For the verbs, not complex was coded as -0.5, and complex as 0.5.

Twenty-two non-V2 learners indicated that one of their L2s was a V2 language (German, Swedish, Norwegian, or Dutch). They did not indicate their proficiency level of these additional L2s. To control whether exposure to a V2 language (other than the L1) improved V2 production in Danish, we carried out a post-hoc test comparing non-V2 learners with and without previous exposure to V2. We ran the same model as mentioned above on a dataset where the V2 learners were excluded. Instead of L1 (V2 vs. non-V2), we included knowledge of V2 as a fixed effect. No knowledge was coded as -0.5, and V2 knowledge as 0.5.

Both models were fitted using the following procedure: We first tried to include all variables mentioned above. Variables were removed one by one to see whether the fit improved. The fit was best fit when all were included. Due to the small amount of data for the V2 learners, an interaction between L1 and CEFR was not included.

5 Results

The first three subsections contain the descriptive statistics for CEFR level (5.1), language background (5.2) and sentence types (5.3). The results of the statistical model are presented in section 5.4. The results are concluded with a further analysis of sentence constituents in 5.5 and an analysis of overuse of V2 in 5.6. The learners had an exceptionally high share of correct V2 for the interrogative clauses (99 %), so the results focus on the use of V2 and V3 in declarative sentences.

5.1 Production patterns from A2-B1

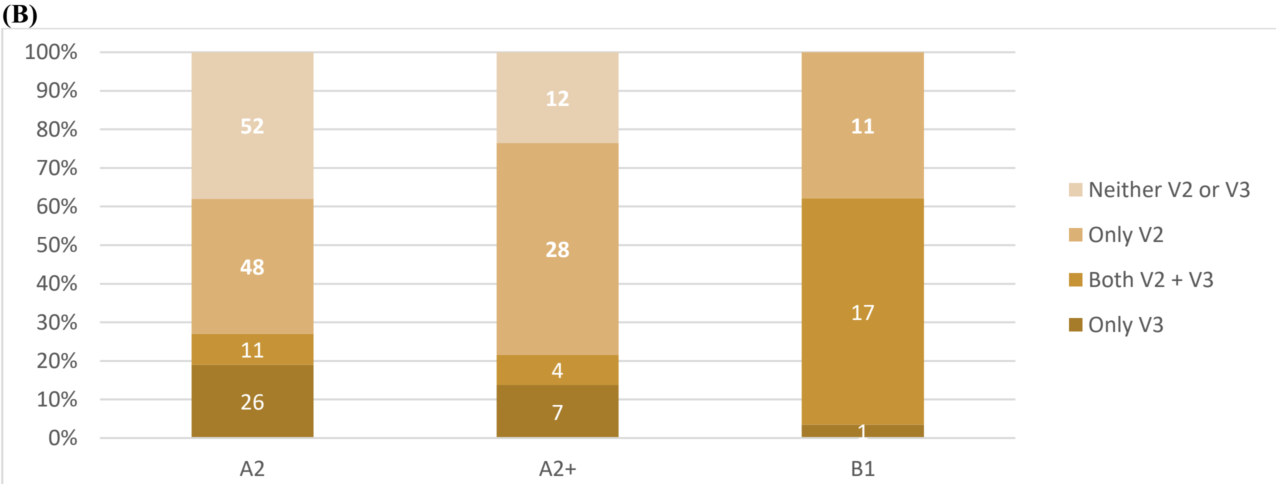
Figure 1.B shows the production patterns for all participants. On A2, there is a relatively high share of participants who do not produce declarative sentences with non-initial subjects (38 %), but the share decreases on A2+ (24 %). On B1, all participants produce XSV/XVS. Of the 217 participants, the most common production pattern is to only produce V2 (87 learners, 40 %). From A2 to A2+, the share increases, but then decreases again from A2+ to B1.

The share of participants who both use V2 and V3 is small on A2 and A2+ (both 8 %), but increases drastically on B1 (59 %), where everyone has started to produce sentences with non-initial subjects. The share of participants who only use V3 is 19 % on A2 and gradually decreases for higher levels. On B1, just one Greek learner only produces V3.

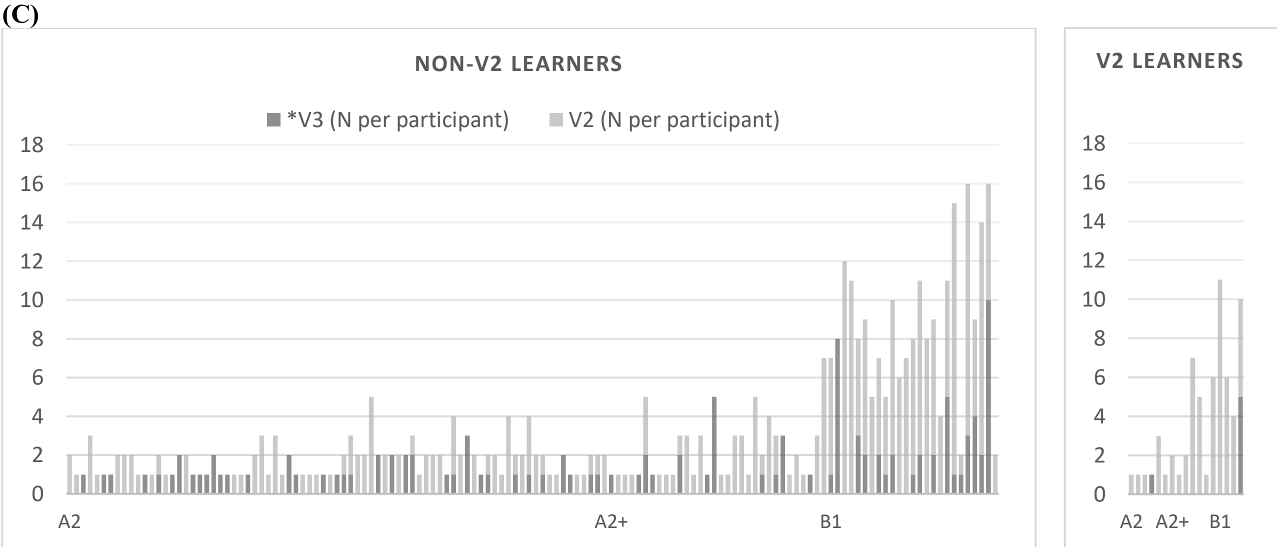
Figure 1. Overview of learners' language background and V2 production patterns. **(A)** provides an overview of learners' Danish Program, module, the corresponding CEFR level and months of teaching. We roughly have data from A2 to B1. Most participants at A2 had 5 months of teaching. In the so-called A2+ group, participants had around 8 months of teaching. B1 learners had around a year of classes. **(B)** shows production patterns for declarative sentences with non-initial subjects for all learners. The y-axis shows the shares in %. The number of learners for each group are shown on the columns. **(C)** shows the number of correct XVS (V2) sentences and incorrect XSV (V3) per participant (one column per participant). The figure only includes the 153 participants who produced declarative sentences with non-initial subjects. The non-V2 learners (N = 136) are to the left; to the right are the V2 learners (N = 17). The number of V3 sentences (if any) appear below (dark grey), while the number of V2 sentences is stacked on top (light grey).

(A)	**Non-V2 learners (N)**	**V2 learners (N)**	**Participants (total)**	**CEFR**	**Months of teaching**
Danish Program 2			**10**		
Module 3	10 (5 %)	-	10	A2	9 months
Danish Program 3			**207**		
Module 2	114 (59 %)	13 (54 %)	127	A2	5 months
Module 3	44 (23 %)	7 (29 %)	51	A2+	8 months
Module 4	25 (13 %)	4 (17 %)	29	B1	12 months
Total	**193 (100%)**	**24 (100 %)**	**217**		

(B)



(C)



5.2 Individual variation and L1 patterns

Figure 1.C shows the number of V2 and V3 sentences for each of the 153 participants who produced declarative sentences with non-initial subjects. For the non-V2 learners, there is individual variation in the number of XVS sentences (and the success rate) per participant, but generally, the number of V2 sentences increases drastically on B1 level.

Seven V2 learners do not produce sentences with non-initial subjects and are not represented in the figure. Of the remaining 17 V2 learners, only two produce V3. As seen in Figure 1.C, the learner on B1 level produces 5 out of 6 of the V3 sentences produced by the entire learner group. The general pattern with a large increase in the number of V2 sentences on B1 is also found for this learner group.

Table 6 (Appendix) shows the distribution on CEFR level and L1 for the 153 learners producing sentences with non-initial subjects. The V2 group has a lower share of A2 learners (35 % vs. 58 %) and a higher share of learners on A2+ (41 % vs. 24 %) and B1 (24 % vs. 18 %).

5.3 Distribution of V2 and V3 on sentence types

Table 2 shows the distribution of V2 and V3, both in declarative clauses and interrogative clauses. As mentioned, the interrogative clauses have consistently high accuracy rates. For the 491 declarative clauses with non-initial subjects, one out of four sentences have ungrammatical V3 word order.

Table 2 also provides an overview of the syntactic functions of the different sentence constituents found in first position in the corpus. The first constituent is almost always an adverbial (similar to Hagen's (1992) findings for Norwegian). Only six sentences have objects in first position, most of them with V2. Finally, the category "After other" contain nine sentences, all with V2. In three of these sentences, the first constituent is the complement of a preposition (N = 3), as shown in Table 2. This category also contains sentence intertwinings (Poulsen, 2008) (N = 2), as seen in (17) where *det* is a topicalized constituent from a subordinate clause. Finally, this category contains sentences in which the syntactic role of the first constituent cannot be decided (N = 4), e.g. (18) in which *Når i skolen* 'when in the school' is anomalous in Danish. It may be intended as a subordinate clause *Når vi er i skolen* 'When we are in the school'.

(17) *Det* ***synes jeg*** *et meget god ide* [...].
It think I a very good idea
'That I think is a very good idea.'

(18) *Når i skolen,* ***laver vi*** *opgaver* [...].
When in school.DEF make we assignments
'When [we are] in the school, we make assignments.'

Table 2. Number of sentences with V2 and V3 word order and share of correct V2

	EXAMPLES OF V2 AND *V3	V2 (N)	TOTAL	SHARE OF V2
Declarative		**370**	**491**	**75 %**
MAIN CLAUSES		366	485	75 %
After adverbial	*Naturligvis* ***er vinteren*** *rigtig cold og mørk* […] Naturally is winter.DEF really cold and dark 'Of course, the winter is really cold and dark'	352	470	75 %
After object	[…] *men min mormors ring* ***får vi*** *ikke tilbage.* but my granny's ring get we not back 'but my granny's ring we will not get back.'	5	6	83 %
After other	[…] *så det* ***er jeg*** *faktisk glad for* so that am I actually happy about 'so I am actually happy about that'	9	9	100 %
SUBORDINATE CLAUSES				
After adverbial	[…] *FORDI NU* ***KAN VI*** *FRIT ÆNDRE ALT* […] because now can we freely change everything 'because now we can change everything as we want'	4	6	67 %
Interrogative		**156**	**158**	**99 %**
Open-ended questions	*Hvordan* ***går det****?* How goes it 'How does it go?'	86	86	100 %
Closed-ended questions	***Savner du*** *mig? :P* Miss you me 'Do you miss me? :P'	70	72	97 %

5.4 Model results: the role of L1, CEFR level and constituent complexity

Table 3.A shows the results for the statistical model. As expected, we found an effect of L1 background (hypothesis 2) ($p < 0.01$), so that V2 learners had higher accuracy than non-V2 learners. Contrary to our expectations, the effect of CEFR level (hypothesis 1) did not reach statistical significance, although it trended ($p = 0.055$). The non-significant effect of CEFR is small compared to the effect of L1, as seen on the estimates. If the estimates are transformed to probabilities, the probability for V2 is 83 % when all fixed effects are set to their baseline. The probability increases to 89 % when CEFR level increases from A2 to B1. The probability increases to 98 % when L1 changes from non-V2 to V2 (with all other fixed effects set to their baseline). Figure 2 (Appendix) provides an overview of the distribution of V2 vs. V3 sentences on CEFR levels and L1s. The effect of L1 is clearly illustrated, as most sentences produced by V2 learners have V2 word order.

As shown in Table 3.A, we also found effects of sentence complexity. We found a small effect of the number of words of the constituent in first position (hypothesis 3.1) ($p < 0.05$). The more words there were in first position, the more V3 we found. To examine whether the effect was carried by subordinate clauses in first position, we ran the model again on a dataset without the 81 constituents containing a subordinate clause. The effect disappeared, suggesting that the length of constituents, besides those which contain a subordinate clause, may not matter (cf. section 6.3) (see model results in Table 7, Appendix). Moreover, we found the expected effect of subject complexity (hypothesis 3.2) ($p < 0.05$), so that when the subject consist of multiple words (vs. one word), the more V3 sentences are produced. The effect of verb complexity (hypothesis 3.3) was small and only trended significance ($p = 0.072$).

Table 3. Effects of L1, CEFR level and constituent complexity. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$. **(A)** shows model results (number of observations = 491, participants = 153). **(B)** shows the share of correct V2 in declarative sentences. **(C)** shows model results for the post-hoc test which only included the non-V2 learners (number of observations = 428, participants = 136).

(A)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2.971	1.724		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	1.56686	0.72734	2.154	0.03122 *
Words in first position (N)	-0.16215	0.06422	-2.525	0.01158 *
Subject (one vs. multiple words)	-1.02377	0.45397	-2.255	0.02412 *
Verb (single vs. complex)	-0.53890	0.29934	-1.800	0.07181 .
CEFR (1-3)	0.49374	0.25780	1.915	0.05547 .
L1 (non-V2 vs. V2)	2.38632	0.88766	2.688	0.00718 **

(B)

L1	**CEFR**	**LEARNERS (N)**	**V2 (N)**	**V3 (N)**	**SHARE OF V2**
NON-V2		**136**	**313**	**115**	**73%**
	A2	79	91	46	66%
	A2+	32	52	19	73%
	B1	25	170	50	77%
V2		**17**	**57**	**6**	**90%**
	A2	6	7	1	88%
	A2+	7	24	0	100%
	B1	4	26	5	84%
TOTAL		**153**	**370**	**121**	**75%**

(C)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2.305	1.518		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	0.47703	0.59602	0.800	0.42351
Words in first position (N)	-0.17242	0.06707	-2.571	0.01015 *
Subject (one vs. multiple words)	-1.23562	0.46500	-2.657	0.00788 **
Verb (single vs. complex)	-0.57057	0.30295	-1.883	0.05965 .
CEFR (1-3)	0.55027	0.24590	2.238	0.02523 *
L2 (no knowledge of V2 vs. knowledge)	0.88106	0.66378	1.327	0.18440

Post-hoc test for L1 and CEFR level
Table 3.B shows the share of V2 for V2 learners vs. non-V2 learners on the different CEFR levels. For non-V2 learners, the share of V2 gradually increases from A2 to B1. For V2 learners, the share of V2 is high in general across levels, but the number of learners per level is small. As Table 3.B showed different patterns for V2 and non-V2 learners, we carried out a post-hoc test on the dataset without the V2 learners. We conducted the post-hoc test for two reasons. First, we suspected that CEFR level and verb complexity had an effect on V2 production for the non-V2 learners, but not for the V2 learners which had high accuracy in general. Second, we wanted to test whether non-V2 learners performed better when they had previous knowledge of a V2 language (e.g. from learning Swedish or German prior to learning Danish). Table 3.C shows the model results for this test. The effect of CEFR reached statistical significance ($p < 0.05$) – the higher the level, the more V2 is produced. The p-value for verb complexity decreased to 0.05965, and is thus still not significant. There was no effect of having previous knowledge of an L2 with V2 word order.

The results of the initial model in Table 3.A show that V2 learners are superior to non-V2 learners when it comes to producing correct V2. Still, these results are not necessarily due to CLI on syntax. The difference between V2 and non-V2 learners could be due to a general benefit of learning a language closely related to one's own (cf. Jarvis 2017). To further examine whether the V2 learners were generally superior to non-V2 learners, we contrasted their performance for other types of grammar anomalies. Figure 3 (Appendix) shows the number of morphological anomalies per 1000 running words for all participants, divided into three groups: non-V2 learners with and without prior knowledge of V2 as well as V2 learners. The plot shows that there is variation in all three groups. For non-V2 learners, the standard variation is 48. For V2 learners and learners with knowledge of V2 the standard variation is 38 and 37 respectively. We therefore find it unlikely that the differences in V2 production are exclusively due to a general benefit of learning a similar language.

5.5 Further analysis of constituents

As seen in section 5.3, the first constituent is often adverbial. In this section, we present results from an analysis of the materials of the first three constituents in declarative clauses and how these are linked to the distribution of V2 vs. V3. We also examine the semantic content of the adverbials in first position.

Complexity of constituents in first position
Table 3.A suggests that it is a challenge to produce V2 after long constituents containing a subordinate clause (N = 81). In (19), the adverbial subordinate clause is followed by V3. Fifty-one of the sentences with a subordinate clause in first position have V2 word order, i.e. 63 %. In comparison, the share of V2 after constituents not containing a subordinate clause is 78 % (N = 404, main clauses only). The V2 learners seemingly do not have challenges here, as only one V2 learner (A2) produces a sentence with V3 in this context, out of 13 examples distributed across levels. CEFR level also seems to play a part, as the share of V2 after subordinate clauses increases on B1 (76 % of 45 sentences), compared to A2 (48 % of 21 sentences) and A2+ (47 % of 15 sentences).

(19) *Selv om det er rigtig sjovt,* ****jeg savner*** *dig!*
Even though it is really funny I miss you
'Even though it is a lot of fun, I miss you!'

Complexity of the subject

Table 4.A provides an overview of the shares of V2 in sentences with subjects consisting of one word vs. multiple words. For the V2 learners, subject complexity, operationalized in this coarse-grained way, does not seem to affect accuracy. Most of the one-word constituents are personal pronouns (410 of 491), so the comparison between one word and multiple words is assumed to be largely correlated with a comparison between pronominal and non-pronominal subjects, though there are also one-word subjects like *Anna, danskere* 'Danes' and *kommunen* 'the municipality'. The non-V2 learners do not produce many multi-word subjects on A2 and A2+, but especially on B1, the share of V2 is lower when subjects are complex. The multi-word subjects range from two-word constituents like *min bror* 'my brother', *mange folkene* 'many people' and *Champions league* to longer constituents like *en god transport systemer* 'a good transportation system', and *min polsk-dansk ordbog, rød plastiken peberfrugter og to nøglen* 'my Polish-Danish dictionary, red plastic peppers and two keys'.

Complexity of the verb

Table 4.B shows the distribution of V2 in sentences with single finite verbs vs. complex verbs, i.e. verbs consisting of a finite and non-finite verb, e.g. present perfect or a modal plus infinitive. In (20), the verb is complex. The finite verb is in second position, and the non-finite verb occurs after the subject. In (21), the subject precedes the finite verb, resulting in V3.

(20) *Til festen* ***skal vi laver*** *løgsuppe* […].
For party.DEF shall we make onion soup
'For the party we will make onion soup.'

(21) *Dereafter* ****vi skal spiser*** *lekkert mad* […].
Afterwards we shall eat delicious food
'Afterwards, we will eat delicious food.'

As seen in Table 4.B, both the V2 learners and non-V2 learners tend to have higher shares of V2 in sentences with single verbs than with complex verbs, but the effect did not reach statistical significance.

Table 4. The share of correct V2 sentences for three types of categorizations. The total shows the number of declarative sentences (i.e. both correct and incorrect). **(A)** shows the share of V2 with one-word vs. multi-word subjects (in %). **(B)** shows the share of V2 with single vs. complex verbs (in %). **(C)** shows the share of V2 for sentences with adverbials in the first position (excluding sentences where the first position contains a subordinate clause).

(A) **Share of correct V2 with one-word vs. multi-word subject**

	A2	A2+	B1	Total
V2 learners				
one word	88 % of 8	100 % of 22	81 % of 27	89 % of 57
more than one word	-	100 % of 2	100 % of 4	100 % of 6
Non-V2 learners				
one word	67 % of 132	73 % of 67	80 % of 188	74 % of 387
more than one word	40 % of 5	75 % of 4	63 % of 32	61 % of 41
Total	**68 % of 145**	**80 % of 95**	**78 % of 251**	**75 % of 491**

(B) **Share of correct V2 with single vs. complex verbs**

	A2	A2+	B1	Total
V2 learners				
single verb	100 % of 1	100 % of 21	86 % of 22	93 % of 44
complex verb	86 % of 7	100 % of 3	78 % of 9	84 % of 19
Non-V2 learners				
single verb	70 % of 84	84 % of 45	80 % of 153	78 % of 282
complex verb	60 % of 53	54 % of 26	70 % of 67	64 % of 146
Total	**68 % of 145**	**80 % of 95**	**78 % of 251**	**75 % of 491**

(C) **Share of correct V2 with adverbials with different semantics**

	A2	A2+	B1	Total
V2 learners				
temporal	100 % of 4	100 % of 6	73 % of 11	86 % of 21
spatial	-	100 % of 1	50 % of 2	67 % of 3
argumentative	100 % of 1	100 % of 5	88 % of 8	93 % of 14
attitudinal	-	100 % of 4	-	100 % of 4
Non-V2 learners				
temporal	80 % of 70	90 % of 31	85 % of 89	84 % of 190
spatial	72 % of 25	89 % of 9	93 % of 14	81 % of 48
argumentative	56 % of 16	83 % of 6	64 % of 50	64 % of 72
attitudinal	25 % of 4	50 % of 8	75 % of 12	58 % of 24

Semantics of adverbials and verbs

Table 4.C shows an explorative analysis of the semantic content of the sentence-initial adverbials (excluding those with subordinate clauses). Typical examples of time/frequency expressions are: *nu* ‘now’, *i dag* ‘today’, *i 2017* ‘in 2017’, *nogle gange* ‘sometimes’, and combinations of *næste/sidste gang/uge/mandag* ‘next/last time/week/Monday’. Place expressions are e.g. *her* ‘here’*, i Norge* ‘in Norway’*, i parken* ‘in the park’, and *i sprogskolen* ‘in the language school’. Argumentative adverbials are those used when arguing for or against something, e.g. *derfor* ‘thus’, *for det første/andet* ‘Firstly/secondly’, and *på den ene/anden side* ‘on the one/other hand’. Attitudinal adverbials denote one’s attitude towards something, such as *heldigvis* ‘luckily’ and *desværre* ‘unfortunately’. For V2 learners, we have little data, and semantic content does not seem to affect V2 production. For non-V2 learners, the share of V2 seems higher when the adverbials denote time/frequency or place, than when argumentative or attitudinal adverbials are used, across CEFR levels.

Bolander (1989) reported to find high shares of V2 after verbs expressing attitude or belief in oral Swedish. We found few examples of these, e.g. there were only 11 examples of *synes* ‘think’ (of which 9 had V2 word order, i.e. 82 %).

5.6 Overuse of V2: V1 word order

Another type of word order anomaly found in the corpus is overuse of V2, or perhaps V1 word order. There are 61 cases in total, all produced by non-V2 learners. Typically, overuse of V2, i.e. verb-before-subject for subject-before-verb, occurs in subordinate clauses (N = 42). In 40 out of these 42 cases, overuse occurs after words functioning as conjunctions. In (22), the learner produces verb-before-subject word order in the subordinate clause after the conjunction *når* ‘when’, but correct subject-before-verb word order after the main clause conjunction *og* ‘and’. There are 19 cases of V1 word order in main clauses, primarily after conjunctions, but in 6 cases sentence-initially, as in (23).

(22) [X personer] *og jeg går i kantinen når* ****har vi*** *pause,*
[X persons] and I go in canteen.DEF when have we break

og vi elsker kantines billige kager!
and we love canteen’s cheap cakes
‘[X persons] and I go to the canteen when we have a break, and we love the canteen’s cheap cakes!

(23) ****Blive jeg*** *taknemmelig hvis du kan.*
Become I grateful if you can
‘I will be grateful if you can.’

6 Discussion

6.1 Summary

Interrogative sentences had a consistently high share of V2, so we focused on declarative sentences. One out of four declarative sentences with non-initial subjects had incorrect V3 word order. On A2 level, more than one third of the learners did not produce sentences with non-initial subjects, but on B1 level all learners produced such sentences. The expected effect of CEFR level (hypothesis 1) was not statistically significant. However, a post-hoc test excluding V2-learners showed a significant effect of CEFR level for the non-V2 learners.

Our study only had 24 V2 learners, but we found a significant effect of L1 on V2 production, so that V2 learners had a lower share of V3 than non-V2 learners (hypothesis 2). We also found that overuse of V2 was restricted to non-V2 learners.

Use of V3 increased for sentences with complex constituents. We found an effect of the number of words in first position (hypothesis 3.1), so that increasing complexity negatively affected the share of V2. We also found an effect of subject complexity (hypothesis 3.2), so that the share of V2 was lower in sentences where the subject consisted of multiple words vs. one word. The effect of verb complexity (hypothesis 3.3) only trended significance.

6.2 Proficiency and L1: Are the effects caused by CLI of V2?

Like previous V2 studies on syntactic CLI (L2 German: Bohnacker, 2006. L2 Norwegian: Johansen, 2008. L2 English: Rankin, 2012; Westergaard, 2003), we found an effect of learners' L1 (V2 vs. non-V2). To our knowledge, the only previous study of L1 influence on V2 production in L2 Danish is Lund (1997), who compared oral and written V2 production for four non-V2 learners to two Dutch learners. Our study contributes to research on syntactic CLI from one V2 language to another by providing the largest amount of participants so far and by using statistical models. Contrary to the findings and hypotheses by Håkansson et al. (2002) and Pienemann (1998), we found an effect of learners' L1, even though most participants were beginners (A2 or A2+).

Our analysis was based on an existing corpus which was not balanced with respect to CEFR level and language background. Also, the corpus did not contain data from modules corresponding to B2 or C1 level, and the CEFR levels were only estimated based on which modules participants attended. This means that the study has limitations. We found that the effect of L1 was larger than the non-significant effect of CEFR for all learners – with the reservation that the V2 learners were generally on higher CEFR levels, although CEFR level did not seem to affect the share of V2 for this learner group. Interestingly, our data indicate that CEFR level plays a minor role compared to L1 background. Future studies based on more balanced corpora concerning L1 and CEFR level, and with better testing of proficiency level, are needed to test this claim. We did not have enough data to include an interaction between L1 and CEFR in our model. This would be possible on a larger and more balanced dataset.

Finally, we cannot rule out the possibility that the higher accuracy for V2 learners compared to non-V2 learners is not due to CLI, but to a general benefit of learning a language closely related to one's L1 (Jarvis, 2017). However, our comparison of morphological anomalies per 1,000 words for

V2 vs. non-V2 learners showed large variation in both groups. If V2 learners have a general benefit, it is clearly expressed for syntax, but not for morphology. We thus find it likely that the differences in V2 production are due to syntactic CLI.

6.3 V2 is not always difficult – complexity of constituents

In line with previous studies of complex constituents in V2 languages (e.g. Bolander, 1989; Hagen, 1992), our study shows that accuracy decreases for complex constituents and gives empirical support to Skehan's (1998) Limited Capacity Hypothesis, which argues that learners are less accurate when they produce complex output, due to limited attentional resources. Even though our study is merely correlational, there may be a causal link between the complexity of the constituents and accuracy rates, as hypothesized by Skehan (1998). Our study contributes to the debate on the role of complexity by examining a new target language, by including more participants and by testing effects of complexity of the first three constituents by means of a statistical model. We also differentiated between V2 and non-V2 learners in the descriptive analysis of constituents' material.

We found an effect of the length of the first constituent (measured as the number of words). This effect disappeared when constituents containing a subordinate clause were excluded. There may not have been enough data to detect an effect of the length of phrases not containing subordinate clauses. Further investigations are needed to test whether the challenges with heavy constituents in first position are driven by subordinate clauses, and not long constituents in general. Our qualitative analysis of the first constituents' material indicated that V2 after subordinate clauses were only challenging to non-V2 learners, and that CEFR level might positively affect V2 production in this context. Previous studies of L2 Swedish have also found indications of sentence-initial subordinate clauses being particularly challenging for V2 production (Bolander, 1989), but not in L2 Norwegian (Brautaset, 1996). An efficient method for investigating this further could be a fill-the-gap task among non-V2 learners contrasting short, and long first constituents (of which the latter both contained subordinate clauses vs. phrases).

Our subject complexity measure distinguished between one word and multiple words. Most of the one-word subjects were personal pronouns, and the comparison therefore resembles the comparison of pronominal subjects and nominal subjects found in previous studies. It may not be the length itself that increases the cognitive load, but the accessibility and content of the subject. Previous studies of pronominal subjects point in different directions. Brautaset (1996) found higher shares of pronominal subjects in sentences with V2, but Hyltenstam (1978) did not find different word order patterns for pronominal vs. nominal subjects. Bolander (1990, p. 289) found the highest share of V2 with nominal subjects, and Hagen (1992) hypothesized that both pronouns and long and heavy subjects favored V2. In our descriptive analysis, only the non-V2 learners were negatively affected by subject complexity, but the V2 learners produced very few complex subjects. Thus, further studies are needed to uncover this potential difference between the two learner groups.

We did not find a significant effect of verb complexity, but it trended, especially for the non-V2 learners in the post-hoc analysis. The qualitative analysis revealed that sentences with complex verbs consisting of a finite + non-finite verb seemingly had lower shares than sentences with single verbs, for both learner groups, across CEFR levels. These results are similar to Brautaset (1996), but not Hyltenstam (1978).

Our study focused on rough complexity measurements, and did not investigate effects of frequency or uniqueness (Johansen, 2008). It may be that both the frequency of the individual constituent and the frequency of the entire "chunk", i.e. how often the three constituents appear together, affect V2 production. For the non-V2 learners, we found that semantics seemed to play a part for the sentence-initial adverbials (subordinate clauses excluded), as temporal and spatial adverbials had higher shares of V2 than argumentative and attitudinal adverbials. This could be related to frequency – if temporal adverbials are frequent in first position (with V2) in learners' input, it may be easier for the learners to produce V2, particularly, the more chunk-like the XVS sequence is.

In the corpus, we found examples of overuse of V2, which could be related to difficulties with distinguishing between adverbials in first position (which must be followed by the finite verb) and conjunctions (which do not occupy the first position). The reason that attitudinal and argumentative adverbials like *derfor* 'thus' had a higher share of V3 for the non-V2 learners could also be that they are interpreted as conjunctions.

6.4 Applications of the study

As mentioned in the introduction, V2 is often described as notoriously difficult to master for L2 learners (Bolander, 1990; Hagen, 1992). The current study provides a more nuanced picture by providing evidence that V2 word order in L2 Danish is not equally challenging for V2 and non-V2 learners, as otherwise claimed in the Processability Theory (Pienemann 1998). Furthermore, factors such as CEFR level and complexity of the sentence constituents affect the accuracy of V2 production. This knowledge could be of didactic value to language instructors in Danish as a second language.

It is common for popular textbooks and teaching materials (e.g. Slotorub & Moreira, 2011, 2014; Thorborg & Riis, 2010, which were used at the language school) to introduce Danish word order as subject-before-verb – with sentences with non-initial subjects as an exception. Often, the term *inversion* is used, indicating that XVS word order is a special case or exception from word order in general, where something needs to be inverted. Instead of seeing XVS as an exception, we propose that it could be beneficial to introduce V2 as the basic Danish word order already on basic A2 level (in Danish Program 3), i.e. teach students the basic principle that the verb must be in second position. When explaining V2 as a basic characteristic of Danish in language classes, it might also be useful to emphasize that constituents in first position vary in complexity (phrases and especially subordinate clauses), especially for non-V2 learners. Likewise, for the subjects, both pronouns and NPs with gradually increasing complexity can be introduced, and for the verbs, both single and complex verbs can be introduced. In relation to overuse of V2, the difference between sentence-initial conjunctions and adverbials could also be a focus point.

7 Conclusion

The study shows that V2 is not difficult *per se* for all learners and not in all contexts. Learners with another V2 language as their L1 had higher accuracy than non-V2 learners, and for non-V2 learners, the share of V2 increased with proficiency. Finally, we found effects of the complexity of the first constituent (measured as the number of words) and subject complexity (one word vs. multiple words), so that the share of V2 decreased significantly with increasing complexity. The study adds knowledge of didactic value to Danish language instructors, by highlighting that V2 is not difficult for all learners, and that the complexity of the constituents involved plays a part. Finally, we argue that explaining the term V2 in language classes (instead of describing Danish word order as SV, but with *inversion* of subject and verb in XVS sentences) could be beneficial for learners.

Data availability statement

The data and code for this study are openly available at https://github.com/ResearchXX/verb-second.

Acknowledgements

The study was financed by Independent Research Fund Denmark. Thanks to student assistants Julie Johanna Hansen, Kasper Rud Jensen, Caroline Ørum-Hansen, and Maja Mittag for tagging the L2 corpus. Special thanks goes to Byurakn Ishkhanyan for statistical support.

References

ASK – The Norwegian Second Language Corpus (2015). https://clarino.uib.no/korpuskel-classic/corpus-list?collection=ASK. Last accessed 22.01.2023.

Andersson, A, Sayehli, S and Gullberg, M (2019) Language background affects online word order processing in a second language but not offline. *Bilingualism (Cambridge, England)* 22(4), 802–825. https://doi.org/10.1017/S1366728918000573

Bates, D, Maechler, M, Bolker, B and Walker, S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beijering, K (2010) The grammaticalisation of mainland Scandinavian MAYBE. *Bergen Language and Linguistics Studies* 1(1), 1–21.

Bohnacker, U (2006) When Swedes begin to learn German: From V2 to V2. *Second Language Research* 22(4), 443–486.

Bohnacker, U and Rosén, C (2008) The clause-initial position in L2 German declaratives: Transfer of information structure. *Studies in Second Language Acquisition* 30(4), 511–538.

Bolander, M (1989) Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In Hyltenstam, K and Obler, L (eds.), *Bilingualism across the lifespan: Aspects of* *acquisition, maturity and loss*. Cambridge: Cambridge University Press, pp. 73–86. https://doi.org/10.1017/CBO9780511611780.006

Bolander, M (1990) On the acquisition of word order rules in swedish as a second language. In Halliday, MAK, Gibbons, J and Nicholas, H (eds.), *Learning, keeping and using language: Selected papers from the Eighth World Congress of Applied Linguistics, Sydney, 16 21 August 1987*. Amsterdam: John Benjamins Publishing Company.

Boye, K (2005) Semantisk analyse og afvigelser mellem sprogbrug og sprogstruktur. *Ny forskning i grammatik* 12, 27–46.

Brautaset, A (1996) *Inversjon i norsk mellomspråk: en undersøkelse av inversjon i stiler skrevet av innlærere med norsk som andrespråk.* Sofiemyr: Novus forlag.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Fabricius-Hansen, C and Solfjeld, K (1994) Deutsche und norwegische Sachprosa im Vergleich: Ein Arbeitsbericht. *Arbeitsberichte des Germanistischen Instituts der Universität Oslo* 6.

Fried, M (2015) Construction grammar. In Kiss, T and Alexiadou, A (eds.)*, Syntax – Theory and analysis: An international handbook (vol. 2)*. Berlin: de Gruyter, pp. 974–1003.

Hagen, JE (1992) Feilinvertering, overinvertering og underinvertering. *NOA,* 15, 27–38.

Holmberg, A (2015). Verb second. In Kiss, T and Alexiadou, A (eds.), *Syntax – Theory and analysis: An international handbook* (vol. 1). Berlin: De Gruyter, pp. 342–383.

Holmen, A (1994) Syntactic development in Danish L2. In Hyltenstam, K and Viberg, A (eds.), *Progression and regression in language: Sociocultural, neuropsychological and linguistic perspectives*. Cambridge: Cambridge University Press, pp. 267–288. https://doi.org/10.1017/CBO9780511627781.012

Hyltenstam, K (1978) Variation in interlanguage syntax. *Working papers* 18. Lund: Lund University, Department of General Linguistics.

Håkansson, G, Pienemann, M and Sayehli, S (2002) Transfer and typological proximity in the context of second language processing. *Second Language Research* 18(3)*,* 250–273.

Jarvis, S (2017) Transfer: An overview with an expanded scope. In Golden, A, Jarvis, S and Tenfjord, K (eds.), *Crosslinguistic influence and distinctive patterns of language learning*. Bristol: Multilingual Matters, pp. 12–28.

Jensen, A (2002) Sætningsknuder i dansk. *NyS – Nydanske Studier & Almen kommunikationsteori* 29, 105–124.

Johansen, H (2008) Inversjon i norsk innlærerspråk. En undersøkelse av variasjonsmønstre i skrevne tekster. *NOA, Norsk som andrespråk* 24(2), 50–71.

Kristensen, LB (2013) *Context, you need: Experimental approaches to information structure processing*. Ph.D. thesis. University of Copenhagen.

Kuznetsova, A, Brockhoff, PB and Christensen RHB (2017) lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lund, K (1997) *Lærer alle dansk på samme måde? En længdeundersøgelse af voksnes tilegnelse af dansk som andetsprog*. Herning: Special-Pædagogisk Forlag.

Ministry of Immigration and Integration (MII) (2019) *Vejledning til visitation og indplacering på danskuddannelser til voksne udlændinge m.fl.* Retrieved from https://www.integrationsviden.dk/filer-til-danskomradet/danskuddannelse/visitation-og-indplacering/vejledning-til-visitation-og-indplacering-paa-danskuddannelser-til-voksne-udlaendinge-m-fl-3.pdf

Mitchell, R, Myles, F and Marsden, E (2013) *Second language learning theories* (3rd edition). London: Routledge.

Pienemann, M (1998) *Language processing and second language development: Processability theory.* Amsterdam: John Benjamins Publishing Company.

Pienemann, M, Di Biase, B, Kawaguchi, S and Håkansson, G (2009) Processing constraints on L1 transfer. In Kroll, JF and De Groot, AM (eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Poulsen, M (2008) Acceptability and processing of long-distance dependencies in Danish. *Nordic Journal of Linguistics* 31(1), 73–107.

Rankin, T (2012) The transfer of V2: inversion and negation in German and Dutch learners of English. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior* 16(1), 139–158. https://doi.org/10.1177/1367006911405578

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ringbom, H (2007) *Cross-linguistic similarity in foreign language learning.* Clevedon: Multilingual Matters.

Sayehli, S, Gullberg, M, Newman, AJ and Andersson, A (2022) Native word order processing is not uniform: An ERP study of verb-second word order. *Frontiers in Psychology* 13*,* article 668276. https://doi.org/10.3389/fpsyg.2022.668276

Schad, DJ, Vasishth, S, Hohenstein, S and Kliegl, R (2020) How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110, article 104038. https://doi.org/10.1016/j.jml.2019.104038

Skehan, P (1998) *A cognitive approach to language learning*. Oxford: Oxford University Press.

Slotorub, F and Moreira, NJ (2011) *Puls 3* (1st edition). Dyssegård: Alfabeta.

Slotorub, F and Moreira, NJ (2014) *Puls 2* (3rd edition). Dyssegård: Alfabeta.

Søby, KF (to appear) *My mistake! Production and perception of grammar anomalies.* Ph.D. thesis. University of Copenhagen.

Søby, KF and Kristensen, LB (2019) Hjælp! Jeg har mistede min yndlings rød taske. Et studie af grammatikafvigelser. *Ny forskning i grammatik* 26, 89–104.

Søby, KF, Ishkhanyan, B and Kristensen, LB (2023) Not all grammar errors are equally noticed: error detection of naturally occurring errors and implications for eye tracking models of everyday texts. *Frontiers in Psychology.*

Søby, KF, Milburn, E, Kristensen, LB, Vulchanov, V and Vulchanova, M (2023) In the native speaker's eye: Online processing of anomalous learner syntax. *Applied Psycholinguistics*, 1–28.

Tenfjord, K, Jarvis, S and Golden, A (2017) Introduction. In Golden, A, Jarvis, S and Tenfjord, K (eds.), *Crosslinguistic influence and distinctive patterns of language learning*. Bristol: Multilingual Matters, pp. 1–11.

Thorborg, L and Riis, MJ (2010) *På vej til dansk – trin for trin* (5th edition). Albertslund: Synope.

Vikør, LS (2015) *Norwegian: Bokmål vs. Nynorsk.* The Norwegian Language Council. Retrieved from http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk

Westergaard, M (2003) Unlearning V2: Transfer, markedness and the importance of input cues in the acquisition of word order in English by Norwegian children. In Foster-Cohen, S and Pekarek-Doehler, S (eds.), *EUROSLA yearbook* (vol. 3). Amsterdam: John Benjamins Publishing Company, pp. 77–101.

Westergaard, M, Vangsnes, Ø and Lohndal, T (2017) Variation and change in Norwegian *wh*-questions: The role of the complementizer *som*. *Linguistic Variation* 17(1).

Westman, M (1974) *Bruksprosa*. Lund: Gleerups.

Yeaton, J (2019) *Visual processing of verb-second (V2) word order in second language acquisition: ERP Evidence from French-Swedish successive bilinguals.* Master thesis. Lund University. https://jeremyyeaton.github.io/papers/Yeaton-2019-MasterThesis-V2_word_order.pdf

Appendix

Table 5. Overview of learners' L1, DP and module. Slashes mark that learners have indicated to have multiple L1s.

L1	**DP2, module 3**	**DP3, module 2**	**DP3, module 3**	**DP3, module 4**	**Total**
English	4	16	11	7	38
Spanish	1	9	4	2	16
German		7	3	3	13
Portuguese	1	6	3	2	12
Russian		6	3	2	11
Italian		7	1	2	10
Polish		6	2	2	10
Dutch		2	4	1	7
Arabic	1	5			6
French		4	2		6
Greek		3	2	1	6
Lithuanian		4	2		6
Romanian		2	4		6
Hungarian		2	2	1	5
Bulgarian		3	1		4
Turkish		2	1	1	4
Bahasa Indonesia		3			3
Finnish		3			3
Malayalam		3			3
Thai		2	1		3
Vietnamese	1		1	1	3
Farsi (Persian)		2			2
Filipino		1		1	2
Icelandic		2			2
Chinese	1			1	2
Korean		2			2
Odia (Oriya)		1	1		2
Serbian		2			2
Tamil	1	1			2
Urdu		2			2
N = 1: Afrikaans (3.2), Albanian/Croatian (3.2), Assamese (3.2), Bangla (Bengali) (3.2), Bisaya (3.2), Cebuano (3.4), English/Italian (3.2), Gujarati (3.2), Hindi (3.2), Cantonese (3.2), Croatian (3.2), Lamnso (3.4), Lettish (3.3), Luganda (3.2), Nandi (3.2), Nepali (3.2), Sindhi (3.2), Somali (3.2), Tagalog (3.2), Telugu (3.2), German/Russian (3.2), Ukranian/Russian (3.3), Hungarian/Slovak/German (3.3), Unknown (3.2)					
Total	**10**	**127**	**51**	**29**	**217**

Table 6. Distribution on CEFR level and L1 (V2 vs. non-V2) for the 153 learners producing declarative sentences with non-initial subjects.

	Non-V2 learners (N) with XVS/XSV	**V2 learners (N) with XVS/XSV**	**Total**
A2	79 (58 %)	6 (35 %)	85
A2+	32 (24 %)	7 (41 %)	39
B1	25 (18 %)	4 (24 %)	29
Total	**136 (100%)**	**17 (100 %)**	**153**

Figure 2. Scatterplot of all declarative sentences (raw data), showing the distribution of V2 vs. V3 sentences (y-axis) on CEFR levels (x-axis) and L1s (V2 = blue triangle, non-V2 = red dot).

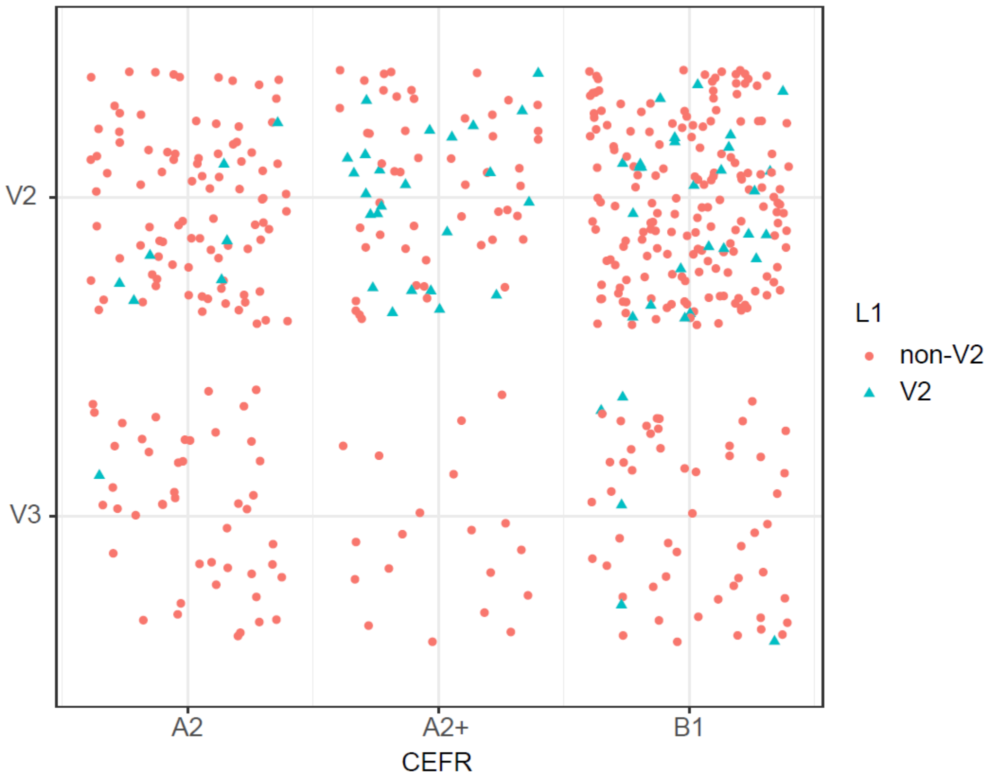
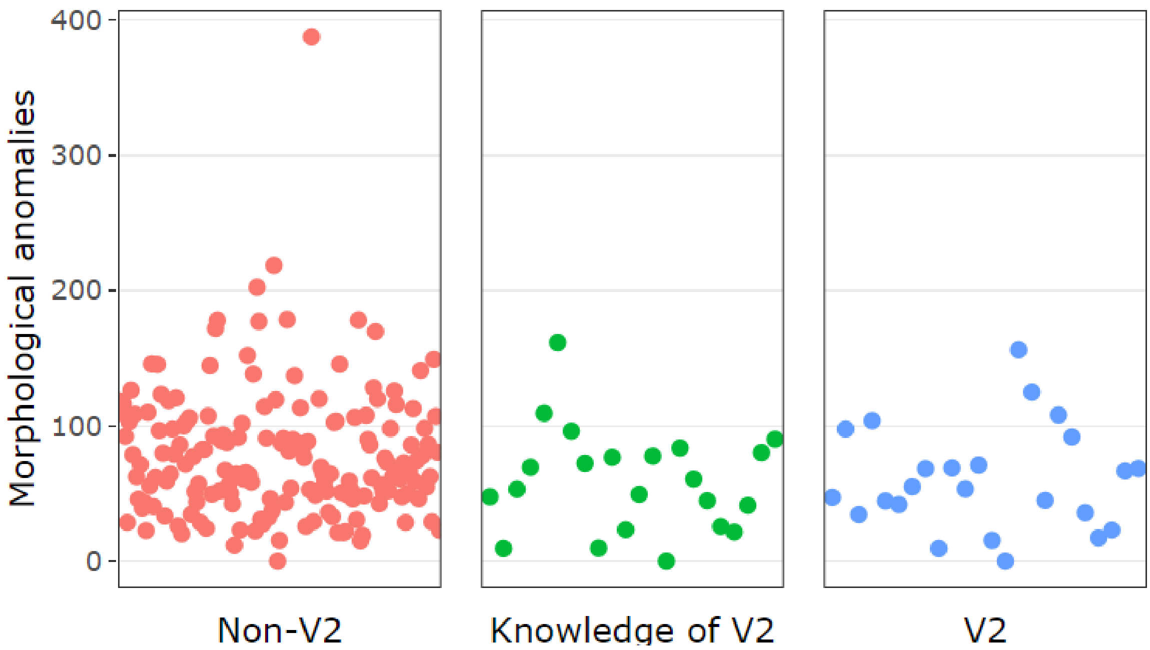


Table 7. Model results when all first constituents containing a subordinate clause are removed (number of observations = 410, participants = 142). Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	3.161	1.778		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	2.0942	0.8999	2.327	0.0200 *
Words in first position (N)	-0.1794	0.1485	-1.208	0.2272
Subject (one vs. multiple words)	-0.8336	0.5062	-1.647	0.0996 .
Verb (single vs. complex)	-0.3566	0.3419	-1.043	0.2968
CEFR (1-3)	0.3572	0.2785	1.283	0.1996
L1 (V2 vs. non-V2)	2.3877	1.0566	2.260	0.0238 *

Figure 3. The number of morphological anomalies per 1000 running words, for each participant in the three learner groups: learners with no previous knowledge of V2 languages (red), learners who have knowledge of V2 from previously acquiring a V2 language (green) and learners with a V2 language as their L1 (blue).



Article 2: Not all grammar errors are equally noticed: error detection of naturally occurring errors and implications for eye-tracking models of everyday texts

Published in *Frontiers in Psychology*, 2023, vol. 14
https://doi.org/10.3389/fpsyg.2023.1124227

The supplementary material for the article is found in Appendix A (section 10).

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 July 2023
DOI 10.3389/fpsyg.2023.1124227

Check for updates

OPEN ACCESS

EDITED BY
Lena Ann Jäger,
University of Potsdam, Germany

REVIEWED BY
Vsevolod Kapatsinski,
University of Oregon, United States
Filip Dechterenko,
Academy of Sciences of the Czech Republic (ASCR), Czechia
Alex Warstadt,
New York University, United States

*CORRESPONDENCE
Katrine Falcon Søby
✉ kafs@kp.dk

RECEIVED 14 December 2022
ACCEPTED 22 May 2023
PUBLISHED 13 July 2023

CITATION
Søby KF, Ishkhanyan B and Kristensen LB (2023) Not all grammar errors are equally noticed: error detection of naturally occurring errors and implications for eye-tracking models of everyday texts.
Front. Psychol. 14:1124227.
doi: 10.3389/fpsyg.2023.1124227

Not all grammar errors are equally noticed: error detection of naturally occurring errors and implications for eye-tracking models of everyday texts

Katrine Falcon Søby*, Byurakn Ishkhanyan and Line Burholt Kristensen

Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark

Grammar errors are a natural part of everyday written communication. They are not a uniform group, but vary from morphological errors to ungrammatical word order and involve different types of word classes. In this study, we examine whether some types of naturally occurring errors attract more attention than others during reading, measured by detection rates. Data from 211 Danish high school students were included in the analysis. They each read texts containing different types of errors: syntactic errors (verb-third word order), morphological agreement errors (verb conjugations; gender mismatches in NPs) and orthographic errors. Participants were asked to underline all errors they detected while reading for comprehension. We examined whether there was a link between the type of errors that participants did not detect, the type of errors which they produce themselves (as measured in a subsequent grammar quiz), and the type of errors that are typical of high school students in general (based on error rates in a corpus). If an error is infrequent in production, it may cause a larger surprisal effect and be more attended to. For the three subtypes of grammar errors (V3 word order, verb errors, NP errors), corpus error rates predicted detection rates for most conditions. Yet, frequency was not the only possible explanation, as phonological similarity to the correct form is entangled with error frequency. Explicit grammatical awareness also played a role. The more correct answers participants had in the grammar tasks in the quiz, the more errors they detected. Finally, we found that the more annoyed with language errors participants reported to be, the more errors they detected. Our study did not measure eye movements, but the differences in error detection patterns point to shortcomings of existing eye-tracking models. Understanding the factors that govern attention and reaction to everyday grammar errors is crucial to developing robust eye-tracking processing models which can accommodate non-standard variation. Based on our results, we give our recommendations for current and future processing models.

KEYWORDS

grammar, error detection, proofreading, production, processing models

1. Introduction

Everyday texts, whether it is an email to a colleague or a high school essay, are rarely edited. Such texts often contain grammar errors like anomalous use of word order and lack of agreement between verb and subject (Lunsford and Lunsford, 2008). Attention to these errors is not uniform. In some cases, readers react to the error. In other cases, the error goes by unnoticed. This variation in the reader's attention and response to errors poses a challenge to existing models of eye movement control in reading, such as E-Z Reader (Reichle et al., 2003) and SWIFT (Engbert et al., 2005). Enhancing our understanding of the factors that govern attention and reaction to everyday grammar errors is necessary for developing robust models of eye movement control (Søby et al., 2023). We need models that take into account variation in the type of naturally occurring grammar anomalies that occur in non-standard language and variation in the reader's grammatical awareness and proficiency, as both these factors may modulate attention and eye movements.

Differential attention to language errors has been examined in previous studies using different methods. Proofreading studies show that attention is not equally distributed between different types of language errors (Hacker et al., 1994; Shafto, 2015). Typos like *toujousr* for *toujours* attract more attention than grammar errors, which again attract more attention than orthographic errors with phonological similarity to the correct form, e.g., *essentiellemment* for *essentiellement* (Larigauderie et al., 2020).

Change blindness studies also provide evidence for differential attention allocation. In this paradigm, a participant reads two almost identical sentences, one after another, and responds to whether the two sentences are identical or not. Only one word is changed from the first display of the sentence to the second. Change blindness studies show that readers attend more to changes in lexical elements (e.g., full verbs and demonstrative pronouns) than to changes in grammatical elements (e.g., auxiliaries and articles; Christensen et al., 2021) and that readers attend more to changes in focused words than in non-focused words (Sturt et al., 2004).

Across EEG and eye-tracking studies, the difference between syntactic and semantic/pragmatic anomalies is well-documented (Ainsworth-Darnell et al., 1998; Ni et al., 1998; Braze et al., 2002; Hahne and Friederici, 2002; Hagoort, 2003). Grammar errors, however, are usually treated as a homogenous group, although grammar errors involve various subtypes (word order errors, verb agreement errors, gender mismatch errors etc.) which are not necessarily noticed to the same degree or not necessarily processed in the same way. With the present study we ask, if sensitivity to different kinds of grammar errors differs too, and what the consequences are for existing models of eye movement control in reading.

Using an error detection paradigm, we study the differences in attention to different types of naturally occurring grammar errors in written Danish. Some error types involve attention to confusion of large elements (e.g., word order errors), while others involve smaller segments at the level of words, suffixes and letters. Some errors appear initially in a sentence. Other errors have a medial or final position. Some grammar errors have phonological similarity with the correct form, and others are distinct. Many of these factors co-vary in naturally occurring errors and cannot be completely disentangled. In our study, we focus on how error type, error frequency in written production and phonological similarity to the correct form affect readers' perception of and attention to grammar errors in Danish. For word order errors, we also consider the position of the misplaced word in the sentence.

Previous studies of writers' spelling accuracy show that exposure to incorrectly spelled words tends to negatively influence later spelling accuracy for those same words (Jacoby and Hollingshead, 1990). Building on these findings, we propose that previous exposure to specific types of incorrectly inflected or misplaced words may also affect attention to this specific type of grammar errors during reading. We also examine the relationship between the type of errors that young readers tend to overlook in texts, the type of errors these young readers produce themselves (when performing a grammar quiz), and the type of errors that are typical of their age group in general (based on corpus studies of naturally occurring texts). Some grammar errors in our study represent types of errors that frequently occur in Danish high school essays. Other errors are less typical of high school students, but characteristic of L2 learners of Danish. We investigate if these typical L2 grammar errors attract more attention than the grammar errors typical of high school students. Our expectation is that attention to a specific type of grammar error is not only a matter of the reader's explicit grammar awareness (as measured in the grammar quiz), but also of whether the specific type of error is common in everyday texts by native speakers. If a specific type is frequent among the peers of the reader, the reader may have more exposure to this type of error and a mental representation of the error. The reader may therefore find it less striking and be less likely to detect it than errors that are infrequent in texts written by peers.

2. Background

Our error detection study does not involve eye-tracking data, but in combination with insights from previous eye-tracking studies on processing of grammar errors, it can address shortcomings in current models of eye movement control during reading. In this section, we present previous eye-tracking studies on processing of grammar errors (section 2.1), and describe the role of grammar errors in existing models of eye movement control in reading (section 2.2). In section 2.3, we describe the error detection paradigm, and how this may contribute to research in attention during reading. We also present the error types chosen for this study. Finally, in section 2.4, we provide an overview of the main factors presumed to influence attention to errors.

2.1. Previous eye-tracking studies on processing of grammar errors

Previous eye-tracking studies of grammar errors differ with respect to language, error types, purpose of the study, and the included reading measurements. Therefore, the findings cannot be easily summarized.

First, the eye-tracking studies have been conducted in different languages (English, Hebrew and Norwegian), making it difficult to

Abbreviations: DEF, Definite; INF, Infinitive; N, Neuter; PRS, Present tense; U, Uter (common gender).

compare across studies. For example, it is difficult to compare Hebrew subject-verb gender agreement to Norwegian word order.

Second, the ungrammatical items are very different, ranging from word order errors such as *The white was cat big* (Huang and Staub, 2021), Norwegian *ASV word order instead of AVS (Søby et al., 2023) to various morphosyntactic agreement errors such as gender agreement (Deutsch and Bentin, 2001; Dank et al., 2015), subject-verb agreement (Pearlmutter et al., 1999; Lim and Christianson, 2015) or modals followed by a progressive form, e.g., *It seems that the cats will not usually eating the food we put on the porch* (Ni et al., 1998), and/or a past tense form (Braze et al., 2002).

Third, previous studies have had different reasons for including ungrammatical items. Their experimental contrasts differ and their results can be difficult to compare. Huang and Staub (2021) examined failure to notice transposition errors to enter a debate about serial vs. parallel processing. Other studies focus on the differences between pragmatic and syntactic processing (Ni et al., 1998; Braze et al., 2002), or the interrelation between semantic and syntactic factors during processing of agreement in Hebrew (Deutsch and Bentin, 2001). Other studies again have investigated the attraction phenomenon, i.e., when a word erroneously agrees with a local distractor noun instead of the head noun, e.g., *The key to the cabinets were rusty from many years of disuse* (Pearlmutter et al., 1999), both in English (Lim and Christianson, 2015) and for subject-predicate agreement in Hebrew (Dank et al., 2015).

Finally, the studies use different reading measurements. While some measure very early effects, such as first fixation duration (Deutsch and Bentin, 2001; Braze et al., 2002; Dank et al., 2015; Lim and Christianson, 2015; Huang and Staub, 2021; Søby et al., 2023); others do not (Ni et al., 1998; Pearlmutter et al., 1999).

Taking these reservations into account, it seems that the different types of grammar errors elicit similar responses in participants' eye movements across languages, with similar time courses. Most of the studies find more regressions out from the error, meaning that participants respond immediately. Most studies also find increased reading times, but the time course varies (see Søby et al., 2023). Very early effects are found on first fixation duration by Deutsch and Bentin (2001), Dank et al. (2015), Huang and Staub (2021), and partly by Søby et al. (2023). Other studies only find increased total durations on the critical region (Pearlmutter et al., 1999) or no reading time effects at all (Ni et al., 1998). Typically the effects of ungrammaticality quickly disappears, either in the critical or subsequent regions.

Only one of the previous eye-tracking studies has explicitly examined whether readers perceived the ungrammatical items as errors or not. Huang and Staub (2021) used readers' grammaticality judgments of each sentence to distinguish between detected and undetected errors. None of the studies have made direct comparisons between different types of grammar errors to examine whether participants elicit stronger or different reactions to some errors than others. Therefore, little is known about the factors that govern attention and reaction to different types of grammar errors. Furthermore, the ecological validity of grammar errors have not been the focus of previous studies. Errors such as transposed words are constructed for the purpose of the experiment, but infrequent in natural language, and therefore may not reflect reading processes for naturally occurring language. Understanding the factors that govern attention and reaction to different types of naturally occurring errors is a necessary prerequisite when developing robust eye-tracking models for reading everyday texts (Søby et al., 2023).

2.2. The role of grammar errors in existing models of eye movement control in reading

Attention to, and processing of, grammar errors have not been a focal point in existing models of eye movement control in reading. Existing models can be divided into two types. *Serial-attention* models share the assumptions that attention is allocated serially, and only to one word at a time, while *attention-gradient* models assume that attention is allocated as a gradient, i.e., to multiple words at a time (Warren, 2011, p. 919). The major models are the influential E-Z Reader (Reichle et al., 2003, 2009; Reichle, 2011), a serial-attention model, and SWIFT, an attention-gradient model (Engbert et al., 2005; Engbert and Kliegl, 2011). Serial-attention models are furthermore described as *cognitive control* models, because they assume that "lexical processing is the 'engine' that determines when the eyes will move from one word to the next during reading" (Reichle, 2011, p. 768), in contrast to models like SWIFT, in which cognition is assumed to play a reduced role for eye movements. For example, the signal to move the eyes forward in SWIFT is provided by an autonomous random timer.

Both E-Z Reader and SWIFT account for effects of lexical processing on eye movements, but a widely acknowledged shortcoming of both models is that they cannot account for effects of higher-level language processing on eye movements (Clifton and Staub, 2011; Warren, 2011). The issue has not been addressed in SWIFT, but for E-Z Reader, Reichle et al. (2009) added a post-lexical integration stage, which is assumed to reflect all of the postlexical processing that is required to integrate a word, *n*, into the higher-level representations which readers construct online. As exemplified by Reichle et al. (2009, p. 5f), this could be to link word *n* into a syntactic structure, to generate a context-appropriate semantic representation, and to incorporate its meaning into a discourse model. Reichle et al. (2009, p. 6) state that "the integration stage […] is a placeholder for a deeper theory of postlexical language processing during reading. Our goal in including this stage is therefore quite modest: to provide a tentative account of how […] postlexical variables might affect readers' eye movements."

In E-Z Reader ver. 10 (Reichle et al., 2009; Reichle, 2011), lexical processing of a word takes place in two stages. First, the early stage of lexical processing (or word identification), also known as L_1 or the *familiarity check,* takes place. This stage corresponds to the identification of the orthographic form of the word, assuming that "this is not full lexical access, as the phonological and semantic forms of the word are not yet fully activated" (Reichle et al., 2003, p. 452). When completed, i.e., when the feeling of familiarity concerning the word exceeds a threshold corresponding to the familiarity check, it triggers the initiation of the programming of a saccade to move the eyes to the next word (Reichle, 2011). The time required to finish the familiarity check depends on the frequency of a word and its cloze probability, defined as the proportion of subjects who are able to guess word *n*, when shown the rest of the sentence (Reichle et al., 2009:3). This predicts that frequent and/or predictable words are processed faster than infrequent and/or unpredictable words (Reichle, 2011).

We assume that the same reasoning applies to frequent and/or predictable errors, but the E-Z Reader model does not explicitly account for input with frequent vs. infrequent errors.

The later stage of lexical processing (L_2) involves the identification of the word's phonological and/or semantic forms, to enable additional linguistic processing (Reichle et al., 2003). This stage corresponds to what is typically referred to as *lexical access,* and with the completion of lexical access, attention shifts to the next word, which can now be processed. Simultaneously, post-lexical processing (i.e., integration) starts on the identified word. This post-lexical processing corresponds to the minimal amount of processing necessary to continue to move attention (and the eyes) forward through the text (Reichle, 2011, p. 776). In most cases, integration is completed without difficulty, meaning that post-lexical processing only has minimal effect on readers' eye movements. Reichle et al. (2009, p. 6) assume that complete incremental post-lexical processing is not always necessary and does not always occur, which is broadly consistent with the "good enough" view of language processing (Ferreira and Patson, 2007). However, integration difficulty may occur. When integration fails, it causes the eyes and attention to pause and/or move backwards (Reichle, 2011). Integration failures happen by default when word $n+1$ is identified before word n is integrated. Rapid integration failure can happen due to severe semantic or syntactic violations (Reichle et al., 2009). If the integration of n fails rapidly, the forward saccade to $n+1$ is canceled, which results in a pause (increasing first fixation duration and gaze duration) and/or a refixation (increasing gaze duration) or an interword regression (Reichle et al., 2009). If the integration failure of n takes place after the eyes have moved to $n+1$, i.e., fails more slowly, a regressive eye movement is made (Clifton and Staub, 2011, p. 904). Thus, the model predicts that problems with integration can have very rapid effects, influencing first-fixation duration on the word that is being integrated. This, however, only happens when the integration failure occurs before the labile stage of saccadic programming (i.e., the stage which can be canceled) to move the eyes forward in the text has completed (Reichle et al., 2009).

The assumption that contextual information (besides cloze probability) only affects postlexical integration is challenged by studies of parafoveal processing, i.e., processing of upcoming words that have been attended, but not yet fixated (Warren, 2011). For example, Veldre and Andrews (2018) used the gaze-contingent boundary paradigm to assess whether parafoveal processing of a word contributes to its subsequent identification. In this paradigm, a target word in a sentence is replaced with another word, until the reader's eyes cross an invisible boundary (e.g., before the space to the left of the target word), after which the word is changed back to the target word. Veldre and Andrews (2018) conducted two experiments, in which they compared contextually plausible previews (which either contained a morphosyntactic agreement violation or not) to implausible previews (either containing a syntactic word class violation or not). The plausible previews were not predictable from the sentence context, as measured in a cloze task. Veldre and Andrews (2018) found that the contextual plausibility and grammatical correctness of an upcoming word can affect processing, early enough to affect skipping of that word. According to the authors, the plausibility effects on skipping rates are unlikely to be reconciled with E-Z Reader's current post-lexical integration mechanisms.[1]

Furthermore, the E-Z Reader model does not address what happens when readers encounter other types of misspellings or grammar errors, besides severe syntactic violations. If the early familiarity check identifies the orthographic form of the word, it should be able to respond to orthographic errors (e.g., *posibility*), but not anomalous use of existing morphological forms (e.g., *eats* for *eat*). The model does not answer the question of why some types of errors are detected while others are not, nor the question of why readers do not always notice the same error.

Finally, Warren (2011) argues that the E-Z Reader model will be incomplete without allowing some role for even higher-level influences, based on research on semantic anomalies. Readers sometimes fail to notice semantic anomalies, suggesting that processing is sometimes shallow (Ferreira et al., 2002). "If different readers, reading for different purposes, perform post-lexical processing more or less quickly or completely […], the precise combination of reader, purpose and motivation will affect the patterns of eye movements to semantic violations" (Warren, 2011, p. 922). In our study, we examine how error detection differs between readers with differences in grammatical awareness and proficiency.

2.3. The error detection paradigm

Both the eye-tracking and error detection paradigms can be used to measure attention during reading. Here we assume that eye-tracking provides a more sensitive measure than error detection. Yet, the exact correlations between the two measures is not well-explored. It may be the case that the error detection paradigm treats two events as the same, while they involve different eye movements. Although we assume that error detection is more rough, there are several advantages to using this paradigm for our purpose. In the previous eye-tracking studies of ungrammaticality, sentences were presented individually. With error detection, we can introduce participants to long, consecutive texts, simulating natural text reading. Furthermore, we can include many different types of grammar errors, unlike previous eye-tracking studies which have included relatively few error types (e.g., pragmatic vs. syntactic). Having many different types of errors in different conditions would result in a long and tiresome eye-tracking experiment. Finally, using error detection, we can get participants' feedback on where errors occur, in a fast way, not having to ask after every trial. Although, error detection can only provide a rough measurement for attention during reading, it can provide insights into which types of errors are more noticed than others, and which other factors than error type is likely to play a role. The results are therefore relevant to future eye-tracking studies and processing models. If differences are found using error detection, they are also likely to be found using a presumably more sensitive measure such as eye-tracking.

1 Veldre and Andrews (2018) also argue that the results cannot be reconciled with the alternative *forced fixation account* of preview effects, proposed by Schotter et al. (2014b).

In our error detection study in Danish, we included one type of syntactic error (*ASV for AVS, see below) and two types of morphological errors (confusion of infinitive and present tense, and gender mismatches between articles or adjectives in NPs), as well as various common orthographic errors. These errors were chosen because they represent a broad range of error types, and they are all attested in natural L1 and/or L2 production, however with different frequencies. For example, ungrammatical verb-third word order (*ASV) instead of grammatical verb-second word order (AVS) is common in L2 Danish (Søby and Kristensen, 2019; Søby and Kristensen, to appear), but rare in L1 Danish, apart from multiethnic urban vernaculars (Quist, 2008). The three types of grammar errors naturally occur in different conditions, varying with respect to error frequency (measured as error rates in L1 production), and/or phonological similarity to the correct form, or placement in the sentence. Since the stimuli is based on naturally occurring errors, error frequency and phonological similarity tend to co-vary.

2.4. Attention to errors during reading

Many potential contributing factors besides error type might influence whether a reader reacts to an error. In this section, we elaborate on why some of the factors we are examining in our study are relevant to include, namely error frequency, phonological similarity to the correct form, and, for word order, placement in the sentence. Finally, we elaborate on the potential role of participants' own production of errors, and individual differences in error perception.

Previous letter detection studies and change-blindness studies review a wide a range of factors which can influence attention during reading (e.g., Smith and Groat, 1979; Sturt et al., 2004; Vinther et al., 2015; Christensen et al., 2021). For example, Smith and Groat (1979) found that the position on the line and in the sentence influenced detection of the letter *e*, so that the outer positions were more prominent than the middle. Using V3 errors with a length manipulation, we examine whether position effects within the sentence are also found for grammar errors.

The main focus of our study is on the role of error frequency. We hypothesize that error frequency, which is tied to the predictability of the error, predicts perception patterns. According to prediction-based approaches to sentence processing, unexpected input attracts attention (Kamide, 2008; Levy, 2008; Christiansen and Chater, 2016). If a reader sees input with common errors, the model will be updated according to the input, meaning that frequent errors should be predicted by the model, and thus should attract less attention than infrequent errors.

Besides error frequency, we expect that phonological similarity to the correct form negatively influences detection rates for grammar errors, in line with Larigauderie et al. (2020) who compared spelling errors which were either phonologically similar to or distinct from the correct form. One example from our stimuli is confusion of homophone verb pairs, such as present tense *kører* and infinitive *køre*, both pronounced [ˈkʰøːɐ]. We expect that confusion of heterophone verb pairs such as *rejser* [ˈʁɑjˀsɐ] and *rejse* [ˈʁɑjsə] will have higher detection rates. When the correct form is homophone to the error, the error is not grammatical in that context, but it is phonologically correct, and may therefore not disturb reading. For such silent errors readers may use all available cues whether they are phonological or orthographic (*cf.* Carassco-Ortiz and Frenck-Mestre, 2014). The E-Z Reader model does not account for homophony effects, but it may predict that the phonological form is more easily identified for homophone compared to heterophone errors in the later stage of lexical processing (L_2). The error frequency and phonological similarity to the correct form tend to co-vary, because L1 speakers of Danish produce more errors when for instance present tense and infinitive forms are homophone. Thus, effects of phonological similarity and frequency are often difficult to disentangle.

On top of that, individual differences are likely to influence error detection. If a type of error is frequent in a person's production, e.g., omitting the-r on verbs in present tense: **han køre* 'he drive.INF,' the rules for verbal inflection may not be fully mastered. It therefore seems likely that this person will overlook this type of error in general. Individual differences in the perception of what constitutes an error in a specific situation could also be a factor: How correct or incorrect on a continuum is an error to a specific reader? How do individual readers differentiate between unusual language and outright errors? And is the perception affected by the context in which it is read, e.g., experimental vs. natural? Our study is not equipped to answer these questions. Studies show that tolerance for various errors can be modulated by participants' perception of the speaker, so that the tolerance and willingness to repair is higher when the speaker is perceived as non-native (Konieczny et al., 1994; Hanulíková et al., 2012; Gibson et al., 2017).

In the public debate and prescriptive literature, some errors are pointed out as typical or basic errors, while other errors are much less debated or accentuated. Publically debated errors may be more prominent to readers (Blom and Ejstrup, 2019b). In Denmark, missing present tense-r is often accentuated in normative discourse. Blom and Ejstrup (2019b) found that readers' intolerance for errors are modulated by the type of error. Their participants were more annoyed with typical and basic grammar/spelling errors than with atypical and complicated errors. The missing present tense *-r* was the most annoying error. The authors also found a correlation between participants' irritation (with a specific item) and the number of errors detected, so that the more errors participants detected in general, the more irritated they were with that item.

2.5. The current study

The current study examines native speakers' attendance to different types of syntactic, morphological and orthographic errors (found in L1 and/or L2 Danish) during reading. We asked Danish high school students to read and comprehend two texts, while underlining all errors they noticed. We also tested their basic grammar skills, using a grammar quiz. The study included one type of syntactic error (V3 word order) and two types of morphological errors (confusion of infinitive and present tense, and gender mismatches between articles or adjectives in NPs), as well as various common orthographic errors. V3 errors are the least frequent, and orthographic errors the most common. In a corpus of 71 high school essays, we found 10 V3 errors, 16 gender mismatches in indefinite articles, 51 gender mismatches in adjectives, 178 confusions of infinitive and present tense, and 1,099 orthographic errors.

The study is designed as a four-in-one study. Each error type (V3, verb, NP, orthographic) constitutes its own subexperiment and appears in different conditions, controlled for a number of variables. We cannot directly compare attention to the four types, as there are too many confound variables, such as their position in the sentences and in the text. Thus, we only indirectly compare the detection rates for the three overall error categories (syntactic, morphological, orthographic) using descriptive statistics.

We examine the relationship between the type of errors that young readers tend to overlook in texts, the type of errors these young readers produce themselves (in the grammar quiz), and the type of errors that are typical of their age group in general (based on corpus studies of high school essays). Our expectation is that attention to a specific type of grammar error is not only a matter of the reader's explicit grammar awareness (as measured in the grammar quiz), but also of whether the specific type of error is common in everyday texts by native speakers. If a specific type is frequent among the peers of the reader, the reader may have more exposure to this type of anomaly and a mental representation of the error, i.e., common errors should be predicted to occur in input, based on prediction theory (Kamide, 2008; Christiansen and Chater, 2016). The reader may therefore find it less striking and be less likely to detect it than errors that are infrequent in texts written by peers, e.g., those found in L2 Danish. This means that for the overall categories of errors (syntactic, morphological and orthographic), we expect that the syntactic errors (V3 errors) have higher detection rates than morphological and orthographic errors, because V3 errors are rare in L1 writing (and are visually large). We also expect readers to overlook orthographic errors the most, because orthographic errors are highly frequent in the L1 writing.

Finally, for the two morphological subtypes of grammar errors (confusion of infinitive and present tense, and gender mismatches between articles or adjectives in NPs), we examine how error frequency and phonological similarity to the correct form may affect attention to errors. For the word order errors, we examine position effects within the sentence. The specific conditions and hypotheses for the three subtypes of grammar errors are presented in the results section where they are treated as three subexperiments. The fourth subexperiment on different types of orthographic errors is primarily included to create variation in the stimuli.

3. Methods

3.1. Participants

The participants were recruited from three different Danish upper secondary education programs (STX, HTX, and HHX).[2] Data were collected in August 2019 at six schools located in and around Copenhagen and Roskilde. Two hundred and forty students from 10 classes participated. We excluded participants with dyslexia (18), with late acquisition of Danish (>6 years, Hyltenstam and Abrahamsson, 2003) (2), or participants who misunderstood or did not finish the reading task (9). This left 211 participants in the analysis (98 women, 113 men), 17–20 years of age (M = 18.31 years; SD = 0.67 years). The majority were part of the STX Program (130), followed by HHX (43), and HTX (38). All participants (or their parents) gave informed written consent prior to the experiment. The study was approved by local research ethics committee at University of Copenhagen, and followed GDPR.

3.2. Experimental tasks and materials

The experimental tasks consisted of a reading task (section 3.2.1) which was followed by a grammar quiz and a questionnaire (section 3.2.2). All test materials are found in Supplementary material (section 3).

3.2.1. Reading task

The reading task consisted of two texts, A (689–692 words) and B (831–832 words). Every participant read both texts. There were four versions of the reading task material to ensure that each participant only saw the same item in one condition. That is, when reading the same sentence in the text, participants reading version 1 were presented with the verb error in one condition, participants reading version 2 were presented with it in another condition, etc. Each participant was presented with a total of 100 errors in text A and B together. Table 1 shows the distribution on subtypes. To avoid priming effects, target items did not occur elsewhere in the texts.

A further description of the stimuli is presented in the sections on each subexperiment. We varied the order of text A and B, so that half of the participants read A before B, and the other half read B before A. Thus, there were eight versions of the reading task in print.

3.2.2. Questionnaire and grammar quiz

The questionnaire addressed the participants' language and dialectal background as well as their attitude to language errors. The purpose of the grammar quiz was to ensure that the participants had the basic grammatical prerequisites to notice errors in the reading task. The grammar quiz included tests on all four types of errors, i.e., verb-second word order after sentence-initial adverbials, verb conjugations in infinitive and present tense, conjugation of adjectives, gender of indefinite articles, and spelling of the four types of target words. Most of the tasks were forced-choice between two options.

3.3. Procedure

The participants were informed that the study was about speed-reading and what readers notice when skimming a text. In the reading task, their task was to underline language errors. Participants had max. 7 min to read each text (A and B). Participants were instructed to skim as fast as possible and finish reading the whole text so they could answer the comprehension questions. Whenever they noticed a language error, they should underline it, but they should avoid going back in the text. Language errors were defined as different types of spelling and grammar errors, but not punctuation. They were instructed to underline the whole word containing the error, or multiple words if they were in the

2 The three education programs (STX, HTX, and HHX) all prepare for higher education, but have different profiles. STX is a general examination program, HTX is a technical examination program with a STEM profile and HHX is a commercial examination program with a business profile (Ministry of Higher Education and Science, 2022).

TABLE 1 Error types, conditions and number of target items in the reading task (text A+B).

Error types		Items
V3 errors (2 conditions, 8 items per condition)		**16[a]**
1) After short adverbial:	*og kl. 14 <u>han ankommer</u> til Berlin* and o'clock 2 he arrive.PRS in Berlin 'and at 2 o'clock, <u>he arrives</u> in Berlin'	8
2) After long adverbial:	*og først ud på eftermiddagen <u>han ankommer</u> til Berlin* and first out on afternoon.DEF he arrive.PRS in Berlin 'and first in the afternoon, <u>he arrives</u> in Berlin'	8
Verb errors (4 conditions, 8 items per condition)		**32**
1) Homophone; Present tense for infinitive:	*han vil <u>kører</u>* [ˈkʰøːɐ] he will <u>drive.PRS</u> 'he'll drive'	8
2) Homophone; Infinitive for present tense:	*han <u>køre</u>* [ˈkʰøːɐ] he <u>drive.INF</u> 'he drives'	8
3) Heterophone; Present tense for infinitive:	*han vil <u>rejser</u>* [ˈʁɑjˀsɐ] he will <u>travel.PRS</u> 'he'll travel'	8
4) Heterophone; Infinitive for present tense:	*han <u>rejse</u>* [ˈʁɑjsə] he <u>travel.INF</u> 'he travels'	8
NP errors (4 conditions, 8 items per condition)		**32**
1) Mismatch ADJ + N; Uter for neuter:	*et <u>dejlig</u> kæledyr* ART.N lovely-U pet.N 'a <u>lovely</u> pet'	8
2) Mismatch ADJ + N; Neuter for uter:	*en <u>dejlig-t</u> undulat* ART.U lovely-N budgie.U 'a <u>lovely</u> budgie'	8
3) Mismatch ART + N; Uter for neuter:	*<u>en</u> dejlig-t kæledyr* ART.U lovely-N pet.N '<u>a</u> lovely pet'	8
4) Mismatch ART + N; Neuter for uter:	*<u>et</u> dejlig undulat* ART.N lovely-U budgie.U '<u>a</u> lovely budgie'	8
Misspellings (4 types — 5 of each type)		**20[b]**
1) Missing double consonant, e.g., *startsku<u>d</u>et* for *startskuddet* 'the starting signal'		5
2) Split compounds, e.g., *by vandring* for *byvandring* 'city walk'		5
3) Missing silent letter, e.g., *siste* [ˈsisd̥ə]/[ˈsisd̥] for *sidste* [ˈsisd̥ə]/[ˈsisd̥] 'last'		5
4) Reduction of syllable, e.g., *virklig* [ˈʋiɐ̯g̊li] for *virkelig* [ˈʋiɐ̯g̊li] 'really'		5
Total		**100**

[a]The V3 errors in version 1 + 2 were identical. The V3 errors in version 3 + 4 were also identical.
[b]The 20 spelling errors were identical in all four versions of the reading task.

wrong order. Underlinings could be canceled with a vertical line. Use of dictionaries and online tools were not allowed.

The researcher registered the starting time and gave statuses on remaining time. When the students finished reading the text, they wrote the finishing time and put the text away (if they did not finish, they marked how far in the text they got). The same procedure was repeated for the second text. Finally, the students completed the comprehension questions for both texts, the questionnaire and the grammar quiz. The whole session lasted around 45 min.

4. Analysis

The error detection data were analyzed with general linear mixed effects models for binomial data in RStudio (R Core Team, 2022, version 2022.07.1), using the lme4 package (Bates et al., 2015, ver. 1.1.30). *p*-values were obtained using the lmerTest package (Kuznetsova et al., 2017, ver. 3.1.3). The dependent variable for all models was detection, i.e., whether the error was detected (=1) or not (=0). We did not penalize false hits. The conditions for each of the four error types were included in the models as fixed effects (p is the probability of correctly detecting an error):

1. Model for V3 errors: log(p/1-p)[3] = Adverbial length [short vs. long] + Total grammar score + (1|Participant) + (1|Item) + Residuals
2. Model for Verb errors: log(p/1-p) = Type [infinitive for present tense vs. present tense for infinitive]*Homophony [homophone vs. heterophone pairs] + Total grammar score + (1|Participant) + (1|Item) + Residuals
3. Model for NP errors: log(p/1-p) = Type [agreement with article vs. adjective]*Gender [uter for neuter vs. neuter for uter] + Total grammar score + (1|Participant) + (1|Item) + Residuals
4. Model for orthographic errors: log(p/1-p) = Type [four different] + Spelling score + (1|Participant) + (1|Item) + Residuals

All models included random intercepts for participant and item. All models also included the scores from the grammar quiz. Participants made few wrong answers in the grammar tasks, so we summarized the results from the individual grammar-related tasks and included a total grammar score as a fixed effect in the models for detection of the three types of grammar errors. The model for orthographic errors included the score from the spelling task in the quiz as a fixed effect.

The models for the four error types did not include random slopes, presentation order (i.e., placement in the text) or irritation scores, as the models failed to converge when they were included. Only one subtype, NP errors, showed an uninterpretable effect of presentation order.

The output of the regression model was in logodds space. To increase interpretability, they were converted back to probabilities and plotted. Thus, the plots for the morphological errors show the models' predicted probabilities of detecting the target.

Finally, we made a general model, collapsing all error subtypes, with accuracy in percentage as the dependent variable, only including irritation scores as a fixed effect (see normal Q-Q plot in Supplementary Figure 3):

5. Model for all errors: accuracy (%) = Irritation score + Residuals.

5. General results

The participants detected 54% of all errors in the two texts (Table 2). As expected, the highest detection rate was found for syntactic errors (71% of all items were detected), followed by the two types of morphological errors (55% detected for NP errors; 59% for verb errors), and the lowest rate was found for orthographic errors (33%). The study is not designed to directly compare these overall categories (syntactic, morphological and orthographic), as there are a number of confounds, such as their position in the sentences and in the text. We therefore do not conduct any statistical tests between them. More detailed results are presented in the sections on each of the four error types (subexperiments).

5.1. Individual variation

As seen in Figure 1, there was individual variation among the participants, with respect to the number of words they underlined, and the share of correct (hits) vs. incorrect underlinings (false alarms). Out of 321,145 words, participants underlined 18,041 words (M = 85,50 words, SD = 31,38 words, range = 9–227 words). Of these only 2,565 words were not part of a target, i.e., false alarms (M = 12,16 words, SD = 13,59 words, range = 0–108). In total, 11,490 targets were underlined, i.e., hits (M = 54,45 words, SD = 21,32 words, range = 1–92 words). Notice that a target can consist of several words (targets are defined in the sections on the subexperiments).

In principle, participants could underline all words in the text and thus detect all errors, resulting in the highest possible score. This, however, was not an issue in general as participants only underlined

3 If p is the probability of detecting an error, 1 − p is the probability of not detecting an error. p/1 − p is the odds of detecting an error and log(p/1 − p) is the logarithm of the odds (logodds).

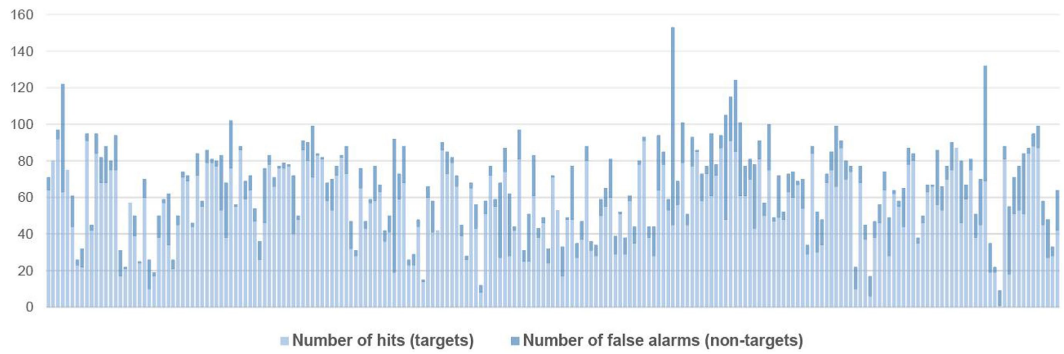


FIGURE 1
Number of underlinings (hits and false alarms) per participant.

0.8% non-target words in the texts (2,565 out of 321,145 words). Figure 1 shows that most participants were relatively exact in their underlinings, apart from 10 participants who had more false alarms than hits.

In the grammar quiz, participants generally made few errors (see sections on subexperiments). In the three grammar tasks (word order, NP agreement and verb conjugations), the highest possible score was 17, one point for each correct answer. Participants' scores had an average of 16.76 (SD = 0.67, range: 11–17). The Supplementary material (section 1.2) include a plot of the total quiz scores (grammar and spelling tasks) and the number of detected errors per participants.

The general model of all error types (5) included the participants' irritation scores (*cf.* Supplementary Table 13). We found a small effect of irritation ($\hat{\beta} = 1.82$, SE = 0.40, $t = 4.57$, $p < 0.001$), so that the more annoyed participants state to be with language errors, the more errors they detected in the reading task (see plots in Supplementary material, section 1.2).

6. Subexperiments

In the following sections, we present the hypotheses, stimuli and results for each of the four subtypes of errors. Sections 6.1–6.3 describe the three subexperiments on grammar errors. Section 6.4 describes the subexperiment on orthographic errors. The Supplementary material show all stimuli (section 2) and model results for the orthographic errors (section 1.1).

For the grammar errors, we start each section with information on error frequencies in L1 production. The error frequencies are based on a corpus of 71 high school essays from a final exam (127,957 words; 71 participants). For the morphological errors, we calculated the error rate by dividing the number of incorrect tokens with the number of correct and incorrect tokens. As an example, when a reader sees a verb in present tense, the error rate reflects how often the verb is incorrect. For the orthographic errors, the error rate is calculated by dividing the number of errors with the number of words in the corpus. For the syntactic errors, we report the absolute number of errors. Since there was a limited number of tokens for certain types of errors, we only use descriptive statistics (not inferential statistics) when accessing differences in error frequency.

TABLE 2 Number of errors in texts and share of detected errors.

Category type	Errors in texts (*N*)	Detected targets (*N*)	Share of detected targets (%)
Syntax			
V3	3,376	2,398	71.03%
Morphology			
Verb errors	6,752	3,992	59.12%
NP errors	6,752	3,719	55.08%
Orthography			
Misspellings	4,220	1,381	32.73%
Total	**21,100**	**11,490**	**54.45%**

6.1. V3 errors

A common word order error in L2 Danish is placing the verb in third position (V3), instead of second (V2; Søby and Kristensen, to appear). In (1a), the adverbial *nu* 'now' is placed in first position, followed by the subject *jeg* 'I' in second position, and the verb *bor* 'live' in third position. In the corrected version of the sentence in (1b), the verb is correctly placed in second position (the mandatory position for finite verbs in Danish main clauses).

(1)	a. [original]	**Men nu* <u>*jeg bor*</u> *i Denmark*
		'but now I live in Denmark'
	b. [corrected]	*Men nu bor jeg i Danmark*
		'but now live I in Denmark'

In the L1 corpus of high school essays, we only found 10 V3 errors. V3 errors are generally not considered typical L1 errors, but may occur in informal texts written by speakers of multiethnic urban vernaculars (Quist, 2008).

We expected these errors to be highly noticed by native speakers for two reasons. First, they are rare in L1 production. Second, large elements, i.e., entire words, are misplaced. In the experiment, the V3 errors were either presented after a short sentence-initial adverbial (1–2 words, consisting of 5–12 characters including spaces) or a long adverbial (4–6 words, 26–39 characters). In L2 Danish, V3 word order most frequently occurs after adverbials, both short and long (Søby and Kristensen, to appear). Examples of the stimuli are shown in Table 3. Previous letter detection studies have found position effects, so that elements in the start or end of a sentence tend to be more prominent than

TABLE 3 Conditions, number of V3 errors in texts and share of detected errors.

Conditions	Errors in texts (*N*)	Detected targets (*N*)	Share of detected targets (%)
Short A *og kl. 14* <u>*han ankommer*</u> *til Berlin* and o'clock 2 he arrive.PRS in Berlin 'and at 2 o'clock, <u>he arrives</u> in Berlin'	1,688	1,200	**71.09%**
Long A *og først ud på eftermiddagen* <u>*han ankommer*</u> *til Berlin* and first out on afternoon.DEF he arrive.PRS in Berlin 'and first in the afternoon, <u>he arrives</u> in Berlin'	1,688	1,198	**70.97%**

TABLE 4 Model (1) estimates for V3 errors.

Random effects	Variance	Std. dev.		
Participant (intercept)	1.7076	1.3068		
Item (intercept)	0.4177	0.6463		
Fixed effects	**Estimate**	**Std. error**	***z*-value**	***p*-value**
(Intercept)	−10.91748	2.70189	−4.041	5.33e-05***
Length	−0.03394	0.08865	−0.383	0.702
Total grammar score (quiz)	0.72652	0.16094	4.514	6.36e-06***

Dependent variable: detection (1 = error detected, 0 = error not detected). Significance code: ***$p < 0.001$.

TABLE 5 Error rates in L1 texts, confusion of present tense and infinitive (N=194).

Type and error rates	Homophone	Heterophone
	e.g., *køre(r)* [ˈkʰøːɐ]	e.g., *rejse* [ˈʁɑjsə], *rejser* [ˈʁɑjˀsɐ]
Target form: present tense 1% errors (12,764 correct present tense verbs[1])	25% (N = 96)	0.30% (N = 35)
Target form: infinitive 1% errors (4,689 correct infinitives[1])	8.60% (N = 37)	1.10% (N = 10)

[1]Found using an automatic POS tagger [Centre for Language Technology, University of Copenhagen (CST), 2022], manually tagged for homophony.

in the middle (Smith and Groat, 1979). We therefore examined whether participants would detect more V3 errors after a short adverbial than a long adverbial.

The target verbs were all in present or perfect tense, and subjects were either pronouns, proper names or nouns in the definite form, with varying lengths. The texts also included 16 similar correct constructions with AVS, i.e., V2 word order (8 after short adverbials; 8 after long). All stimuli can be seen in Supplementary material (section 2).

The V3 errors were considered detected when either the adverbial, subject or verb was underlined by a participant, since the order of subject and verb would be correct if the adverbial was placed elsewhere. In Table 3, the number and share of detected targets are seen. There were no effects of adverbial length ($\hat{\beta}$ = −0.03, SE = 0.09, $z = -0.38$, $p = 0.70$), but there was an effect of total grammar score ($\hat{\beta}$ = 0.73, SE = 0.16, $z = 4.51$, $p < 0.001$; *cf.* Table 4). The higher grammar score in the quiz, the more V3 errors were detected. In the grammar quiz, participants had to place words in the correct order after conjunctions and adverbials. Out of 633 answers, only 3 were wrong (0.5%), confirming that V3 is not a typical L1 error.

6.2. Verb errors

Confusion of finite and infinite verb forms is the most frequent morphological error in the L1 corpus. More specifically, there are 181 cases of confusion of infinitive and present tense in the L1 corpus. When examining these, the error frequency seems influenced by phonological similarity (Table 5). L1 speakers produce more errors when the two verb forms are homophone (e.g., infinitive *køre* [ˈkʰøːɐ] and present tense *kører* [ˈkʰøːɐ]) than when the verb forms are heterophone (e.g., infinitive *rejse* [ˈʁɑjsə] and present tense *rejser* [ˈʁɑjˀsɐ]). This is both the case when examining the total number of errors and the error rates. For example, the error rate for using infinitive for present tense (homophone verb pairs) is 25%, i.e., out of all correct verbs in present tense (with the same pronunciation in infinitive) plus the cases where infinitive is used for a homophone present tense form, 25% are erroneous. L1 speakers also produce more errors of the type infinitive for present tense (132) than present tense for infinitive (49), i.e., they leave out an *-r* in writing. However, the error rates for the two types of confusion are both 1%, because there are more verbs in present tense in the corpus.

Based on error rates (which are entangled with phonological similarity), we expected that participants would detect more errors in the heterophone than homophone conditions. We did not expect differences between the two types of target forms (whether the target was infinitive or present tense), as there was no difference in error rates. Finally, the error rates in Table 5 also show a larger difference between the homophone and heterophone conditions when the target is present tense, compared to when the target form is infinitive. This predicts an interaction between homophony and type.

Table 6 shows the four experimental conditions for the verb errors. We used a 2 (heterophone vs. homophone) × 2 (target infinitive vs. present tense) design. Notice, that there is a visual difference between the two types of errors, because in one condition (present tense for infinitive), an extra *-r* is added, while an *-r* is missing in the other condition (infinitive for present tense). The heterophone vs. homophone verb pairs were controlled for length (number of letters in infinitive) and frequency. *T*-tests (correlated samples) showed no significant differences in length or frequency [Det Danske Sprog- og Litteraturselskab (DSL), 2022] for the homophone vs. heterophone verbs. The texts also included a minimum of 32 correct verbs (other lexemes), 8 in each condition. All stimuli can be seen in the Supplementary material (section 2).

Table 6 also shows the number and share of detected targets. In the condition present tense for infinitive, a target is considered detected if either the modal and/or the main verb is underlined.

As expected (based on error rates and phonological similarity), we found an effect of homophony ($\hat{\beta}$ = −1.21, SE = 0.09, $z = -13.38$, $p < 0.001$), so that participants detected more errors in heterophone than homophone pairs. Counter to the expectation based on error rates, we found an effect of type, so that more errors of the type infinitive for present tense were found, than for present tense for infinitive ($\hat{\beta}$ = −0.20, SE = 0.09, $z = -2.23$, $p < 0.05$). There was no interaction, contrary to the predictions based on error rates (*cf.* Table 7).

Figure 2 shows the model's predicted probability of responding correctly (i.e., detecting the error) in the different conditions. The probability of a correct answer (a detected error) is much higher in the heterophone than homophone conditions. Although, the effect of type

TABLE 6 Conditions, number of verb errors in texts and share of detected errors.

Conditions	Errors in texts (*N*)	Detected targets (*N*)	Share of detected targets (%)
HETEROPHONE PAIRS	**3,376**	**2,306**	**68.31%**
INFINITIVE FOR PRESENT TENSE: *han* <u>*rejse*</u> [ˈʁɑjsə] he <u>travel.INF</u>	1,688	1,178	69.79%
PRESENT TENSE FOR INFINITIVE: *han vil* <u>*rejser*</u> [ˈʁɑjˀsɐ] he will <u>travel.PRS</u>	1,688	1,128	66.82%
HOMOPHONE PAIRS	**3,376**	**1,686**	**49.94%**
INFINITIVE FOR PRESENT TENSE: *han* <u>*køre*</u> [ˈkʰøːɐ] he <u>drive.INF</u>	1,688	867	51.36%
PRESENT TENSE FOR INFINITIVE: *han vil* <u>*kører*</u> [ˈkʰøːɐ] he will <u>drive.PRS</u>	1,688	819	48.52%
Total	**6,752**	**3,992**	**59.12%**

TABLE 7 Model (2) estimates for verb errors.

Random effects	Variance	Std. dev.		
Participant (intercept)	2.7768	1.6664		
Item (intercept)	0.2204	0.4695		
Fixed effects	**Estimate**	**Std. error**	***z*-value**	***p*-value**
(Intercept)	−10.75433	2.99431	−3.592	0.000329***
Homophony	−1.20594	0.09014	−13.378	<2e-16***
Type	−0.20145	0.09025	−2.232	0.025614*
Homophony*type (Interaction)	0.01925	0.12440	0.155	0.877033
Total grammar score (quiz)	0.71993	0.17850	4.033	5.5e-05***

Dependent variable: detection (1 = error detected, 0 = error not detected). Significance codes: ***$p < 0.001$, *$p < 0.05$.

was significant, the plot shows that it is small. Also, according to the predictions based on error rates, the column with *han køre* should have been the smallest.

Finally, we found an effect of total grammar score ($\hat{\beta}$ = 0.72, SE = 0.18, $z = 4.03$, $p < 0.001$), so that the higher total grammar score in the quiz, the more verb errors were detected. The grammar quiz contained 8 sentences where participants made a forced choice between infinitive or present tense for a missing verb. Out of 1.688 answers, there were only 25 errors (1.5%), made by 16 students. Twenty-two of 25 errors were in homophone verb pairs, supporting the role of phonological similarity on error production.

6.3. NP errors

In Danish, nouns are either uter (most common) or neuter gender. There are two indefinite articles, *en* (uter) and *et* (neuter) 'a.' Adjectives are inflected for gender, definiteness, and number. Typically, the suffix *-t* 'neuter,' *-e* 'definite,' or *-e* 'plural,' can be added to the uninflected basic form, corresponding to singular, indefinite, uter gender (Becker-Christensen, 2010). The most common adjective error in the L1 corpus is to leave out a suffix (*-t* or *-e*). Table 8 shows error rates for gender mismatches in adjectives and indefinite articles. Confusing the two indefinite articles is less common than missing gender agreement in adjectives, as seen in the error rates. Using uter for neuter is slightly more common than using neuter for uter.

Based on the error rates, we expected higher detection rates for mismatching articles than for mismatching adjectives, and higher detection rates for neuter for uter more than uter for neuter. The error rates in Table 8 show a slightly larger gender difference for adjectives than for articles, and we therefore predicted an interaction between word class and gender.

The four experimental conditions for the NP errors are seen in Table 9 (2 × 2 design). In continuous speech, there is phonological similarity between the correct and incorrect form in the condition mismatch with adjective, uter for neuter (where the suffix is missing). Notice, that there are also visual differences between the two word class conditions: when manipulating the adjectives, an element (*-t*) is either added or left out. When manipulating the articles, a *t* or an *n* is replaced with each other.

The neuter and uter nouns were controlled for length and frequency. The target items did not have the same syntactic function (e.g., object, subject complement or part of an adverbial) and thus were not in the same position in the sentences. The text also contained a minimum of 32 control items (16 uter NPs; 16 neuter NPs), which were inflected adjectives not already used as targets.

Table 9 shows the number and share of detected targets. Targets were considered detected if min. one of the three words in the NP was underlined.

As predicted based on error rates, we found an effect of word class ($\hat{\beta}$ = 0.90, SE = 0.08, $z = 11.30$, $p < 0.001$), so that mismatches with articles were detected more than mismatches with adjectives. As expected based on error rates, we found an effect of gender ($\hat{\beta}$ = 0.72, SE = 0.08, $z = 9.08$, $p < 0.001$), so that participants detected more neuter for uter than uter for neuter in general (*cf.* Table 10). We also found the expected interaction ($\hat{\beta}$ = −0.70, SE = 0.11, $z = -6.23$, $p < 0.001$), which can be seen in Figure 3. It shows the model's predicted probability of responding correctly (detecting the error) in the different conditions. For the articles, the effect of gender is less pronounced than for the adjectives. The lowest detection rates were found for *et* <u>*dejlig*</u> *kæledyr* (mismatch with adjective; uter for neuter), as expected. However, the interaction might also be explained by the phonological similarity to the correct form in this condition, or visual differences between conditions. Perhaps, it is harder to spot a missing *-t* than an extra *-t* or to spot a *t* which is replaced with an *n*. Finally, we found an effect of total grammar score ($\hat{\beta}$ = 0.42, SE = 0.12, $z = 3.38$, $p < 0.001$), so that the higher total grammar score in the quiz, the more NP errors were detected. In the grammar quiz, participants were given an adjective and asked to insert it before both an uter and a neuter noun. The article task was forced choice, and participants had to choose between uter or

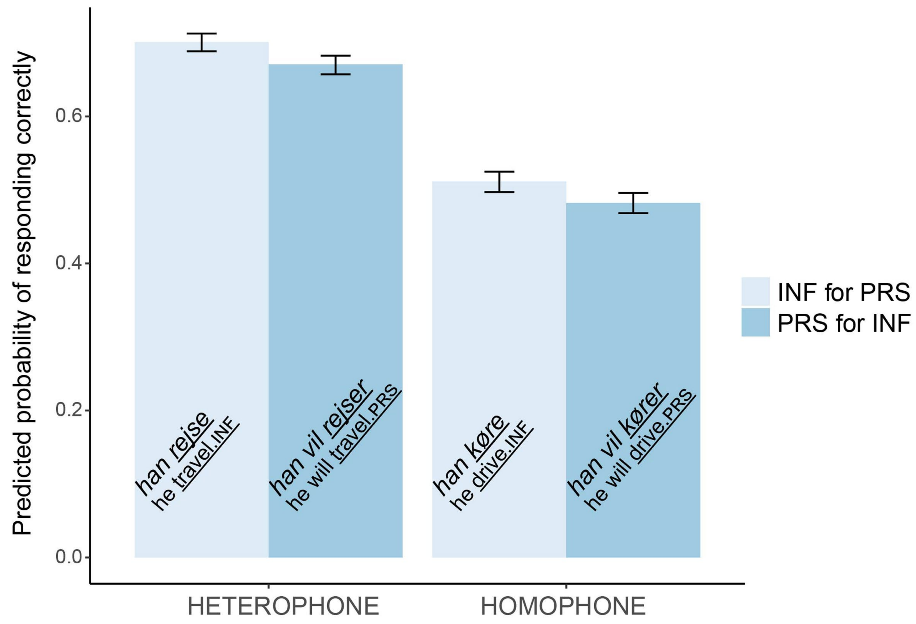


FIGURE 2
The model's predicted probabilities of detecting verb errors. Error bars show SDs.

TABLE 8 Error rates in L1 texts, gender mismatch between indefinite articles or adjectives with noun.

	N errors	N correct	Error rate (%)
Indefinite articles	**16**	**3,132**	**0.51%**
Uter for neuter (*en* for *et*)	6	984[1]	0.61%
Neuter for uter (*et* for *en*)	10	2,178	0.46%
Adjectives	**51**	**2798[2]**	**1.79%**
Uter for neuter (Ø for -*t*)	29	1,368	2.08%
Neuter for uter (-*t* for Ø)	22	1,430	1.49%

[1]Number of correct occurrences of et 'a' (neuter), found with a POS tagger [Centre for Language Technology, University of Copenhagen (CST), 2022].
[2]The number of correct adjectives with a correct -Ø or -*t* suffix. Found with a POS tagger [Centre for Language Technology, University of Copenhagen (CST), 2022]. Manually, the following were removed: adjectives with no/optional gender conjugations (ending with -*sk*, -*vis*), indeclinable adjectives (e.g., *ekstra* 'extra'), and adjectives ending with a -*t* (e.g., *stolt* 'proud').

neuter indefinite articles for four nouns. There were only 6 errors for the 844 articles (0.7%) and no errors for the 422 adjectives.

6.4. Orthographic errors

In general, we expected common types of misspellings to be noticed less than syntactic and morphological errors. In the high school corpus, orthographic errors are the most common type of error (0.86% of all words are misspelled). The 20 target items were created based on four types of misspellings which others have found to be common in L1 writing (e.g., Blom et al., 2017). Examples can be seen in Table 11. Table 11 also shows the number and shares of detected errors. Most of the errors are phonologically similar to the correct form. Some are entirely homophone (e.g., the error *virklig*), while other errors could be prosodically different, e.g., with respect to vowel length or stress.

The only significant effect of type was that reduced syllables were detected more often than missing double consonants, which were noticed the least ($\hat{\beta} = 1.40$, SE = 0.55, $z = 2.56$, $p < 0.05$). Finally, there was a significant effect of the score in the spelling task in the quiz, so that the more correct answers participants had in the spelling task, the more orthographic errors participants found in the reading task ($\hat{\beta} = 0.50$, SE = 0.08, $z = 6.52$, $p < 0.001$). In the spelling task, participants had to determine whether 8 words were spelled correctly. If not, they should write the correct form. There were 196 errors out of 1.688 answers (12% errors), made by 115 participants (1–5 errors per participant).

7. Discussion

Section 7.1 is a summary and discussion of the general findings of the study. In section 7.2, we discuss the relation between error detection rates and two seemingly dominant (and co-varying) factors in our study: the frequency of the error and its phonological similarity to the correct form. Section 7.3 discusses challenges for current and future models of eye movement control in reading and presents our recommendations based on the study.

7.1. General findings and effects of explicit grammar awareness

The present study examined the relationship between the type of errors young readers tend to overlook in texts, the type of errors these young readers produce themselves in the grammar quiz, and the type of errors that are typical of their age group in general (based on corpus error rates). When examining attention to naturally occurring grammar anomalies, some factors co-vary. Still, to use ecological stimuli is necessary if future models of language processing are to be able to accommodate naturally occurring, non-standard grammar.

TABLE 9 Number of NP errors in texts and share of detected errors.

Conditions	Errors in texts (*N*)	Detected targets (*N*)	Share of detected targets (%)
MISMATCH ART + N	**3,376**	**2034**	**60.25%**
Neuter for uter: *et* *dejlig* *undulat* ART.N lovely-U budgie.U 'a lovely budgie'	1,688	1,021	60.49%
Uter for neuter: *en* *dejlig-t* *kæledyr* ART.U lovely-N pet.N 'a lovely pet'	1,688	1,013	60.01%
MISMATCH ADJ + N	**3,376**	**1,685**	**49.91%**
Neuter for uter: *en* *dejlig-t* *undulat* ART.U lovely-N budgie.U 'a lovely budgie'	1,688	959	56.81%
Uter for neuter: *et* *dejlig* *kæledyr* ART.N lovely-U pet.N 'a lovely pet'	1,688	726	43.01%
Total	**6,752**	**3,719**	**55.08%**

TABLE 10 Model (3) estimates for NP errors.

Random effects	Variance	Std. dev.		
Participant (intercept)	1.260	1.1226		
Item (intercept)	0.192	0.4381		
Fixed effects	**Estimate**	**Std. error**	***z*-value**	***p*-value**
(Intercept)	−7.37728	2.07508	−3.555	0.000378***
Word class	0.90480	0.08013	11.292	<2e-16***
Gender	0.72108	0.07945	9.076	<2e-16***
Word class*Gender (interaction)	−0.70256	0.11276	−6.231	4.64e-10***
Total grammar score (quiz)	0.41706	0.12357	3.375	0.000738***

Dependent variable: detection (1 = error detected, 0 = error not detected). Significance code: ***$p < 0.001$.

In our study, grammar errors seem to attract more attention than orthographic errors. This finding is in line with Larigauderie et al. (2020) who studied attention to grammatical and orthographic errors in French. Their grammar errors were comparable to ours, as they related to number and gender agreement and misuse of the past participle form in French. Their orthographic errors (like most of ours) did not affect the phonology of the word. Previous proofreading studies of English (Hacker et al., 1994; Shafto, 2015), however, found the opposite pattern, as orthographic errors attracted more attention than grammar errors in their studies. It is likely that this discrepancy stems from differences in what is understood by a grammar error vs. an orthographical error. In Shafto (2015), the grammar errors were heterogeneous ranging from errors in verb agreement and number agreement to punctuation and capitalization errors, thus grouping types of errors which are quite distinct. The orthographic errors also included typos such as letter switches which resulted in an incorrect phonological form, and which are therefore also qualitatively different from the orthographic errors in our study. Larigauderie et al. (2020) found that typos were the most frequently detected type of error. In Hacker et al. (1994), the error categories were not clearly defined. Their grammar errors included errors in verb agreement as well as confusion of word classes (e.g., *affects* for *effects*). Altogether, these differences in the definitions of grammar vs. orthography may explain the seemingly contradictory results.

Error detection is not entirely explained by explicit grammar awareness. In the grammar quiz, the general performance was almost at ceiling with error rates ranging from 0.5% to 1.5% per task. Yet, all readers overlooked errors in the proofreading study.

Although there were generally few errors in the responses to the grammar quiz, the participants' total score in the grammar quiz did explain some of the variance in the detection rates. For the three types of grammar errors (V3 word order, verb errors, NP errors), we found an effect of the total grammar score, so that the more correct answers participants had in the three grammar tasks in the quiz, the more errors they detected. Similarly, the more correct answers participants had in the spelling task, the more orthographic errors they detected. Finally, we found that the more annoyed with language errors participants reported to be, the more errors they detected.

Unlike most previous psycholinguistic studies which either group many different types of grammar errors into one experimental condition (Hacker et al., 1994; Shafto, 2015) or only investigate one specific type as representative of all grammar errors (often using the cover term *syntactic violations*), our study distinguishes between different types of grammar errors. The descriptive statistics showed differences in detection rates between syntactic and morphological errors in our study, which seems to suggest that not all grammar errors are treated alike. Future eye-tracking studies may determine if this pattern is not just due to quantitative differences (degree of attention), but also due to qualitative differences (differences in how they are processed and attended to).

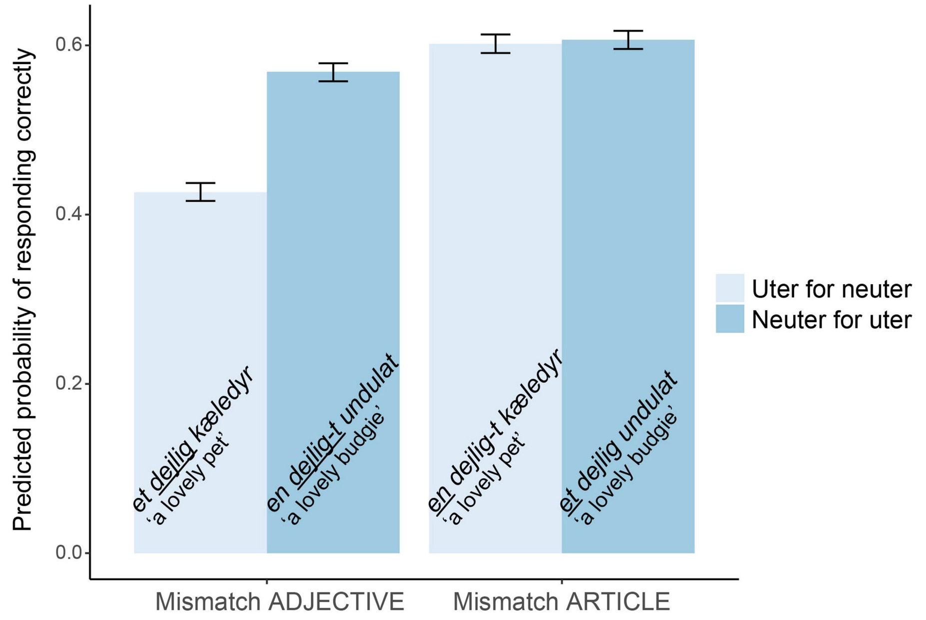


FIGURE 3
The model's predicted probabilities of detecting NP errors. Error bars show SDs.

TABLE 11 Types of orthographic errors, number of errors in texts and share of detected errors.

Types of orthographic errors (four types — five of each type)	Errors in texts (*N*)	Detected targets (*N*)	Share of detected targets (%)
Missing double consonant, e.g., *startskudet* for *startskuddet* 'the starting signal'	1,055	224	21.23%
Split compounds, e.g., *by vandring* for *byvandring* 'city walk'	1,055	342	32.42%
Missing silent letter, e.g., *siste* [ˈsisd̥ə]/[ˈsisd̥] for *sidste* [ˈsisd̥ə]/[ˈsisd̥] 'last'	1,055	359	34.03%
Reduction of syllable, e.g., *virklig* [ˈʋiɡ̊li] for *virkelig* [ˈʋiɡ̊li] 'really'	1,055	456	43.22%
Total	**4,220**	**1,381**	**32.73%**

7.2. The relation between what students typically produce and what they notice

Models of natural reading processing must deal with naturally occurring errors. Yet, a complication of using naturally occurring errors is that several factors co-vary between conditions. In the following sections, we discuss two main potential contributing factors when it comes to readers' perception of and attention to grammar errors in Danish: the frequency of the error (section 7.2.1) and the phonological similarity between the error and the correct form (section 7.2.2).

7.2.1. Error frequency

Our study suggests that the frequency of grammar errors is a relevant factor to include in future models of eye movements during reading. Attention to a specific type of grammar error is not only a matter of the reader's explicit grammar awareness (as measured in the grammar quiz). If a specific type is frequent among the peers of the reader, the reader may have more exposure to this type of error and a mental representation of it. The reader may therefore find it less striking and be less likely to detect it compared to errors that are infrequent in texts written by peers. According to the descriptive statistics in our study, the error detection rates for the three overall error categories (syntactic > morphological > orthographic) were inversely proportional with the error rates in L1 writing. Syntactic errors have the lowest error rates in L1 writing and the highest detection rates. Orthographic errors have the highest error rates and the lowest detection rates. Within the three grammar subexperiments, we also found that error types with relatively high error rates (errors in homophone verb pairs, mismatching adjectives in NPs, overuse of uter in NPs) had lower detection rates than errors with lower error rates (errors in heterophone verb pairs, mismatching articles in NPs, overuse of neuter in NPs).

Yet, frequency is not the only possible explanation to these results. The higher share of detected syntactic errors could be influenced by differences in size (manipulating word order vs. letters). The homophony effect for verb errors is closely tied to the phonological similarity to the correct form (section 7.2.2.). In the subexperiment on NPs, phonological similarity to the correct form may also explain the interaction between word class and gender (section 7.2.2). Furthermore, frequency and word class co-varied. Also, the effect of word class could be influenced by differences in the placement of the error within the NP. It may be that phrase-initial errors (such as the article errors) attract more attention than errors placed in the middle of a phrase (such as the adjective errors). Thus, future studies are needed, in which effects of position in the phrase and frequency can be distinguished — and if possible, in which effects of frequency can be distinguished from phonological similarity to the correct form.

These reservations aside, it seems likely that frequency plays an important part in error detection, and that the role of frequency is worth studying in future studies with more controlled and less confounded stimuli. Frequency is, as mentioned in the introduction, tied to predictability. According to prediction-based approaches to sentence processing, unexpected input attracts attention (Kamide, 2008; Levy, 2008; Christiansen and Chater, 2016). If a reader sees input with frequent errors, the model will be updated according to the input, meaning that frequent errors should be predicted by the model, and thus should attract less attention than infrequent errors. The error rates in our study were based on texts written by high school students. We do not assume that high school students read each other's essays, but the errors they produce in school essays are likely to occur in their writing in general, including informal text directed at their peers. Furthermore, we assume that the error production patterns found in high school texts to a large extent reflect the error types found in the media.

Frequency does not explain all findings and it seems to be interacting with other factors in our study. Not all predictions based on error rates were confirmed: we did not expect an effect of type for the verb errors, but found higher detection rates for infinitive for present tense than vice versa. In the public debate and prescriptive literature, missing present tense *-r* is often accentuated as a typical or basic error (Blom and Ejstrup, 2019b), and in the study by Blom and Ejstrup (2019a), participants rated the missing present tense *-r* as the most annoying error of all included errors. This special status of the missing *-r* in present tense might explain why this error type was noticed more than the superfluous *-r* on infinitives, although the frequency in production (as measured by error rates) does not differ between the two. If looking at occurrences per 1,000 words, omitting the -r is, in fact, more frequent in written texts. Counter to our expectations, we did not find an interaction between homophony and type. The surprising result might also be explained by the great prescriptive focus on the most frequent error type (homophone; infinitive for present tense).

In our study, frequency measures were based on error rates in a small corpus of naturally occurring L1 texts. For erroneous use of gender in articles, the error rates were based on only 16 article errors, and the distribution between uter and neuter gender in errors may well be different in a larger corpus. Future studies with a larger corpus may use inferential statistics for a more adequate calculation and assessment of differences in error rates. They may also consider the pros and cons of using error rates vs. raw frequency (errors per 1,000 running words) as the basic measure. In most cases, these measures lead to the same predictions, but in one case, type for verb errors, our frequency-based predictions would have been different if we had based them on occurrences per 1,000 running words, instead of error rates. Homophony set aside, there are more errors per 1,000 words where the target form is present tense (1.02) than when it is infinitive (0.37). Thus, infinitive for present tense should be least noticed. This was, however, not the case, and this frequency measurement therefore does not seem better at predicting error detection than error rates.

To conclude, frequency (measured by error rates) in most cases predicted detection rates of different types of errors. Due to the confounded nature of the highly ecological error types in the stimuli, we cannot determine the exact nature of the interplay with other contributing factors.

7.2.2. Phonological similarity to the correct form

In naturally occurring language we often find errors that intersect grammar and phonology. Since we aimed to study error detection of naturally occurring grammar errors, our stimuli included such intersectional errors. We contrasted grammar errors where the confused forms were phonologically identical (homophone) with errors where the two forms were clearly distinct in pronunciation (heterophone). Our study showed significantly lower detection rates for verb errors in the homophone condition compared to the heterophone condition. These results suggest that phonology interferes with grammatical processing during error detection. Yet, the difference between homophone and heterophone forms may also be due to differences in frequency, as error rates in L1 writing are higher when the present tense and infinitive are homophone. In the verb error subexperiment, we therefore cannot disentangle the effect of phonological interference from that of frequency. Still, we find it plausible that phonological interference constitutes a separate effect when taking into account the findings from the subexperiment on NP errors. For NP errors, detection rates were low when the adjective was inflected in uter instead of the correct neuter form (e.g., *dejlig* instead of *dejligt*). This error with a missing *-t* is not only visually similar to the correct form (*cf.* section 7.3), but also phonologically similar. In distinct speech the final [d̪] in *dejligt* may be pronounced, but in running speech there is usually no audible difference. This similarity between forms may explain why we found an interaction between gender and word class. Frequency differences in error rates may also account for this effect. Yet, the differences in frequency are small. It therefore seems more likely that phonological similarity plays a key role in explaining the low detection rates for uter for neuter in adjectives.

Errors that intersect the boundary between grammar and phonology are not unique to Danish. "Silent suffix" errors with confusion of homophone verb forms are also frequent in other languages. In Dutch the 1st person verb *word* and the 3rd person verb *wordt* have the same pronunciation and are commonly confused (Sandra et al., 2004). In French, there is no audible difference between the verb forms *mange*, *manges* and *mangent*, and ERP studies show that responses to confusion of such homophone verb forms differs from responses to confusion of heterophone verb forms like *mange* vs. *mangez* (Carassco-Ortiz and Frenck-Mestre, 2014). This finding is in line with Larigauderie et al. (2020) who found that typographical errors (i.e., incorrect successions of letters resulting in incorrect phonology) are more frequently detected than orthographic errors which did not affect the phonology of the word. Potential interference from phonology is not limited to confusion of verb forms. The confusion of English *its* and *it's* is a prime example. Although our study cannot disentangle effects of phonological similarity from error frequency, we recommend that future eye-tracking models of reading and sentence processing models in general consider the possible role of phonological resemblance of errors to correct forms.

7.3. Challenges for current and future models of eye movement control in reading

Presumably, the error detection measure is less sensitive than eye-tracking. Although the degree of correlation between the two measures is uncertain, we assume that the overall results could be replicated using eye-tracking, which is a natural next step. More fine-grained differences may also be detected using eye-tracking, e.g., it may be that eye movements are affected, though errors are not underlined by the participant. This was, however, not found in the eye-tracking study by Huang and Staub (2021). Disruption in eye movement measures caused by transposition errors were only found in those sentences participants judged to be ungrammatical. The majority of previous eye-tracking studies of ungrammaticality did not ask participants whether they noticed and perceived the individual errors as ungrammatical or not. Using the error detection paradigm, we collected this information without interrupting participants' reading excessively and found that attention to different types of naturally occurring errors is not uniform. This variation in the reader's attention and response to errors poses a challenge to the major present models of eye movement control in reading (Reichle et al., 2003; Engbert et al., 2005). The E-Z Reader model (Reichle et al., 2009) addresses reactions to severe syntactic violations, but does not address what happens when readers encounter misspellings or other types of grammar errors. Results from previous eye-tracking studies of ungrammaticality indicate that different types of grammar errors (e.g., V3 and morphological agreement errors) elicit similar responses in participants' eye movements across languages, with similar time courses (*cf.* section 2.1) — including the very early effects, which E-Z Reader explicitly predicts for syntactic violations. If attention to different types of errors should be integrated in the E-Z Reader model, a first step could be to integrate detection of orthographic errors as part of the early familiarity check, and to account for both morphological and syntactic errors.

The E-Z Reader model does not explain why some errors are detected while others go by unnoticed, and why different readers do not always notice the same error. Also, as Warren (2011) points out, the model does not consider the precise combination of reader and the purpose or motivation for the reading. Our study both shows an effect of participants' explicit grammar awareness and general irritation with errors on detection rates.

In our study, we have demonstrated the complexity of measuring error frequency and determining when there is phonological similarity. It is therefore challenging to integrate these factors in models of eye movement control during reading. Still, the two factors are entangled, and even a rough measure of error frequency would improve current and future models when dealing with reading of everyday texts.

Previous letter detection experiments (Smith and Groat, 1979) have found position effects, e.g., that elements in the start or end of a line or within a sentence tend to be more prominent than elements in the middle. Our study on V3 errors manipulated the length of the sentence-initial adverbial, but we found no effects of the placement in the sentence (close to the start vs. further toward the middle). This lack of an effect of position was confirmed in an eye-tracking study where Norwegian readers read similar types of V3 with long and short adverbials (Søby et al., 2023). Smith and Groat (1979) did not consider different sentence structures in their analysis, only numerical order of the words, and the position effects varied between items. Further studies are needed to test the potential role of error position within the sentence.

For the verb and NP errors, there were visual differences between elements that were deleted, added and replaced with other elements. The NP data suggest that replacing two elements with another (i.e., *-t* and *-n* in indefinite articles) is noticed more than when an element is added or missing (*-t* in adjectives). However, for verb errors, a missing *-r* was more noticed than an extra *-r*. It therefore seems that other factors than visual differences are more important, e.g., word class or error frequency.

In this study, we have examined outright errors which both deviated from the norms defined by the Danish Language Council and from most participants' own answers in the grammar quiz. Language norms, however, are subject to language change and sociolinguistic variation. Natural texts therefore both contain outright errors and language anomalies in the gray zone between language errors and language variation. For instance, the inflection of Danish modal verbs seem to be subject to language change. In written production most high school students do not inflect the Danish modal verb *måtte* according to the norms defined by the Danish language council (Kristensen et al., 2023). These anomalies should also be considered in future studies.

Our study only included one type of task, i.e., proofreading while reading for comprehension. Using eye-tracking, Schotter et al. (2014a) found that the task (proofreading for letter transpositions vs. reading for comprehension) affected processing patterns. The patterns when reading for comprehension may therefore differ from what we find in our study. Still, based on our study, we recommend that future models take the following factors into account, as they may all modulate attention and eye movements:

1. Variation in the type of naturally occurring grammar errors that occur in non-standard language (e.g., syntactic errors compared to morphological errors, and different subtypes within these categories).
2. Variation in error frequencies as a general predictor, and importantly, when present: phonological similarity with the correct form (which tends to be entangled with error frequency).
3. Variation in the reader's grammatical awareness and proficiency.

Data availability statement

The dataset for this study and code for analyses can be found in an online repository: http://github.com/ResearchXX/ErrorDetection.

Ethics statement

The study involving human participants was reviewed and approved by the Faculty of Humanities' Research Ethics Committee, University of Copenhagen. Written informed consent to participate in this study was provided by participants (if above 18 years old) or by the participants' legal guardian/next of kin.

Author contributions

KS and LK contributed to designing the study. KS was responsible for making the test material and collecting data. BI wrote the code for the analyses, which KS used. The first draft was written by KS. LK and BI commented and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was financed by Independent Research Fund Denmark, grant number: 7023-00131B.

Acknowledgments

The authors would like to thank student assistants Julie Johanna Brink Hansen for digitizing the data and Maja Mittag for summarizing L1 corpus data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1124227/full#supplementary-material

References

Ainsworth-Darnell, K., Shulman, H. G., and Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: evidence from event-related potentials. *J. Mem. Lang.* 38, 112–130. doi: 10.1006/jmla.1997.2537

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Becker-Christensen, C. (2010). *Dansk syntaks. Indføring i dansk sætningsgrammatik og sætningsanalyse*. København: Samfundslitteratur.

Blom, J. N., and Ejstrup, M. (2019a). "Læsernes holdninger til journalisters stavefejl i de digitale medier" in *17. Møde om Udforskningen af Dansk Sprog*. eds. Y. Goldshtein, I. S. Hansen and T. T. Hougaard (Aarhus: Aarhus Universitet), 113–132.

Blom, J. N., and Ejstrup, M. (2019b). BRÆJKING NEWS: En eksperimentel undersøgelse af fejlobservante læseres holdninger til ukorrekte og korrekte stavemåder opfattet som stavefejl i nyhedsrubrikker på nettet. *NyS* 57, 10–46. doi: 10.7146/nys.v1i57.117114

Blom, J. N., Rathje, M., le Fevre Jakobsen, B., Holsting, A., Hansen, K. R., Svendsen, J. T., et al. (2017). Linguistic deviations in the written academic register of Danish university students. *OSLa* 9, 169–190. doi: 10.5617/osla.5855

Braze, D., Shankweiler, D., Ni, W., and Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *J. Psycholinguist. Res.* 31, 25–44. doi: 10.1023/A:1014324220455

Carassco-Ortiz, H., and Frenck-Mestre, C. (2014). Phonological and orthographic cues enhance the processing of inflectional morphology. ERP evidence from L1 and L2 French. *Front. Psychol.* 5, 1–14. doi: 10.3389/fpsyg.2014.00888

Christensen, M. H., Kristensen, L. B., Vinther, N. M., and Boye, K. (2021). Grammar is background in sentence processing. *Lang. Cogn.* 13, 128–153. doi: 10.1017/langcog.2020.30

Christiansen, M. H., and Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39, e62–e72. doi: 10.1017/S0140525X1500031X

Clifton, C., and Staub, A. (2011). "Syntactic influences on eye movements during reading" in *The Oxford handbook of eye movements*. eds. S. P. Liversedge, I. Gilchrist and S. Everling (Oxford: Oxford University Press), 896–909.

Centre for Language Technology, University of Copenhagen (CST) (2022). Online Part-Of-Speech tagger. Available at: https://cst.dk/online/pos_tagger/index.html (Accessed November 24, 2022).

Dank, M., Deutsch, A., and Bock, K. (2015). Resolving conflicts in natural and grammatical gender agreement: evidence from eye movements. *J. Psycholinguist. Res.* 44, 435–467. doi: 10.1007/s10936-014-9291-9

Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224. doi: 10.1006/jmla.2000.2768

Det Danske Sprog-og Litteraturselskab (DSL) (2022). KorpusDK. Available at: https://ordnet.dk/korpusdk (Accessed 24 November 2022).

Engbert, R., and Kliegl, R. (2011). "Parallel graded attention models of reading" in *The Oxford handbook of eye movements*. eds. S. P. Liversedge, I. Gilchrist and S. Everling (Oxford: Oxford University Press), 788–800.

Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during Reading. *Psychol. Rev.* 112, 777–813. doi: 10.1037/0033-295X.112.4.777

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Curr. Dir. Psychol. Sci.* 11, 11–15. doi: 10.1111/1467-8721.00158

Ferreira, F., and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83. doi: 10.1111/j.1749-818x.2007.00007.x

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., et al. (2017). Don't underestimate the benefits of being misunderstood. *Psychol. Sci.* 28, 703–712. doi: 10.1177/0956797617690277

Hacker, D. J., Plumb, C., Butterfield, E. C., Quathamer, D., and Heineken, E. (1994). Text revision: detection and correction of errors. *J. Educ. Psychol.* 86, 65–78. doi: 10.1037/0022-0663.86.1.65

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/089892903322370807

Hahne, A., and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Brain Res. Cogn. Brain Res.* 13, 339–356. doi: 10.1016/S0926-6410(01)00127-6

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103

Huang, K., and Staub, A. (2021). Using eye-tracking to investigate failure to notice word transpositions in reading. *Cognition* 216:104846. doi: 10.1016/j.cognition.2021.104846

Hyltenstam, K., and Abrahamsson, N. (2003). "Maturational constraints in SLA" in *The handbook of second language acquisition*. eds. C. J. Doughty and M. H. Long (Oxford: Blackwell Publishing), 538–588.

Jacoby, L. L., and Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: effects of reading incorrectly and correctly spelled words. *Can. J. Psychol.* 44, 345–358. doi: 10.1037/h0084259

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Linguist. Compass* 2, 647–670. doi: 10.1111/j.1749-818X.2008.00072.x

Konieczny, L., Scheepers, C., and Hemforth, B. (1994). "Reanalyses vs. internal repairs: non-monotonic processes in sentence perception" in *First Analysis, Reanalysis and Repair*. eds. B. Hemforth, L. Konieczny, C. Scheepers and G. Strube, vol. *8* (IIG-Berichte: University of Freiburg), 2–22.

Kristensen, L. B., Schack, J., and Søby, K. F. (2023). Om unge der har skulle bøje modalverber, men ikke har turde, ikke har kunne eller ikke har ville. *NfG* 30, 74–90. doi: 10.7146/nfg.v1i29.132901

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). Lmer test package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Larigauderie, P., Guignouard, C., and Olive, T. (2020). Proofreading by students: implications of executive and non-executive components of working memory in the detection of phonological, orthographical, and grammatical errors. *Read. Writ.* 33, 1015–1036.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006

Lim, J. H., and Christianson, K. (2015). Second language sensitivity to agreement errors: evidence from eye movements during comprehension and translation. *Appl. Psycholinguist.* 36, 1283–1315. doi: 10.1017/S0142716414000290

Lunsford, A. A., and Lunsford, K. J. (2008). 'Mistakes are a fact of life': a National Comparative Study. *Coll. Compos. Commun.* 59, 781–806.

Ministry of Higher Education and Science. (2022). Upper secondary education. Available at: https://ufm.dk/en/education/the-danish-education-system/upper-secondary-education (Accessed 24 November 2022).

Ni, W., Fodor, J. D., Crain, S., and Shankweiler, D. (1998). Anomaly detection: eye-movement patterns. *J. Psycholinguist. Res.* 27, 515–539. doi: 10.1023/A:1024996828734

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: language variety and stylistic practice. *Int. J. Biling.* 12, 43–61. doi: 10.1177/13670069080120010401

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reichle, E. D. (2011). "Serial-attention models of reading" in *The Oxford handbook of eye movements*. eds. S. P. Liversedge, I. Gilchrist and S. Everling (Oxford: Oxford University Press), 767–786.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behav. Brain Sci.* 26, 445–476. doi: 10.1017/s0140525x03000104

Reichle, E. D., Warren, T., and McConnell, K. (2009). Using E-Z reader to model the effects of higher level language processing on eye movements during reading. *Psychon. Bull. Rev.* 16, 1–21. doi: 10.3758/PBR.16.1.1

Sandra, D., Frisson, S., and Daems, F. (2004). Still errors after all those years ...: Limited attentional resources and homophone frequency account for spelling errors on silent verb suffixes in Dutch. *Writ. Lang. Lit.* 7, 61–77. doi: 10.1075/wll.7.1.07san

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., and Rayner, K. (2014a). Task effects reveal cognitive flexibility responding to frequency and predictability: evidence from eye movements in reading and proofreading. *Cognition* 131, 1–27. doi: 10.1016/j.cognition.2013.11.018

Schotter, E. R., Reichle, E. D., and Rayner, K. (2014b). Rethinking parafoveal processing in reading: serial-attention models can explain semantic preview benefit and N + 2 preview effects. *Vis. Cogn.* 22, 309–333. doi: 10.1080/13506285.2013.873508

Shafto, M. A. (2015). Proofreading in young and older adults: the effect of error category and comprehension difficulty. *Int. J. Environ. Res. Public Health* 12, 14445–14460. doi: 10.3390/ijerph121114445

Smith, P. T., and Groat, A. (1979). "Spelling patterns, letter cancellation and the processing of text. Processing of visible language" In *Proceedings of the first conference on processing of visible language, September 5–8, 1977*, ed. P. A. Koolers, M. E. Wrolstad, and H. Bouma (New York: Plenum), 309–324.

Søby, K. F., and Kristensen, L. B. (2019). Hjælp! Jeg har mistede min yndlings rød taske. Et studie af grammatikafvigelser. *NfG* 26, 89–104. doi: 10.7146/nfg.v0i26.115995

Søby, K. F., and Kristensen, L. B. (to appear). V2 is not difficult to all learners in all contexts - a cross-sectional study of L2 Danish.

Søby, K. F., Milburn, E., Kristensen, L. B., Vulchanov, V., and Vulchanova, M. (2023). In the native speaker's eye: Online processing of anomalous learner syntax. *Appl. Psycholinguist* 44, 1–28. doi: 10.1017/S0142716422000418

Sturt, P., Sanford, A. J., Stewart, A., and Dawydiak, E. (2004). Linguistic focus and good-enough representations: an application of the change-detection paradigm. *Psychon. Bull. Rev.* 11, 882–888. doi: 10.3758/BF03196716

Veldre, A., and Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during Reading. *J. Mem. Lang.* 100, 1–17. doi: 10.1016/j.jml.2017.12.002

Vinther, N. M., Boye, K., and Kristensen, L. B. (2015). Grammatikken i baggrunden – opmærksomhed under læsning. *NyS* 47, 99–139. doi: 10.7146/nys.v47i47.19877

Warren, T. (2011). "The influence of implausibility and anomaly on eye movements during reading" in *The Oxford handbook of eye movements*. eds. S. P. Liversedge, I. Gilchrist and S. Everling (Oxford: Oxford University Press), 912–923.

Article 3: In the native speaker's eye: Online processing of anomalous learner syntax

Published in *Applied Psycholinguistics*, 2023, vol. 44, issue 1.
https://doi.org/10.1017/S0142716422000418

The supplementary material for the article is found in Appendix B (section 10).

Applied Psycholinguistics (2023), **44**, 1–28
doi:10.1017/S0142716422000418

ORIGINAL ARTICLE

In the native speaker's eye: Online processing of anomalous learner syntax

Katrine Falcon Søby[1,*], Evelyn Arko Milburn[2,3], Line Burholt Kristensen[1], Valentin Vulchanov[3] and Mila Vulchanova[3]

[1]Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark, [2]Department of Psychology, Mills College, Oakland, CA, USA and [3]Language Acquisition and Language Processing Lab, Department of Language & Literature, Norwegian University of Science & Technology, Trondheim, Norway
*Corresponding author. Email: kafs@kp.dk

(Received 27 January 2022; revised 17 October 2022; accepted 20 October 2022; first published online 12 January 2023)

Abstract
How do native speakers process texts with anomalous learner syntax? Second-language learners of Norwegian, and other verb-second (V2) languages, frequently place the verb in third position (e.g., *Adverbial-Subject-Verb), although it is mandatory for the verb in these languages to appear in second position (Adverbial-Verb-Subject). In an eye-tracking study, native Norwegian speakers read sentences with either grammatical V2 or ungrammatical verb-third (V3) word order. Unlike previous eye-tracking studies of ungrammaticality, which have primarily addressed morphosyntactic anomalies, we exclusively manipulate word order with no morphological or semantic changes. We found that native speakers reacted immediately to ungrammatical V3 word order, indicated by increased fixation durations and more regressions out on the subject, and subsequently on the verb. Participants also recovered quickly, already on the following word. The effects of grammaticality were unaffected by the length of the initial adverbial. The study contributes to future models of sentence processing which should be able to accommodate various types of "noisy" input, that is, non-standard variation. Together with new studies of processing of other L2 anomalies in Norwegian, the current findings can help language instructors and students prioritize which aspects of grammar to focus on.

Keywords: eye-tracking; verb-second; verb-third; syntactic anomalies; sentence processing; learner language; word order violation; Norwegian

Writing in a second language (L2) often comes with production of syntactic anomalies. Although there is extensive research on learners' production of syntactic anomalies, surprisingly little is known about how these anomalies are processed by native speakers, and to what extent they may disrupt processing. This is specifically relevant in the context of increased global mobility, where native speakers of a language need to accommodate anomalies produced by immigrant adult L2 learners.

In the present eye-tracking study, we investigated how native speakers process anomalous L2 syntax. We presented native Norwegian speakers with written sentences with syntactic anomalies in order to elicit their responses to typical non-native word order.

The study focuses on verb-second (V2) word order, which is common in most Germanic languages (apart from English). In V2 languages, the finite verb occurs in the second position of a declarative main clause, preceded only by a single first constituent. In the Norwegian examples below, sentence (1a) is grammatical, as the verb *spiller* 'plays' is correctly placed in second position, preceded by one constituent, the fronted adverbial *på torsdager* 'on Thursdays.' The subject *gutten* 'the boy' is placed after the main inflected verb. However, (1b) is ungrammatical in Norwegian, since two constituents, both the adverbial *på torsdager* 'on Thursdays' and the subject *gutten* 'the boy,' precede the verb, which is in third position. Thus, (1b) is an example of ungrammatical V3 word order.

1. a. *På torsdager spiller gutten fotball*. 'On Thursdays plays the boy football.'
   b. **På torsdager gutten spiller fotball*. 'On Thursdays the boy plays football.'

Typologically, V2 word order is a rare phenomenon. It is notoriously difficult to master fully for L2 speakers whose L1 does not feature V2 (e.g., Bolander, 1990). A common trait in L2 production is the use of V3 where V2 is required, as in (1b) (for Norwegian, Hagen, 1992; Johansen, 2008; for Swedish, Bolander, 1990; Bohnacker, 2006; for Danish, Lund, 1997; Søby & Kristensen, 2019). Even learners whose native language features V2 may produce V3 word order in a V2 second language, possibly due to influence from another L2, for example, English (Bohnacker, 2006).

The ungrammatical sentence with V3 in (1b) does not express a different propositional content than the grammatical sentence with V2 in (1a). Sentences with V3 are found in multiethnic urban vernaculars in Sweden (Kotsinas, 2000), Denmark (Quist, 2008), Norway (Hårstad & Opsahl, 2013), and Germany (Freywald et al., 2015), and they are used with the same meaning as an equivalent sentence with V2, but as part of a different stylistic practice (see Quist, 2008). Language attitude experiments document that V3 may be associated either with immigrant status or with multiethnic youth varieties (Freywald et al., 2015; Quist, 2008).

Though ungrammatical V3 word order is common in L2 production (and in urban vernaculars), there are only a few studies on the perception of V3. Generally, there is little research on native speakers' processing of non-native or non-standard syntax, which is surprising given the prevalence of this type of "noisy" and non-standard variation. This research is likewise critically important for developing models of sentence processing that can accommodate said variability. The current study contributes valuable input to such models for two reasons. Firstly, the word order anomalies in the study are naturally occurring in both oral and written production, rather than consisting of randomly scrambled words, as in previous

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

eye-tracking studies on word order (Huang & Staub, 2021). Secondly, previous eye-tracking studies of ungrammaticality have primarily addressed morphosyntactic anomalies. We cannot a priori know whether word order anomalies elicit the same effects as anomalies involving morphological changes. According to some neurolinguistic models (e.g., Friederici, 2002), initial syntactic structure building and morphosyntactic processes differ in timing.

The current findings can thus inform future models of processing of naturally occurring word order anomalies which are part of everyday communication in multi-lingual and multiethnic societies leading to more robust models which accommodate noisy input from non-proficient language users and other types of non-standard variation.

Background

In this section, we review results from EEG studies on the processing of V3 word order. Given the lack of eye-tracking studies on V3, we review results from the relatively few eye-tracking studies that have investigated other types of ungrammaticality, that is, morphosyntactic anomalies or transposed words. The review focuses on the time course of the effects of ungrammaticality, which has varied in previous studies and which we return to in the discussion. We expect that all types of ungrammaticality will result in a surprisal effect when predictions about the morphological form of words, or the order of words, are not met, consistent with prediction-based approaches to sentence processing (Christiansen & Chater, 2016; Kamide 2008; Levy, 2008). However, given the different nature of word order anomalies versus morphosyntactic anomalies, their eye-tracking record may differ.

Processing of V3 – evidence from EEG

Three studies on Swedish have examined online processing of ungrammatical V3 after sentence-initial adverbials, measured by event-related potentials (ERPs) (Andersson et al., 2019; Yeaton, 2019; Sayehli et al., 2022). Andersson et al. and Yeaton manipulated the order of subject and verb, as shown in (2).

2. a. *Idag läste hon tidningen*. 'Today read she the paper.'
   b. **Idag hon läste tidningen*. 'Today she read the paper.'

Both studies found a P600 effect, an ERP component often elicited by syntactic violations and considered a later response, typically related to an effort to integrate anomalous input into the context of the sentence. The P600 occurred for the processing of anomalous compared to correct sentences, both in native Swedish speakers and in L2 learners (with German, English, or French as L1). Despite similar patterns for this late effect, only the native speakers showed a left anterior negativity (LAN) effect, which may reflect more automatic processing (Andersson et al., 2019). The stimuli in these studies had little variation in the choice of adverbials. Sentences always started with the adverbs *idag* ('today') or *hemma* ('at home'). Sayehli et al. (2022) also included sentences with V3 after *kanske* 'maybe,' which were

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

judged to be more acceptable than sentences with the other two adverbials. Accordingly, the ERP analyses showed stronger effects for V3 after *hemma* and *idag*, especially for the P600. The authors suggest V3 with *kanske* "is processed differently than V3 with other adverbials where the V2 norm is stronger" (Sayehli et al., 2022, p. 1). Swedish and Norwegian are closely related languages and may show similarities in the processing of V3.

Effects of syntactic processing difficulty reflected in the eye movements

Syntactic processing difficulty[1] has been examined in a number of eye-tracking studies (for an overview, see Clifton et al., 2007). Typically, such studies have employed grammatical structures that result in ambiguous sentences or garden-paths (e.g., Frazier & Rayner, 1982), structures that disconfirm expectations (e.g., Staub & Clifton, 2006), non-canonical word order (Gattei et al., 2021), and structures that violate rules of grammar, both in the form of real and "seeming" violations (e.g., Pearlmutter et al., 1999). Effects of syntactic processing difficulty differ from study to study and are seen at various points in the eye-tracking record, thus leaving it open which factors determine the observed patterns of effects (Clifton et al., 2007; Clifton & Staub, 2011).

There are relatively few eye-tracking studies of ungrammaticality, and, to our knowledge, only one manipulating word order. Huang and Staub (2021) investigated readers' tendency to overlook random transposition errors like *The white was cat big*. Transpositions were less likely to be noticed when both words were short, and when readers' eyes skipped one of the two words, instead of directly fixating on both. The transpositions caused early and sustained disruption on the critical word *cat* (see Table 1), but only on trials that participants judged to be ungrammatical.

Eye-tracking studies with ungrammatical items in their manipulations mostly examine morphosyntactic anomalies (Braze et al., 2002; Dank et al., 2015; Deutsch & Bentin, 2001; Lim & Christianson, 2015; Ni et al., 1998; Pearlmutter et al., 1999). Most of these studies find increased regressions out from the site of the morphosyntactic anomaly and from subsequent words, often, but not always, combined with longer reading times (see Hallberg & Niehorster, 2021). Thus, there are systematic effects, but the results differ regarding *when* the effect of the anomaly first appears in the eye movements.

Ni et al. (1998) compared reading patterns for sentences where the verb was morphosyntactically anomalous (3a) to non-anomalous sentences (3b).

3. a. *It seems that the cats won't usually <u>eating</u> the food we put on the porch.*
   b. *It seems that the cats won't usually <u>eat</u> the food we put on the porch.*

The authors did not find significant differences between the baseline and the morphosyntactically anomalous version at any sentence position regarding either first-pass reading times (i.e., the sum of all fixations in a region from first entering it until leaving it again, a.k.a. *gaze duration*) or residual reading times.[2] However, morphosyntactically anomalous sentences induced significantly more regressions than baseline sentences in the region containing the anomalous progressive verb

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Table 1. Overview of eye-tracking studies using ungrammatical items (transpositions and morphosyntactic anomalies)

Reference	Type of anomaly	Example sentence	Effects of anomaly found	No effects found
Huang and Staub (2021)	Transposition	*The white was cat big.*	More *regressions in* (to the critical 4th word) and *out.* Increased *first fixation duration, gaze duration, regression path duration,* and *total time.*	
Ni et al. (1998)	Modal verb + progressive verb in English	*It seems that the cats won't usually eating the food we put on the porch.*	More *regressions out* in the verb region and subsequent region.	*First-pass reading times***, *total reading time* (residual reading time)
Braze et al. (2002)	Modal verb + progressive or past tense form of a verb in English	*The wall will surely cracking after a few years in this harsh climate./The engine will softly whined while it is running at low capacity.*	More *regressions out* in the verb region and increased *first-pass reading times*** (but only by subject).	
Pearlmutter et al. (1999)	Subject-verb number agreement in English (the attraction phenomenon)	*The key to the cabinet(s) were rusty from many years of disuse.*	More *regressions out* of and increased *total reading times* in the verb region.	*First-pass reading times***
Lim and Christianson (2015)	Subject-verb number agreement in English (the attraction phenomenon).	*The teacher who instructed the student(s) were very strict.*	Increased *gaze duration, regression path duration,* and *total reading time.*	*Regressions out, first fixation duration*
Dank et al. (2015)	Subject-predicate gender agreement in Hebrew (the attraction phenomenon).	English example: *The newspaper wrote that the stew*(masc.) *of the (male) cook*(masc.) *had become famous*(fem.) *throughout the state.*	More *regressions in* (to the head noun), more *regressions out* (from the predicate). Increased *first fixation duration, gaze duration, total time* on the predicate.	
Deutsch and Bentin (2001)	Subject-verb gender agreement in Hebrew.	English example: *I enjoyed seeing how the actors*(masc.pl.)*/* the actresses*(fem.pl.) *were enchanting*(masc.pl.) *the tired audience*	For morphologically marked verbs: *first fixation duration, gaze duration,* and *second-pass duration.*	

**The term *gaze duration* is used in the current study.

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

form (*eating the*), as well as in the subsequent region (*food we*). Thus, the increase in regressions was "immediate, but short-lived" (Ni et al., 1998, p. 532). A study by Braze et al. (2002) used similar materials (but also including anomalies in past tense inflection, cf. Table 1) and found similar effects, as well as increased first-pass reading times in the verb region, for example, *cracking after*. It is worth of notice that both studies tested morphosyntactic anomalies which are typically not attested in natural speech.

Another strand of studies using ungrammatical items have investigated so-called attraction phenomena, for example, when a word erroneously agrees with a local distractor noun instead of the head noun (Hallberg & Niehorster, 2021). Attraction errors have been investigated in subject-verb number agreement in English (Lim & Christianson, 2015; Pearlmutter et al., 1999) and in subject-predicate gender agreement in Hebrew (Dank et al., 2015). In general, these studies report higher regression ratios and increased total times on the anomalous word in ungrammatical sentences without a distractor, compared to anomalous sentences with a distractor, and to correct control sentences (Hallberg & Niehorster, 2021). However, the results, especially regarding early measurements, differ (cf. Table 1).

Based on the previous studies of morphosyntactic anomalies, Hallberg and Niehorster (2021, p. 32) conclude that syntactic anomalies "reliably produce increased regressions out from the site of the anomaly and from subsequent words, and often also longer reading time." Readers respond immediately, as they make more regressions. However, the time course regarding reading times is less clear. Ni et al. (1998) and Pearlmutter et al. (1999) do not find increased first-pass reading times. Dank et al. (2015) and Deutsch & Bentin (2001) find very early effects on first fixation duration, but Lim and Christianson (2015) do not. Finally, readers relatively quickly recover from the anomalies (e.g., compared to pragmatic counterparts, see Braze et al. (2002); Ni et al. (1998)).

The present study

In the present study, we investigated native readers' online responses to sentences with anomalous word order. The aim of the study was to test whether there was an expected slow-down in processing of the ungrammatical V3 sentences, compared to grammatical V2 baselines. According to the E-Z Reader model of eye movement control in reading (Reichle et al., 2009), severe syntactic violations can result in rapid integration failure of a word n. If the integration of n fails rapidly, the forward saccade to $n + 1$ is canceled. This results in a pause (increasing first fixation duration and gaze duration on n) and/or a refixation (increasing gaze duration) or an interword regression. Thus, the model predicts that "problems with postlexical integration can sometimes have very rapid effects" (Reichle et al., 2009, p. 10). Rather than assuming that integration only happens after the input is presented, the prediction-based approaches to sentence processing (Christiansen & Chater, 2016; Kamide 2008; Levy, 2008) assume that readers make predictions about the input before it is presented, for example, about the word order of upcoming sentences. When these predictions are not met, extra resources are spent, reflected in increased reading times (Kristensen & Wallentin, 2015). Based on previous eye-tracking studies of ungrammaticality, we expect to find similar surprisal effects on the subject and verb (the critical regions), manifested as longer fixation durations and more regressions out in the ungrammatical condition, and both manifested in reading measurements reflecting early (first fixation duration, gaze duration, first-pass regression ratio, regression path duration) and later stages of processing (total duration). Because previous studies (e.g., Braze et al., 2002; Huang and Staub, 2021; Pearlmutter et al., 1999) document that readers recover relatively quickly, we did not expect to see effects of ungrammaticality in the post-critical or wrap-up region. The results may give insights into how L1 readers react to different types of non-standard variation, by comparing the time course of V3 processing to results from previous eye-tracking studies of morphosyntactic anomalies and to eye-tracking studies of non-canonical, but grammatical, word order.

We also manipulated the length of the sentence-initial adverbials, which vary greatly in sentences with V3 in L2 production (Søby & Kristensen, to appear), in order to examine whether long sentence-initial constituents increase the severity of the ungrammaticality effect (inspired by Braze et al., 2002). Finally, we expected an adaptation effect for all trials, including the ungrammatical sentences, such that participants generally became faster and regressed less over time.

Method

Participants

Fifty-two native speakers of Norwegian participated in the study, primarily students and employees from the Norwegian University of Science and Technology. Participants were monolingual until starting school (with a wide variety of dialectal backgrounds) and had normal/corrected to normal vision and no reading deficits. None of them participated in the norming of the test stimuli. In compensation for participation, they chose between a gift voucher (160 NOK) and a lab t-shirt.

Data from four of these 52 participants were identified as outliers and were not entered in the analysis. Three of these participants were excluded because more than 33% of experimental trials had track losses or blinks in the critical regions. Track losses may indicate poor data quality (Staub & Goddard, 2019). A fourth participant was excluded due to a significantly high average sentence reaction time (>6.8 *SD* from group mean) (Weiss et al., 2018). This participant also read all sentences twice, a possible indicator of reading difficulties. No participants were excluded due to poor accuracy on comprehension questions (see results section). This left 48 participants in the analysis (18 males, 30 females; aged 19–36 years, M = 23.7 years, *SD* = 3.7 years).

Apparatus

Participants' right eyes were tracked using an EyeLink 1000 eye tracker (SR Research Ltd., Ontario, Canada) with a sampling rate of 1000 Hz. Stimuli were displayed in a fixed-width-font (Courier New, size 27) in black, on a light gray background. All sentences were displayed on a single line. Participants viewed stimuli binocularly on a monitor around 68 cm from their eyes so that approximately three

Table 2. Example of the four types of experimental stimuli

	Region				
	Pre-critical	Critical 1	Critical 2	Post-critical	Wrap-up
Grammatical (V-S) Short adverbial	*På tirsdager* 'On Tuesdays'	*tilbyr* 'offers'	*biblioteket* 'the library'	*høytlesning* 'a read-aloud'	*for barn og unge.* 'for children and adolescents.'
Ungrammatical (S-V) Short adverbial	*På tirsdager* 'On Tuesdays'	*biblioteket* 'the library'	*tilbyr* 'offers'	*høytlesning* 'a read-aloud'	*for barn og unge.* 'for children and adolescents.'
Grammatical (V-S) Long adverbial	*Klokken halv sju på tirsdager* 'Half past six on Tuesdays'	*tilbyr* 'offers'	*biblioteket* 'the library'	*høytlesning* 'a read-aloud'	*for barn og unge.* 'for children and adolescents.'
Ungrammatical (S-V) Long adverbial	*Klokken halv sju på tirsdager* 'Half past six on Tuesdays'	*biblioteket* 'the library'	*tilbyr* 'offers'	*høytlesning* 'a read-aloud'	*for barn og unge.* 'for children and adolescents.'

characters equaled 1 degree of visual angle. Head movements were minimized by using a chin rest and (when possible) a forehead rest. The experiment was written in Experiment Builder (SR Research Ltd., version 2.2.61).

Materials

Instructions and stimuli were written in *Bokmål*, the most commonly used standard for written Norwegian (Vikør, 2015). There were 40 items with four conditions in a 2×2 design: grammatical (V2) vs. ungrammatical (V); short adverbial vs. long adverbial.

All experimental items consisted of sentences with five regions[3] (cf. Table 2), and each region contained at least five characters. To avoid confounds, we compared exactly the same words or phrases to each other (besides from the long-short distinction), that is, they had the same length, shape, or frequency.

The **pre-critical region** contained either a short temporal adverbial or a long temporal adverbial. The last word(s) in the adverbial phrases (i.e., *på tirsdager*) were identical. "Short" adverbials consisted of 1–2 words between 5 and 12 characters including spaces (mean length = 9.1 characters). "Long" adverbials were at least twice as long, consisting of 4–7 words between 25 and 38 characters (mean length = 30.58 characters). A *t* test (correlated samples, one-tailed) showed a significant difference in number of characters between the two groups, $p < .0001$. All 40 adverbials in the short condition were different. However, in order to create the 40 long adverbials, reusing adverbials was necessary. Structurally, all adverbials can be considered a unit since they can be topicalized together in the sentence.

The **two critical regions** (critical region 1 and critical region 2) contained either the subject followed by the verb (the ungrammatical condition) or the verb followed by the subject (the grammatical condition). All verbs were frequent (defined as having > 15.000 occurrences of the lemma in the HaBiT Norwegian Web Corpus, 2015)

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

and referred to typical everyday activities; they were in the present tense and all were transitive. Most of the verbs were reused once. In order to create some variation, many different subjects were used: typical Norwegian first names, nouns (*gutten* 'the boy,' *jenta* 'the girl'), kinship terms (*bestemor* 'grandmother'), occupations (*sjefen* 'the boss'), non-human subjects (*kattene* 'the cats'), and inanimate subjects (*kommunen* 'the municipality'). The length of the subjects was 5–13 characters (mean = 6.45).

The **post-critical region** contained a syntactic object, which referred to a physical object, an animal, or a human.

The **wrap-up region** contained another adverbial, primarily prepositional phrases like *på kjøkkenet* ('in the kitchen'). This region made it possible to distinguish between spill-over effects (i.e., when a region is "swamped by processing continuing from the (immediately) preceding region" (Vasishth, 2006, p. 97)) from the critical regions and sentence wrap-up effects.

The Appendix contains a list of all experimental items. Conditions were counterbalanced across four lists in a Latin square design, so each participant only saw each item in one of the four conditions. All participants were exposed to 10 items from each of the four conditions. Each list of stimuli was presented in four blocks, so that conditions were balanced across blocks. The presentation order was randomized, both of the blocks and of the trials in each block.

All lists also contained 40 filler sentences (see online-only Supplementary materials A) with various kinds of syntactic constructions (e.g., passives and cleft constructions). Half of the fillers contained morphological anomalies such as agreement errors, incorrect use of gender, or definite vs. indefinite form, which occurred in many different sentence positions. Ten fillers had a structure similar to the target items with a locative or temporal sentence-initial adverbial (some also containing a morphological anomaly). The purpose of the fillers was to avert participant expectations of V3 when a sentence-initial adverbial was presented.

Thirty items (50% targets and 50% fillers) were followed by a simple yes-no comprehension question about the content of the sentence (50% yes/no) in order to keep participants' attention and make them read for comprehension. All comprehension questions can be seen in the Supplementary materials A.

Norming

Prior to the eye-tracking experiment, a judgment task was carried out with 44 grammatical sentences, each in a short and long version, distributed in two lists. Forty-two participants, who did not later participate in the eye-tracking experiment, rated the naturalness of the sentences on a five-point Likert scale from 1 *very unnatural* to 5 *very natural.* In a second "correction" task, participants saw two incorrect sentences with V3 and were asked to state whether or not the sentences were grammatically correct in Norwegian and, if not, where something was wrong. In the judgment task, all items with an average score below three were either discarded (4 items) or changed and re-normed (3 items). In the correction task, 95% of the anomalies were discovered, indicating that this type of anomaly is noticed by native speakers.

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Procedure

Participants provided informed consent and various background information, for example, about handedness and dialect. Participants were instructed to read for comprehension in a natural manner and to avoid blinking while reading. A break screen appeared three times during the experiment, but breaks could be taken whenever needed. The experiment lasted around 15 min.

The eye tracker was calibrated using a nine-point calibration grid. Re-calibrations were performed during the experiment, if necessary. A short (two-trial) practice session followed the calibration. Participants responded to the questions by pressing buttons on the keyboard. Corrective feedback was given on the screen.

Analysis

Response accuracy

To ensure that all participants had read the sentences for comprehension, we analyzed the accuracy of comprehension questions. The group mean was >90% in all four experimental conditions, and all participants had at least 76% correct answers.

Data cleaning

The experimental trials were inspected visually in the EyeLink Data Viewer software package (SR Research Ltd., version 4.1.1). Trials with track losses and blinks in the two critical regions (subject and verb) were removed (following the procedure of e.g., Frisson et al., 2017; Micai, 2018; Warren et al., 2015). As noted above, data from four subjects were excluded. For the remaining 48 participants, track losses or blinks led to removal of 72 trials (3.6%), leaving 1,848 trials that were included in the analysis. Data were cleaned using the automatic Four-stage Fixation Cleaning in Eyelink Data Viewer (SR Research): Short fixations (<80 ms) within one character position of a preceding or following fixation longer than 80 ms were collapsed. Other fixations less than 80 ms in duration were removed, as were fixations greater than 1500 ms in duration (following Frisson et al., 2017; Milburn, 2018).

Reading measurements

We conducted analyses over the five regions (cf. Table 2). The following four standard fixation duration measures were computed:

- *First fixation duration*: the duration of the first fixation on a region during first-pass reading.
- *Gaze duration*: the total duration of all first-pass fixations on a region until leaving it in either direction.
- *Regression path duration*[4]: the total duration of all fixations from entering a region during first-pass reading until leaving it to the right, including any refixations on previous text.
- *Total duration*: of all fixations on a region.

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Furthermore, the following fixation ratio measures were computed:

- *First-pass regression ratio*: the proportion of fixations following fixation on a region that are regressive relative to that region, considering first-pass reading only.
- *First-pass skipping ratio*: the proportion of times when the target region is skipped during first-pass reading.

We included standard measures that both reflect early (first fixation duration, gaze duration, first-pass regression ratio) and later processing (total duration). Both total duration and regression path duration include gaze duration and cannot be independent of it. Regression path duration is sometimes categorized as a later processing measure because it includes re-reading. However, we consider it to reflect early processing, even though it includes re-reading, since it indicates how long it takes to move past a certain region during first-pass reading (Warren et al., 2015).

Statistical models

Data were analyzed using linear mixed effects models in RStudio (R Core Team, 2019, version 1.2.1335), using the lme4 package (Bates et al., 2015, ver. 1.1.21). *P*-values were obtained using the lmerTest package (Kuznetsova et al., 2017, ver. 3.1.1). All models included the following fixed effects: Grammaticality (grammatical vs ungrammatical), length (short vs long), trial order, and an interaction between grammaticality and length as well as between grammaticality and trial order. Models also included random effects of participant and item. Random slopes were not included in the models, as they either resulted in a "singular fit" or failed to converge (even in very simple models). Comparisons were coded using sum contrasts (Schad et al., 2020), so that short and grammatical were coded as −0.5 and ungrammatical and long were coded as 0.5.

For binominal data (skips and regressions), a generalized linear mixed model was used to carry out logistic regressions (Frisson et al., 2017). Trial order was rescaled to a scale from 0 to 1. In one case (skipping ratio in the pre-critical region), we used BOBYQA (Powell, 2009), an optimizer that allows more iterations for attempting to reach convergence.

Results

Model results for total reading time and for all eye-tracking measures in the five regions are found in the online-only Supplementary materials B. An overview of all main effects and interactions in the different regions is shown in Table 3 (unexpected effects are in italic writing). Because of the nature of our stimuli, syntactic subjects appear in different critical regions depending on whether the sentence is grammatical or ungrammatical (i.e., ungrammatical subjects are presented in the first critical region, and grammatical subjects are presented in the second critical region). However, we compare subjects to subjects regardless of sentence position.

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Table 3. Main effects and interactions

Region	Pre-critical	Critical regions		Post-critical	Wrap-up
Constituent	Adverbial	Subject	Verb	Object	Adverbial
Main effects	*På tirsdager* 'On Tuesdays'	**biblioteket* 'the library'	*tilbyr* 'offers'	*høytlesning* 'a read-aloud'	*for barn og unge.* 'for children and adolescents.'
Gram	**UG>G** TD	**UG>G** GD, RR, RPD, TD	**UG>G** FFD, GD, RR, RPD, TD	(no effects)	(no effects)
Length of adverbial	**LONG>SHORT** GD, RPD, TD ***SHORT>LONG*** *FFD*	**LONG>SHORT** FFD ***SHORT>LONG*** *RPD*	**LONG>SHORT** FFD (no effects)	(no effects) ***SHORT>LONG*** *GD, RPD*	**LONG>SHORT** RR, RPD
Trial order	**EARLY>LATE** SR, GD, RPD, TD	**EARLY>LATE** GD, RPD, TD	**EARLY>LATE** RR, RPD, TD	**EARLY>LATE** RPD, TD	**EARLY>LATE** GD, RR, RPD, TD
Interactions					
Gram*Length	(no effects)	(no effects)	(no effects)	*RPD*	(no effects)
Gram*Trial	TD	RPD, TD	RPD, TD	(no effects)	(no effects)

Subjects in ungrammatical conditions (in Critical region 1) are compared to subjects in grammatical conditions (in Critical region 2). Verbs in ungrammatical conditions (in Critical region 2) are compared to verbs in grammatical conditions (in Critical region 1). Unexpected effects are in italic writing. SR, first-pass skipping ratio. FFD, first fixation duration. GD, gaze duration. RR, first-pass regression ratio. RPD, regression path duration. TD, total duration.

Likewise, verbs in the grammatical conditions (in the first critical region) are compared to verbs in the ungrammatical conditions (in the second critical region).

Table 3 shows that we found the expected effects of grammaticality, with longer fixation durations and more regressions out, in the critical regions. The effects were found on several early measurements, such as first fixation duration (FFD) (only on the verb, though), gaze duration (GD), first-pass regression ratio (RR), regression path duration (RPD), which includes re-reading, and on the only late measurement, total duration (TD). This confirms that V3 causes immediate disturbance on the subject and subsequently on the verb. There were no reliable effects of grammaticality after the critical regions, confirming that V3 causes local disturbance, that is, participants recover quickly. An unexpected interaction between grammaticality and length was found on the object for regression path duration, but this effect is doubtful due to several factors. It only arises in one measure, is not localized in the critical region, and is accompanied by an unexpected main effect of length.[5]

The results of the length manipulation are mixed. If the length of the adverbial prior to the anomaly influenced processing of the anomaly, we should see crossing interactions between grammaticality and length in the critical regions. This is not the case, and thus, it seems that the effects of V3 are stable across contexts with short or long adverbials. Obviously, length effects were found in the pre-critical region, which was either short or long. Here, fixation durations, as expected, were longer in

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Table 4. Sentence reading times. Mean reading times (and standard deviations) are reported in ms

	Grammatical	Ungrammatical	Total
Short Adverbial	2103 (±838)	2292 (±973)	2197 (±912)
Long Adverbial	2859 (±1161)	3036 (±1350)	2946 (±1260)
Total	**2479 (±1080)**	**2659** (±1231)	**2569** (±1160)

the long conditions for three measurements, both early and late. We assumed that working memory load would be higher after long adverbials, so the length effects on first fixation duration in the critical regions and on two measurements in the wrap-up region were expected. However, since unexpected length effects, with longer fixation durations in the short conditions, were also found in the pre-critical, one of the critical and in the post-critical region, the results regarding length are uncertain.

Effects of trial, with shorter fixation durations and less regressions (or skips) for later trials, were found for several measurements in all regions, always for regression path duration and total duration, and often for gaze duration and first-pass regression ratio (in one case also for first-pass skipping ratio (SR). Crossing interactions between grammaticality and trial were also found in the pre-critical and critical regions for regression path and total duration, so that adaptation seemed greater in the ungrammatical conditions (but see the discussion on adaptation).

In the following subsections, we present the results in more details, first for total reading times and then for the five regions of the sentence.

Total sentence reading time

Table 4 shows total sentence reading time. As expected, there is an effect of grammaticality on total sentence reading time, which is longer in the ungrammatical conditions ($\hat{\beta} = 409.93$ ms, SE = 69.75, $t = 5.88$, $p < .001$), see Figure 1. There is also an obvious effect of length ($\hat{\beta} = 749.94$ ms, SE = 34.04, $t = 22.03$, $p < .001$). Furthermore, we find an increased reading speed for later trials, that is, an effect of trial order ($\hat{\beta} = -15.01$ ms, SE = 1.53, $t = -9.81$, $p < .001$). Finally, there is a crossing interaction between grammaticality and trial order ($\hat{\beta} = -11.02$ ms, SE = 3.08, $t = -3.58$, $p < .001$), as seen in Figure 2: The slope is much steeper for ungrammatical conditions, which seems to indicate a larger adaptation effect here.

Region-by-region eye movement measures

Table 5 shows means and standard deviations for all eye movement measures in the individual regions and is presented in more detail in the following sections.

Pre-critical region: Adverbial

In the pre-critical region, we did not expect grammaticality to affect any measures besides the total duration (the only late measurement). This pattern was confirmed. Participants had longer total durations in the ungrammatical conditions ($\hat{\beta} = 149.47$ ms, SE = 37.92, $t = 3.94$, $p < .001$), meaning that they regressed more to the sentence-initial adverbial from other regions.

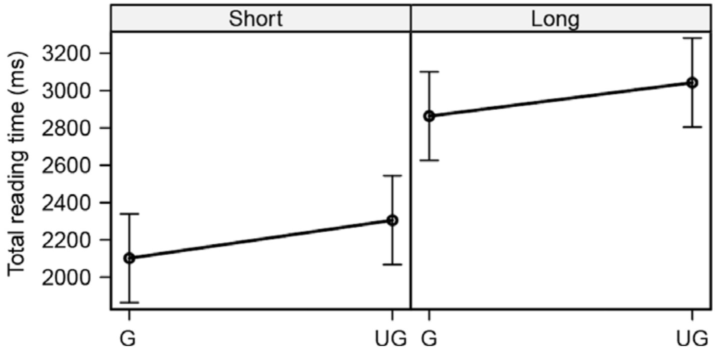


Figure 1. Effect Plot of Total Sentence Reading Time in ms.

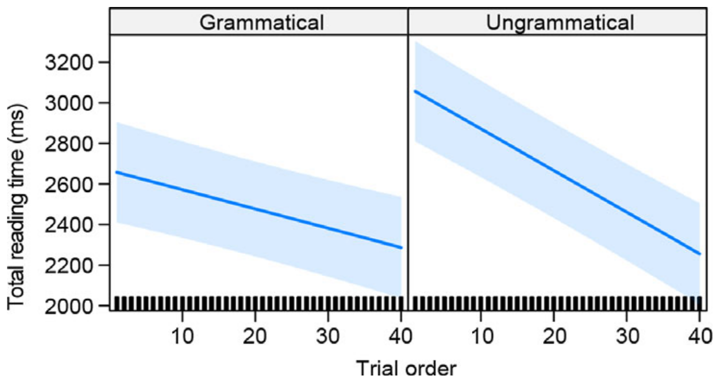


Figure 2. Interaction Between Grammaticality and Trial Order: Effect Plot of Total Sentence Reading Time in ms.

This region was either short or long, and we found an effect of length with increased durations in long conditions for several measurements: Gaze duration ($\hat{\beta} = 649.98$ ms, SE = 11.80, $t = 55.07$, $p < .001$), regression path duration ($\hat{\beta} = 650.13$ ms, SE = 11.79, $t = 55.15$, $p < .001$), and total duration ($\hat{\beta} = 780.92$ ms, SE = 18.53, $t = 42.15$, $p < .001$). Furthermore, there was an unexpected effect of length on first fixation duration ($\hat{\beta} = -8.23$ ms, SE = 3.07, $t = -2.68$, $p < .01$), with *shorter* fixations on long adverbials.

There were effects of trial order for first-pass skipping ratio ($\hat{\beta} = 1.06$, SE = 0.48, $z = -2.23$, $p < .05$), gaze duration ($\hat{\beta} = -1.37$ ms, SE = 0.53, $t = -2.57$, $p < .05$), regression path duration ($\hat{\beta} = -1.41$ ms, SE = 0.53, $t = -2.66$, $p < .01$), and total duration ($\hat{\beta} = -3.76$ ms, SE = 0.83, $t = -4.52$, $p < .001$). Participants became faster during the course of the experiment. They also made fewer skips in this region for later trials – perhaps because they discover that the information provided can be relevant for answering questions.

A crossing interaction between grammaticality and trial order was found for total duration ($\hat{\beta} = -4.31$ ms, SE = 1.67, $t = -2.58$, $p < .01$), showing a larger adaptation effect in ungrammatical conditions (see plot in Supplementary materials D).

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Table 5. Mean eye movement measures in all analysis regions (SD). Reading times in ms, skipping, and regression ratios in percentages (all reading times are rounded to the nearest millisecond)

	Condition			
Region	G, short	UG, short	G, long	UG, long
Pre-critical region: Adverbial				
First-pass skipping ratio (%)	9	7.8	0	3
First fixation duration	181 (68)	188 (80)	176 (63)	176 (62)
Gaze duration	373 (174)	388 (201)	1028 (414)	1023 (407)
First-pass regression ratio (%)	DNA	DNA	DNA	DNA
Regression path duration	373 (174)	388 (201)	1028 (414)	1023 (406)
Total duration	425 (237)	489 (305)	1209 (606)	1269 (692)
Critical regions: Subject				
First-pass skipping ratio (%)	7.3	7.5	9.7	7.1
First fixation duration	218 (72)	227 (79)	222 (67)	241 (89)
Gaze duration	252 (110)	293 (176)	253 (106)	296 (153)
First-pass regression ratio (%)	8.5	14	6.7	17.9
Regression path duration	288 (172)	403 (330)	273 (136)	368 (305)
Total duration	286 (180)	411 (287)	286 (189)	385 (252)
Critical regions: Verb				
First-pass skipping ratio (%)	6	5.4	6.5	4.9
First fixation duration	224 (75)	238 (94)	237 (79)	250 (102)
Gaze duration	265 (126)	277 (131)	270 (115)	284 (124)
First-pass regression ratio (%)	9.8	21.2	6.2	20.5
Regression path duration	313 (192)	393 (300)	301 (184)	403 (312)
Total duration	320 (188)	352 (211)	313 (188)	360 (194)
Post-critical region: Object				
First-pass skipping ratio (%)	2.8	3.4	3.4	4.4
First fixation duration	247 (82)	243 (80)	251 (106)	236 (89)
Gaze duration	356 (203)	357 (193)	342 (196)	320 (167)
First-pass regression ratio (%)	8.6	10.9	9.2	9.3
Regression path duration	406 (264)	433 (324)	398 (281)	374 (275)
Total duration	436 (284)	420 (256)	430 (280)	394 (247)
Wrap-up region: Adverbial				
First-pass skipping ratio (%)	2.8	3.4	3.4	4.4
First fixation duration	269 (127)	267 (128)	268 (124)	265 (109)
Gaze duration	570 (337)	569 (329)	550 (349)	548 (364)

(*Continued*)

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Table 5. (*Continued*)

Region	Condition			
	G, short	UG, short	G, long	UG, long
First-pass regression ratio (%)	26.9	23.7	32.8	30.7
Regression path duration	807 (597)	778 (553)	924 (770)	943 (898)
Total duration	637 (363)	621 (359)	621 (370)	629 (392)

DNA: In the pre-critical region, there is no previous text to look at and hence no regressions out.

Critical region: Subject

We compared eye movement measures on sentential subjects in the ungrammatical conditions (critical region 1) and subjects in the grammatical conditions (critical region 2). In the ungrammatical conditions, durations were longer and participants made more regressions (early measurements: Gaze duration ($\hat{\beta} = 46.61$ ms, SE = 11.77, $t = 3.96$, $p < .001$), first-pass regression ratio ($\hat{\beta} = 1.13$, SE = 0.32, $z = 3.54$, $p < .001$), regression path duration ($\hat{\beta} = 149.50$ ms, SE = 22.11, $t = 6.76$, $p < .001$), late measurement: Total duration ($\hat{\beta} = 170.72$ ms, SE = 18.07, $t = 9.45$, $p < .001$), likely reflecting increased processing difficulty in ungrammatical conditions.

First fixation duration was longer for long adverbial phrases than for short adverbial phrases ($\hat{\beta} = 8.16$ ms, SE = 3.39, $t = 2.41$, $p < .05$). Assuming that increased first fixation duration reflects processing difficulties, this indicates that participants paid more attention to subjects after long sentence-initial adverbials. However, this pattern was reversed for regression path duration ($\hat{\beta} = -27.98$ ms, SE = 10.80, $t = -2.59$, $p < .01$); durations decreased in the long conditions.

There were effects of trial order for gaze duration ($\hat{\beta} = -0.66$ ms, SE = 0.26, $t = -2.57$, $p < .05$), regression path duration ($\hat{\beta} = -2.21$ ms, SE = 0.48, $t = -4.56$, $p < .001$), and total duration ($\hat{\beta} = -2.25$ ms, SE = 0.40, $t = -5.69$, $p < .001$). This is reflected in shorter durations for later trials.

Crossing interactions between grammaticality and trial order were found for regression path duration ($\hat{\beta} = -2.28$ ms, SE = 0.97, $t = -2.35$, $p < .05$) and total duration ($\hat{\beta} = -2.92$ ms, SE = 0.80, $t = -3.67$, $p < .001$), showing larger adaptation effects in ungrammatical conditions (see Supplementary materials D).

Critical region: Verb

When comparing data for verbs in the grammatical vs. ungrammatical conditions, patterns similar to the subject regions were found. There were effects of grammaticality on all measurements (besides first-pass skipping ratio as verbs are not often skipped) with longer durations and more regressions in ungrammatical conditions (early measurements: First fixation duration ($\hat{\beta} = 22.59$ ms, SE = 7.95, $t = 2.84$, $p < .01$), gaze duration ($\hat{\beta} = 32.60$ ms, SE = 11.10, $t = 2.94$, $p < .01$), first-pass regression ratio ($\hat{\beta} = 1.11$, SE = 0.29, $z = 3.80$, $p < .001$), regression path duration $161.60 = \hat{\beta}$) ms, SE = 23.62, $t = 6.84$, $p < .001$), late measurement: Total duration ($\hat{\beta} = 88.06$ ms, SE = 16.49, $t = 5.34$, $p < .001$)).

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

As found for the subjects, there was an effect of length on first fixation duration ($\hat{\beta} = 12.08$ ms, SE = 3.87, $t = 3.12$, $p < .01$), so that durations increased in the long conditions.

There were effects of trial order for regression path duration ($\hat{\beta} = -2.61$ ms, SE = 0.52, $t = -5.04$, $p < .001$), first-pass regression ratio ($\hat{\beta} = -0.97$, SE = 0.28, $z = -3.48$, $p < .001$), and total duration ($\hat{\beta} = -1.60$ ms, SE = 0.36, $t = -4.42$, $p < .001$), reflected in shorter durations and fewer regressions for later trials.

Similar to the subjects, crossing interactions between grammaticality and trial order were also found for regression path duration ($\hat{\beta} = -3.50$ ms, SE = 1.04, $t = -3.37$, $p < .001$) and total duration ($\hat{\beta} = -2.40$ ms, SE = 0.73, $t = -3.31$, $p < .001$), showing larger adaptation effects in ungrammatical conditions (see Supplementary materials D).

Post-critical region: Object

In the post-critical region, there were no effects of grammaticality.

Unexpected effects of length were found for gaze duration ($\hat{\beta} = -26.70$ ms, SE = 7.25, $t = -3.69$, $p < .001$) and regression path duration ($\hat{\beta} = -34.36$ ms, SE = 12.06, $t = -2.85$, $p < .01$), with shorter durations in the long conditions.

A crossing interaction between grammaticality and length was found for regression path duration ($\hat{\beta} = -57.35$ ms, SE = 24.15, $t = -2.37$, $p < .05$) (see plot in Supplementary materials D). In the short conditions, regression path duration increased in the ungrammatical versions, but for the long conditions, it decreased in the ungrammatical versions.

Regression path duration ($\hat{\beta} = -1.15$ ms, SE = 0.54, $t = -2.12$, $p < .05$) and total duration ($\hat{\beta} = -2.16$ ms, SE = 0.44, $t = -4.97$, $p < .001$) showed shorter fixations and fewer regressions for later trials.

Wrap-up region: Adverbial

In the wrap-up region, there were no effects of grammaticality on any measures.

Length effects were found for first-pass regression ratio ($\hat{\beta} = 0.40$, SE = 0.12, $z = 3.41$, $p < .001$) and regression path duration ($\hat{\beta} = 138.71$ ms, SE = 26.92, $t = 5.15$, $p < .001$); participants had longer durations and made more regressions in the long adverbial conditions.

For later trials, there were decreased durations and fewer regressions for the early measurements gaze duration ($\hat{\beta} = -3.79$ ms, SE = 0.54, $t = -6.95$, $p < .001$), first-pass regression ratio ($\hat{\beta} = -0.67$, SE = 0.21, $z = -3.13$, $p < .01$), regression path duration ($\hat{\beta} = -8.72$ ms, SE = 1.21, $t = -7.21$, $p < .001$), and the late measurement total duration ($\hat{\beta} = -5.62$ ms, SE = 0.56, $t = -9.34$, $p < .001$).

Post hoc analysis with combined critical regions

Since word order is V-S in the grammatical conditions and S-V in the ungrammatical conditions, we compared constituents in different sentence positions. In order to check whether this confounded the results, we carried out a post hoc analysis on a unified subject-verb region. The only reading measurement which we could

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

calculate for the combined subject-verb region post hoc was total duration. The model results of total durations for the combined region showed the same effects as the original analyses of the two regions (see model results in the Supplementary materials C), that is, no indication of a confound.

Discussion

In sum, how do the eyes move in response to anomalous V3 word order?

In the pre-critical region (the short vs. long adverbial, for example, *På tirsdager* 'On Tuesdays'/*Klokken halv sju på tirsdager* 'Half past six on Tuesdays'), participants displayed longer total durations in the ungrammatical conditions, as expected. This is because participants regressed more to the sentence-initial adverbial from other regions. Because this region was either short or long, length effects on several measurements were expected and found. However, there was also an unexpected effect of length on first fixation duration, so that fixations were shorter in the long conditions.

Results for the two critical regions, the subject (e.g., *biblioteket* 'the library') and the verb (e.g., *tilbyr* 'offers'), were quite similar. There were effects of grammaticality on most measurements besides first-pass skipping ratio (and first fixation duration on the subject). Fixation durations were longer, and more regressions were made in the ungrammatical conditions. In both regions, first fixation duration (assumed to reflect processing difficulties) was longer after long adverbials, indicating that participants paid more attention in this condition. However, this effect of length was not echoed in other measurements – on the subject, a reversed effect of length was found for regression path duration, which decreased in the long conditions.

In the post-critical region, the object (e.g., *høytlesning* 'a read-aloud'), no main effects of grammaticality were found. An unexpected interaction between grammaticality and length was found for regression path duration. It was only found for one measurement and was furthermore accompanied by an unexpected effect of length (which was also found for gaze duration), with durations *decreasing* in the long conditions.

In the wrap-up region, the second adverbial (e.g., *for barn og unge* 'for children and adolescents'), no effects of grammaticality were found. There were effects of length on regression path duration and first-pass regression ratio, with more regressions and longer durations in the long conditions. This could be explained by a heavier load on working memory – the need for regressing to previous parts in the sentence is likely greater when sentences are long.

In sum, participants responded immediately to the V3 anomalies, as reflected in longer fixation durations and more regressions out on the subject and subsequently the verb. Participants recovered quickly, already on the word after the misplaced subject and verb, that is, the object. The effects of V3 were stable across contexts with short or long sentence-initial adverbials. Finally, participants generally read faster and regressed less for later trials.

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

Effects of V3 – a prominent anomaly

Our results are in line with previous EEG studies of Swedish V3 (e.g., Andersson et al., 2019), as we also found a reaction to V3 after temporal adverbials on online processing.

Previous eye-tracking studies with ungrammatical items have addressed morphosyntactic anomalies, for example, agreement errors (Dank et al., 2015; Deutsch & Bentin, 2001; Lim & Christianson, 2015; Pearlmutter et al., 1999), anomalous verb conjugations (Braze et al., 2002; Ni et al., 1998), and randomly transposed words (Huang & Staub, 2021). Their results varied regarding the time course of the effects found. As expected, our results were similar to those of Huang and Staub (2021), whose word order manipulation caused early and sustained disruption on the critical word. Furthermore, our results are similar to the studies of gender agreement in Hebrew (Dank et al., 2015; Deutsch & Bentin, 2001) as they both found effects on early (including first fixation duration) and later measurements. The only other study which included first fixation duration was Lim and Christianson (2015), who surprisingly did not find effects of missing subject-verb agreement on regressions out or first fixation duration in English. Based on Huang and Staub (2021), our study of V3, and the studies of Hebrew (Dank et al., 2015; Deutsch & Bentin, 2001), it seems that word order anomalies and morphosyntactic anomalies elicit the same responses, with similar time courses. However, as our experiment does not directly compare the two, it remains uncertain whether there are differences in prominence when reading. A behavioral error detection study in Danish shows that there are indeed differences in prominence. High school students underlined different anomalies (syntactic, morphological, and orthographic) in texts under time pressure. As much as 71% of the V3 anomalies were discovered, compared to 59% of anomalous verb conjugations and 55% of gender mismatches in NP-s (Søby et al., to appear). Behavioral data from our eye-tracking study confirm that V3 is a prominent anomaly. In a post-experimental interview, all participants either reported or confirmed (if they did not mention it initially) to have noticed the word order anomalies. Also, a different set of participants, who carried out a correction task when norming the stimuli, corrected 95% of sentences with V3.

The reaction to V3 anomalies in our study was immediate, as reflected in effects on early measurements on the subject. Previous eye-tracking studies that compared grammatical, but non-canonical OVS word orders to canonical SVO word orders in Spanish (e.g., Gattei et al., 2021) primarily found effects on later measurements. The early effects in our study and in Huang and Staub (2021) therefore seem unique to ungrammatical, not just atypical, word order. This suggests that the degree of acceptability for non-standard variation has consequences for the reactions seen in the eye-tracking record. Similarly, Sayehli et al. (2022) suggested, based on their EEG study, that V3 after *kanske* 'maybe' (which is a more acceptable construction) was processed differently than V3 after other adverbials.

Adverbial length does not affect processing of V3

To test whether the length of the preceding constituent affected anomaly processing, we manipulated the length of the first constituent. However, the manipulation did

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

not result in crossing interactions in the critical regions, suggesting that the effects of V3 are stable across contexts with short or long adverbials. Instead, we found main effects of length on a few measurements for the subject, verb, and second adverbial which could be explained by a heavier working memory load in the long conditions. However, since these were accompanied by unexpected effects of length for a few measurements on the first adverbial, subject, and object, the interpretation is uncertain.

Braze et al. (2002) also examined whether readers' sensitivity to anomaly detection and anomaly processing is affected by variation in processing load prior to the anomaly. They hypothesized that "[i]mposing a decoding challenge prior to the anomaly might plausibly reduce a reader's capability to cope with the anomaly" (Braze et al. 2002, p. 4). They varied the length and frequency of the subject nouns preceding the anomalous verbs, and length and frequency were correlated, so that long nouns (mean length: 9.94 letters) were reliably lower in frequency than short ones (mean length: 5.39 letters), but found no consistent effects of length, possibly due to a relatively small difference in length between the nouns. The difference between short and long conditions in our experiment was larger. We initially assumed that longer (and less common) adverbial phrases are more demanding on working memory until the point of the anomaly than short (and frequent) ones. Thus, we expected a (larger) effect of length for the ungrammatical sentences (i.e., an interaction), manifested as longer fixation durations and more regressions in the critical regions after long adverbials compared to short. Yet, length effects could also manifest as *less* disturbance after long adverbials. Participants might overlook more anomalies in the long condition, that is, increased processing load prior to the anomaly might camouflage its presence.

A corpus study of learners' production of written Danish by Søby and Kristensen (to appear) found that V3 anomalies occur most frequently after subordinate clauses, for example, *Selv om det er rigtig sjovt, jeg* [S] *savner* [V] *dig!* 'Even though it is a lot of fun, I miss you!' Although the length of the adverbial did not affect the processing of the anomalies in our study, there may be differences between processing the long adverbials in our study and the even lengthier and more structurally complex subordinate clauses in naturally occurring V3 anomalies.

The (non)finding regarding sentence-initial adverbial length is supported by data from the Danish error detection study (Søby et al., to appear) who found no significant differences in the probabilities of discovering V3 anomalies after short vs. long adverbials. Furthermore, although the EEG studies of Swedish V3 (Andersson et al., 2019; Yeaton, 2019) included a length manipulation of the sentence-initial adverbials, they did not report results regarding length effects.

Effects of trial: Task adaptation or syntactic adaptation to V3?

It is well documented that participants can adapt to the experimental task and perform faster and better during an experiment (e.g., Kristensen et al., 2014; Prasad & Linzen, 2021). An interesting question is whether participants also adapt to word order anomalies, such as V3. According to prediction theory, language users constantly update their expectations to language input (Kristensen & Wallentin, 2015; Levy, 2008). Therefore, it may be that the first occurrence of a word order anomaly results in a surprisal effect and disrupted eye movements, but that, for later occurrences, readers update their expectations for language input, and adapt to the anomaly at hand.

In this study, we found that participants in general read sentences faster for later trials, that is, an adaptation effect. This effect was seemingly larger for ungrammatical sentences. In the analysis of the five sentence regions, we also found effects of trial in all regions and on several measurements (see Table 3), as well as crossing interactions between grammaticality and trial order for total duration (the first three regions), and for regression path duration (the critical regions), suggesting that adaptation seemingly is greater in ungrammatical sentences. However, as an anonymous reviewer noted, due to the current study design, we cannot know whether the effects of trial are the result of syntactic adaptation to V3, or simply task adaptation. The speed-up in processing time could reflect a shift in task-related strategies. Participants might lose focus toward the end of the experiment and read faster or learn that the comprehension questions can be answered correctly with less re-reading. As pointed out by the reviewer, task-related effects might not reliably affect early processing measurements, such as first fixation duration and gaze duration, but task-related effects are likely to affect regression strategies (Weiss et al., 2018), and thus the late measurement, total duration, as well as regression path duration, which includes refixations on previous text. Task adaptation predicts a main effect of trial, but could also predict an interaction between grammaticality and trial, if the grammatical conditions have floor-level regressions to begin with. The fact that first fixation duration is never affected by trial order, as well as the fact that the interactions between grammar and trial are only observed in regression path duration and total duration, speaks in favor of the adaptation effect simply being due to task adaptation rather than satiation towards V3 (or a combination of the two).

Going forward, better-suited study designs could examine adaption to V3. However, finding a task that is less vulnerable to strategic processing is difficult. V3 sentences do not express different propositional content, and therefore one cannot ask control questions where the anomaly is crucial. One option is to use a between-group design like Prasad & Linzen's (2021) and compare V3 effects in two groups of participants: one exposed to V3 sentences prior to the actual experiment, and one exposed to filler sentences. In this way, it could be clarified whether there is syntactic adaptation "over and above" task adaptation (Prasad & Linzen, 2021, p. 19). Also, one could test participants with great exposure to V3, for example, from a spouse with L2 Norwegian or with friends speaking the multiethnic urban vernacular, to see whether they react less to V3. Adaptation to non-standard syntax after great exposure, that is, change in predictions based on non-standard input, speaks in favor of prediction-based approaches to sentence processing (e.g., Christiansen & Chater, 2016).

Applications of the study

There is surprisingly little research on native speakers' processing of non-native or non-standard syntax. The current study used manipulations based on naturally occurring anomalies typical of L2 learners, increasing the ecological validity.

Thus, the results can be valuable to research on processing of non-standard language varieties, including future models of sentence processing which should be able to accommodate "noisy" input from non-proficient language users and other types of non-standard variation. It may also contribute to research on L2 processing, being a useful baseline for comparison. The study could, for example, be repeated with two groups of L2 speakers of Norwegian (one whose L1 features V2, one whose L1 does not) to examine crosslinguistic influence, as in Andersson et al. (2019). Furthermore, our study is a first step in helping language instructors prioritize which aspects of grammar to focus on in an often tight curriculum. The behavioral data from the Danish proofreading study (Søby et al., to appear) indicate that V3 is noticed more than other common L2 anomalies. However, future studies on online processing of other L2 anomalies in Norwegian are needed to make a direct comparison with processing of V3 in this study.

Norwegian, Danish, and Swedish are to a great extent mutually intelligible (Vikør, 2015). Compared to Danes and Swedes, Norwegians are described as being more receptive to linguistic variation (Torp, 2004). In the Norwegian "polylectal" language situation, dialect use is well-accepted, with dialects used widely in all registers and contexts, and no officially codified spoken standard variety of the language (Havas & Vulchanova, 2018; Røyneland, 2009). Furthermore, Norwegian has two distinct written standards: *Bokmål* ('Book Language') and *Nynorsk* ('New Norwegian'), both taught in school. Even in this context, with active diglossia at both the spoken and written level, including grammar, we find clear responses and sensitivity to syntactic anomalies. Therefore, we expect that native speakers of other V2 languages will show the same – or an even larger – degree of sensitivity to V3 anomalies. Indeed, Andersson et al. (2019) found ERP effects in the processing of V3 in Swedish. Interestingly, that study, which also included learners of Swedish, found that effects were more native-like for German learners whose L1 also features V2 than for English learners. Thus, future controlled comparisons between native speakers and L2 learners' sensitivity to syntactic anomalies, and the impact of learner proficiency and language background, are in order.

Tolerance for various anomalies can be modulated by participants' perception of the speaker or experimenter, so that the tolerance and willingness to repair is higher for non-native speakers (Gibson et al., 2017; Hanulíková et al., 2012; Konieczny et al., 1994). We do not know if the participants in our study perceived the author of the stimuli as a non-native speaker, but due to the association between V3 and immigrant status (Freywald et al., 2015), combined with the relatively high amount of anomalies in the stimuli, including the fillers, it seems likely. The study was conducted by a Danish experimenter in Danish. This might have affected the participants – at the first appearance of an anomaly, some participants asked if the experimenter was aware that there was a mistake. However, even if they were affected by non-nativeness of either the experimenter or the stimuli, they still responded to the V3 anomalies. Whether tolerance towards V3 can be modulated, could for example, be tested in an EEG paradigm similar to Hanulíková et al. (2012), where the P600 effects of Dutch gender agreement errors disappeared when presented in a foreign accent. If such morphological processing and syntactic processing are similar, we would expect a similar decrease in response to ungrammatical V3 in Norwegian for speakers with foreign accent and speakers of multiethnic urban vernacular.

Conclusion

The present study demonstrates the consequences of using non-native syntax in written production aimed at native speakers. The study contributes new knowledge to the relatively unexplored field of native speaker responses to naturally occurring anomalies, for example, those produced by L2 learners of the language. Hopefully, this knowledge can be used to create more robust sentence processing models in the future, which can accommodate various types of "noisy" input from non-proficient language users and other types of non-standard variation.

Our results show that native speakers react immediately to V3 word order, as reflected in longer fixation durations and more regressions out on the subject and subsequently on the verb (for reading measurements reflecting both early and later stages of processing). Participants appear to recover from seeing the anomaly equally fast, however. The effects of grammaticality on fixation durations and regressions out are stable across contexts with short or long sentence-initial adverbials.

We argue that V3 is a prominent anomaly in V2 languages, to which native speakers show sensitivity and which negatively affects processing. This first step in a line of potential future studies of online processing of other L2 anomalies in Norwegian can help teachers and learners at language schools prioritize which aspects of grammar to focus on.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0142716422000418

Acknowledgements. This study was financed by Independent Research Fund Denmark. Thanks to Heming Strømholt Bremnes, Charlotte Sant, and Randi Alice Nilsen, NTNU, for help with translation and proofreading of stimuli. Thanks to Byurakn Ishkhanyan for helping with tricky code in R.

Conflict of interest. The authors declare none.

Notes

1 Syntactic processing difficulty can, among other things, be caused by ungrammatical items. We distinguish between word order anomalies (only manipulating the order of words) and morphosyntactic anomalies (involving morphological changes). Most other studies do not make this distinction and use the term syntactic anomaly as a cover term.

2 Due to length differences in the verbs, this measure was used instead of total fixation durations.

3 A reviewer pointed out that a portion of the V3 sentences are temporarily compatible with an analysis where the critical regions are inside a zero-relative clause: *Hver onsdag kveld Svein danser folkedans (blir Marit sjalu)* 'Every Wednesday night (that) Svein dances folk dance, Marit get jealous'. However, this is a very infrequent structure, and readers will not likely expect to see it. Furthermore, if participants parsed the ungrammatical sentences as grammatical until *folkedans*, there should be a slow down here when they realize that it is not a zero-relative clause. This is not the case.

4 Also referred to as *go past* (*duration*) – a term that has two definitions and is therefore avoided here.

5 We thank an anonymous reviewer for this interpretation.

References

Andersson, A., Sayehli, S., & Gullberg, M. (2019). Language background affects online word order processing in a second language but not offline. *Bilingualism (Cambridge, England)*, **22**, 802–825. https://doi.org/10.1017/S1366728918000573

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Bohnacker, U. (2006). When Swedes begin to learn German: From V2 to V2. *Second Language Research*, **22**, 443–486. https://doi.org/10.1191/0267658306sr275oa

Bolander, M. (1990). On the acquisition of word order rules in Swedish as a second language. In M. A. K. Halliday, J. Gibbons, & H. Nicholas (Eds.), *Learning, keeping and using language: Selected papers from the Eighth World Congress of Applied Linguistics, Sydney, 16–21 August 1987* (pp. 287–298). John Benjamins.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, **31**, 25–44. https://doi.org/10.1023/A:1014324220455

Clifton, C., Jr., & Staub, A. (2011). Syntactic influences on eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 895–910). Oxford University Press.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 341–371). Elsevier.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *The Behavioral and Brain Sciences*, **39**, 1–72. https://doi.org/10.1017/S0140525X1500031X

Dank, M., Deutsch, A., & Bock, K. (2015). Resolving conflicts in natural and grammatical gender agreement: Evidence from eye movements. *Journal of Psycholinguistic Research*, **44**, 435–467. https://doi.org/10.1007/s10936-014-9291-9

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. https://doi.org/10.1006/jmla.2000.2768

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Freywald, U., Cornips, L., Ganuza, N., Nistov, I., & Opsahl, T. (2015). Beyond verb second: A matter of novel information structural effects? Evidence from Norwegian, Swedish, German and Dutch. In N. Jacomine & B. Svendsen (Eds.), *Language, youth and identity in the 21st century: Linguistic practices across Urban spaces* (pp. 73–92). Cambridge: Cambridge University Press.

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

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, **95**, 200–214. https://doi.org/10.1016/j.jml.2017.04.007

Gattei, C. A., París, L. A., & Shalom, D. E. (2021). Information structure and word order canonicity in the comprehension of Spanish texts: An eye-tracking study. *Frontiers in Psychology*, **12**, Article 629724. https://doi.org/10.3389/fpsyg.2021.629724

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, **28**, 703–712.

HaBiT Norwegian Web Corpus. (2015). Web corpus for both bokmål (1.18 billion words) and nynorsk (55 million words). Last visited November 2019.

Hagen, J. E. (1992). Feilinvertering, overinvertering og underinvertering. *NOA. Norsk som andrespråk*, **15**, 27–38.

Hallberg, A., & Niehorster, D. C. (2021). Parsing written language with non-standard grammar: An eye-tracking study of case marking in Arabic. *Reading & Writing*, **34**, 27–48. https://doi.org/10.1007/s11145-020-10040-6

Hanulíková, A., Van Alphen, P. M., Van Goch, M. 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**, 878–887.

Hårstad, S., & Opsahl, T. (2013). *Språk i byen: Utviklingslinjer i urbane språkmiljøer i Norge*. Fagbokforlaget.

Havas, V., & Vulchanova, M. (2018). The nature and effects of Norwegian Diglossia. Å Skrive Nynorsk og Bokmål. Nye Tverrfaglege Perspektiv. Oslo: Samlaget.

Huang, K., & Staub, A. (2021). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition*, **216**, 104846–104846 https://doi.org/10.1016/j.cognition.2021.104846

Johansen, H. (2008). Inversjon i norsk innlærerspråk. En undersøkelse av variasjonsmønstre i skrevne tekster. *NOA norsk som andrespråk*, **24**, 50–71.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, **2**, 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Konieczny, L., Scheepers, C., & Hemforth, B. (1994). Reanalyses vs. internal repairs: Non-monotonic processes in sentence perception. *IIG-Berichte*, **8**, 2–22.

Kotsinas, U.-B. (2000). *Kontakt, variation och förändring: Studier i Stockholmsspråk*. Almqvist & Wiksell International.

Kristensen, L. B., Engberg-Pedersen, E., & Poulsen, M. (2014). Context improves comprehension of fronted objects. *Journal of Psycholinguistic Research*, **43**, 125–140. https://doi.org/10.1007/s10936-013-9241-y

Kristensen, L. B., & Wallentin, M. (2015). Putting Broca's region into context: fMRI evidence for a role in predictive language processing. In R. M. Willems (Ed.), *Towards a cognitive neuroscience of natural language use* (pp. 160–181). Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**, 1–26. https://doi.org/10.18637/jss.v082.i13

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, **36**, 1283–1315. https://doi.org/10.1017/S0142716414000290

Lund, K. (1997). Lærer alle dansk på samme måde? En længdeundersøgelse af voksnes tilegnelse af dansk som andetsprog. Special-pædagogisk forlag.

Micai, M. (2018). Reading comprehension mechanisms in autism spectrum disorder: Metacognitive processes and executive function [Doctoral dissertation, University of Seville]. Depósito de Investigación Universidad de Sevilla. https://idus.us.es/handle/11441/75576

Milburn, E. (2018). The effects of meaning dominance and meaning relatedness on ambiguity resolution: Idioms and ambiguous words [Doctoral dissertation, University of Pittsburgh). D-Scholarship. http://d-scholarship.pitt.edu/33234/

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye-movement patterns. *Journal of Psycholinguistic Research*, **27**, 515–539. https://doi.org/10.1023/A:1024996828734

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, **41**, 427–456. https://doi.org/10.1006/jmla.1999.2653

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Department of Applied Mathematics and Theoretical Physics, Cambridge England, technical report NA2009/06.

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **47**, 1156–1172. https://doi.org/10.1037/xlm0001046

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, **12**, 43–46.

R Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, **16**, 1–21.

Røyneland, U. (2009). Dialects in Norway: Catching up with the rest of Europe? *International Journal of the Sociology of Language*, **196/197**, 7–30.

Sayehli, S., Gullberg, M., Newman, A. J., & Andersson, A. (2022). Native word order processing is not uniform: An ERP study of verb-second word order. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.668276

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, **110**, Article 104038. https://doi.org/10.1016/j.jml.2019.104038

Søby, K. F., Ishkhanyan, B., & Kristensen, L. B. (to appear). Not all grammar errors are equally noticed. Error detection of naturally occurring errors and implications for eye tracking models of everyday texts.

Søby, K. F., & Kristensen, L. B. (2019). Hjælp! Jeg har mistede min yndlings rød taske. Et studie af grammatikafvigelser. *Ny forskning i grammatik*, **26**, 89–104.

Søby, K. F., & Kristensen, L. B. (to appear). Mother tongues matter! A cross-sectional study of V2 in L2 Danish.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either . . . or. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **32**, 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Staub, A., & Goddard, K. (2019). The role of preview validity in predictability and frequency effects on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **45**, 110–127. https://doi.org/10.1037/xlm0000561

Torp, A. (2004). Nordiska språk i urtid och forntid – språklikhet och språkskillnad, språkfamiljer och språksläktskap. In I. S. Sletten (Ed.), *Nordens språk med rötter och fötter* (pp. 19–74). Nordiska ministerrådet.

Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories. In *Proceedings of the international conference on linguistic evidence*, Tübingen, Germany, 96–100.

Vikør, L. S. (2015). Norwegian: Bokmål vs. Nynorsk. The Norwegian Language Council. http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk

Warren, T., Milburn, E., Patson, N. D., & Dickey, M. W. (2015). Comprehending the impossible: What role do selectional restriction violations play? *Language, Cognition and Neuroscience*, **30**, 932–939. https://doi.org/10.1080/23273798.2015.1047458

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology*, **71**, 198–210. https://doi.org/10.1080/17470218.2017.1307862

Yeaton, J. (2019). Visual processing of verb-second (V2) word order in second language acquisition: ERP Evidence from French-Swedish successive bilinguals [Master thesis, University of Lund]. https://jeremyyeaton.github.io/papers/Yeaton-2019-MasterThesis-V2_word_order.pdf

Appendix A

Stimuli

The stimuli appear in the grammatical conditions with Verb-Subject word order (V2). In ungrammatical versions (V3), the order is Subject-Verb. In long conditions, words in italics are displayed. In short conditions, words in parentheses are omitted.

1. *(Tidlig om morgenen) i helgen* leser pappa avisen på sofaen. ‘Early in the morning in the weekends, dad reads the newspaper on the sofa’
2. *(Før klokken halv åtte) hver dag* lufter mannen hunden sin i parken. ‘Before 7.30 every day, the man walks his dog in the park.’
3. *(Minst to ganger i uken) i 2020* holder kommunen nynorskkurs for offentlig ansatte. ‘At least twice a week in 2020, the municipality holds a Nynorsk course for public employees.’
4. *(Etter middag hver) lørdag kveld* spiser gutten gelato på Solsiden. ‘After dinner every Saturday night, the boy eats gelato at Solsiden.’
5. *(Etter klokken ett) om natten* løser Marit kryssord på mobilen. ‘After 1 AM, Marit solves crossword puzzles on her phone.’
6. *(Hver eneste ettermiddag) i jula* baker jenta pepperkaker hos bestemor. ‘Every single afternoon during Christmas, the girl bakes cookies at grandmother’s house.’

7. *(Før juleferien i) desember* sender bestefar julekort til alle barnebarna sine. 'Before the Christmas holidays in December, grandfather sends Christmas cards to all his grandchildren.'
8. *(Om ettermiddagen) på torsdager* spiller gutten fotball med vennene sine. 'In the afternoon on Thursdays, the boy plays football with his friends.'
9. *(Veldig tidlig) om morgenen* drikker hunden vann fra toalettet. 'Very early in the morning, the dog drinks water from the toilet.'
10. *(Om formiddagen) på søndager* synger Julie salmer i kirken. 'In the morning on Sundays, Julie sings hymns in the church.'
11. *(Omtrent klokken ni) om kvelden* skriver storesøster dagbok på soverommet. 'At around 9 PM, big sister writes in her diary in the bedroom.'
12. *(Nesten hver søndag) i januar* renser damen teppene sine i snøen. 'Almost every Sunday in January, the woman cleans her rugs in the snow.'
13. *(Etter kveldsmat) på mandager* vasker Harald sokker i vaskemaskinen. 'After dinner on Mondays, Harald washes socks in the washing machine.'
14. *(Hvert eneste år) den 17. mai* feirer Gunnar nasjonaldagen i Trondheim. 'Every single year on the 17th of May, Gunnar celebrates the National Day in Trondheim.'
15. *(På dager med snø) om vinteren* bygger Anders snømann på jordet. 'On days with snow in the winter, Anders builds a snowman on the ground.'
16. *(Hver onsdag kveld) om høsten* danser Svein folkedans til tradisjonell musikk. 'Every Wednesday evening in the fall, Svein dances folk dance to traditional music.'
17. *(Rett før daggry) en julidag* føder hesten et føll på gresset. 'Just before dawn a day in July, the horse gives birth to a foal on the grass.'
18. *(Klokken halv sju) på tirsdager* tilbyr biblioteket høytlesning for barn og unge. 'At 6.30 on Tuesdays, the library offers reading aloud to children and adolescents.'
19. *(På nesten alle kvelder) før jul* strikker Kristin gensere til hele familien. 'Almost every evening before Christmas, Kristin knits sweaters for the whole family.'
20. *(Hver mandag kveld) klokken seks* lager Håkon middag til kollektivet sitt. 'Every Monday evening at six o'clock, Håkon cooks dinner for his shared house.'
21. *(På triste gråværsdager) i april* leser bestemor magasiner i hagestuen. 'On sad overcast days in April, grandmother reads magazines in the garden room.'
22. *(Hele onsdag formiddag) før påske* maler barna påskeegg i barnehagen. 'All Wednesday morning before Easter, the children paint Easter eggs in the kindergarten.'
23. *(En gang om formiddagen) hver uke* vasker gutten sykkelen med såpevann. 'Once in the morning every week, the boy washes the bike with soapy water.'
24. *(Hver eneste dag) i ferien* bygger Helge terrasse i hagen. 'Every single day of the holidays, Helge builds a terrace in the garden.'
25. *(På lune solskinnsdager) i mars* besøker pensjonistene Botanisk hage inne i byen. 'On warm sunny days in March, the pensioners visit the Botanical Garden in the city.'
26. *(På sene ettermiddager) om våren* føder kattene ungene sine ute i stallen. 'On late afternoons in the spring, the cats give birth to their cubs in the stable.'
27. *(I oddetallsuker) i totiden* henter Astrid tvillingene på skolen. 'In odd weeks at two o'clock, Astrid picks up the twins from school.'
28. *(På alle hverdager) i november* strikker Helene strømper på bussen. 'On every weekday in November, Helene knits socks on the bus.'
29. *(På sensommerdager) i august* selger Eirik blomster på torget. 'On late summer days in August, Eirik sells flowers on the market square.'
30. *(Etter klokken ni) hver kveld* tilbyr restauranten middag til knallpriser. 'After nine o'clock every evening, the restaurant offers dinner at great prices.'
31. *(Fra 1. september) neste år* skriver Marius avhandling på universitetet. 'From the 1st of September next year, Marius writes his thesis at the university.'
32. *(De fleste dager) etter skolen* sender jenta meldinger på Snapchat. 'Most days after school, the girl sends messages on Snapchat.'
33. *(Før filmkveld) på fredager* kjøper vennene godteri på butikken. 'Before movie night on Fridays, the friends buy candy at the store.'

https://doi.org/10.1017/S0142716422000418 Published online by Cambridge University Press

34. *(På allehelgensaften) i oktober* lager Hilde gresskarlykter med datteren sin. 'On Halloween in October, Hilde makes jack-o'-lanterns with her daughter.'
35. *(Klokken halv elleve) før lunsj* spiser sjefen en kanelbolle på kontoret. 'At 10.30 before lunch, the boss eats a cinnamon bun in the office.'
36. *(Omtrent klokken ni) i kveld* synger Berit karaoke på puben. 'At around nine o'clock tonight, Berit sings karaoke in the pub.'
37. *(I partallsuker) om sommeren* selger Monica smykker på vikingmarkedet. 'In even weeks in the summer, Monica sells jewelry at the viking market.'
38. *(Hver ettermiddag) i februar* smører Trond skiene sine med voks. 'Every afternoon in February, Trond lubricates his skis with wax.'
39. *(På alle hverdager) etter jobb* baker Ingrid rundstykker på kjøkkenet. 'All weekdays after work, Ingrid bakes buns in the kitchen.'
40. *(Rundt klokken fire) på lørdag* treffer Hanne venninnen sin på kafé. 'Around four o'clock on Saturday, Hanne meets her friend at a café.'

Cite this article: Søby, KF., Milburn, EA., Kristensen, LB., Vulchanov, V., and Vulchanova, M. (2023). In the native speaker's eye: Online processing of anomalous learner syntax. *Applied Psycholinguistics* **44**, 1–28. https://doi.org/10.1017/S0142716422000418

Part III

Discussion and conclusion

8 Discussion

This section summarizes the main findings of the thesis with respect to the two research questions, addresses limitations and points to future directions. The two research questions of the thesis are (section 1.1):

RQ1, PRODUCTION:
How does the language background of the writer and linguistic context affect V2 production in written L2 Danish?

RQ2, PERCEPTION:
How do L1 users process V3 anomalies compared to other types of grammar anomalies?

Section 8.1 summarizes and discusses the findings in relation to RQ1 on production (Article 1), while section 8.2 does the same in relation to RQ2 on perception (Article 2 and Article 3). Section 8.3 presents the implications for current and future processing models, and section 8.4 presents the applied implications of the thesis. Finally, section 8.5 concludes the thesis.

8.1 Main findings: Production of V2 and V3

In the following, the main findings from the corpus study in Article 1 are summarized (section 8.1.1), and a perspective (i.e. the role of frequency) which is not covered in relation to production of V2 and V3 in the thesis is discussed (section 8.1.2). Lastly, the limitations of the study as well as future directions are discussed (section 8.1.3).

8.1.1 Article 1 (PRODUCTION)

The first article examined how the language background of the writer affected V2 production in L2 Danish. Specifically, the roles of learners' L1 (V2 vs. non-V2) and their proficiency level (CEFR levels A2-B1) were investigated. Furthermore, I examined how the linguistic context influenced V2 production; firstly, by including rough measures of the complexity of the three first constituents in the statistical model, and, secondly, by making an explorative, qualitative analysis of the sentence-initial adverbials, verbs and subjects.

Main findings (Article 1):

- ➢ V3 anomalies are frequent in L2 Danish (A2-B1).
    - In the corpus, one of four declarative sentences with non-initial subjects had V3 word order.
- ➢ V2 word order is, however, not difficult for all learners.
    - Learners with another V2 language as their L1 had a significantly higher share of V2 word order than non-V2 learners.
    - For the non-V2 learners, the share of V2 significantly increased with CEFR level (A2 to B1).
    - In this dataset, overuse of V2 was restricted to non-V2 learners.
- ➢ V2 word order is not equally challenging in all linguistic contexts.
    - There were significant effects of the complexity of the sentence-initial constituent (number of words) and the subject (one word vs. multiple words), so that accuracy decreased with the length of the first constituent and for subjects consisting of multiple words.
    - The effect of verb complexity (single vs. complex verbs) on V2 production trended statistical significance.
- ➢ V2 and non-V2 learners do not seem to be challenged by the same linguistic contexts.
    - In the qualitative analysis, we found different patterns for V2 vs. non-V2 learners, apart from the verbs. Both learner groups had lower shares of V2 in sentences with complex verbs consisting of a finite + non-finite verb vs. sentences with single verbs.
    - Only non-V2 learners seemed to find V2 after subordinate clauses in first position challenging.
    - The semantic content of the adverbials seemed to play a part for the non-V2 learners, as temporal and spatial adverbials had higher shares of V2 than argumentative and attitudinal adverbials (subordinate clauses excluded).
    - Only the non-V2 learners seemed to be negatively affected by subject complexity. However, the V2 learners produced very few complex subjects in this dataset.

8.1.2 Perspectives not covered in the thesis (PRODUCTION)

The role of anomaly frequency is central in the error detection study in Article 2, but the role of frequency in anomaly production is not examined in Article 1. In line with previous studies which have investigated the role of the sentential context for L2 Swedish and L2 Norwegian (Bolander 1989, Brautaset 1996, Hagen 1992, Hyltenstam 1978, Johansen 2008), I focused on the complexity of the constituents. However, the frequency of the constituents, or the entire XVS sequence, may also play a part in V2 vs. V3 production. Highly frequent and/or more chunk-like XVS sequences may be more entrenched in the minds of the L2 learners – in line with high frequency expressions (Olofsson & Prentice 2020) – and thus more likely to be retrieved as a unit with a word order according to norms. An example could be the highly frequent semi-schematic construction *Det tror jeg* 'I think so' (cf. Bolander 1989). Johansen (2008) did examine patterns related to the frequency of the constituent in first position in a qualitative analysis (highly frequent vs. somewhat frequent vs. unique) and concluded that constituents which are infrequent in first position tend to be followed by V3. However, frequency and complexity were not kept apart in the analysis, as most subordinate clauses were categorized as unique and thus low in frequency. Going forward, it would be relevant to examine the role of frequency by including the frequencies of all three constituents (i.e. how often those three occur together in a corpus) in the statistical model.

8.1.3 Limitations of the study and future directions (PRODUCTION)

This section first points to limitations related to using the L2 corpus for the scope of the thesis and discuss future directions. Second, I discuss whether the effects of L1 found in Article 1 are likely to be due to crosslinguistic influence.

The texts for the L2 corpus were collected before the scope of the thesis was narrowed down, i.e. the data were not collected specifically for studying crosslinguistic influence in relation to V2 word order. As a consequence, the dataset in Article 1 was not balanced with respect to CEFR level and language background. The results thus come with the reservation that the V2 learners were generally on higher CEFR levels, although proficiency level did not seem to affect V2 production for the V2 learners. Furthermore, the CEFR levels were only estimated based on which modules the learners attended, and the corpus does not contain data from modules corresponding to B2 or C1. The data in Article 1 indicate that CEFR level (on proficiency levels A2 to B1) plays a minor role compared to L1 background concerning V2 production, but there was not enough data to include an interaction between L1 and CEFR level in the statistical model. With future studies based on more balanced corpora concerning L1 and CEFR levels (e.g. from A2-C1), and with better estimates of learners' proficiency levels (e.g. with assistance from experienced language instructors), it would be possible to test interactions between L1 and CEFR level. If I were to collect data for the L2 corpus again, I would also consider collecting L1 texts from the learners as well, as these can be useful when examining crosslinguistic influence in general (cf. section 3.2.3). L1 texts can help "to determine the extent to which the L2 patterns really do reflect L1 tendencies" (Tenfjord et al. 2017:9).

An alternative to using a text-based corpus for the study in Article 1 could be to elicit sentences using a fill-the-gap task like Hyltenstam (1978). This could be an efficient way of creating a balanced dataset regarding language background and CEFR level, and an efficient way of reproducing results from Article 1, as well as testing new hypotheses based on the descriptive analysis. For example, it would be relevant to test whether there is an effect of the length of sentence-initial phrases not containing a subordinate clause (for non-V2 learners), to find out whether the challenges with heavy constituents in first position are only driven by subordinate clauses, or if they can be driven by other types of long constituents as well (Article 1). Furthermore, it would be possible to control for differences in constituent material when examining effects of subject complexity. The comparison between one word and multiple word subjects in Article 1 was likely to be largely correlated with a comparison between pronominal and non-pronominal subjects. Future studies could help determine the separate contributions of word class (pronominal vs. non-pronominal) and length.

Even though significant effects of L1 background (V2 vs. non-V2) were found on correct V2 production in Article 1, it cannot be ruled out that the higher accuracy of the V2 learners is not due to crosslinguistic influence, but to a general benefit of learning a language closely related to one's L1 (Jarvis 2017). Based on a comparison of morphological anomalies per 1,000 words for V2 vs. non-V2 learners showing large variation within both groups, I argued that if the V2 learners have a general benefit, it is clearly expressed for syntax, but not for morphology. Thus, is seems likely that the differences in V2 production are due to CLI. This claim may have been strengthened, as suggested by Jarvis (2017), by examining L1 texts from the learners to verify that these learners do in fact use XVS word order in a similar way in their L1. As the L2 corpus did not contain such texts, I argued that clauses with non-initial subjects are common in e.g. Danish, Swedish and German, based on corpus studies (Danish: Kristensen 2013, German: Fabricius-Hansen & Solfjeld 1994, Swedish: Westman 1974). However, analysing L1 texts from the same learners would be preferable.

Another way to strengthen the claim about CLI being likely is to supplement corpus findings with psycholinguistic experiments, using various experimental paradigms. Andersson et al. (2019) used EEG to compare processing of V3 after sentence-initial adverbials by L1 users of Swedish and L2 learners, with either German (V2), French or English (non-V2) as L1. The eye-tracking study in Article 3 could be repeated with three similar participant groups: one group of L1 users (e.g.

Danish or Norwegian), one group of L2 users whose L1 features V2 (V2 learners), and one group of L2 users whose L1 does not (non-V2 learners), in order to examine whether V2 learners' eye movement patterns are more L1-like than non-V2 learners' in response to V3 anomalies.

8.2 Main findings: Perception of V3

This section summarizes the main findings of the two perception studies in Article 2 and Article 3 (section 8.2.1). Section 8.2.2 informs on the persepctives in relation to perception of V3 which are not covered in the thesis, while section 8.2.3 discusses the limitations of the perception studies and possible future directions.

8.2.1 Article 2 and Article 3 (PERCEPTION)

The second article investigated whether some types of naturally occurring anomalies attract more attention than others during reading. In an error detection study, Danish high school students read texts with different types of errors: syntactic errors (V3), morphological agreement errors (verb inflections; gender mismatches in NPs) and orthographic errors (all presented in different conditions). Furthermore, the study examined if there was a link between the type of errors that participants did not detect, the type of errors which participants produced themselves in a grammar quiz, and the type of errors which are frequent in highs school essays in general.

Main findings (Article 2):

- ➢ Different overall error categories (syntactic, morphological vs. orthographic) do not seem to attract the same amount of attention.
  - The highest detection rate was found for syntactic errors (71 % of all V3 errors were detected), followed by the two types of morphological errors (55 % detected for NP errors; 59 % for verb errors), and the lowest rate was found for orthographic errors (33 %). The study was not designed to directly compare these overall categories, as there are a number of confounds (e.g. their position in the sentences and in the text), and thus no statistical tests were conducted between error categories. Still, based on the descriptive statistics, V3 is a prominent (easily detected) error type, and grammar errors seem to attract more attention than orthographic errors.
- ➢ Attention to various subtypes of grammar errors is not uniform.
  - For the verb errors, participants detected more errors in heterophone than homophone pairs. They also detected more errors of the type infinitive for present tense than of the type present tense for infinitive.

- For the NP errors, there was an effect of word class, so that mismatching articles + nouns were detected more than mismatching adjectives + nouns. There was an effect of gender, so that participants detected more neuter for uter than uter for neuter. An interaction between word class and gender was found, so that the effect of gender was less pronounced for the articles than for the adjectives.
- For the V3 errors, there were no effects of the length of the sentence-initial adverbial (short vs. long). The probability of detection was high in both conditions.
- A complication of using naturally occurring errors is that potentially contributing factors co-vary between conditions. Two important, but entangled, factors seem to be the frequency of the error and the phonological similarity between the error and the correct form.

➢ Frequency (measured by error rates in high school essays) can in most cases be used (in the hypotheses) to predict detection rates of different types of errors, so that frequent errors are noticed less than infrequent errors. This result comes with the reservation that we only used descriptive statistics when accessing differences in error frequency.

- The detection rates for the three overall error categories (syntactic > morphological > orthographic) were inversely proportional with the error rates in L1 writing (according to the descriptive statistics). The syntactic errors have the lowest error rates in L1 writing and the highest detection rates. Contrarily, orthographic errors have the highest error rates and the lowest detection rates.
- In the subexperiments on grammar errors, the error types with relatively high error rates (errors in homophone verb pairs, mismatching adjectives in NPs, overuse of uter in NPs) had lower detection rates than errors with lower error rates (errors in heterophone verb pairs, mismatching articles in NPs, overuse of neuter in NPs).

➢ Phonology seems to interfere with grammatical processing during error detection.

- The higher detection rates for heterophone vs. homophone verb pairs may also be explained by phonological interference. This effect cannot be disentangled from that of frequency.
- For the NP errors, phonological similarity between forms may explain the interaction between gender and word class. Frequency differences may also account for the effect, however, they are small. Thus, it seems more likely that phonological similarity plays an important part in explaining the low detection rates for uter for neuter in adjectives.

- ➢ High scores in the grammar quiz and high levels of self-reported irritation with errors both affect error detection positively.
  - Although participants' performance in the grammar quiz was almost at ceiling, they all overlooked errors in the detection task (54 % of all errors were detected). Their quiz scores explained some of the variance in the detection rates. For the V3, verb, and NP errors, there was an effect of the total grammar score, so that the more correct answers participants had in the grammar tasks in the quiz, the more grammar errors they detected. Likewise, the more correct answers participants had in the spelling task, the more orthographic errors they detected.
  - The more irritated with language errors participants reported to be, the more errors they detected.
- ➢ Current models of eye movement control in reading (Reichle et al. 2009) only address what happens when readers encounter severe syntactic violations, or do not address grammar errors at all (Engbert et al. 2005). Future models should be able to accommodate that attention to different types of naturally occurring errors is not uniform.
  - Based on the study, we recommend that future models take the following factors into account, which may modulate attention: 1) Variation in the type of grammar errors, 2) variation in error frequencies, and when present: phonological resemblance to the correct form, 3) variation in the reader's grammatical awareness and proficiency.

The third article examined L1 Norwegian users' online processing of V3 vs. V2 in an eye-tracking study of reading. The length of the sentence-initial adverbial was also manipulated (short vs. long), similarly to the study in Article 2. The results were compared to previous eye-tracking studies of ungrammaticality, which have primarily examined morphosyntactic anomalies.

Main findings (Article 3):

- ➢ L1 Norwegian users react immediately to V3 word order, but recover quickly.
  - We found longer fixation durations and more regressions out on the subject, and subsequently on the verb. The effects were found on several early measurements: first fixation duration (only on the verb, though), gaze duration, first-pass regression ratio, regression path duration, and on the only late measurement: total duration.
  - Participants recovered quickly, already on the word after the misplaced subject and verb (the object).
- ➢ The effects of V3 are stable across contexts with short vs. long sentence-initial adverbials.

- V3 affects the total reading time of the sentence negatively.
    - Participants did, however, generally read faster and regress less for later trials. We argue that this is likely due to task adaptation rather than syntactic adaptation to V3 (or a combination of the two).
- Word order anomalies and morphosyntactic anomalies seem to elicit the same responses with similar time courses.
    - This conclusion comes with reservations, as eye-tracking studies of ungrammaticality differ with respect to language, anomaly type, purpose of the study, and the reading measurements included.

Section 6.4 described how the studies in the thesis provide an empirical foundation for developing theories of production and perception in combination by pointing to some of the similarities between production and perception patterns. In sum, the thesis has shown that very different aspects of production are relevant to include in perception studies, ranging from anomaly frequency, phonological similarity to the correct form, and individuals' grammatical awareness (Article 2). Based on Article 1 and Andersson et al. (2019), it also seems that learners' language background (V2 vs. non-V2) may contribute to both production and perception of anomalies. However, more systematic and direct methods for examining the relation between production and perception can be developed, e.g. investigating the same language user's production and perception (cf. section 8.2.3). Examining L2 users' perception of anomalies, not only L1 users' as in this thesis, is also relevant, as mentioned in the next section.

8.2.2 Perspectives not covered in the thesis (PERCEPTION)

This section discusses aspects of anomaly perception not covered in the thesis. These include perception of other types of V3 anomalies than *ASV word order, sentence comprehension, adaptation to anomalies, irritation with errors, participants' perception of the sender, perception in different contexts, and L2 users' perception.

The focus of the thesis is on V3 anomalies. In the perception studies, only the processing of adverbial-subject-verb sentences was examined (and the adverbials did not include subordinate clauses). Processing of sentences with other sentence-initials constituents, such as objects, remains unexplored. This thesis does also not examine perception of overuse of V2, i.e. verb-before-subject for subject-before-verb, which was also found in L2 production (Article 1).

Neither the error detection nor the eye-tracking study examined if the anomalies affected participants' comprehension of the sentences. In the error-detection study, it would be difficult to incorporate this aspect, as participants read long text passages containing multiple types of anomalies. In the eye-tracking study, it was difficult to create questions actually testing participants' comprehension (and not just short-term memory) of the short sentences presented. Examining if and how different types of grammar anomalies affect comprehension could provide important knowledge to current and future processing models.

The perception studies in this thesis were not designed to enlighten processing models concerning adaptation to grammar anomalies. In the eye-tracking study, effects of trial were found, but due to the design of the study, it could not be concluded whether the effects of trial were indeed caused by syntactic adaptation to V3 or simply task adaptation. In Article 3, it is argued that the adaptation effects are more likely to be due to task adaptation. Using a better-suited design (cf. Prasad & Linzen 2021), it would be relevant to examine adaptation to grammar anomalies. As argued in Article 3: If language users adapt to anomalies after excessive exposure, i.e. change their predictions based on the anomalous input, it would speak in favour of prediction-based approaches to sentence processing (Kuperberg & Jaeger 2016, Christiansen & Chater 2016, Kamide 2008, Levy 2008).

In the error detection study (Article 2), it was found that high levels of irritation with errors affected error detection positively. Here, participants scored their general irritation with errors. As Danish readers' irritation with errors are modulated by the type of error (Blom & Ejstrup 2019b), including more specific irritation scores could be relevant in future error detection and eye-tracking studies. None of the perception studies examine participants' irritation with V3 word order (or any other emotions which L1 users of Danish or Norwegian might have towards this anomaly). Another question which is out of scope of this thesis is whether tolerance towards V3 can be modulated by participants' perception of the speaker? That is, whether receivers' tolerance and willingness to repair is higher when the sender is an L2 user (Gibson et al. 2017, Hanulíková et al. 2012, Konieczny et al. 1994). As suggested in Article 3, this could be tested using an EEG paradigm similar to the study by Hanulíková et al. (2012), in which P600 effects of gender agreement anomalies in Dutch disappeared when they were presented in a foreign accent. If processing of these morphological anomalies is similar to processing of word order anomalies such as V3 (cf. Article 3), a similar effect of accent would be expected for V3 anomalies. It may also be the case that tolerance increases when V3 anomalies are presented by a speaker of multiethnic urban vernacular (Hårstad & Opsahl 2013, Kotsinas 2000, Quist 2008). From a prediction-based point of view, the reason may not be that receivers are more tolerant towards L2 users, but that receivers have other predictions for L2 users.

As described in Article 2, language norms are subject to language change and sociolinguistic variation, and thus, natural texts both contain outright errors (according to the norms of The Danish Language Council) and anomalies in the grey-zone between "error" and "variation". In this thesis (Article 2), I have argued that the grammar anomalies under investigation were all outright errors in the context of a formal experiment conducted in a school setting, meaning that they deviated from the norms of the Danish Language Council (which e.g. schools are obliged to follow by law (Dansk Sprognævn 2023)), as well as from the majority of participants' answers in the grammar quiz. Although, this is likely to be true in a formal experiment in an educational setting, it may be less clear-cut in other contexts that especially V3 word order is considered anomalous by participants – and if so *how* anomalous. Participants' perception of what constitutes variation vs. outright errors may be affected by the context in which it is read, for example an experimental setting or educational setting vs. a natural setting such as reading text messages from a friend. The studies conducted in this thesis are not equipped to answer how individual readers differentiate between "erroneous" and "unusual" language, or equipped to answer whether this perception is modulated by the context. In future perception studies, grey-zone anomalies should be included, and the formality of the setting should be manipulated as well. In Article 2, it is suggested to examine the inflection of Danish modal verbs which may seemingly be subject to language change (Kristensen et al. in press).

Finally, this thesis does not examine L2 users' perception of grammar anomalies. In relation to the error detection study, this perspective could be relevant to examine: Do L2 users notice other anomalies compared to L1 users, and if so: Could this inform us further about the roles of error frequency and phonology? If different error rates in L1 vs. L2 Danish are reflected in different detection patterns for L1 and L2 users, it may inform us further on the role of frequency. For example, V3 anomalies are more frequent in L2 Danish. Does this mean that L2 users notice these less than L1 users (provided that they know they are anomalous)? L1 Danish users have a strong tendency to mainly confuse homophone verb pairs (present tense vs. infinitive), but this tendency is not found among L2 users with L1 English (Hansen et al. 2019). It would be interesting to see if this is reflected in perception. In relation to the eye-tracking study, comparing processing of L1 users with different L1s can also be used to examine CLI (cf. section 8.1.3).

8.2.3 Limitations of the studies and future directions (PERCEPTION)

In this section, I describe limitations in the designs of the two perception studies and discuss how the conclusions of the thesis may have been strengthened, either by designing the studies differently or by conducting additional experiments. The following themes are described: conducting the studies in two different languages, limitations in the comparisons of different anomaly types, limitations in relation to using the L1 corpus and the grammar quiz, the potential confound of position in the eye-tracking study, and the length manipulation used in both studies.

The two perception studies in the thesis are conducted in different languages: Danish and Norwegian. Although, the two languages are mutually intelligible (Vikør 2015), there may be differences between L1 Danish users' and L1 Norwegian users' attention to and processing of V3. As argued in Article 3, Norwegians are often considered more receptive to linguistic variation (Torp 2004), and therefore it may be that L1 Danish speakers show a larger degree of sensitivity to V3 anomalies. Conducting an eye-tracking study of L1 Danish speakers' processing of various grammar anomalies, including V3, would thus be relevant.

A limitation in the design of the error detection study in Article 2 is that detection rates for the overall error categories (syntactic, morphological and orthographic) could not be directly compared, because there were a number of confounds, such as the position of the errors in the sentences and in the text. Thus, no statistical tests were conducted between them. With a different design, more comparable contexts could be created across error categories. For example, errors could be placed in the same sentences and sentence slots: **Ofte <u>han</u> spiller skak* 'Often he plays.PRS chess' (syntactic error: V3) vs. **Ofte <u>spille</u> han skak* 'Often he play.INF chess' (morphological error: infinitive for present tense) vs. **Ofte <u>spiler</u> han skak* 'Often he plays.PRS chess' (orthographic error: single for double consonant). As there were already many conditions in the error detection study, this would complicate the design further and result in a very long experiment. The lack of a direct comparison of processing of morphological and syntactic anomalies is also a limitation in Article 3. In hindsight, the eye-tracking study would have been strengthened by including stimuli with morphosyntactic anomalies, instead of just comparing to previous eye-tracking studies of these.

In Article 2, the L1 error rates are based on a small corpus, and the differences between error rates are not tested using inferential statistics. This limits what can be concluded about the relationship between production and perception, and how closely the two are tied together. Going forward, a larger corpus of L1 texts is needed to test whether the differences in error rates are statistically significant. Also, the texts used to build the L1 corpus were collected by the Danish Language Council and shared with the BGB project. The Danish Language Council did not collect information about the high school students' language background. Thus, it is not known when the participants started learning Danish. If corpus data for Article 2 were to be collected again, I would 1) build a larger corpus, 2) test error rates using inferential statistics, and 3) collect information about participants' language background.

If the error detection study in Article 2 were to be replicated, an improvement would be to use a more challenging grammar quiz to avoid ceiling effects. This could e.g. be done by choosing other, more challenging error types than e.g. V3, and by using other tasks than fill-the gap or multiple choice. If students' scores in the quiz varied more, the link between production data and perception data could be examined more directly, as the grammar scores for each grammatical error type could be included in the model, instead of using a total score across tasks, as done in Article 2. A more directly, yet, more cumbersome, way of examining individual production and perception would be to collect essays from all participating students and calculate individual error rates (or another score).

A potential confound in the eye-tracking study in Article 3 is that constituents in different sentence positions were compared, e.g. subjects in ungrammatical sentences (ASV) were compared to subjects in grammatical sentences (AVS). I checked whether position was a confound in a post hoc analysis on a unified subject-verb region. However, only one reading measurement could be calculated post hoc, without access to the Eyelink Data Viewer software (SR Research). In an improved version of the study, data could be extracted from a unified subject-verb region as well. An alternative method to avoiding the potential position confound could be to include sentences with initial *kanskje* 'maybe' followed by subject-verb order as well. This adverb can both be succeeded by verb-subject and subject-verb word order (Beijering 2010, Boye 2005), and ERP analyses have shown stronger effects for V3 after spatial and temporal adverbials than after Swedish *kanske* 'maybe' (Sayehli et al. 2022). That is, sentences with *ASV (anomalous V3) could be compared to *kanskje*+SV (acceptable V3). If effects are found in the eye-tracking record between V3 after *kanskje* vs. other adverbials, they cannot be confounded by constituent position.

Both perception studies (Article 2, Article 3) included a length manipulation of the sentence-initial adverbial prior to the V3 anomalies, which in either case did not affect attention to, or processing of, the anomalies significantly. The initial plan was to analyze the data from the error detection study before choosing stimuli for the eye-tracking experiment. Due to a delay in the data collection, the results from the error detection study were, however, not ready before I went to NTNU to create the eye-tracking experiment. The idea that the length of the initial adverbial might affect attention to V3 was based on position effects found in letter detection studies (Smith and Groat 1979), so that elements in the beginning or end of a sentence tend to be more prominent than in the middle. The hypothesis was that participants would detect more V3 anomalies after a short adverbial than a long adverbial. In Article 3, it was examined whether the length of the preceding constituent affected anomaly processing, inspired by Braze et al. (2002), who hypothesized that anomaly detection and processing could be affected by variation in processing load prior to the anomaly. We hypothesized that longer and less common adverbial phrases may be more demanding on working memory than short (and frequent) ones, until the point of the anomaly, but based on the eye-tracking data, it seemed that the effects of V3 were stable across contexts with short or long adverbials (Article 3). In the discussion of the paper, we suggest that effects may be found using the even lengthier and more structurally complex subordinate clauses found in naturally occurring V3 anomalies (Article 3). This could have been done in the error detection study, but it may be more difficult in an eye-tracking study, as it is preferable to have sentences which can be presented in one line (cf. section 6.3).

Retrospectively, it may have been more relevant to include the length manipulation in a study where participants were L2 learners. In L2 production, there was an effect of the length of the sentence-initial constituent, measured as the number of words, but possibly being carried by subordinate clauses in first position (Article 1). Using a length manipulation, it could be tested whether sentence-internal factors in production also play a part in perception. Generally, further explorations of the role of sentence-internal factors, or context, on both production and perception would be relevant going forward.

8.3 Implications for current and future processing models

Although grammar anomalies are a natural part of everyday written communication, L1 user responses to naturally occurring anomalies is a relatively unexplored field (Article 2, Article 3). Generally within psycho- and neurolinguistics, language comprehension models are either based on processing studies with no anomalies (e.g. Ferreira & Patson 2007) or studies which include constructed anomalies (e.g. Bornkessel & Schlesewsky 2006, Friederici 2002, Hagoort 2005) (cf. Kristensen 2017:1). Within previous eye-tracking studies of ungrammaticality, surprisingly little research on L1 users' processing of anomalous word order exists, besides Huang & Staub (2021). They examined online processing of random transposition errors (**The white was cat big*), which may potentially be produced by language users when typing. However, this type of anomaly only has few occurrences in our L1 corpus. In contrast, V3 anomalies, which are examined in Article 3, are found in both oral and written production by L2 users (and potentially by speakers of multi-ethnic urban vernaculars). The lack of previous eye-tracking studies focusing on naturally occurring word order anomalies is surprising in the context of a globalized world with increased mobility, where L1 users need to accommodate to anomalies produced by adult L2 learners.

Attention to, and processing of, different types of grammar anomalies have not been a focal point in the current major models of eye movement control in reading (Reichle 2003, Reichle et al. 2009, Engbert et al. 2005). As discussed in Article 2, the E-Z Reader model (Reichle et al. 2009) makes predictions about how the eyes move in response to severe syntactic violations, but other types of anomalies are not explicitly mentioned. Grammar anomalies are often treated as a homogenous group with the cover term *syntactic violations* (Article 2, Reichle 2003, Reichle et al. 2009), but as found in Article 2, attention to different types of naturally occurring anomalies is not uniform, and the interplay between potential contributing factors is complex.

This thesis draws attention to the need for improving our understanding of the factors which govern attention and reaction to different types of grammar anomalies found in everyday texts. Furthermore, the thesis points to the general need within the fields of psycho- and neurolinguistics for building robust processing models in the future which can accommodate naturally occurring non-standard variation. Specifically, future models of eye movement control in reading should be able to accommodate that attention to different types of anomalies is not uniform. Finally, several factors which may modulate attention and thus should be taken into account in future eye-tracking models have been presented (cf. section 8.2.1 or Article 2).

8.4 Applied perspectives

To master V2 word order is often described as notoriously difficult for L2 learners (Bolander 1990, Hagen 1992), with different theoretical explanations of why V2 should be hard to acquire (e.g. Pienemann 1998, Lund 1997). Based on the findings in Article 1, this thesis provides a more nuanced picture: V2 is not equally challenging for all learners and is not equally challenging in all contexts. This more nuanced picture is consistent with language instructors' impressions in the online survey in section 4 (Gosselke Berthelsen & Søby, in prep.). According to the instructors, learners e.g. have more difficulties producing V2 in oral vs. written production, and learners in DP1 generally find V2 more difficult than learners in DP2, who again find it more difficult than in DP3. In DP3, which most of the learners of the L2 corpus attended, instructors rated learners' difficulty with V2 in writing as being neither particularly difficult or easy. Having empirical research finding that both L1, proficiency level and the complexity of the sentence can influence V2 production could be of didactic value to language instructors; for example when developing new teaching materials and when working with word order in classes. In Article 1, I also argued that introducing V2 as the basic Danish word order, instead of introducing Danish word order as subject-before-verb (with exceptions), might be beneficial. This could be tested empirically, e.g. in a practice-oriented intervention study.

The most frequent grammar anomalies in the L1 and L2 corpora were presented in section 3.3 (and Article 2). High school teachers and language instructors in L2 Danish are most likely aware of which types of anomalies their students frequently produce and can thus plan lessons focusing on grammar accordingly. This thesis takes a first step in providing knowledge about which anomalies are most prominent to readers (Article 2), adding a new aspect that may help teachers prioritize grammatical focus areas in an often tight curriculum. It may also be beneficial for writers to know which types of anomalies to focus on avoiding, if they want to ensure smooth communication (i.e. processing). Going forward, future error detection studies and online processing studies are needed to directly compare different types of other naturally occurring anomalies.

8.5 Conclusion

This thesis has examined the production and perception of naturally occurring grammar anomalies, focusing on V3 anomalies. The thesis was based on two RQs. The first was related to production: *How does the language background of the writer and linguistic context affect V2 production in written L2 Danish?* The cross-sectional corpus study in Article 1 showed that many factors play a part in relation to whether learners produce V2 or V3. Having an L1 which is also V2 facilitates V2 production. For learners with non-V2 L1s, increasing proficiency level facilitates V2 production. Contrarily, complex constituents (measured as the number of words) affects V2 production negatively. The second RQ was related to perception: *How do L1 users process V3 anomalies compared to other types of grammar anomalies?* The error detection study in Article 2 showed that V3 is a prominent anomaly compared to morphological and orthographic anomalies. The eye-tracking study in Article 3 showed that L1 users of Norwegian react immediately to V3 word order, as reflected in longer fixation durations and more regressions out on the subject and subsequently on the verb. V3 also increases the total reading time of sentences. Seemingly, V3 anomalies elicit the same responses in the eye-tracking record as previously tested morphological anomalies.

Methodologically, the thesis is novel by including production data (anomaly frequencies in L1 texts) in perception studies, thus combining production and perception. The studies in this thesis provide an empirical foundation for developing theories by pointing to some of the similarities between production and perception patterns. Anomaly frequencies in L1 texts seem to be a relevant factor in perception concerning allocation of attentional resources (Article 2). The thesis provides a preliminary model of how the two are linked (section 5.1.2). Individual differences such as the grammatical awareness (measured in a grammar quiz) (Article 2) and the language background of the participant (Article 1, Andersson et al. 2019) also seem to contribute to both production and perception of anomalies. Finally, the phonological similarity between the anomaly and the correct form (although often correlating with frequency) also seems important (Article 2). The thesis presents a tentative model of all potentially contributing factors in anomaly detection, which includes *language variables*, *situational variables*, and *cultural variables* (section 5.2). A deeper understanding of the relation between production and perception of anomalies is an issue for future research to explore.

Previous research in Danish as a second language has predominantly been qualitative and grounded in sociolinguistic theories, with little focus on anomalies or crosslinguistic influence. This thesis contributes with new perspectives by combining Danish as a second language with psycholinguistics. Results from psycholinguistic experiments can be used to inform teachers of Danish as L1 and L2 on how to prioritize grammatical focus areas in their classrooms. The studies in the thesis suggest that it is useful for L2 writers to prioritize avoiding V3 to ensure smooth communication, as V3 is prominent to L1 users.

Finally, the thesis raises awareness of the need for enhancing our understanding of the factors governing attention and reaction to different types of grammar anomalies found in everyday written communication. This is crucial to build robust processing models which can accommodate non-standard variation, not just within eye-tracking research, but generally within psycho- and neurolinguistics.

9 References

This list contains references mentioned in the extended introduction and in the extended discussion. References within the three research articles appear at the end of each article.

Andersson, A., Sayehli, S., & Gullberg, M. (2019). Language background affects online word order processing in a second language but not offline. *Bilingualism*, *22*(4), 802-825. https://doi.org/10.1017/S1366728918000573

Appelman, A. & Schmierbach, M. (2018). Make no mistake? Exploring cognitive and perceptual effects of grammatical errors in news articles. *Journalism & Mass Communication Quarterly*, *95*(4), 930-947. https://doi.org/10.1177/1077699017736040

ASK – The Norwegian Second Language Corpus (2015). https://clarino.uib.no/korpuskel-classic/corpus-list?collection=ASK. Last accessed 22.01.2023.

Beijering, Karin. (2010). The Grammaticalisation of Mainland Scandinavian MAYBE. *Bergen Language and Linguistics Studies*, *1*(1), 1-21.

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition*, *38*(6), 740-752.

Blom, J. N., & Ejstrup, M. (2019a). Læsernes holdninger til journalisters stavefejl i de digitale medier. In Y. Goldshtein, I. S. Hansen, & T. T. Hougaard (Eds.), *17. Møde om Udforskningen af Dansk Sprog* (pp. 113-132). Aarhus: Aarhus Universitet.

Blom, J. N., & Ejstrup, M. (2019b). BRÆJKING NEWS: En eksperimentel undersøgelse af fejlobservante læseres holdninger til ukorrekte og korrekte stavemåder opfattet som stavefejl i nyhedsrubrikker på nettet. *NyS*, *57*, 10-46.

Blom, J. N., Rathje, M., Jakobsen, B. le F., Holsting, A., Hansen, K. R., Svendsen, J. T., Vildhøj, T. W., & Lindø, A. V. (2017). Linguistic Deviations in the Written Academic Register of Danish University Students. *Oslo Studies in Language*, *9*(3), 169-190.

Bohnacker, Ute. (2006). When Swedes Begin to Learn German: From V2 to V2. *Second Language Research*, *22*(4), 443-486.

Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the Lifespan: Aspects of Acquisition, Maturity and Loss* (pp. 73-86). Cambridge: Cambridge University Press. https://doi.org/:10.1017/CBO9780511611780.006

Bolander, M. (1990). On the Acquisition of Word Order Rules in Swedish as a Second Language. In M. A. K. Halliday, J. Gibbons, & H. Nicholas (Eds.), *Learning, Keeping and Using Language: Selected papers from the Eighth World Congress of Applied*

Linguistics, Sydney, 16-21 August 1987 (pp. 287-298). John Benjamins Publishing Company.

Boye, K. (2005). Semantisk analyse og afvigelser mellem sprogbrug og sprogstruktur. *Ny forskning i grammatik, 12*, 27-46.

Boye, K., & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. *Language, 88*(1), 1-44. https://doi.org/10.1353/lan.2012.0020

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, *113*(4), 787-821.

Brandt, S. (1996). Stavefejl i aviserne. *Mål & Mæle, 19*(4), 16-19.

Brautaset, A. (1996). *Inversjon i norsk mellomspråk: en undersøkelse av inversjon i stiler skrevet av innlærere med norsk som andrespråk.* Novus forlag.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31*(1), 25-44. https://doi.org/10.1023/A:1014324220455

Brink, E. T., Elbro, C. & Johannsen, S. (2014). Sproglige fejl og karakterer i danske gymnasiestile. *Mål og Mæle, 36*(2), 5-9.

Cambridge Dictionary (2023). https://dictionary.cambridge.org/dictionary/english/first-language. Last accessed 22.08.2023.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences, 39*, 1-72. https://doi.org/10.1017/S0140525X1500031X

Christensen, M. H., Kristensen, L. B., Vinther, N. M., & Boye, K. (2021). Grammar is background in sentence processing. *Language and Cognition, 13*(1), 128-153.

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*(3), 453-467. https://doi.org/10.1177/0267658316637401

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Dank, M., Deutsch, A., & Bock, K. (2015). Resolving conflicts in natural and grammatical gender agreement: Evidence from eye movements. *Journal of Psycholinguistic Research*, *44*(4), 435-467. https://doi.org/10.1007/s10936-014-9291-9

Dansk Sprognævn (2023). Lovgrundlag. https://dsn.dk/organisation-og-lovgivning/opgaver/lovgrundlag/. Last accessed 25.08.2023.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121. https://doi.org/10.1038/nn1504

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*(2), 200-224. https://doi.org/10.1006/jmla.2000.2768

Diderichsen, P. (1974). *Elementær dansk grammatik* (3rd ed.). København: Gyldendal.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological review*, *112*(4), 777-813. https://doi.org/10.1037/0033-295X.112.4.777

Fabricius-Hansen, C., & Solfjeld, K. (1994). Deutsche und norwegische Sachprosa im Vergleich: Ein Arbeitsbericht. *Arbeitsberichte des Germanistischen Instituts der Universität Oslo, 6*.

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language & Linguistics Compass*, *1*, 71-83.

Freywald, U., Cornips, L., Ganuza, N., Nistov, I., & Opsahl, T. (2015). Beyond verb second: A matter of novel information structural effects? Evidence from Norwegian, Swedish, German and Dutch. In N. Jacomine, & B. Svendsen (Eds.), *Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces* (pp. 73-92). Cambridge: Cambridge University Press.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78-84.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological science*, *28*(6), 703-712.

Glahn, E., Håkansson, G., Hammarberg, B., Holmen, A., Hvenekilde, A., & Lund, K. (2001). Processability in Scandinavian second language acquisition. *Studies in second language acquisition*, *23*(3), 389-416. https://doi.org/10.1017/S0272263101003047

Gosselke Berthelsen, S., & Søby, K. F. (manuscript in preparation). Bønder jeg har spist i går. Prioritering af udtale og grammatik i undervisningen i dansk som andetsprog.

Gujord, A.-K. (2017). The 'Perfect Candidate' for Transfer: A Discussion of L1 Influence in L2 Acquisition of Tense-Aspect Morphology. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 29-63). Bristol: Multilingual Matters.

Hacker, D. J., Plumb, C., Butterfield, E. C., Quathamer, D., & Heineken, E. (1994). Text Revision: Detection and Correction of Errors. *Journal of Educational Psychology, 86*(1), 65-78. https://doi.org/10.1037/0022-0663.86.1.65

Hagen, J. E. (1992). Feilinvertering, overinvertering og underinvertering. *Norsk som andrespråk*, *15*, 27-38.

Hagoort Peter. 2005. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9*(9), 416-423.

Hansen, J. J., Kristensen, L. B., & Søby, K. F. (2019, May 22-24). The effects of homophony on verbal inflection in L1 and L2 Danish [Conference presentation]. Seventh Conference of the Scandinavian Association for Language and Cognition (SALC7), Aarhus, Denmark.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. 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.

Harder, P. (2010). *Meaning in Mind and Society. A Functional Contribution to the Social Turn in Cognitive Linguistics*. Berlin: Mouton de Gruyter.

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society open science*, *5*(8), 180502-180502. https://doi.org/10.1098/rsos.180502

Holmberg, A. (2015). Verb second. In T. Kiss & A. Alexiadou (Eds.), *Syntax – Theory and analysis: An international handbook* (vol. 1, pp. 342-383). Berlin: De Gruyter.

Holmen, A. (1990). *Udviklingslinier i tilegnelsen af dansk som andetsprog*. Danmarks Lærerhøjskole, Institut for dansk sprog og litteratur.

Holmen, A. (1994). Syntactic development in Danish L2. In K. Hyltenstam & A. Viberg (Eds.), *Progression and Regression in Language: Sociocultural, Neuropsychological and Linguistic Perspectives* (pp. 267-288). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627781.012

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Huang, K., & Staub, A. (2021). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition, 216,* 104846-104846. https://doi.org/10.1016/j.cognition.2021.104846

Hyltenstam, K. (1978). Variation in Interlanguage Syntax. *Working papers, 18*, 1-79. Lund: Department of General Linguistics.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 538-588). Oxford: Blackwell Publishing.

Håkansson, G. (1988). Hungry I Am – Breakfast I Want. On the Acquisition of Inverted Word Order in Swedish. *Lund University Department of Linguistics Working papers, 33*, 123-130.

Håkansson, G., Pienemann, M., & Sayehli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research*, *18*(3)*,* 250-273.

Hårstad, S., & Opsahl, T. (2013). *Språk i byen: Utviklingslinjer i urbane språkmiljøer i Norge*. Fagbokforlaget.

Janik, M. O. (2017). Positive and Negative Transfer in the L2 Adjective Inflection of English-, German- and Polish-speaking Learners of L2 Norwegian. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 84-109). Bristol: Multilingual Matters.

Jarvis, S. (2017). Transfer: An overview with an expanded scope. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 12-28). Bristol: Multilingual Matters.

Jensen, T. J., & Christensen, T. K. (2013). Promoting the demoted: The distribution and semantics of "main clause word order" in spoken Danish complement clauses. *Lingua*, *137,* 38-58. https://doi.org/10.1016/j.lingua.2013.08.005

Jervelund, A. Å. (2007). *Sådan staver vi – om ortografi og stavefejl*. København: Dansk Sprognævn, Dansklærerforeningen.

Jervelund, A. Å., Schack, J., Jensen, J. N., & Andersen, M. H. (2012). *Retskrivningsordbogen* (4^{th} ed.). København: Dansk Sprognævn, Alinea.

Johannsen, S. (2012). Fejl og karaktergivning i danske gymnasiestile. *Mål og Mæle, 35*(1), 18-23.

Johansen, H. (2008). Inversjon i norsk innlærerspråk. En undersøkelse av variasjonsmønstre i skrevne tekster. *Norsk som andrespråk*, *24*(2), 50-71.

Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Language and Linguistics Compass*, *2*(4), 647-670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Konieczny, L., Scheepers, C., & Hemforth, B. (1994). Reanalyses vs. internal repairs: Non-monotonic processes in sentence perception. *IIG-Berichte, 8*(94), 2-22.

Kotsinas, U.-B. (2000). *Kontakt, variation och förändring: Studier i Stockholmsspråk*. Almqvist & Wiksell International.

Kristensen, L. B. (2013). *Context, you need: Experimental approaches to information structure processing*. Ph.D. thesis. University of Copenhagen.

Kristensen, L. B. (2017). Grant application for Sapere Aude, Independent Research Fund Denmark. Grant number: 7023-00131B.

Kristensen, L. B., & Boye, K. (2016). Svært ved Grammatikken – del 3. *Logos, 77,* 27-31.

Kristensen, L. B., Ibholt, T. B., & Nielsen, A. P. (2007). Avisernes fejl er en gammel nyhed. *Mål & Mæle, 30*(3), 7-11.

Kristensen, L. B., Schack, J., & Søby, K. F. (in press). Om unge der har skulle bøje modalverber, men ikke har turde, ikke har kunne eller ikke har ville. *Ny Forskning i Grammatik, 30.*

Kristensen, L. B., & Søby, K. F. (2022, June 9-10). Gymnasiegrammatik – i produktion og perception [Conference presentation]. GRAMINO 3: Grammatik i Norden 3, Copenhagen, Denmark.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Langgaard, B. (2011). *Pæredansk*. København: Gyldendal.

Larigauderie, P., Guignouard, C., & Olive, T. (2020). Proofreading by students: implications of executive and non-executive components of working memory in the detection of phonological, orthographical, and grammatical errors. *Reading and Writing*, *33*, 1015-1036.

Lennon, P. (1991a). Error and the Very Advanced Learner. *International Review of Applied Linguistics in Language Teaching, 29*(1), 31-44.

Lennon, P. (1991b). Error: Some Problems of Definition, Identification, and Distinction. *Applied Linguistics, 12*(2), 180-196.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, B. A., Di Persio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 957-971. https://doi.org/10.1037/0278-7393.18.5.957

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283-1315. https://doi.org/10.1017/S0142716414000290

Lüdeling, A., & Hirschmann, H. (2015). Error annotation systems. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 135-158). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.007

Lund, K. (1997). *Lærer alle dansk på samme måde? En længdeundersøgelse af voksnes tilegnelse af dansk som andetsprog*. Herning: Special-Pædagogisk Forlag.

Martin-Lacroux, C., & Lacroux, A. (2016). Do employers forgive applicants' bad spelling in résumés? *Business and Professional Communication Quarterly*, *80*(3), 321-335. https://doi.org/10.1177/2329490616671310

McKoon, G., Ratcliff, R., Ward, G., & Sprout, R. (1993). Syntactic prominence effects on discourse processes. *Journal of Memory and Language*, *32*(5). 593-607.

Messerschmidt, M., Boye, K., Overmark, M., Kristensen, S., & Harder, P. (2018). Sondringen mellem grammatiske og leksikalske præpositioner. *Ny forskning i grammatik*, *25*, 89-106.

Ministry of Higher Education and Science (2022). Upper secondary education. https://ufm.dk/en/education/the-danish-education-system/upper-secondary-education. Last accessed 24.11.2022.

Ministry of Immigration and Integration (MII) (2019). Vejledning til visitation og indplacering på danskuddannelser til voksne udlændinge m.fl. https://www.integrationsviden.dk/filer-til-danskomradet/danskuddannelse/visitation-og-indplacering/vejledning-til-visitation-og-indplacering-paa-danskuddannelser-til-voksne-udlaendinge-m-fl-3.pdf. Last accessed 20.09.2022.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second Language Learning Theories* (3rd ed.). New York: Routledge.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye-movement patterns. *Journal of Psycholinguistic Research*, *27*(5), 515-539. https://doi.org/10.1023/A:1024996828734

Olofsson, J., & Prentice, J. (2020). För tre enorma öl sedan: Befästning av semi-schematiska konstruktioner i L2-svenska. *Språk och stil*, *30*, 91-116. https://doi.org/10.33063/DIVA-427676

Pawlak, M. (2013). *Error correction in the foreign language classroom, reconsidering the issues*. Heidelberg: Springer.

Pienemann, M. (1998). *Language processing and second language development: Processability theory.* Amsterdam: John Benjamins Publishing Company.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427-456. https://doi.org/10.1006/jmla.1999.2653

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1156-1172. https://doi.org/10.1037/xlm0001046

Queen, R., & Boland, J. E. (2015). I think your going to like me: Exploring the role of errors in email messages on assessments of potential housemates. *Linguistics Vanguard*, *1*(1), 283-293. https://doi.org/10.1515/lingvan-2015-0011

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, *12*(1-2), 43-46.

Ragnhildstveit, S. (2017). Gender Assignment and L1 Transfer in Norwegian Second Language Learners' Written Performance. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 110-154). Bristol: Multilingual Matters.

Rankin, T. (2012). The transfer of V2: inversion and negation in German and Dutch learners of English. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *16*(1), 139-158. https://doi.org/10.1177/1367006911405578

Rathje, M., & Kjærgaard, A. (2009). *Sproget i DR. En undersøgelse af problematisk sprogbrug, engelske lån og groft sprog i et udvalg af DR's produktioner*. København: DR.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *The Behavioral and Brain Sciences*, *26*(4), 445-526. https://doi.org/10.1017/s0140525x03000104

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1-21.

Ringbom (1985). The Influence of Swedish on the English of Finnish Learners. In H. Ringbom (Ed.), *Foreign language learning and bilingualism*. Åbo Akademi.

Sayehli, S., Gullberg, M., Newman, A. J., & Andersson, A. (2022). Native Word Order Processing Is Not Uniform: An ERP Study of Verb-Second Word Order. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.668276

Shafto, M. A. (2015). Proofreading in young and older adults: The effect of error category and comprehension difficulty. *International journal of environmental research and public health*, *12*(11), 14445-14460. https://doi.org/10.3390/ijerph121114445

Slotorub, F., & Moreira, N. J. (2011). *Puls 3* (1st edition). Dyssegård: Alfabeta.

Slotorub, F., & Moreira, N. J. (2014). *Puls 2* (3rd edition). Dyssegård: Alfabeta.

Smith, P. T., & Groat, A. (1979). Spelling patterns, letter cancellation and the processing of text. In P. A. Koolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language. Proceedings of the first Conference on Processing of Visible Language, September 5-8, 1977* (pp. 309-324). New York: Plenum.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context Effects in Syntactic Ambiguity Resolution: Discourse and Semantic Influences in Parsing Reduced Relative Clauses. *Canadian Journal of Experimental Psychology*, *47*(2), 276-309. https://doi.org/10.1037/h0078826

SR Research 2023. https://www.sr-research.com/about-eye-tracking/. Last accessed 27.06.2023..

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic bulletin & review*, *11*(5), 882-888.

Szymańska, O. (2017). On how Polish Learners of Norwegian Render Spatial Prepositions in L2: A Corpus-Based Study of *i* and *på*. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 64-83). Bristol: Multilingual Matters.

Søby, K. F., & Kristensen, L. B. (2019). Hjælp! Jeg har mistede min yndlings rød taske. Et studie af grammatikafvigelser. *Ny forskning i grammatik*, *26*, 89-104.

Tenfjord, K., Jarvis, S., & Golden, A. (2017). Introduction. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning* (pp. 1-11). Bristol: Multilingual Matters.

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, *97*(S1), 77-101. https://doi.org/10.1111/j.1540-4781.2012.01422.x

Thewissen, J. (2015). *Accuracy across Proficiency Levels. A Learner Corpus Approach*. Louvain-LaNeuve: Presses Universitaires de Louvain.

Thorborg, L., & Riis, M. J. (2010). *På vej til dansk – trin for trin* (5th edition). Albertslund: Synope.

Torp, A. (2004). Nordiska språk i urtid och forntid – språklikhet och språkskillnad, språkfamiljer och språksläktskap. In I. S. Sletten (Ed.), *Nordens språk med rötter och fötter* (pp. 19-74). Nordiska ministerrådet.

Vikør, L. S. (2015). *Norwegian: Bokmål vs. Nynorsk.* The Norwegian Language Council. http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk. Last accessed 27.06.2023.

Vinther, N. M., Boye, K., & Kristensen, L. B. (2015). Grammatikken i baggrunden – opmærksomhed under læsning. *NyS, 47*, 99-139. https://doi.org/10.7146/nys.v47i47.19877

Westergaard, M. (2003). Unlearning V2: Transfer, markedness and the importance of input cues in the acquisition of word order in English by Norwegian children. In S. Foster-Cohen & S. Pekarek-Doehler (Eds.), *EUROSLA yearbook* (vol. 3, pp. 77-101). Amsterdam: John Benjamins Publishing Company.

Westman, M. (1974). *Bruksprosa*. Lund: Gleerups.

Yeaton, J. (2019). Visual processing of verb-second (V2) word order in second language acquisition: ERP Evidence from French-Swedish successive bilinguals [Master thesis, University of Lund]. https://jeremyyeaton.github.io/papers/Yeaton-2019-MasterThesis-V2_word_order.pdf

10 Appendix

Appendix A: Supplementary materials, Article 2

Supplementary Material

1 Supplementary Figures and Tables

1.1 Supplementary Tables – model results

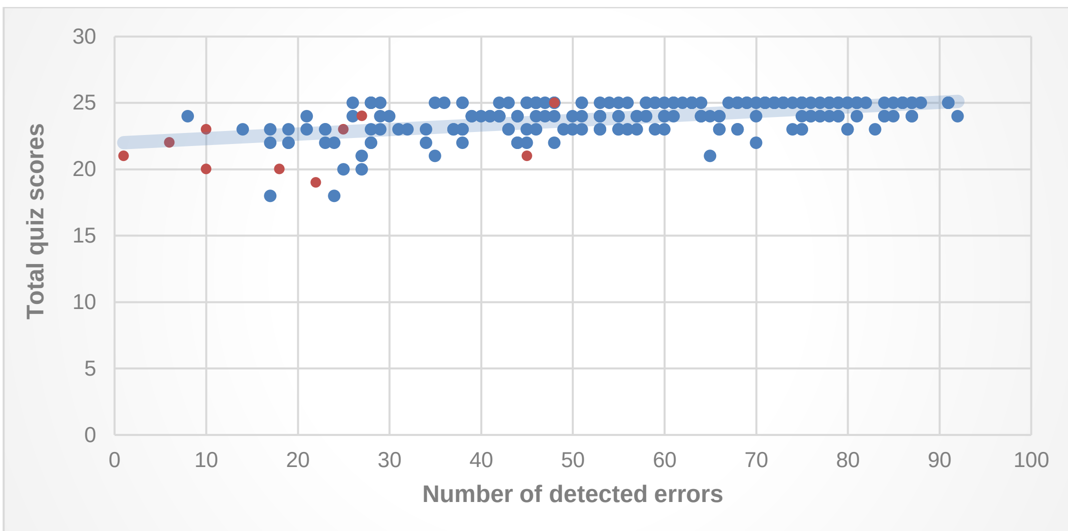
Table 12. Model (4) estimates for orthographic errors. Dependent variable: Detection (1 = error detected, 0 = error not detected)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	1.0563	1.0277		
Item (Intercept)	0.7147	0.8454		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-5.32930	0.67971	-7.841	4.49e-15 ***
TYPE: Reduction of syllable	1.39987	0.54755	2.557	0.0106 *
TYPE: Compounds written in two	0.81022	0.54737	1.480	0.1388
TYPE: Missing silent letter	0.81582	0.54792	1.489	0.1365
Spelling quiz score	0.50287	0.7715	6.518	7.11e-11 ***

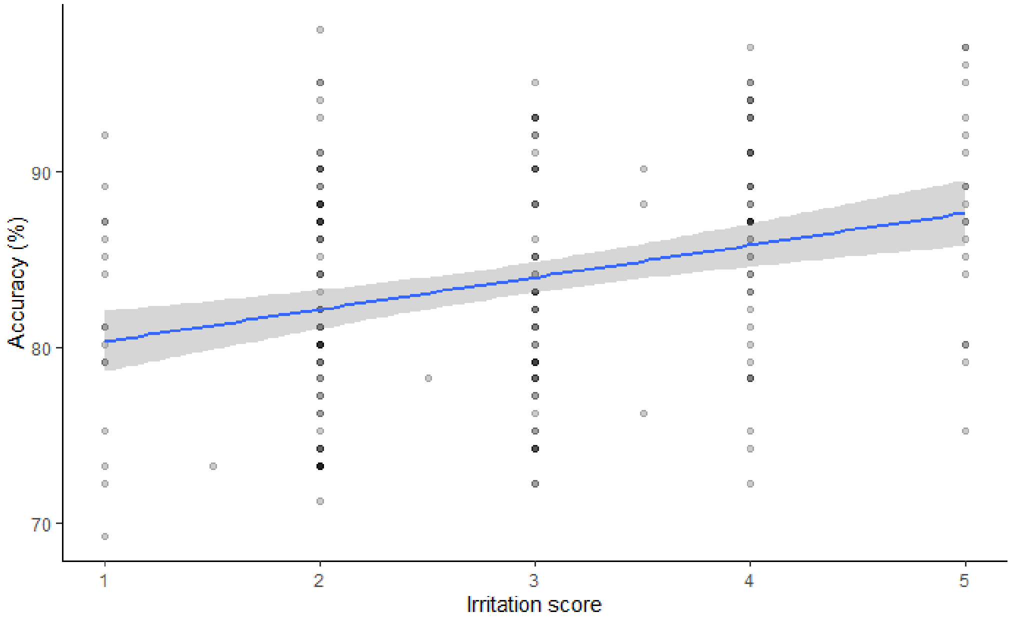
Table 13. Model (5) estimates for all error types (collapsed). Dependent variable: Accuracy in percentage

Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	78.5258	1.2311	63.787	< 2e-16 ***
Irritation	1.8184	0.3982	4.566	8.48e-06 ***

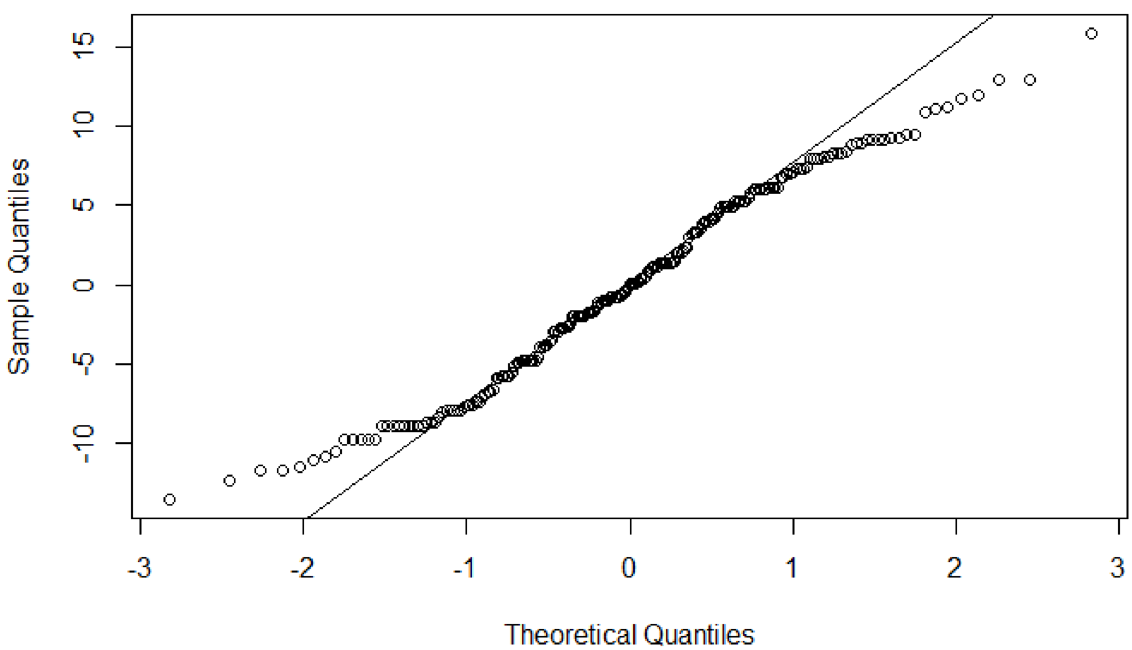
1.2 Supplementary Figures



Supplementary Figure 1. Total quiz scores (N of correct answers) and number of detected errors per participant (10 outliers with more false alarms than hits marked with orange)



Supplementary Figure 2. Effect of reported irritation with language errors on accuracy in error detection (%), model 5.



Supplementary Figure 3. Normal Q-Q plot of model 5.

2 Stimuli

We tried to use controls already present in the texts, or words that could easily be altered from the texts, to avoid making the texts longer and the content even stranger. This means that controls are not perfectly matched to target items. The aim was first to make the controls as similar to targets as possible, with an even distribution in the texts. If more possible control words were available, the ones with a frequency (KorpusDK) closest to the target items were chosen.

2.1 V3 errors

No.	Adverbial (a, short; b, long)	*Subject	Finite verb	Context
1a	*Heldigvis*	*han*	*har*	*lavet en aftale:*
	Luckily	he	has	made a deal:
1b	*Til Peters kæmpestore held*			
	To Peters great luck			
2a	(og) *kl. 14*	*han*	*ankommer*	*til Berlin –*
	(and) at 2	he	arrives	in Berlin –
2b	(og) *først ud på eftermiddagen*			
	(and) first in the afternoon			
3a	*Om aftenen*	*han*	*finder*	*sin computer frem.*
	In the evening	he	takes	his computer out
3b	*Omkring kl. 21 om aftenen*			
	Around 9 p.m. in the evening			
4a	*I spillet*	*man*	*er*	*en ridder, der skal smadre et væmmeligt monster med et gevaldigt hoved.*
	In the game	you	are	a knight, who has to bash a nasty monster with a huge head.
4b	*I det relativt syrede spil*			
	In the relatively weird game			
5a	*(og) kort efter*	*Peter*	*har*	*det skidt.*
	(and) shortly after	Peter	is	feeling bad.
5b	*(og) allerede efter ganske kort tid*			
	(and) already after a fairly short period of time			
6a	*I parken*	*han*	*falder*	*til ro igen.*
	In the park	he	calms	down again.
6b	*I den smukke, grønne park*			
	In the beautiful, green park			
7a	*Lidt derfra*	*dørmanden*	*står*	*og stirrer uhæmmet.*
	A bit from there	the doorman	stands	and stares without restraint.
7b	*Lidt derfra tæt ved udgangen*			
	A bit from there, close to the exit			
8a	*Pludselig*	*de*	*får*	*øjenkontakt,*
	Suddenly	they	make	eye contact
8b	*Pludselig efter lang tids stirren*			

	Suddenly after a long time of staring			
9a	*Efter festen*	*de*	*skal*	*på en overdådig bryllupsrejse på et luksuriøst hotel på Maldiverne –*
	After the party	they	are going	on a lavish honeymoon at a luxurious hotel in the Maldives –
9b	*Efter den ekstravagante bryllupsfest*			
	After the extravagant wedding			
10a	*Ved kiosken*	*børnehaven*	*ligger.*	
	By the kiosk	the kindergarten	lies.	
10b	*Ved siden af den lille kiosk*			
	Next to the small kiosk			
11a	*Derfor*	*Lars*	*har*	*travlt på arbejdet,*
	Therefore	Lars	is	busy at work,
11b	*Derfor og af flere andre grunde*			
	Therefore and for multiple other reasons			
12a	*Desværre*	*han*	*slipper*	*ikke så let:*
	Unfortunately	he	does	not get off that easily:
12b	*Desværre for den travle mand*			
	Unfortunately for the busy man			
13a	*På studiet*	*underviserne*	*har*	*fokus på, at de studerende kan beherske mange forskellige genrer:*
	At the university	the teachers	focus	on the students being able to master many different genres:
13b	*På det tidskrævende universitetsstudie*			
	During the time-consuming university study			
14a	*I dag*	*det*	*driller*	*også.*
	Today	it	is tricky	too.
14b	*Denne ellers hyggelige lørdag formiddag*			
	This otherwise pleasant Saturday morning			
15a	*Derefter*	*det*	*dur*	*heldigvis igen!*
	Afterwards	it	works	again, fortunately!
15b	*Efter det fantastisk smarte trick*			
	After the amazingly clever trick			
16a	*Indenfor*	*ekspedienten*	*hilser*	*venligt.*
	Inside	the shop assistant	says hello	with kindness.
16b	*Inde i det hyggelige konditori*			
	In the cosy patisserie			

2.2 Verb errors

No.	HOMOPHONE VERB PAIRS			HETEROPHONE VERB PAIRS		
	Lexeme	**Characters (N)**	**Frequency (infinitive + present tense)**	**Lexeme**	**Characters (N)**	**Frequency (infinitive + present tense)**
1	*køre* ‘drive’	4	14234	*rejse* ‘travel’	5	8087
2	*motionere* ‘exercise’	9	120	*styrketræne* ‘weight-lift’	11	26
3	*fodre* ‘feed’	5	336	*passe* ‘look after’	5	8960
4	*føre* ‘guide’	4	10703	*lede* ‘lead’	4	4306
5	*sludre* ‘talk’	6	130	*chatte* ‘chat’	6	116
6	*gennemføre* ‘complete’	10	3542	*afslutte* ‘finish’	8	769
7	*vurdere* ‘evaluate’	7	5438	*overveje* ‘consider’	8	3363
8	*præstere* ‘achieve’	8	705	*formå* ‘be able to’	5	894
9	*ændre* ‘change’	5	6526	*skifte* ‘switch’	6	3604
10	*forklare* ‘explain’	8	7993	*berette* ‘recount’	7	1069
11	*angre* ‘regret’	5	47	*fortryde* ‘regret’	8	639
12	*manipulere* ‘manipulate’	10	301	*fordreje* ‘twist’	8	67
13	*medføre* ‘entail’	7	3531	*forårsage* ‘cause’	9	497
14	*score* ‘score’	5	1177	*drikke* ‘drink’	6	3896
15	*observere* ‘observe’	9	253	*iagttage* ‘watch’	8	520
16	*indikere* ‘indicate’	8	179	*antyde* ‘hint’	6	1145
17	*undre* ‘surprise’	5	2483	*glæde* ‘please’	5	5159
18	*invitere* ‘invite’	8	1241	*indbyde* ‘invite’	7	307
19	*forsvare* ‘defend’	8	2143	*beskytte* ‘protect’	8	2463
20	*arrestere* ‘arrest’	9	115	*anholde* ‘arrest’	7	213
21	*aktivere* ‘activate’	8	292	*tænde* ‘switch on’	5	1446
22	*aflevere* ‘deliever’	8	1355	*hente* ‘fetch’	5	5374
23	*skure* ‘scrub’	5	37	*skrubbe* ‘scrub’	7	84
24	*lære* ‘learn’	4	9887	*skrive* ‘write’	6	17241
25	*sondre* ‘distinguish’	6	56	*skelne* ‘distinguish’	6	1264
26	*præsentere* ‘present’	10	2196	*fremstille* ‘depict’	10	2073
27	*argumentere* ‘argue’	11	836	*kæmpe* ‘fight’	5	2662
28	*fungere* ‘function’	7	6155	*virke* ‘work’	5	9749
29	*formatere* ‘format’	9	5	*gendanne* ‘restore’	8	33
30	*arrangere* ‘organize’	9	1136	*forberede* ‘prepare’	9	1704
31	*brillere* ‘shine’	8	95	*optræde* ‘perform’	7	3720
32	*kreere* ‘create’	6	82	*bage* ‘bake’	4	592

2.3 NP errors

All target items included regularly inflected adjectives, where inflection is obligatory and not optional (e.g. adjectives ending with *-vis*). Furthermore, we chose adjectives without changes in pronunciation of the stem when the neuter suffix was added, e.g. *sød* [ˈsøˀð] vs. *sødt* [ˈsød̥] ‘sweet’. These restrictions meant that almost half of the target adjectives ended with *-ig*, which might be striking to the participants. Neuter nouns beginning with *d-* or *t-* were avoided, because it might affect participants’ detection of a missing *-t* [d̥] on the adjective that the sound is represented elsewhere, e.g. *et *dejlig tehus* for *et dejligt tehus* ‘a lovely tea house’. The criteria were less strict for the 32 control items (16 uter NPs; 16 neuter NPs); we used inflected adjectives which had not already been used as targets.

No.	**Adjective**	**Neuter noun**		**Uter noun**	
		Lexeme	**Characters (N)**	**Lexeme**	**Characters (N)**
1	*dejlig* ‘lovely’	*kæledyr* ‘pet’	7	*undulat* ‘budgie’	7
2	*enorm* ‘huge’	*kinderæg* ‘Kinder Egg’	8	*slikpose* ‘bag of sweets’	8
3	*fin* ‘nice’	*center* ‘mall’	6	*butik* ‘store’	5
4	*brun* ‘brown’	*bælte* ‘bælte’	5	*rygsæk* ‘rucksack’	6
5	*lækker* ‘smashing’	*hotelværelse* ‘hotel room’	12	*lejlighed* ‘apartment’	9
6	*mærkelig* ‘weird’	*univers* ‘universe’	7	*verden* ‘world’	6
7	*solid* ‘solid’	*morgenmåltid* ‘morning meal’	12	*morgenmad* ‘breakfast’	9
8	*utrolig* ‘unbelievable’	*held* ‘luck’	4	*chance* ‘opportunity’	6
9	*hård* ‘tough’	*løb* ‘run’	3	*dag* ‘day’	3
10	*stærk* ‘strong’	*bolsje* ‘hard candy’	6	*lakrids* ‘liquorice’	7
11	*lokal* ‘local’	*sted* ‘place’	4	*bar* ‘bar’	3
12	*iskold* ‘ice-cold’	*shot* ‘shot’	4	*drink* ‘drink’	5
13	*ren* ‘clean’	*bord* ‘table’	4	*stol* ‘chair’	4
14	*hvid* ‘white’	*stearinlys* ‘candle’	10	*olielampe* ‘oil lamp’	9
15	*venlig* ‘friendly’	*smil* ‘smile’	4	*gestus* ‘gesture’	6
16	*uforglemmelig* ‘unforgettable’	*arrangement* ‘function’	11	*begivenhed* ‘event’	10
17	*vanvittig* ‘insane’	*antal* ‘number’	5	*mængde* ‘amount’	6
18	*rummelig* ‘spacious’	*hus* ‘house’	3	*villa* ‘villa’	5
19	*grøn* ‘green’	*areal* ‘area’	5	*park* ‘park’	4
20	*skøn* ‘great’	*barn* ‘child’	4	*dreng* ‘boy’	5
21	*frygtelig* ‘terrible’	*kaos* ‘chaos’	4	*uro* ‘stir’	3
22	*dyr* ‘expensive’	*system* ‘system’	6	*løsning* ‘solution’	7
23	*vidunderlig* ‘wonderful’	*arbejde* ‘job’	7	*kæreste* ‘girlfriend’	7
24	*lykkelig* ‘happy’	*ægtepar* ‘married couple’	7	*familie* ‘family’	7

25	*lang* 'long'	*indlæg* 'article'	6	*artikel* 'article'	7
26	*ubehagelig* 'uncomfortable'	*forløb* 'process'	6	*affære* 'affair'	6
27	*besværlig* 'difficult'	*program* 'program'	7	*proces* 'process'	6
28	*elendig* 'lousy'	*product* 'product'	7	*opfindelse* 'invention'	10
29	*fabelagtig* 'fabulous'	*uddrag* 'extract'	6	*novelle* 'short story'	7
30	*smuk* 'beautiful'	*slør* 'veil'	4	*kjole* 'dress'	5
31	*hel* 'whole'	*vindue* 'window'	6	*sektion* 'section'	7
32	*saftig* 'juicy'	*rosinbrød* 'raisin bread'	9	*frugtkage* 'fruit cake'	9

2.4 Orthographic errors

Error	**Target form**	**Translation**
	MISSING DOUBLE CONSONANTS	
startskudet	*startskuddet*	'the starting signal'
parets	*parrets*	'the couple's'
kroniker	*kronikker*	'feature articles'
butiken	*butikken*	'the shop'
girafen	*giraffen*	'the giraffe'
	SPLIT COMPOUNDS	
æble cider	*æblecider*	'apple cider'
by vandring	*byvandring*	'city walk'
drømme bryllup	*drømmebryllyp*	'dream wedding'
chokolade smag	*chokoladesmag*	'chocolate taste'
prøve smagning	*prøvesmagning*	'sample tasting'
	MISSING SILENT LETTERS	
egenlig	*egentlig*	'actually'
sjælent	*sjældent*	'rarely'
(i det) minste	*(i det) mindste*	'at least'
siste	*sidste*	'last'
selsikre	*selvsikre*	'confident'
	REDUCED SYLLABLES	
temlig	*temmelig*	'rather'
endlig	*endelig*	'finally'
spør	*spørger*	'ask' (present tense)
virklig	*virkelig*	'really'
ærgligt	*ærgerligt*	'annoyingly'

3 Test materials

3.1 Reading task and comprehension questions (in Danish)

Errors are marked with colors: V3 errors, verb errors, NP errors and orthographic errors.

Peter i Berlin

Peter er 33 år og en ivrig maratonløber. I dag er en vigtig dag. Han køre til Tyskland, nærmere bestemt Berlin. Her skal han deltage i et populært løb.

Peter er typen, der sætter ambitiøse mål for alt. Han skal styrketræner meget og gerne hver dag. Han risikerer gerne liv og helbred for en rigtig udfordring, og han vil sejre for enhver pris. Imidlertid har han også en blød side: Han er en rigtig dyreven! Peter har en dejligt kæledyr, som ærgligt nok ikke kan komme med til Berlin.

Heldigvis han har lavet en aftale: Naboens børn passe Piphans, som fuglen hedder. Som tak for nabobørnenes hjælp vil han købe en enormt slikpose med hjem.

Peter tager afsted tidligt om morgenen, og først ud på eftermiddagen han ankommer til Berlin – dagen inden løbet. Stemningen i byen er helt i top.

Peter har tid til lidt sightseeing, så han har meldt sig til en by vandring. Guiden vil fører turisterne gennem det 19.000 m^2 store mindesmærke for Europas myrdede jøder, som består af 2.711 betonpiller. De når også at gå forbi et fin center, hvor Peter bl.a. køber et brun rygsæk. Peter har booket en lækkert lejlighed i centrum. Omkring kl. 21 om aftenen han finder sin computer frem. Han vil chatter med en gammel ven over Skype. Vennen fortæller, at der er kommet et sjovt computerspil fra deres yndlingsspiludviklere, og Peter downloader det straks! Spillet foregår i en mærkeligt univers med prinsesser og zombier. I spillet man er en ridder, der skal smadre et væmmeligt monster med et gevaldigt hoved. Peter gennemføre første bane. Det bliver desværre sent. Peter spiser også et stort marcipanbrød og slubrer et par dåser æble cider og øl i sig. Det er ikke så godt.

Næste morgen er Peter ikke helt frisk. Han overveje situationen: Kan han levere et godt resultat i dag? Eller er det egenlig bedre at droppe at løbe? Han bliver så skuffet, hvis han ikke kan præsterer at få en god tid. Efter et solid morgenmåltid beslutter han sig for at løbe.

Startskudet lyder, og allerede efter ganske kort tid Peter har det skidt. Dagsformen er ikke god, og han kan ikke holde et højt tempo, som han plejer. Han må skifter taktik. Et sted i starten af ruten kan man skyde genvej gennem en smal sidegade. Sikke et utrolig chance!

Peter kommer i mål med en bedre tid, end han plejer. Men nogen har set ham snyde. Peter er opdaget, og han forklare det hele, da de løbsansvarlige spør. Han burde angrer det, han har gjort. Sikke en åndssvag bommert.

Han håber godt nok ikke, at det kommer frem i medierne – de kan fordrejer alt. Det ville være temlig pinligt, hvis kammeraterne i løbeklubben derhjemme fik noget at vide! Peter går hen og sætter sig i en park for at samle tankerne. I parken han falder til ro igen. I det minste har han lært noget vigtigt: Uærlighed medføre mange problemer.

Efter en hårdt dag vil Peter tage i byen. Først spiser han lige en stærkt bolsje for at tage noget af den dårlige ånde. Han orker ikke proppede turistfælder, så han finder et lokal bar. Ifølge Peter og hans venner er det ikke en ordentlig bytur, hvis man ikke kan drikker løs.

Han går op i baren og bestiller en iskoldt drink. Bardisken er lidt for klistret for Peters smag, så han forsøger at finde en rent bord. Mens han kigger, får han øje på noget interessant: en ældre, men ret flot, tysk kvinde. Peter sætter sig. Han iagttage hende fra sin plads og tænker, at hun faktisk ser rigtig flot ud, som hun sidder der i lyset fra et hvid stearinlys. Peter opdager, han ikke er den eneste, der kigger på hende. Lidt derfra tæt ved udgangen dørmanden står og stirrer uhæmmet.

Peter må rykke snart. Han vil indikerer, at han er interesseret. Den selsikre Peter tænker, at hun undre sig over, at en ung mand er interesseret.

Pludselig de får øjenkontakt, og kvinden sender Peter en venligt gestus. Nu tør Peter godt gå derhen! Han tager en siste tår og går hen mod hende …

Lars og Lone i Hvidovre

Der sker store ting for Lars og Lone. Efter at have sparet op i årevis er tiden endlig kommet: Det er i år, de indbyde hele familien til bryllup. Det bliver et uforglemmelig begivenhed, for Lone kan organisere de vildeste fester. Lars og Lone elsker fransk champagne på elegante flasker, og de har allerede indkøbt en vanvittigt antal. Efter festen de skal på en overdådig bryllupsrejse på et luksuriøst hotel på Maldiverne – dét ser de frem til!

Lone og Lars bor sammen med sønnen Storm i et rummelig hus i Hvidovre. Kvarteret er hyggeligt, med mange, gode faciliteter – fx en herlig legeplads, et grøn park og en kiosk.

Ved siden af den lille kiosk børnehaven ligger. Her går Storm. Han er altid glad, og pædagogerne i børnehaven synes, han er en skønt dreng.

Langt størstedelen af tiden plaprer Storm bare løs om girafen Ingolf – et gammelt tøjdyr, han fandt i glemmekassen. Lars er ansat i PET, som lige nu skal forsvarer staten mod et stort cyberangreb, der har skabt et frygtelig kaos i hele samfundet. Derfor Lars har travlt på arbejdet, men de er heldigvis langt i sagen.

Politiet, som lige nu skal patruljere i civil i et skummelt område, regner med, at de kan anholder den første mistænkte, inden ugen er omme. Lars går meget op i sikkerhed og har installeret et nyt alarmsystem i parets hus. Han tænde alarmen, hver gang de forlader huset – også hvis de bare går en hurtig tur i villakvarteret. Det synes Lone er åndssvagt.

Hun synes også, det er åndssvagt at bruge så mange penge på alarmer – det var en dyrt system. De penge kunne have været brugt på det kommende drømme bryllup! Normalt styrer Lone ellers familiens økonomi med hård hånd, og hun vil notere alle større udgifter i et særligt regneark. Lars vil hellere sikre huset end at bruge penge på et tåbeligt champagnespringvand til brylluppet.

Lars synes, han har et vidunderlig kæreste, selvom Lone er lidt sur over, at han arbejder så meget. Hun synes, Lars skal prioritere familien højere. Hun synes heller ikke, han hjælper nok til derhjemme, så Lars lover, at han aflevere Storm på mandag. Desværre for den travle mand han slipper ikke så let: Lone svarer, at hun først bliver tilfreds, hvis Lars også vil skurer gulvet, inden Lones mor kommer på besøg om aftenen. Ellers er de nu en lykkeligt familie det meste af tiden.

Lone har også travlt med mange ting. Hun læser journalistik på RUC. Hun skal skriver en masse i løbet af et semester. På det tidskrævende universitetsstudie underviserne har fokus på, at de studerende kan beherske mange forskellige genrer: Man sondre mellem mange forskellige typer artikler. Alt fra baggrundsartikler, ledere, features, kroniker, reportager og klummer til interviews.

Lone synes ikke, det er nemt. Hun arbejder på en langt indlæg, som gerne skal udkomme i et nyt magasin. Artiklen handler om Lones nabo, Monika, som påstår, hun har været udsat for chikane på sin arbejdsplads og er blevet fyret uden grund. Lone vil præsenterer Monika som en troværdig kilde. Det baserer Lone bl.a. på, at man under retssagen ser, hvor godt Monika kæmpe for sin sag. Lone kan også mærke på Monika, at hun virklig synes, det har været et ubehagelig affære.

I de ellers tiltrængte weekender sidder Lone tit og producerer eller omstrukturerer sine tekster, men der er ofte problemer med Word. Det er et besværlig program. Lone synes, det er lidt for sjælent, at programmet fungere ordentligt. I dag det driller også.

Hun skal gendanner dokumentet. Efter det fantastisk smarte trick det dur heldigvis igen! Lone (og faktisk også Lars) synes i dén grad, at Word er en elendigt produkt.

Lone er også med i en aktiv gruppe på studiet, hvor de forberede en hyggelig aften med oplæsning af skønlitteratur. Lone kan allerede se for sig, hvordan hun vil brillerer med et af sine egne værker. Hun synes selv, hun har skrevet et fabelagtig uddrag. Inden Lone er færdig med sine dagdrømmerier, afbryder Lars. Han skal altid forstyrre.

De skal gå nu, hvis de skal nå til prøve smagning af bryllupskager hos byens bedste konditor, som bage nogle helt ekstravagante kager, og som har vundet mange priser. Og Lars skal jo først lige nå at slå alarmen til!

På vej hen til konditoriet ser Lone et smuk kjole i en eksklusiv brudebutik. Lars er glad for, at butiken er lukket i dag, så de kan komme hurtigt videre. Allerede på meget lang afstand kan de se et hel vindue med de yndigste bryllupskager. Indenfor ekspedienten hilser venligt. Hun serverer fire slags kage for dem. Lone synes, det er et vigtigt valg, de skal træffe. Lars vil faktisk hellere hjem og ligge på sofaen, så han stemmer bare på chokolade smag, og Lone er heldigvis enig.

I konditoriet er der mange lækre fristelser, og de køber også en saftigt frugtkage til at tage med hjem. Storm kan næsten ikke gå efter at have spist så meget kage!

Spørgsmål til teksterne

Peter i Berlin

Sæt kryds ved det rigtige svar:

1	*Skal Peter til Berlin for at løbe maraton?*	☐ *ja*	☐ *nej*
2	*Bliver Peters snyderi opdaget?*	☐ *ja*	☐ *nej*
3	*Tager Peter i Berlin Zoo?*	☐ *ja*	☐ *nej*

Lars og Lone i Hvidovre

Sæt kryds ved det rigtige svar:

1	*Har Lars og Lone en datter?*	☐ *ja*	☐ *nej*
2	*Arbejder Lars for PET?*	☐ *ja*	☐ *nej*
3	*Læser Lone til tandlæge?*	☐ *ja*	☐ *nej*

3.2 Questionnaire (in Danish)

Spørgeskema

A. Generel information

A1 Hvilket køn har du? Sæt kryds: ☐ mand ☐ kvinde ☐ andet

A2 Hvor gammel er du? Sæt kryds: ☐ 17 år ☐ 18 år ☐ 19 år ☐ andet: _____ år

A3 Hvilken studieretning har du? Sæt kryds:

STX: ☐ naturvidenskab ☐ samfundsvidenskab ☐ sprog ☐ kunst

HTX: ☐ anvendt naturvidenskab ☐ teknologi ☐ kommunikationsteknik

HHX: ☐ økonomi og marked ☐ økonomi og sprog ☐ sprog

B. Sproglig baggrund

A1 Er du ordblind? ☐ nej ☐ ja ☐ måske (men jeg er ikke blevet testet)

A2 Hvilke sprog taler du, hvornår begyndte du ca. at lære dem, og hvor godt taler du dem? Husk at nævne, hvornår du begyndte at lære dansk – var det fx som barn eller senere i livet?

Hvilke sprog taler du?	**Hvor gammel var du ca., da du startede med at lære sproget?** *Fx ”0 år”*	**Hvor godt taler du sproget?** *Fx ”flydende”, ”avanceret niveau”, ”mellem”, ”begynder”*	**Taler du normalt sproget med din mor eller far derhjemme?** *”Ja”/”nej”*
Dansk			
Engelsk			

A3 Har du tidligere boet i andre kommuner (eller lande), end hvor du bor nu?

☐ nej ☐ ja

Hvis ja: Hvor? Hvor længe boede du der, og hvor gammel var du ca., mens du boede der? Skriv svarene i tabellen:

Kommune eller land	**Hvor længe boede du der?** (Hvor mange måneder/år?)	**Hvor gammel var du ca., mens du boede der?**

C. Holdning til sprogfejl

C1 Hvor irriteret bliver du, hvis der er sprogfejl (dvs. stavefejl og grammatikfejl) i en tekst, du læser? Markér dit svar ved at sætte en ring om én af nedenstående bokse.

slet ikke irriteret	*en smule irriteret*	*irriteret*	*meget irriteret*	*ekstremt irriteret*

C2 Her kan du uddybe dit svar. Kommer det fx an på, hvilken type tekst der er tale om? Er der fx nogle særlige sprogfejl, du især bliver irriteret over? Eller er du helt ligeglad med, om der er sprogfejl i en tekst?

____
____
____
____
____

3.3 Questionnaire (in English)

Questionnaire

A. General information

A1 What is your gender? Check the box: ☐ man ☐ woman ☐ other

A2 How old are you? Check the box: ☐ 17 y/o ☐ 18 y/o ☐ 19 y/o ☐ other: _____ y/o

A3 What is your study program? Check the box:

STX: ☐ natural science ☐ social science ☐ language ☐ art

HTX: ☐ applied natural sciences ☐ technology ☐ communication and IT

HHX: ☐ economy and market ☐ economy and language ☐ language

B. Linguistic background

A1 Are you dyslexic? ☐ no ☐ yes ☐ maybe (but I haven't been tested)

A2 Which languages do you speak, approximately when did you start to learn them, and how well do you speak them? Remember to mention when you started learning Danish – e.g. was it as a child or later in life?

Which languages do you speak?	**How old were you approximately, when you started learning the language?** *E.g. "0 y/o"*	**How well do you speak the language?** *E.g."fluently", "advanced level", "intermediate", "beginner"*	**Do you normally speak the language with your mom or dad at home?** *"Yes"/"no"*
Danish			
English			

A3 Have you previously lived in other municipalities (or countries) than where you currently reside?

☐ no ☐ yes

If yes: Where? How long did you live there, and how old were you approximately, while you lived there? Write your answers in the table:

Municipality or country	**How long did you live there?** (How many months/years?)	**How old were you approximately, while you lived there?**

C. Attitude to language errors

C1 How annoyed are you, if there are language errors (i.e. spelling mistakes and grammar mistakes) in a text you are reading? Mark your answer by circling one of the boxes below.

not at all annoyed	*a bit annoyed*	*annoyed*	*very annoyed*	*extremely annoyed*

C2 Here, you can elaborate on your answer. Does it depend on what type of text it is? Are there certain types of language mistakes that you are particularly annoyed by? Or are you completely indifferent to the presence of language errors in a text?

____

____

____

____

____

3.4 Grammar quiz (in Danish)

Grammatikquiz

FORMÅL: Nogle af spørgsmålene i denne quiz kan måske virke nemme: Vi skal bare tjekke, om du kender helt almindelig dansk grammatik. Andre spørgsmål er måske lidt sværere. Det er vigtigt, at du ikke slår svaret op i en ordbog eller på nettet. Vi er nemlig interesserede i, hvad du gør, når du ikke har mulighed for at slå noget op.

1 Ordstilling

Her skal du indsætte "*han ankommer til Berlin*" på linjerne, sådan at ordene kommer til at stå i rigtig rækkefølge.

1.1 *Kl. 14* ____

1.2 *Og* ____

1.3 *Hvis alt går godt,* ____

2 Adjektiver og artikler

Indsæt den korrekte form af tillægsordet (adjektivet) *lække*r på nedenstående linjer

2.1 *en* __________________ *kage*

2.2 *et* __________________ *brød*

Hedder det *en* eller *et*? Sæt kryds:

2.3 ☐ *en* ☐ *et* *shot*

2.4 ☐ *en* ☐ *et* *undulat*

2.5 ☐ *en* ☐ *et* *bolsje*

2.6 ☐ *en* ☐ *et* *lakrids*

3 Verber

Sæt kryds ved den korrekte form af verbet:

3.1 *Nu* ______ *han hurtigt.* ☐ *køre* ☐ *kører*

3.2 *Han vil _______ hurtigt.* ☐ *køre* ☐ *kører*

3.3 *Hun skal _______ i morgen.* ☐ *rejse* ☐ *rejser*

3.4 *Hun _______ senere i dag, hvis piloterne ikke strejker.* ☐ *rejse* ☐ *rejser*

3.5 *Han _______ mange nyttige ting.* ☐ *lære* ☐ *lærer*

3.6 *Han skal _______ at danse brudevals inden brylluppet.* ☐ *lære* ☐ *lærer*

3.7 *De skal _______ ekstreme mængder kage i weekenden.* ☐ *spise* ☐ *spiser*

3.8 *Om lidt _______ de en sandwich.* ☐ *spise* ☐ *spiser*

4 Stavning

Er disse ord stavet korrekt? Hvis ikke, så skriv den korrekte form:

4.1 *butiken* ☐ korrekt ☐ Nej, det staves: ________________________

4.2 *by vandring* ☐ korrekt ☐ Nej, det staves: ________________________

4.3 *virkelig* ☐ korrekt ☐ Nej, det staves: ________________________

4.4 *æblecider* ☐ korrekt ☐ Nej, det staves: ________________________

4.5 *sjælent* ☐ korrekt ☐ Nej, det staves: ________________________

4.6 *temlig* ☐ korrekt ☐ Nej, det staves: ________________________

4.7 *startskuddet* ☐ korrekt ☐ Nej, det staves: ________________________

4.8 *egentlig* ☐ korrekt ☐ Nej, det staves: ________________________

Appendix B: Supplementary materials, Article 3

SUPPLEMENTARY MATERIALS

A. FILLER SENTENCES AND COMPREHENSION QUESTIONS

Below is a list of all filler items. Half of the fillers contain various types of morphological anomalies (*underlined).

1. Nede ved vannet ser Morten *<u>et</u> elg. ‘Down by the water, Morten sees an elk.’ (*The article is neuter gender instead of masculine, *en*).
2. På skolen syr jenta en jakke til hunden. ‘At school, the girl makes a jacket for the dog.’
3. I Trondheim kjører Linda på sparkesykkel hver helg. ‘In Trondheim, Linda rides a kick scooter every weekend.’
4. I skogen løper Karin ofte med venninner. ‘In the forest, Karin often jog with friends.’
5. I *<u>en sjøen</u> fanger Øystein masse laks. ‘In a/the (?) lake, Øystein catches lots of salmon.’ (*Both the indefinite article *en* and the suffix *–en* denoting definite form is used simultaneously).
6. Mandag den 8. november åpner den nye *<u>butikk</u> i sentrum. ‘Monday the 8th of November, the new store opens in the city centre.’ (*The suffix *-en* denoting definite form is missing).
7. Etter skitur bestiller skitreneren takeaway til kveldsmat. ‘After the ski tour, the ski instructor orders take away for dinner.
8. Stort sett *<u>hvert</u> helg drikker Steinar cider på et utested. ‘Almost every weekend, Steinar drinks cider in a bar.’ (The determiner has a suffix *-t* denoting neuter gender instead of masculine, *hver*).

9. På onsdager kjøper mannen kaker til hele familien *<u>sitt</u>. 'On Wednesdays, the man buys cakes for his whole family.' (*The determiner is neuter gender, and not masculine: *sin*).
10. Neste uke feirer jenta bursdag hos en venninne. 'Next week, the girl celebrates her birthday at a friend's house.'
11. Hvis det regner, sykler ikke Bjørn til universitetet. 'If it rains, Bjørn does not bike to the university.'
12. Når klassen skal på tur, planlegger læreren turen godt på forhånd. 'When the class is going on excursion, the teacher plans the trip well in advance.'
13. Selv om pappa mosjonerer hver *<u>dager</u>, går han ikke ned i vekt. 'Even though dad exercises every days, he does not lose weight.' (*The plural form of the noun is used instead of singular *dag*).
14. Podcast hører Espen alltid på om morgenen. 'Podcasts, Espen always listens to in the morning.'
15. Sjokoladekake fra bakeriet elsker mannen. 'Chocolate cake from the bakery, the man loves.'
16. Maraton løper Einar *<u>hver</u> år i juni. 'Marathon, Einer runs every year in June. (*The determiner, now in its masculine/feminine form, is missing the neuter suffix -*t*).
17. Bilen blir angrepet av *<u>et</u> grizzlybjørn 100 meter fra hytta. 'The car is attacked by a grizzly bear 100 meters from the cabin.' (*The article is neuter gender instead of masculine, *en*).
18. Kanelbollen blir spist av gutten på *<u>kjøkkenen</u>. 'The cinnamon bun is eaten by the boy in the kitchen.' (*The masculine suffix -*en* is used instead of the neuter -*et* to denote definite form of the noun).
19. Kaffen blir drukket av en hipster på kaféen. 'The coffee is drunk by a hipster in the cafe.'
20. Hunden blir sett på gata i Oslo. 'The dog is seen in the street in Oslo.'
21. Jenta blir sur fordi faren hennes er *<u>kleint</u> på bussen. 'The girl gets angry because her father is annoying on the bus.' (*The neuter form of the adjective, with the suffix -*t*, is used instead of the masculine form *klein*).
22. Turid blir glad fordi solen skinner utenfor. 'Turid becomes happy because the sun is shining outside.'
23. Mekanikeren reparerer kun bilen sin når været er *<u>god</u>. 'The mechanic only fixes his car when the weather is nice.' (*The adjective, now in its masculine/feminine form, is missing a neuter suffix -*t*).
24. Inger hekler om *<u>kveldet</u> selv om hun er trøtt. 'Inger crochets in the evening even though she is tired.' (*The neuter suffix -*et* is used instead of the masculine -*en* to denote definite form of the noun).
25. Tanten hennes, som bor i Trondheim, elsker opera. 'Her aunt, who lives in Trondheim, loves opera.'
26. Terje, som er glad i mat, arbeider som *<u>kokker</u> på en restaurant. 'Terje, who loves food, works as a chefs in a restaurant.' (*The plural form of the noun is used instead of the singular *kok*).

27. Jenta, som går i barnehage, er alltid *glade. 'The girl, who goes to kindergarten, is always happy.' (*The plural form of the adjective is used instead of the singular form *glad*).
28. Anita, som er meget aktiv, investerer penger i aksjer hver måned. 'Anita, who is very active, invests money in stocks every month.'
29. Det er kaféen som er populær. 'It is the cafe which is popular.'
30. Det er lillesøster som tegner på bordet. 'It is little sister who is drawing on the table.'
31. Det er gutten som reparerer den *fin stolen på verkstedet. 'It is the boy who is repairing the nice chair in the workshop.' (*The indefinite form of the adjective is used instead of the definite form with -*e*).
32. Det er Bjørg som går ned en halv kilo i uken. 'It is Bjørg who is losing half a kilo a week.'
33. Bestemor syr *flott klær til hele familien. 'Grandmother makes nice clothes for the whole family.' (*The singular form of the adjective is used instead of the plural form with -*e*).
34. Bente mosjonerer i skogen på onsdager. 'Bente exercises in the forest on Wednesdays.'
35. Frode planlegger *et hyttetur med vennene sine. 'Frode is planning a cottage trip with his friends.' (*The article is neuter gender instead of masculine, *en*).
36. Legen går på skøyter to ganger i uken. 'The doctor goes ice skating twice a week.'
37. Maria går på kino i helgen. 'Maria goes to the cinema in the weekend.'
38. Bjørnen fisker ørret i *et vannet i skogen. 'The bear catches trout in a/the (?) lake in the forest.' (*Both the indefinite article *et* and the suffix -*et* denoting definite form is used simultaneously).
39. Gartneren luker *ugresser hver måned om sommeren. 'The gardener weeds every month during the summer.' (*The plural form of the noun is *ugress*, without the plural suffix -*er*).
40. Plenklipperen klipper automatisk *gressen i hagen hver uke. 'The lawn mowerautomatically cuts the grass in the garden every week.' (*The masculine suffix -*en* is used instead of the neuter -*et* to denote definite form of the noun).

Comprehension questions

Below are all comprehension question for the target and filler sentences.

Target sentences

1. Leser pappa romaner på sofaen? 'Does dad read novels on the sofa?' No.
2. Lufter mannen hunden sin i parken? 'Does the man walk his dog in the park?' Yes.
7. Sender bestefar julekort i november? 'Does grandfather send Christmas cards in November?' No.
8. Spiller gutten fotball med faren sin? 'Does the boy play football with his father?' No.
13. Vasker Harald på mandager? 'Does Harald wash on Mondays?' Yes.
16. Danser Svein folkedans om høsten? 'Does Svein dance folk dance in the fall?' Yes.

20. Hekler Kristin vesker til familien? ‘Does Kristin crochet bags for the family?’ No.
22. Maler barna påskeegg i barnehagen? ‘Do the children paint Easter eggs in the kindergarten?’ Yes.
25. Besøker pensjonistene en pub i mars? ‘Do the pensioners visit a pub in March?’ No.
27. Henter Astrid tvillingene klokken fem i oddetallsuker? ‘Does Astrid pick up the twins at five o’clock in odd weeks?’ No.
29. Selger Eirik blomster i august? ‘Does Eirik sell flowers in August?’ Yes.
33. Stjeler vennene godteri på butikken? ‘Do the friends steal candy from the store?’ No.
35. Spiser sjefen en kanelbolle i bilen før lunsj? ‘Does the boss eat a cinnamon bun in the car before lunch?’ No.
36. Synger Berit karaoke på puben i kveld? ‘Does Berit sing karaoke in the pub tonight?’ Yes.
39. Baker Ingrid rundstykker etter jobb? ‘Does Ingrid bake buns after work?’ Yes.

Filler sentences

1. Ser Morten en rein ved vannet? ‘Does Morten see a reindeer by the water?’ No.
3. Kjører Linda på sparkesykkel? ‘Does Linda ride a kick scooter?’ Yes.
6. Åpner butikken den 8. november? ‘Does the store open on the 8th of November?’ Yes.
9. Kjøper mannen grønnsaker? ‘Does the man buy vegetables?’ No.
11. Sykler Bjørn til universitetet i regnvær? ‘Does Bjørn bike to the university in rainy weather?’ No.
13. Går pappa ned i vekt? ‘Does dad lose weight?’ No.
15. Elsker mannen sjokoladekake fra bakeriet? ‘Does the man love chocolate cake from the bakery?’ Yes.
19. Drikker hipsteren kaffen? ‘Does the hipster drink coffee?’ Yes.
24. Strikker Inger om kvelden? ‘Does Inger knit in the evening?’ No.
26. Er Terje glad i mat? ‘Does Terje love food?’ Yes.
30. Er det gutten som tegner på bordet? ‘Is it the boy who is drawing on the table?’ No.
32. Går Bjørg ned en halv kilo i uken? ‘Does Bjørg lose half a kilo a week?’ Yes.
35. Planlegger Frode en hyttetur? ‘Does Frode plan a cottage trip?’ Yes.
37. Går Maria på kino i helgen? ‘Does Maria go to the cinema in the weekend?’ Yes.
38. Fisker bjørnen gjedde i et vann i skogen? ‘Does the bear catch pike in a lake in the forest?’ No.

B. MODEL RESULTS

Model estimates for total sentence reading times

Table 6. Model estimates for total sentence reading times in ms

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	558973	747.6		
Item (Intercept)	76336	276.3		
Residual	533671	730.5		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	2875.377	121.477	23.670	< 2e-16 ***
GRAM	409.925	69.746	5.877	4.97e-09 ***
LENGTH	749.937	34.036	22.033	< 2e-16 ***
TRIAL	-15.012	1.530	-9.812	< 2e-16 ***
GRAM:LENGTH	-24.326	68.214	-0.357	0.721431
GRAM:TRIAL	-11.022	3.075	-3.584	0.000347 ***

Model estimates for eye tracking measures in the pre-critical region (adverbial)

Table 7. Model estimates, first fixation duration in ms (pre-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	479.37	21.89		
Item (Intercept)	35.04	5.92		
Residual	4131.69	64.28		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	177.7608	4.5557	39.020	< 2e-16 ***
GRAM	6.4307	6.2977	1.021	0.30734
LENGTH	-8.2289	3.0684	-2.682	0.00739 **
TRIAL	0.1178	0.1381	0.853	0.39378
GRAM:LENGTH	-6.4615	6.1343	-1.053	0.29233
GRAM:TRIAL	-0.1437	0.2761	-0.520	0.60285

Table 8. Model estimates, gaze duration (pre-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	36796	191.82		
Item (Intercept)	6562	81.01		
Residual	60948	246.88		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	726.870	32.822	22.146	< 2e-16 ***
GRAM	29.803	24.330	1.225	0.2208
LENGTH	649.976	11.802	55.073	< 2e-16 ***
TRIAL	-1.368	0.532	-2.571	0.0102 *
GRAM:LENGTH	-17.877	23.599	-0.758	0.4488
GRAM:TRIAL	-1.233	1.069	-1.154	0.2489

Table 9. Model estimates, regression path duration (pre-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	36737	191.7		
Item (Intercept)	6528	80.8		
Residual	60809	246.6		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	728.0229	32.7855	22.206	< 2e-16 ***
GRAM	32.2441	24.3017	1.327	0.18474
LENGTH	650.1281	11.7885	55.150	< 2e-16 ***
TRIAL	-1.4108	0.5314	-2.655	0.00801 **
GRAM:LENGTH	-17.5830	23.5715	-0.746	0.45580
GRAM:TRIAL	-1.3255	1.0674	-1.242	0.21448

Table 10. Model estimates, total duration (pre-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	77812	278.95		
Item (Intercept)	9404	96.97		
Residual	158157	397.69		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	924.2121	47.0358	19.649	< 2e-16 ***
GRAM	149.4711	37.9209	3.942	8.41e-05 ***
LENGTH	780.9166	18.5271	42.150	< 2e-16 ***
TRIAL	-3.7648	0.8323	-4.524	6.49e-06 ***
GRAM:LENGTH	-1.0636	37.1209	-0.029	0.97715
GRAM:TRIAL	-4.3122	1.6711	-2.580	0.00995 **

Table 11. Model estimates, first-pass skipping ratio (pre-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2.5903	1.6095		
Item (Intercept)	0.6542	0.8088		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-8.4905	32.0664	-0.265	0.7912
GRAM	7.9819	64.1256	0.124	0.9009
LENGTH	-10.9765	64.1277	-0.171	0.8641
TRIAL (rescaled)	-1.0604	0.4762	-2.227	0.0259 *
GRAM:LENGTH	16.3876	128.2511	0.128	0.8983

Since the model would not converge, the interaction between grammar and trial was removed. The optimizer BOBYQA was used as well.

Model estimates for eye tracking measures in the critical region I (subject)

Table 12. Model estimates, first fixation duration in ms (critical region I, subject)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	1001.9	31.65		
Item (Intercept)	147.8	12.16		
Residual	4854.7	69.68		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	230.2834	6.0435	38.104	< 2e-16 ***
GRAM	6.7504	6.9281	0.974	0.3300
LENGTH	8.1628	3.3909	2.407	0.0162 *
TRIAL	-0.2346	0.1520	-1.544	0.1228
GRAM:LENGTH	11.4659	6.7839	1.690	0.0912
GRAM:TRIAL	0.3604	0.3046	1.183	0.2368

Table 13. Model estimates, gaze duration in ms (critical region I, subject)

Random effects	***Variance***	***Std. Dev.***		
Participant (Intercept)	2845	53.34		
Item (Intercept)	2632	51.30		
Residual	13934	118.04		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	283.2806	12.6286	22.432	< 2e-16 ***
GRAM	46.6087	11.7717	3.959	7.84e-05 ***
LENGTH	-0.2837	5.7484	-0.049	0.9607
TRIAL	-0.6614	0.2579	-2.565	0.0104 *
GRAM:LENGTH	3.2586	11.5013	0.283	0.7770
GRAM:TRIAL	-0.2153	0.5180	-0.416	0.6778

Table 14. Model estimates, regression path duration in ms (critical region I, subject)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	7516	86.69		
Item (Intercept)	4935	70.25		
Residual	49229	221.88		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	372.5903	20.0359	18.596	<2e-16 ***
GRAM	149.4993	22.1085	6.762	1.89e-11 ***
LENGTH	-27.9815	10.8021	-2.590	0.00967 **
TRIAL	-2.2105	0.4845	-4.563	5.42e-06 ***
GRAM:LENGTH	-21.5756	21.6121	-0.998	0.31828
GRAM:TRIAL	-2.2828	0.9727	-2.347	0.01904 *

Table 15. Model estimates, total duration in ms (critical region I, subject)

Random effects	***Variance***	***Std. Dev.***		
Participant (Intercept)	9321	96.55		
Item (Intercept)	7573	87.02		
Residual	35819	189.26		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	387.4995	21.5467	17.984	< 2e-16 ***
GRAM	170.7236	18.0732	9.446	< 2e-16 ***
LENGTH	-13.0142	8.8177	-1.476	0.140146
TRIAL	-2.2524	0.3963	-5.683	1.54e-08 ***
GRAM:LENGTH	-29.2804	17.6731	-1.657	0.097742
GRAM:TRIAL	-2.9220	0.7969	-3.667	0.000253 ***

Table 16. Model estimates, first-pass skipping ratio (critical region I, subject)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	1.5101	1.2289		
Item (Intercept)	0.2826	0.5316		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-3.0275	0.3013	-10.048	< 2e-16 ***
GRAM	0.2371	0.3725	0.637	0.524
LENGTH	0.1259	0.1855	0.679	0.497
TRIAL (rescaled)	-0.2294	0.3339	-0.687	0.492
GRAM:LENGTH	-0.4488	0.3760	-1.194	0.233
GRAM:TRIAL	-0.8462	0.6716	-1.260	0.208

Table 17. Model estimates, first-pass regression ratio (critical region I, subject)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	0.3944	0.6280		
Item (Intercept)	0.2018	0.4492		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-2.07715	0.20200	-10.283	< 2e-16 ***
GRAM	1.12744	0.31894	3.535	0.000408 ***
LENGTH	-0.28370	0.16219	-1.749	0.080256
TRIAL (rescaled)	-0.45080	0.28771	-1.567	0.117151
GRAM:LENGTH	-0.06884	0.32393	-0.213	0.831694
GRAM:TRIAL	-0.54672	0.57543	-0.950	0.342054

Model estimates for eye tracking measures in the critical region II (verb)

Table 18. Model estimates, first fixation duration in ms (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	1242.05	35.243		
Item (Intercept)	33.84	5.817		
Residual	6514.73	80.714		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	238.1546	6.5218	36.517	< 2e-16 ***
GRAM	22.5873	7.9474	2.842	0.00454 **
LENGTH	12.0798	3.8742	3.118	0.00185 **
TRIAL	-0.1508	0.1746	-0.864	0.38780
GRAM:LENGTH	0.2275	7.7457	0.029	0.97657
GRAM:TRIAL	-0.4458	0.3491	-1.277	0.20187

Table 19. Model estimates, gaze duration in ms (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2588.7	50.88		
Item (Intercept)	247.1	15.72		
Residual	12643.1	112.44		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	278.8768	9.5329	29.254	< 2e-16 ***
GRAM	32.6015	11.1027	2.936	0.00337 **
LENGTH	5.1584	5.3988	0.955	0.33948
TRIAL	-0.3599	0.2436	-1.477	0.13979
GRAM:LENGTH	2.8138	10.7962	0.261	0.79441
GRAM:TRIAL	-0.9459	0.4883	-1.937	0.05288

Table 20. Model estimates, regression path duration in ms (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	5344.9	73.11		
Item (Intercept)	452.4	21.27		
Residual	57486.2	239.76		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	401.4608	16.1897	24.797	< 2e-16 ***
GRAM	161.6043	23.6185	6.842	1.09e-11 ***
LENGTH	-1.7611	11.5067	-0.153	0.878375
TRIAL	-2.6121	0.5186	-5.037	5.24e-07 ***
GRAM:LENGTH	22.1857	23.0086	0.964	0.335069
GRAM:TRIAL	-3.4970	1.0379	-3.369	0.000771 ***

Table 21. Model estimates, total duration in ms (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	6978.6	83.54		
Item (Intercept)	902.5	30.04		
Residual	29997.5	173.20		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	368.4372	15.3435	24.013	< 2e-16 ***
GRAM	88.0624	16.4911	5.340	1.05e-07 ***
LENGTH	0.3660	8.0677	0.045	0.96382
TRIAL	-1.6012	0.3622	-4.421	1.04e-05 ***
GRAM:LENGTH	15.9832	16.1593	0.989	0.32275
GRAM:TRIAL	-2.4024	0.7264	-3.307	0.00096 ***

Table 22. Model estimates, first-pass skipping ratio (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2.9490	1.7173		
Item (Intercept)	0.1249	0.3534		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-3.81898	0.41509	-9.200	< 2e-16 ***
GRAM	-0.46246	0.44455	-1.040	0.298
LENGTH	-0.01809	0.22274	-0.081	0.935
TRIAL (rescaled)	-0.38926	0.40199	-0.968	0.333
GRAM:LENGTH	-0.38665	0.45299	-0.854	0.393
GRAM:TRIAL	0.52301	0.81229	0.644	0.520

Table 23. Model estimates, first-pass regression ratio (critical region II, verb)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	0.26693	0.5166		
Item (Intercept)	0.03104	0.1762		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-1.5538	0.1692	-9.181	< 2e-16 ***
GRAM	1.1061	0.2914	3.796	0.000147 ***
LENGTH	-0.2724	0.1542	-1.767	0.077250
TRIAL (rescaled)	-0.9734	0.2801	-3.475	0.000511 ***
GRAM:LENGTH	0.4479	0.3084	1.452	0.146476
GRAM:TRIAL	0.1825	0.5596	0.326	0.744366

Model estimates for eye tracking measures in the post-critical region (object)

Table 24. Model estimates, first fixation duration in ms (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	1138.2	33.74		
Item (Intercept)	326.9	18.08		
Residual	6657.5	81.59		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	244.12372	6.88529	35.456	< 2e-16 ***
GRAM	-3.01105	7.92019	-0.380	0.704
LENGTH	-2.13152	3.87091	-0.551	0.582
TRIAL	-0.03601	0.17429	-0.207	0.836
GRAM:LENGTH	-10.21189	7.74877	-1.318	0.188
GRAM:TRIAL	-0.33795	0.35000	-0.966	0.334

Table 25. Model estimates, gaze duration in ms (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	5701	75.70		
Item (Intercept)	7805	88.34		
Residual	23313	152.68		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	349.7868	19.1919	18.226	< 2e-16 ***
GRAM	-6.2397	14.8570	-0.420	0.674550
LENGTH	-26.6994	7.2452	-3.685	0.000236 ***
TRIAL	-0.5049	0.3266	-1.546	0.122270
GRAM:LENGTH	-26.7907	14.5076	-1.847	0.064969
GRAM:TRIAL	-0.2497	0.6570	-0.380	0.703943

Table 26. Model estimates, regression path duration in ms (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	7454	86.34		
Item (Intercept)	10790	103.88		
Residual	64631	254.23		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	422.4931	23.9980	17.605	< 2e-16 ***
GRAM	18.7818	24.7191	0.760	0.44747
LENGTH	-34.3612	12.0618	-2.849	0.00444 **
TRIAL	-1.1531	0.5434	-2.122	0.03398 *
GRAM:LENGTH	-57.3454	24.1506	-2.374	0.01768 *
GRAM:TRIAL	-0.8285	1.0929	-0.758	0.44851

Table 27. Model estimates, total duration in ms (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	14619	120.9		
Item (Intercept)	13784	117.4		
Residual	43283	208.0		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	464.1927	27.3271	16.987	< 2e-16 ***
GRAM	-12.0881	19.8730	-0.608	0.543
LENGTH	-15.7605	9.6932	-1.626	0.104
TRIAL	-2.1637	0.4357	-4.965	7.51e-07 ***
GRAM:LENGTH	-27.2042	19.4291	-1.400	0.162
GRAM:TRIAL	-0.5737	0.8764	-0.655	0.513

Table 28. Model estimates, first-pass skipping ratio (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2.453	1.566		
Item (Intercept)	2.104	1.451		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-5.6856	0.6190	-9.185	< 2e-16 ***
GRAM	0.2087	0.6436	0.324	0.7457
LENGTH	0.2541	0.2927	0.868	0.3852
TRIAL (rescaled)	1.0681	0.5451	1.959	0.0501
GRAM:LENGTH	0.4985	0.6276	0.794	0.4270
GRAM:TRIAL	-0.1362	1.0616	-0.128	0.8979

Table 29. Model estimates, first-pass regression ratio (post-critical region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	0.2056	0.4534		
Item (Intercept)	0.1427	0.3778		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-2.11818	0.18676	-11.341	< 2e-16 ***
GRAM	0.24894	0.31588	0.788	0.4307
LENGTH	-0.05718	0.16217	-0.353	0.7244
TRIAL (rescaled)	-0.57119	0.29355	-1.946	0.0517
GRAM:LENGTH	-0.30082	0.32542	-0.924	0.3553
GRAM:TRIAL	-0.20932	0.58877	-0.356	0.7222

Model estimates for eye tracking measures in the wrap-up region (adverbial)

Table 30. Model estimates, first fixation duration in ms (wrap-up region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	2296.4	47.92		
Item (Intercept)	423.2	20.57		
Residual	12269.5	110.77		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	270.8234	9.2838	29.172	< 2e-16 ***
GRAM	-8.6469	10.5794	-0.817	0.414
LENGTH	-1.8581	5.1741	-0.359	0.720
TRIAL	-0.2279	0.2323	-0.981	0.327
GRAM:LENGTH	-0.0949	10.3644	-0.009	0.993
GRAM:TRIAL	0.3049	0.4661	0.654	0.513

Table 31. Model estimates, gaze duration in ms (wrap-up region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	32051	179.0		
Item (Intercept)	18518	136.1		
Residual	67280	259.4		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	631.6596	35.8227	17.633	< 2e-16 ***
GRAM	-20.7221	24.8454	-0.834	0.4044
LENGTH	-20.2776	12.1192	-1.673	0.0945
TRIAL	-3.7880	0.5448	-6.952	5.04e-12 ***
GRAM:LENGTH	-5.6171	24.2921	-0.231	0.8172
GRAM:TRIAL	0.8911	1.0957	0.813	0.4162

Table 32. Model estimates, regression path duration in ms (wrap-up region)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	149799	387.0		
Item (Intercept)	22218	149.1		
Residual	332035	576.2		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	1037.623	66.547	15.592	< 2e-16 ***
GRAM	72.722	55.111	1.320	0.187
LENGTH	138.705	26.920	5.152	2.86e-07 ***
TRIAL	-8.721	1.209	-7.210	8.25e-13 ***
GRAM:LENGTH	39.346	53.940	0.729	0.466
GRAM:TRIAL	-3.660	2.429	-1.507	0.132

Table 33. Model estimates, total duration in ms (wrap-up region)

Random effects	Variance	Std. Dev.		
Participant (Intercept)	41910	204.7		
Item (Intercept)	20299	142.5		
Residual	72238	268.8		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	731.3359	39.2874	18.615	<2e-16 ***
GRAM	12.4040	25.6733	0.483	0.629
LENGTH	-2.6183	12.5227	-0.209	0.834
TRIAL	-5.2596	0.5630	-9.343	<2e-16 ***
GRAM:LENGTH	15.3464	25.1004	0.611	0.541
GRAM:TRIAL	-0.7581	1.1321	-0.670	0.503

Table 34. Model estimates, first-pass skipping ratio (wrap-up region)

Random effects	Variance	Std. Dev.		
Participant (Intercept)	4.0888	2.0221		
Item (Intercept)	0.2737	0.5232		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-6.97366	1.34224	-5.196	2.04e-07 ***
GRAM	0.88255	0.70241	1.256	0.209
LENGTH	0.45213	0.66106	0.684	0.494
TRIAL (rescaled)	-0.01681	1.14651	-0.015	0.988

The model would not converge, when including interactions. Both interactions were thus removed.

Table 35. Model estimates, first-pass regression ratio (wrap-up region)

Random effects	Variance	Std. Dev.		
Participant (Intercept)	1.52616	1.2354		
Item (Intercept)	0.02937	0.1714		
Fixed effects	**Estimate**	**Std. Error**	***z*-value**	***p*-value**
(Intercept)	-0.83171	0.21641	-3.843	0.000121 ***
GRAM	0.07071	0.23384	0.302	0.762363
LENGTH	0.40196	0.11800	3.407	0.000658 ***
TRIAL (rescaled)	-0.66588	0.21266	-3.131	0.001741 **
GRAM:LENGTH	0.08641	0.23547	0.367	0.713648
GRAM:TRIAL	-0.45728	0.42446	-1.077	0.281332

C. POST-HOC ANALYSIS OF COMBINED CRITICAL REGIONS

Model estimates for post-hoc analysis of combined subject-verb region

Table 36. Model estimates, total duration in ms (critical regions I+II)

Random effects	**Variance**	**Std. Dev.**		
Participant (Intercept)	31162	176.5		
Item (Intercept)	10714	103.5		
Residual	76250	276.1		
Fixed effects	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	755.6856	32.9995	22.900	< 2e-16 ***
GRAM	258.0955	26.3659	9.789	< 2e-16 ***
LENGTH	-13.0605	12.8685	-1.015	0.31
TRIAL	-3.8282	0.5785	-6.618	4.83e-11 ***
GRAM:LENGTH	-13.5326	25.7897	-0.525	0.60
GRAM:TRIAL	-5.2642	1.1629	-4.527	6.39e-06 ***

D. FIGURES

Figure 3. Total duration on the sentence-initial adverbial: Effect plot of grammaticality and trial order

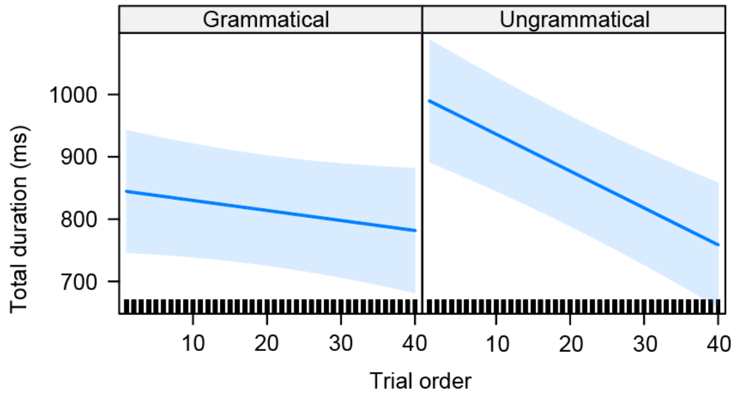


Figure 4. Regression path duration on the subject: Effect plot of grammaticality and trial order

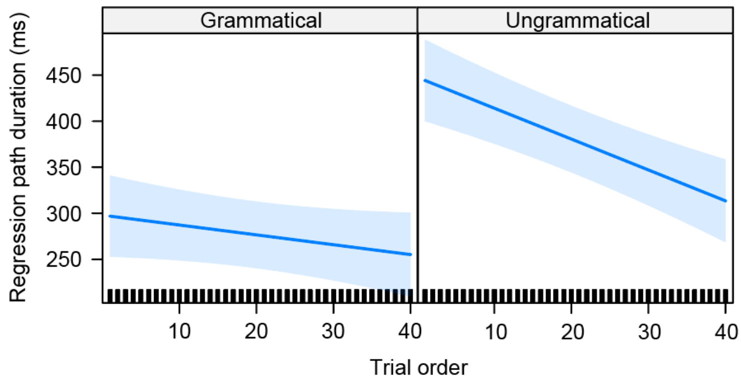


Figure 5. Total duration on the subject: Effect plot of grammaticality and trial order

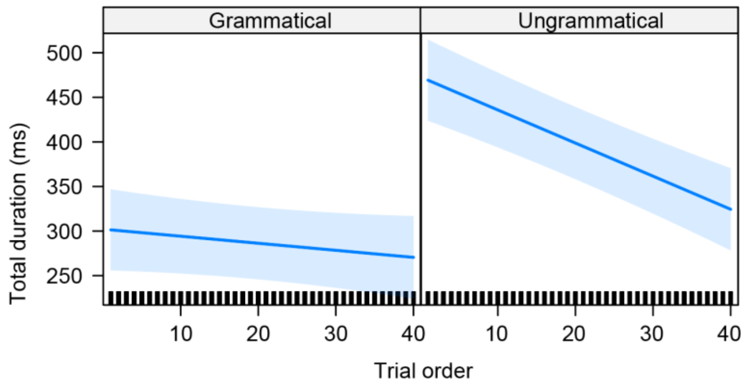


Figure 6. Regression path duration on the verb: Effect plot of grammaticality and trial order

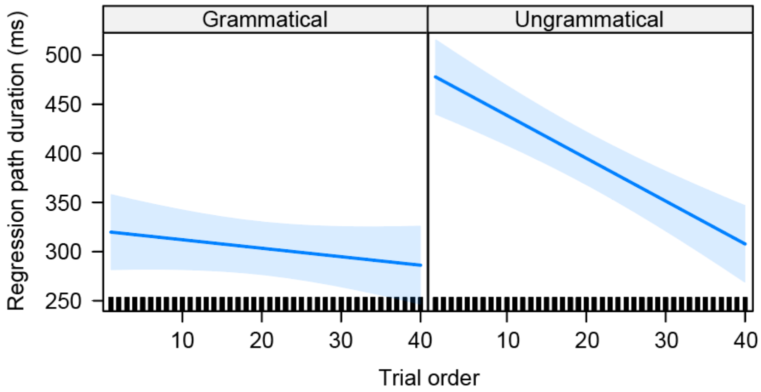


Figure 7. Total duration on the verb: Effect plot of grammaticality and trial order

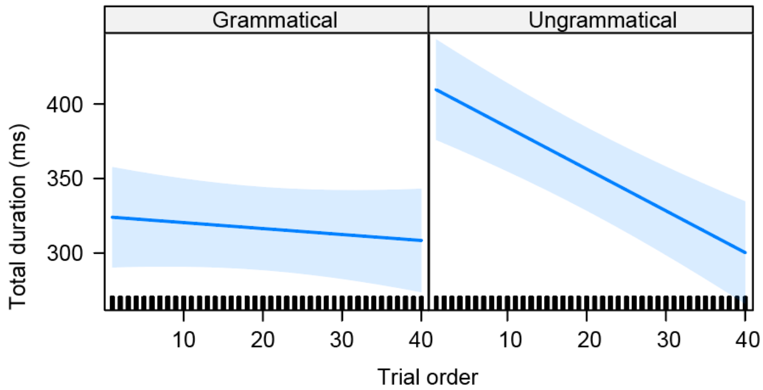
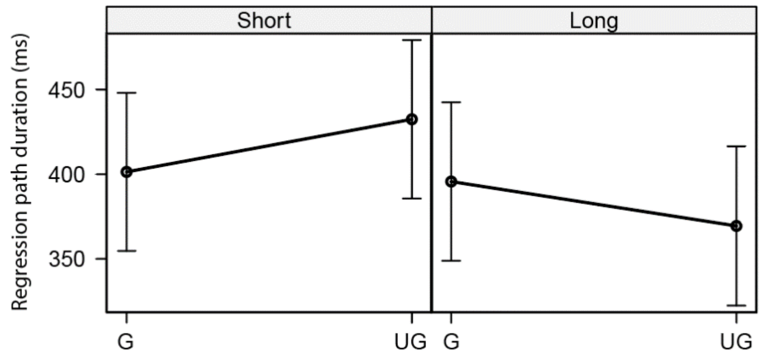


Figure 8. Effect plot of regression path duration in ms in the post-critical region (object)



ISBN 978-82-326-7802-0 (printed ver.)
ISBN 978-82-326-7801-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

UNIVERSITY OF COPENHAGEN

NTNU
Norwegian University of Science and Technology