


ORIGINAL ARTICLE

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

Katrine Falcon Søby<sup>1\*</sup> , Evelyn Arko Milburn<sup>2,3</sup>, Line Burholt Kristensen<sup>1</sup>,  
Valentin Vulchanov<sup>3</sup> and Mila Vulchanova<sup>3</sup> 

<sup>1</sup>Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark,

<sup>2</sup>Department of Psychology, Mills College, Oakland, CA, USA and <sup>3</sup>Language Acquisition and Language Processing Lab, Department of Language & Literature, Norwegian University of Science & Technology, Trondheim, Norway

\*Corresponding author. Email: [kafs@kp.dk](mailto: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

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

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

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

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<sup>1</sup> 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 eating the food we put on the porch.*
- b. *It seems that the cats won’t usually eat 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.<sup>2</sup> However, morphosyntactically anomalous sentences induced significantly more regressions than baseline sentences in the region containing the anomalous progressive verb

**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	<i>The white was <u>cat</u> big.</i>	More regressions in (to the critical 4 <sup>th</sup> 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	<i>It seems that the cats won't usually <u>eating</u> the food we put on the porch.</i>	More regressions out in the verb region and subsequent region.	First-pass reading times <sup>**</sup> , total reading time (residual reading time)
Braze et al. (2002)	Modal verb + progressive or past tense form of a verb in English	<i>The wall will surely <u>cracking</u> after a few years in this harsh climate./The engine will softly <u>whined</u> while it is running at low capacity.</i>	More regressions out in the verb region and increased first-pass reading times <sup>**</sup> (but only by subject).	
Pearlmutter et al. (1999)	Subject-verb number agreement in English (the attraction phenomenon)	<i>The key to the cabinet(s) <u>were</u> rusty from many years of disuse.</i>	More regressions out of and increased total reading times in the verb region.	First-pass reading times <sup>**</sup>
Lim and Christianson (2015)	Subject-verb number agreement in English (the attraction phenomenon).	<i>The teacher who instructed the student(s) <u>were</u> very strict.</i>	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: <i>The newspaper wrote that the stew<sub>(masc.)</sub> of the (male) cook<sub>(masc.)</sub> <u>had become famous</u><sub>(fem.)</sub> throughout the state.</i>	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>I enjoyed seeing how the actors<sub>(masc.pl.)</sub> the actresses<sub>(fem.pl.)</sub> <u>were enchanting</u> the tired audience<sub>(masc.pl.)</sub></i>	For morphologically marked verbs: first fixation duration, gaze duration, and second-pass duration.	

<sup>\*\*</sup>The term *gaze duration* is used in the current study.

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	<i>På tirsdager</i> 'On Tuesdays'	<i>tilbyr</i> 'offers'	<i>biblioteket</i> 'the library'	<i>høytlesning</i> 'a read-aloud'	<i>for barn og unge.</i> 'for children and adolescents.'
Ungrammatical (S-V) Short adverbial	<i>På tirsdager</i> 'On Tuesdays'	<i>biblioteket</i> 'the library'	<i>tilbyr</i> 'offers'	<i>høytlesning</i> 'a read-aloud'	<i>for barn og unge.</i> 'for children and adolescents.'
Grammatical (V-S) Long adverbial	<i>Klokken halv sju på tirsdager</i> 'Half past six on Tuesdays'	<i>tilbyr</i> 'offers'	<i>biblioteket</i> 'the library'	<i>høytlesning</i> 'a read-aloud'	<i>for barn og unge.</i> 'for children and adolescents.'
Ungrammatical (S-V) Long adverbial	<i>Klokken halv sju på tirsdager</i> 'Half past six on Tuesdays'	<i>biblioteket</i> 'the library'	<i>tilbyr</i> 'offers'	<i>høytlesning</i> 'a read-aloud'	<i>for barn og unge.</i> '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  $2 \times 2$  design: grammatical (V2) vs. ungrammatical (V); short adverbial vs. long adverbial.

All experimental items consisted of sentences with five regions<sup>3</sup> (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)



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 counter-balanced 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.

### **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*<sup>4</sup>: the total duration of all fixations from entering a region during first-pass reading until leaving it to the right, including any rerefixations on previous text.
- *Total duration*: of all fixations on a region.

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.

**Table 3.** Main effects and interactions

Region	Pre-critical	Critical regions		Post-critical	Wrap-up
Constituent	Adverbial	Subject	Verb	Object	Adverbial
<b>Main effects</b>	<i>På tirsdager</i> 'On Tuesdays'	<i>*biblioteket</i> 'the library'	<i>tilbyr</i> 'offers'	<i>høytlesning</i> 'a read-aloud'	<i>for barn og unge.</i> 'for children and adolescents.'
Gram	<b>UG&gt;G</b> TD	<b>UG&gt;G</b> GD, RR, RPD, TD	<b>UG&gt;G</b> FFD, GD, RR, RPD, TD	(no effects)	(no effects)
Length of adverbial	<b>LONG&gt;SHORT</b> GD, RPD, TD <b>SHORT&gt;LONG</b> FFD	<b>LONG&gt;SHORT</b> FFD <b>SHORT&gt;LONG</b> RPD	<b>LONG&gt;SHORT</b> FFD (no effects)	(no effects) <b>SHORT&gt;LONG</b> GD, RPD	<b>LONG&gt;SHORT</b> RR, RPD
Trial order	<b>EARLY&gt;LATE</b> SR, GD, RPD, TD	<b>EARLY&gt;LATE</b> GD, RPD, TD	<b>EARLY&gt;LATE</b> RR, RPD, TD	<b>EARLY&gt;LATE</b> RPD, TD	<b>EARLY&gt;LATE</b> GD, RR, RPD, TD
<b>Interactions</b>					
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.<sup>5</sup>

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

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

	Grammatical	Ungrammatical	Total
Short Adverbial	2103 ( $\pm 838$ )	2292 ( $\pm 973$ )	2197 ( $\pm 912$ )
Long Adverbial	2859 ( $\pm 1161$ )	3036 ( $\pm 1350$ )	2946 ( $\pm 1260$ )
<b>Total</b>	<b>2479 (<math>\pm 1080</math>)</b>	<b>2659 (<math>\pm 1231</math>)</b>	<b>2569 (<math>\pm 1160</math>)</b>

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.

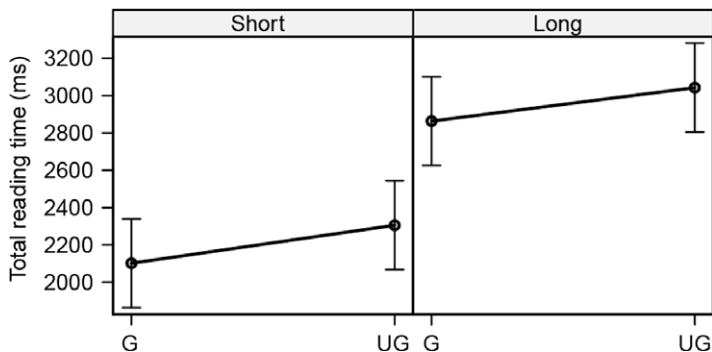


Figure 1. Effect Plot of Total Sentence Reading Time in ms.

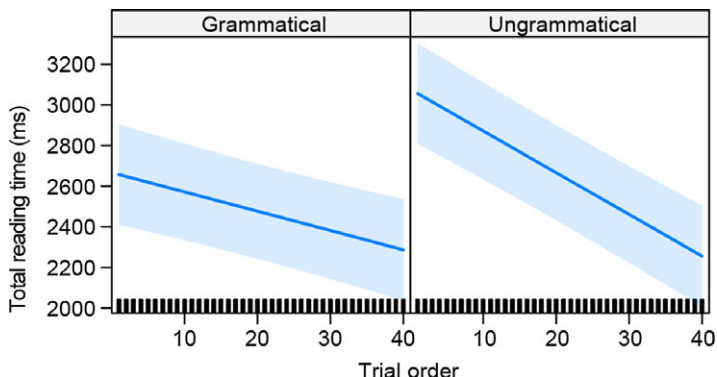


Figure 2. Interaction Between Grammaticality and Trial Order: Effect Plot of Total Sentence Reading Time in ms.

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).

**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)

Region	Condition			
	G, short	UG, short	G, long	UG, long
<b>Pre-critical region: Adverbial</b>				
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)
<b>Critical regions: Subject</b>				
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)
<b>Critical regions: Verb</b>				
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)
<b>Post-critical region: Object</b>				
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)
<b>Wrap-up region: Adverbial</b>				
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)

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 ( $\hat{\beta} = 161.60$  ms, SE = 23.62,  $t = 6.84$ ,  $p < .001$ ), late measurement: Total duration ( $\hat{\beta} = 88.06$  ms, SE = 16.49,  $t = 5.34$ ,  $p < .001$ )).



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

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.

### **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

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 Soby 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 (Soby *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 (Soby *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](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., Paris, 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 Stockholmspråk*. 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.
- Røynealand, 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>

- Soby, 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.
- Soby, 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.
- Soby, 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 språkläktenskap. 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://jeremyeaton.github.io/papers/Yeaton-2019-MasterThesis-V2\\_word\\_order.pdf](https://jeremyeaton.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.

- (*Tidlig om morgenen*) i helgen leser pappa avisen på sofaen. ‘Early in the morning in the weekends, dad reads the newspaper on the sofa’
- (*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.’
- (*Minst to ganger i uken*) i 2020 holder kommunen nynorsk kurs for offentlig ansatte. ‘At least twice a week in 2020, the municipality holds a Nynorsk course for public employees.’
- (*Etter middag hver*) lørdag kveld spiser gutten gelato på Solsiden. ‘After dinner every Saturday night, the boy eats gelato at Solsiden.’
- (*Etter klokken ett*) om natten løser Marit kryssord på mobilen. ‘After 1 AM, Marit solves crossword puzzles on her phone.’
- (*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 rensar 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 17<sup>th</sup> 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 1<sup>st</sup> 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.'

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*) for 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.'