

ARTICLE

Control in a Norwegian grammar maze

Tori Larsen  and Christer Johansson 

University of Bergen, Department of Linguistic, Literary and Aesthetic Studies, Bergen, 5007, Norway
Emails: tori.larsen@uib.no; christer.johansson@uib.no

(Received 1 October 2021; revised 22 August 2022; accepted 25 August 2022; first published online 30 September 2022)

Abstract

Coreference processing of Control constructions and their pronoun-containing counterparts can be studied experimentally using priming or interference paradigms. We replicate findings in a priming study on non-finite Control constructions in Norwegian (Larsen & Johansson, 2020) and contrast them with their finite counterparts using interference effects in a grammatical maze (G-maze) design. We asked participants to read sentences word-by-word and to select the grammatically correct continuation from two options. When the ungrammatical option was a potential antecedent from within the sentence, we predicted interference, i.e., longer reaction times compared to an unrelated baseline. We observed a trend towards significant interference effects when a participant was presented with either of the potential noun phrase (NP) antecedents of PRO in competition with the infinitive marker (test position zero) during the processing of a Control sentence. This indicates reactivation of potential antecedents at the infinitive marker, and a reactivation position (PRO) near or at the infinitive marker. We also observed significant differences between Control constructions and their pronoun counterparts. A significant interference effect was recorded for Subject Pronoun constructions when either potential NP antecedent of the pronoun was presented in competition with the pronoun itself. A similar trend was recorded for Object Pronoun sentences.

Keywords: Control; PRO; G-maze; language processing; coreference; reaction time

1. Introduction

Coreference assignment and how it is achieved by means of syntax, semantics, and pragmatics is a topic of debate among linguistic theorists as well as researchers concerned with language processing. Depending on the theory and the type of referent-dependent item in discussion, different dependencies seem to have varying effects on coreference processing. Some studies investigate antecedent reactivation using reaction time experiments predicting facilitation by priming (e.g. Nicol & Swinney, 1989; Hestvik et al., 2005, 2010; Larsen & Johansson, 2020), while others analyze interference effects (e.g. Parker et al., 2015; Sturt & Kwon, 2015). Priming occurs when some item, whether it be a word, image, or part of a syntactic

representation, *facilitates* the processing of an upcoming item or word. This process allows for an early activation of said item by preparing or pre-activating it for faster processing. Interference takes place in a similar manner, but it has the opposite effect on participant reaction times, for example by activating a competitor. Slower reaction times are explained by *interference* with the processing of an upcoming item. This can be due to semantic or syntactic competition. Both effects generally occur undetected by the participant.

Our research focuses on non-finite Control constructions and the null pronoun (PRO) that fills the subject position of Control clauses. We aim to investigate coreference processing in Norwegian non-finite clauses and complement clauses containing a pronoun. Our experimental design takes into consideration the position and pattern of antecedent reactivation. We investigate NP reactivation interference effects when in competition with the infinitive marker/complementizer and the subordinate clause verb. We then combine our findings with previous research on Control in Norwegian to discuss the theoretical implications of real-time language processing data.

1.1 Non-finite clauses and PRO

In non-finite clauses, PRO occupies the subject position of the clause as a covert pronoun (Chomsky, 1981). Example 1.1 presents a Control sentence in English in which PRO is coreferent with the subject of the matrix clause. This is called Subject Control. Depending on the sentence, PRO can be controlled by the subject or object of the matrix clause, or by a noun phrase from outside of the sentence.

- (1) [Mary]_i promised [PRO]_i to give an example.

Control constructions are found in all languages that contain non-finite clauses (Bisang, 2008); however, that leads to the presumption that PRO is the same across these languages. The established characteristics of PRO have evolved throughout the years (Reed, 2014), yet the underlying assumption that these characteristics remain the same regardless of language has become increasingly difficult to support. Sigursson (2008) argues that PRO in Icelandic can be assigned dative case. If one of the *established* characteristics of PRO is that it is always assigned null case (Chomsky & Lasnik, 1993), we already see an issue with claiming the existence of a universal PRO.

Our research investigates Control constructions in Norwegian. Much of the research on Control has been done using English. English and Norwegian have arguably similar surface structures when it comes to non-finite Control clauses; however, the standard analysis of Norwegian non-finite clauses has been to consider the infinitive marker *å* as a complementizer, as shown in (2a) (Faarlund et al., 1997; Åfarli et al., 2003; Faarlund, 2007). In English, though, the infinitive marker *to* is commonly represented in the position of the tense head, T, as shown in (2b) (See Carnie, 2013).

- (2) a. Mary lovet [_{CP} å [_{TP} PRO gi et eksempel]].
 Mary promised [_{CP} to [_{TP} PRO give an example]].
 b. Mary promised [_{CP} [_{TP} PRO to give an example]].

Support for the infinitive marker being positioned in C in Norwegian relies on the sentence position of adverbials in non-finite clauses being flexible. This positioning allows finite and non-finite clauses to maintain similar syntactic structures (Faarlund, 2015).

There are other theories of Control that propose an analysis of non-finite clauses that does not contain PRO. One theory of Control suggests that they are formed by movement in a manner similar to Raising, and this would make the subject of the non-finite clause a regular NP trace (Hornstein, 1999). Another theory maintains a similar claim and states that Control should be handled by A-movement but treated differently than Raising (Hornstein & Polinsky, 2010). Yet another approach accounts for Control using an alternative theory of theta roles. Theta role percolation (Neeleman & van de Koot, 2002) is used to reinterpret the subject properties of controlled infinitives, which allows theta roles to regulate case on the predicate without necessitating PRO (Janke, 2003, 2007, 2008).

These competing theories generate more questions. Are Control constructions processed similarly to other constructions containing referent-dependent items? If so, which ones? Can we determine the syntactic properties of PRO using experimental methods? Do constructions with similar syntactic structures but different dependencies, such as finite complement clauses, display effects for coreference processing? We examine these questions by comparing the processing effects of PRO and pronouns in non-finite and finite clauses in Norwegian. The effects are contrasted with current theories of Control to consider which proposed syntactic properties of PRO are supported by our results.

1.2 Finite clauses and pronouns

PRO has been called a null pronoun, having properties similar to both anaphors and pronouns. PRO has been compared to NP traces both theoretically (Hornstein, 1999; Hornstein & Polinsky, 2010) and empirically (e.g. Nicol & Swinney, 1989; Walenski, 2002; Larsen & Johansson, 2020). PRO and NP traces have the common characteristic of being phonologically null. Since another characteristic of PRO is that it is a pronoun, it should also be studied in relation to overt pronouns.

The umbrella term '*complement clause*' encompasses both finite and non-finite clauses. Non-finite clauses contain PRO while finite clauses are less selective. Since theories of Generative Grammar allow Norwegian finite and non-finite clauses to hold a similar syntactic structure, it enables these minimal differences in structure to be tested and manipulated experimentally. A pronoun can be placed in a finite clause in the same position in which PRO would appear in a non-finite clause.

- (3) a. Mary lovet [_{CP} å [_{TP} PRO aldri gi et eksempel]].
 Mary promised [_{CP} to [_{TP} PRO never give an example]].
 b. Mary lovet [_{CP} at [_{TP} hun ville gi et eksempel]].
 Mary promised [_{CP} that [_{TP} she would give an example]].
 c. Mary lovet [_{NP}John]_[NPsjokolade].
 Mary promised [_{NP}John]_[NPchocolate].

We explore this in our study, assuming minimal differences based on the input provided to the parser before the test positions. Looking at (3), it is apparent that the continuation of the sentence cannot be predicted prior to the presentation of the infinitive marker or complementizer. If PRO behaves like a pronoun, we would expect to observe the same pattern of antecedent reactivation.

1.3 Anaphor interpretation and coreference processing

There is a plethora of theories on anaphora and pronoun interpretation. One prominent theory is Centering Theory (CT) (Grosz et al., 1995; Johshi & Prasad, 2006), which states that particular antecedents are more central in discourse than others. The ranking of antecedents is defined by morphogrammatical properties, particularly grammatical role (Kehler & Rohde, 2013). Resolution may occur immediately or be delayed, depending on when the information is available. A competing theory by Hobbs (1979, 1990) focuses on the role that world knowledge and coherence relations play on pronoun interpretation. This theory may more easily account for referent-swapping constructions, such as the one given in Example 1.3, since it is not strictly based on factors such as grammatical form and information structure. In 1.3, PRO is able to swap from being coreferent with the subject of the matrix clause (a) to being coreferent with an NP from outside of the sentence (b). Recent theories combine the two into a model that can account for incremental interpretation and contextual influences (Kehler & Rohde, 2013).

- (4) [Mary]_i promised [PRO]_i to have . . .
 a. an example ready.
 b. the opportunity [PRO]_k to give an example.

Experimental data supports CT in its claims that grammatical information is available to the parser during anaphor processing; however, the same data shows that some sentences must be fully processed prior to antecedent assignment. Previous research on antecedent assignment and pronouns has found that *the set of all potential antecedents* are reactivated at the position of the pronoun during on-line processing (Corbett & Chang, 1983; Nicol, 1988; Nicol & Swinney, 1989; MacDonald & MacWhinney, 1990; Nicol & Swinney, 2002, a.o.). Evidence shows that the activation of potential antecedents occurs in compliance with binding constraints (Nicol, 1988; Clifton et al., 1997; Sturt, 2003). This means that pronouns do not reactivate NPs contained within their binding domain.

Other information has been found to affect pronoun antecedent assignment and the time course of resolution. Research shows that information richness (Karimi & Ferreira, 2016) and implicit causality play a role in pronoun interpretation during integration (Stewart et al., 2000; Järviö et al., 2017; Weatherford & Arnold, 2021). This suggests that predictability influences coreference processing.

Though some results indicate that semantic information is utilized as soon as an attempt at pronoun resolution is made, other research demonstrates that syntactic information is vital to this process. There is evidence that number and grammatical gender are accessed during on-line processing (Ehrlich & Rayner, 1983; Arnold, 1998; Arnold et al., 2000; Aoshima et al., 2009). It seems that the ability of the parser

to integrate a variety of information at first pass leads to the delay of pronoun resolution (Ehrlich & Rayner, 1983; Stewart et al., 2007). Arnold et al. (2000) suggests that this supports a dynamic model of language processing where information contributes to probabilistic pronoun resolution. This is similar to the combined theory of anaphora and pronoun interpretation proposed by Kehler & Rohde (2013).

The exact role of all linguistic elements in the process of pronoun resolution is still unknown. There have been various suggestions as to what influences this process in the literature throughout the years (See Cowles et al., 2007, for an overview). Though preferences such as those for antecedents with the same thematic role or grammatical relation may affect pronoun processing, this is not something we explore in this article. One interesting hypothesis that could affect our empirical findings is posited by Nicol & Swinney (2002). They suggest that the occurrence of a pronoun only triggers immediate reactivation of a set of potential antecedents when presented auditorily. Though they mention that this might be an issue of participant engagement rather than presentation method. Reading requires active participation in the task.

Previous research demonstrates the importance of pinpointing and understanding the factors that affect coreference processing. For our study, we attempt to control as many of these factors as possible by limiting outside information such as context. Prior knowledge concerning the NPs included in the sentence stimuli is considered as well. We focus primarily on investigating morphogrammatical properties and how the parser incorporates them during processing and interpretation.

1.4 Control processing

Nicol & Swinney (1989) present a compilation of research on English investigating the reactivation of various referent-dependent items during sentence processing. This research examines the time course of coreference processing that occurs in English in relation to *wh*-traces, NP-traces, overt anaphors, pronouns, and PRO. The experimental designs are motivated by the Trace Reactivation Hypothesis which states that a reference-dependent item causes reactivation of its antecedent NP. They discuss studies that focus on cross-modal priming, an online psycholinguistic method developed by Swinney (1979), that measures the activation of lexical and syntactic information as participants read or listen to sentences in real-time. They find that syntactic constraints play a key role in restricting the reactivation of potential antecedents. When the grammar specifies a *unique* antecedent then *only* that antecedent is immediately reactivated at the position of the referent-dependent item; however, they found that pronouns and PRO reactivate *multiple* potential antecedents. Nicol and Swinney hypothesize that the final antecedent selection occurs later for pronouns and PRO than for *wh*- and NP traces because more information is typically needed for the parser to make the correct selection.

Parker et al. (2015) investigated the processing of adjunct Control dependencies in English by testing interference effects during memory retrieval. Prior research on interference effects suggested that anaphors subject to structural constraints are immune to interference (Sturt, 2003; Xiang et al., 2009; Dillon et al., 2013). To examine this claim, Parker et al. (2015) ran a self-paced reading experiment

to explore whether retrieval of null subject licensing in English was susceptible to interference from distractors in structurally inappropriate locations. They focused on measuring a memory retrieval error called facilitatory interference, which arises when a structurally inappropriate but feature matching item facilitates the processing of an ill-formed linguistic dependency. Their results showed that PRO in adjunct Control (though they do not use this specific term for the name of the null pronoun) was subject to facilitatory interference at two sentence positions, at the gerundive verb and the reflexive. See (5) for an example of two grammatical adjunct Control conditions with these positions marked. This suggests that adjunct Control dependencies are indeed susceptible to facilitatory interference, contrary to previous claims. This implies that PRO might also be vulnerable to other types of interference effects.

- (5) (a) **Grammatical, distractor:** The doctor that the *researcher* described meticulously was certified after debunking the urban myth himself in the new scientific journal.
- (b) **Grammatical, no distractor:** The doctor that the *report* described meticulously was certified after debunking the urban myth himself in the new scientific journal.

Witzel & Witzel (2016) examined whether antecedents are assigned in Japanese Control sentences before verb information becomes available and whether there are biases that drive the process. They used a lexicality maze (L-maze) task to test preverbal analysis of empty subjects. In maze tasks, participants are presented with two alternative continuations of a sentence and must select the correct choice between the two. The L-maze task is similar to that of G-maze, discussed further in Sections 1.4 and 2, but it uses decisions between a real word and a legal nonword. The results showed an Object Control bias. The main-clause object was initially assigned as the antecedent of PRO before reanalysis. It is suggested that this might be due to a preference for the most recent antecedent.

Kwon & Sturt (2016) conducted an eye-tracking experiment with a gender match/mismatch paradigm to study the processing of nominal Control in English. The goal was to investigate the initial access and use of nominal Control information during on-line processing. Participants read sentence stimuli from a screen while their eye movements were tracked. The results supported a recency preference for the processing of nominal Control and the claim that Control information constrains dependency formation during on-line interpretation.

Green et al. (2020) performed an eye-tracking study on adjunct Control in English. These constructions differ from complement Control because they contain a participial clause that is not selected by any element of the matrix clause. The study used an onset-contingent analysis to examine how quickly reference was resolved for PRO and overt pronoun conditions. They found no difference in resolution speed between PRO and pronoun conditions at the cue position (non-finite verb or pronoun). They claim that this means that listeners use structural information as quickly as morphological features when searching for an antecedent.

Delgado et al. (2021) focus their research on the processing of object and subject relatives in comparison to Object and Subject Control in European Portuguese.

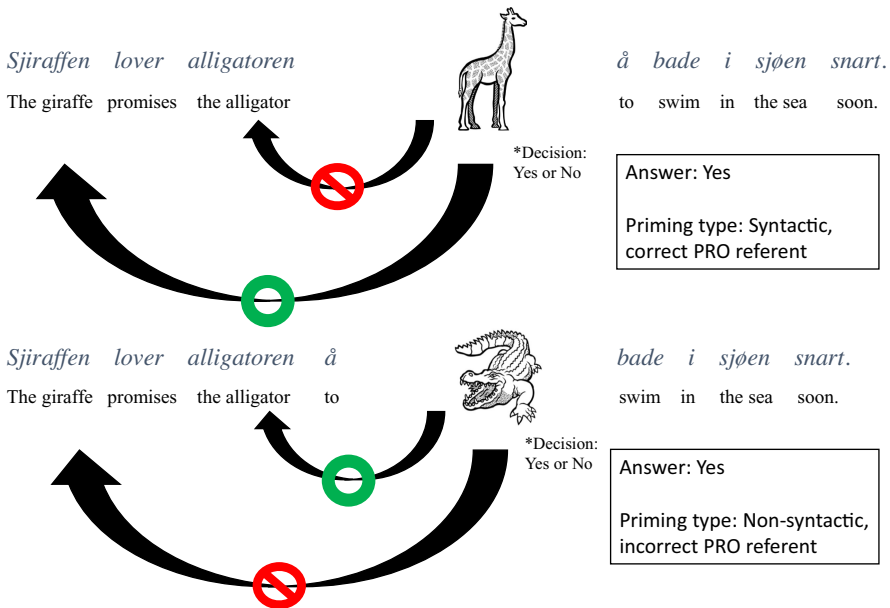


Figure 1. An example of the two priming types (baseline is not included) and the two test positions in the study done by Larsen & Johansson (2020).

They designed a self-paced reading task with a moving-window display. Participants were asked to complete additional tasks during the study to measure resistance to interference, lexical knowledge, working memory capacity, and lexical access ability. They found that object relatives were harder to process than subject relatives. This was not the case for Object and Subject Control. They state that this is evidence against movement accounts of Control, where PRO is treated similarly to NP traces (Hornstein, 1999).

Our earlier research on Control includes two reaction time experiments, using a word-to-image priming design (Larsen & Johansson, 2020). Participants were asked to read a sentence word by word and, when presented with an image, to decide whether or not that image had been presented previously in the sentence in word form. We predicted that the presentation of the correct antecedent in an image format at the position of PRO would facilitate a quicker response. Figure 1 illustrates two priming conditions and the two test positions, directly before and after the infinitive marker. We detected significant facilitation effects for all potential PRO antecedents in Subject Control sentences after the infinitive marker in the non-finite clause in Norwegian. These results were obtained in comparison to a baseline taken from responses to the condition in which the other mentioned, but incorrect, image of a potential antecedent was shown to the participant. This isolates the effect of reactivation from the effect of being mentioned previously, leaving a measure of *purely syntactic reactivation*.

The research that has been conducted thus far on Control and non-finite clauses in various languages suggests that PRO is processed differently than other referent-dependent items, such as NP traces and pronouns. It also shows that at least some

types of Control are susceptible to interference during sentence processing and that this can be investigated using self-paced reading and maze tasks. Eye tracking has been used to investigate what type of information is available during on-line processing and when it becomes available, but more research needs to be done in this area. There is evidence that suggests Control is processed differently across languages as well. This could have repercussions for universal theories of Control.

1.5 Our current research: Hypotheses and predictions

In this article, we present our most recent research on the processing of Control in Norwegian. We opted to use a grammatical-maze task (G-maze) as an alternative to a moving window self-paced reading task. This means that the participant reads each sentence word by word; however, the participant is presented with a choice between two different words as possible sentence continuations. The participant was asked to choose the grammatical alternative from each pair of alternatives in order to continue the sentence. One of the alternatives was obviously ungrammatical given the previous context. For a more detailed description of the experimental design, see the methods section. Methodologically, we wanted to test the maze design's sensitivity to anaphoric dependencies and its ability to detect them via interference between mentioned, ungrammatical alternatives and correct alternatives. If the antecedent was not activated the decision would be easy. Therefore, it was important to establish a baseline using similar NPs that were not mentioned in the sentence.

The second aim of this experiment was also to make close comparisons between two sentence construction types. We investigated non-finite Control clauses and the position of antecedent reactivation during on-line processing. We also investigated what we call their *pronoun counterparts*, containing a finite complement clause and an overt pronoun instead of PRO. These sentence types can be matched down to a minimal distinction between an infinitive marker and a complementizer prior to the first test position. An example of one of these comparisons is included in Example 1.5. Example 1.5 is a Subject Control sentence and (6-b) is its pronoun counterpart.

- (6) a. [Den snille apekatten]_i tilbyr [elefanten] å PRO_i bære [posen]
 [The kind monkey]_i offers [the.elephant] to PRO_i carry [the.bag]
 hjem.]].
 home.
- b. [Den raske leoparden]_i tilbyr [elefanten] at han_i betaler for
 [The fast leopard]_i offers [the.elephant] that he_i pays for
 [taxien].
 [the.taxi].

We predict that if there is an interference effect for either condition or test position, we should expect slower reaction times in comparison to an unrelated baseline. We have two test positions for each sentence, but only one of these contains a referent-dependent item. We predict that an effect will occur when we have competition with an antecedent (see Section 2.3). This can be expected based on the G-maze design.

If there is immediate reactivation of the grammatically correct antecedent of PRO and pronoun, as proposed by the Trace Reactivation Hypothesis

(See Nicol & Swinney, 1989; Sekerina, 2019), we should expect interference effects and longer reaction times for only the correct antecedent of PRO in competition with the infinitive marker. We should also see interference effects and longer reaction times for the correct antecedent of pronouns in competition with the pronoun in the complement clause. If the coreference processing of PRO and pronouns in Norwegian patterns similarly to English, interference effects and longer reaction times should be recorded for all potential antecedents in both Control and Pronoun conditions compared to an unrelated baseline (Nicol & Swinney, 1989). Differences in interference effects and reaction time length across test positions for Control and Pronoun conditions should be obtained if PRO and pronouns undergo antecedent assignment differently.

2. Methods

The selection of the G-maze design was motivated by our previous work on Control (Larsen & Johansson, 2020). Since our results using a priming design showed that antecedent reactivation was facilitated, we wanted to explore the possibility that the dependencies involved in coreference assignment could undergo interference, inhibiting reactivation, if the antecedent was activated in competition with the correct continuation (for example the infinitive marker). If there is no relation the choice should be easy, but if there is a pre-activated word it will be harder to ignore it and choose the correct continuation. This G-maze design forces the participant into an incremental processing mode in which all information must be processed before moving on to the next word in the sentence (Freedman & Forster, 1985; Forster et al., 2009). We recognized that this method can be used to investigate coreference processing. It enables us to more accurately analyze the timeline of antecedent (re)activation and assignment during on-line processing. Due to the task demands, we can assume that all context up to the test position(s) has been processed if the participant can make the correct choices (Witzel & Witzel, 2016). This is not always possible with other self-paced reading designs.

2.1 Participants

A total of 56 participants voluntarily took part in the experiment and signed consent forms prior to the start of the experiment. The consent forms were not connected in any way to the data collected from the experiment. This was done in order to maintain anonymity. An experimenter was present outside of the testing room at all times so that the participants were able to ask questions before and after the experiment. Participants were also given notice that they had the right to leave the room and end their participation in the experiment at any time for any reason.

All the participants were native speakers of Norwegian and residents of the city of Bergen in Norway. Participants received a movie ticket as compensation for their participation. Table 1 shows the self-reported statistics concerning gender, age, education, and written language preference of the participants. The selected preference did not affect the language of the experiment (Norwegian - bokmål), but it was included as a factor in the data analysis.

Table 1. The 56 participants described by gender, age, education, and written language preference. NA means that the participant(s) declined to give an answer

Gender	Age	Education	Written Language Preference
Male: 22	18-25: 38	High School: 11	Bokmål: 47
Female: 33	26-35: 12	Bachelors: 21	Nynorsk: 8
NA: 1	36-45: 1	Masters: 23	NA: 1
	46+: 5	NA: 1	

2.2 Experimental conditions

The design of the experiment contained four test sentence constructions: Object Control (OC) and Subject Control (SC) sentences and a pronoun counterpart of each Control type (Object pronoun (OPN) and Subject pronoun (SPN)). We used three test sentences of each type for a total of 12 test sentences. Each test sentence was presented in six different conditions. They were then repeated twice in each condition with the correct sentence continuation appearing once as the top selection and once as the bottom, making in total 144 unique test presentations ($12 \times 6 \times 2$). There were three levels of type of activation in relation to PRO (or pronoun): *mentioned*, *intended* (correct antecedent), and *unrelated*. Each of these sentence conditions were presented four times. Two independent presentations showed the correct answer either on the top or on the bottom. This was done for both positions.

There were also 12 unrelated filler sentences, which were presented once each. The participants had to make a decision on every word, so adding more fillers would have made the experiment much longer. This leads to increased fatigue, while not providing any more data. The main goal was to reduce the impact of the marked constructions, which were assumed to be the Control sentences. The Control and pronoun sentences were also considered different constructions. We assumed one would not facilitate the other.

In addition to the filler sentences, each participant was presented with 6 training sentences at the beginning of the experiment in order to get them familiar and comfortable with the task. A total of 162 sentence presentations were shown to each participant, with 144 of those being test conditions. For a full list of the test, filler, and training sentences used, see Appendix A, B, and C.

We used a repeated measures design, where the participants were their own controls. This allows for a more precise estimate of individual variance for each of the conditions. It permits the use of a smaller number of participants since more data can be obtained from each individual. Reaction time studies are often planned as factorial designs where the time of presentation or presentation order are considered handled by randomization (Stowe & Kaan, 2006). The idea is that since conditions are randomly distributed across the presentation order, the main effect is not affected; however, this disregards the effect on the variance.

Baayen & Milin (2010) state that the inclusion of control predictors in the model “not only helps satisfy to a better extent the independence assumption of the linear model, but also contributes to a more precise model with a smaller residual error.

Simply stated, these predictors allow a more precise estimation of the contributions of the other, theoretically more interesting, predictors.” There is no reason that the first presentations would be more accurate for the task at hand. Rather, on the contrary, the participants are initially less confident in their decisions (as mirrored in the slower reaction times). The solution to include more participants rather than presenting more items would thus not necessarily mean more accurate measurements of the phenomenon of interest

Even so, the learning effect is rarely handled using the statistical model. Baayen & Milin (2010) is an exception, and we have not seen it used previously for studies of linguistic coreference. This could be because some researchers may hesitate to add more factors for fear of making it harder for the model to converge since more parameters have to be estimated. Nonetheless, additional variables, such as those that control the effect of learning, often make it easier for the model to converge. Variables that control effects allow the model to account for more of the variance.

To account for the possible influence of a learning effect, we included the order of individual word presentation as a factor in our statistical analysis. Including the order of word presentation is an effective way to account for individual variance, across the entire experiment as well as the sentence stimuli. Additionally, this allows all conditions to be presented to each participant since we do not rely only on the randomization of sentence stimuli.

When constructing the sentence stimuli, we had to consider the role that the distance between a context-dependent item (pronoun, PRO, or trace) and its antecedent can play in regards to storage and computational demands. Previous studies have shown that the shorter the distance, the more readily available the antecedent is in working memory (Garrod & Sanford, 1977; Clark & Sengul, 1979; Ehrlich, 1980; Ehrlich & Rayner, 1983; Blanchard, 1987). For example, evidence has been found that object relatives are more difficult to process than subject relatives (e.g. Holmes, 1973; Wanner & Maratsos, 1978; King & Just, 1991). One possibility proposed by these studies is that this increase in difficulty is due to the predictions in object relatives being maintained across a greater linear distance in comparison to subject relatives in which the gap predictions are satisfied immediately. Others have also suggested that the distance between predictions and their fulfillment leads to greater storage demands and, ultimately, greater processing difficulty (Kaan & Stowe, 2002). With this in mind, we used modifying clauses throughout our test sentence stimuli in order to maintain a consistent linear distance between the PRO NP antecedent creating these predictions and PRO itself, where the predictions should be satisfied. This does not change the hierarchical distance between the two items, but it does provide the parser with additional units to process before encountering PRO in Object Control constructions.

Each of the 12 test sentences had three ungrammatical alternative templates. The top three conditions in Table 2 provide three examples of these templates. They were used twice in the six conditions so that the ungrammatical choices did not become obvious to the participants over time. The sentence-initial ungrammatical alternative was always the infinitive form of a verb. The correct and incorrect alternatives were presented equally as top and bottom choices across the experiment, but we show the correct choices on top in these examples so that it is easier to distinguish between the two.

Table 2. “The kind monkey offers [the elephant] to carry [the bag] home.” An example of a Subject Control sentence ((7-a)) appearing in the six different experimental conditions. The critical words in each condition are highlighted. The top three conditions show examples of ungrammatical alternatives. ‘****’ is used to represent the ungrammatical alternatives in the other conditions. The predicted points of interference are highlighted

Condition	One:	Intended,			Zero			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
Aktivere	oppsett	kortet	sko	vurdere	apekatt	kjøttdeig	rope	medium.
Condition	Two:	Intended,			One			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
Stave	gester	overskrift	resepsjon	rikke	influert	apekatt	rugby	apati.
Condition	Three:	Mentioned,			Zero			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
Separere	likheter	regninger	seieren	galakse	elefant	suiten	gispa	sekvens.
Condition	Four:	Mentioned,			One			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
****	***	***	***	***	***	elefant	***	***.
Condition	Five:	Unrelated,			Zero			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
****	***	***	***	***	isbjørn	***	***	***.
Condition	Six:	Unrelated,			One			
Den	snille	apekatten	tilbyr	elefanten	å	bære	posen	hjem.
****	***	***	***	***	***	isbjørn	***	***.

Example 2.2. provides a Subject Control sentence from the study. Table 2 shows this sentence in each of the six conditions (three levels of *activation* and two levels of *position*). A similar table is also included for Example 2.2, a Subject Pronoun sentence from the study (Table 3). A line of three asterisks is used to represent an ungrammatical alternative. It should be obvious to native speakers of Norwegian that the root form of a noun is ungrammatical at all test positions, even though the definite form of a noun could be grammatical in the pronoun conditions in position one.

- (7) a. [Den snille apekatten]_i tilbyr [elefanten] å PRO_i bære [posen]
 [The kind monkey]_i offers [the.elephant] to PRO_i carry [the.bag]
 hjem.]].
 home.
- b. [Den raske leoparden]_i tilbyr [elefanten] at han_i betaler for
 [The fast leopard]_i offers [the.elephant] that he_i pays for
 [taxien].
 [the.taxi].

Table 3. “The fast leopard offers [the elephant] that he pay for [the taxi].” An example of a Subject Pronoun sentence appearing in the six different experimental conditions. The predicted points of interference are highlighted. ‘****’ is used to represent the ungrammatical alternatives

Condition	One:	Intended,			Zero				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	leopard	***	***	***	***.
Condition	Two:	Intended,			One				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	***	leopard	***	***	***.
Condition	Three:	Mentioned,			Zero				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	elefant	***	***	***	***.
Condition	Four:	Mentioned,			One				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	***	elefant	***	***	***.
Condition	Five:	Unrelated,			Zero				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	grevling	***	***	***	***.
Condition	Six:	Unrelated,			One				
Den	raske	leoparden	tilbyr	elefanten	at	han	betaler	for	taxien.
****	***	***	***	***	***	grevling	***	***	***.

Another factor that was controlled for during the design phase was NP animacy features since they have been found to influence antecedent assignment (Bickel et al. 2015a; Bickel et al. 2015b; Bornkessel-Schlesewsky & Schlewsky, 2009; Dahl & Fraurud, 1996; Larsen & Johansson, 2007, 2009; Primus, 2006, a.o.). The sentences were designed so that each one contained two potential NP antecedents. Considering the fact that animacy can affect coreference resolution in a variety of ways, our study was purposefully designed so that animacy was not a factor. All potential NP antecedents were animate and non-local, exotic animals. We used uncommon animal NPs that the participants presumably did not have daily interactions with so that they would not be able to use familiarity, world knowledge, or episodic memory to aid their decisions. The antecedents were matched for number of syllables and length. There were six potential antecedent NPs and three unrelated NPs used in the study. All pronouns were “han”(he), and all the animals had no obvious gender. This strategy for NP selection prevented gender and semantic associations from creating participant biases.

Finally, anthropomorphized stories are easy to process and often used in research on children. We assumed that adults would have no problem interpreting the sentences, and none of our participants expressed any such concerns in debriefings.

2.3 Predicting relevant positions: Zero and One

Our experimental design investigates the position in which inference effects occur during coreference processing. Based on the theoretical discussions concerning the position of the infinitive marker in Norwegian Control constructions, we selected two test positions. We also considered the possibility that the parser needs information from the subordinate clause verb prior to beginning the process of antecedent selection. For an illustration of these test positions, see Tables 2 and 3.

For PRO:

1. Position Zero – in competition with the *infinitive* marker.
2. Position One – in competition with the subordinate clause *verb*.

For pronoun counterparts:

1. Position Zero – in competition with the *complementizer*.
2. Position One – in competition with the *pronoun* of the finite clause.

2.4 Procedure

Participants were asked to sit in a sound-proof recording studio and use a computer to take part in the experiment. The experiment lasted for approximately 15 minutes, depending on the participant. Participant decisions were recorded using a Cedrus RB-540 Response Pad. The participant was asked to begin by reading the instructions on the screen. The first part of the experiment involved four short questions regarding the participant's gender, age, level of education, and written language preference. The participant could choose to either answer the questions or to not respond.

The second part of the experiment began with an additional set of instructions. The participant was told that they would be asked a couple of comprehension questions at the end of the experiment. This ensured their attention would be held throughout the duration of the task. We asked the participant to read the sentences presented on the screen word by word. Each time, they had a choice between a grammatical or ungrammatical continuation. The participant was asked to choose the correct grammatical continuation of the sentence. The possible sentence continuations were presented with a top and a bottom alternative such that the participant could press the marked top or bottom button to select the correct continuation. The screen position of the alternatives was balanced with an equal number of correct top and bottom alternatives.

The sentences were displayed linearly across the computer screen. We first presented the participant with a series of training sentences. Participants were not informed that the first six sentence presentations were for training purposes. If the participant selected the incorrect answer, they were shown the word *Feil* (*Wrong*) on the screen and the sentence began again from the beginning.

After going through the training sentences, the experiment proceeded immediately to the presentation of test and filler sentences. There was no feedback after the

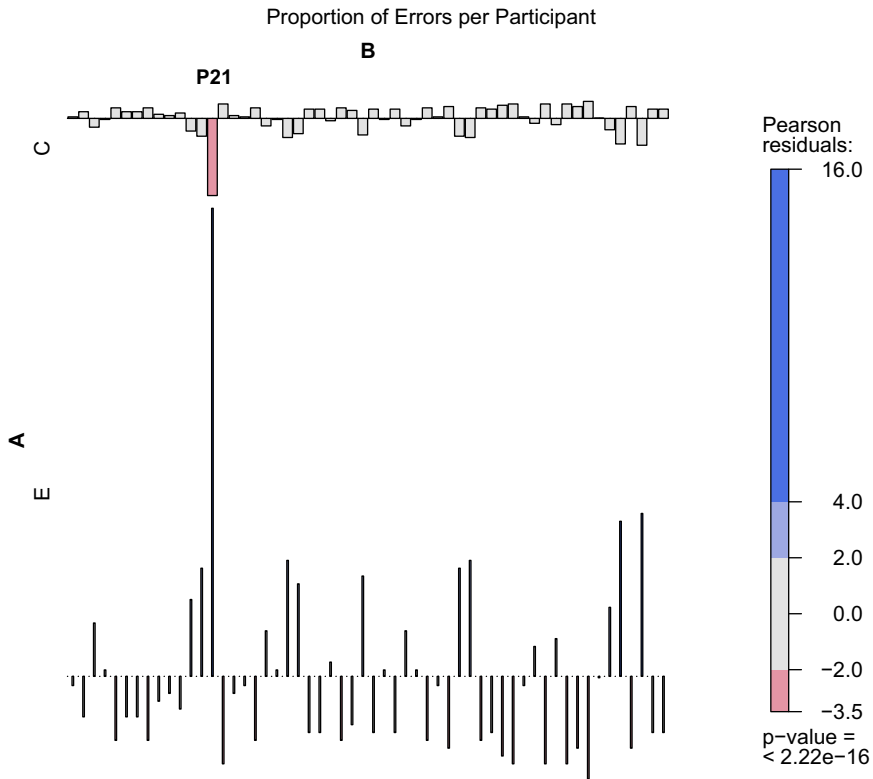


Figure 2. Error analysis per participant. Participant P21 has a significantly higher error rate.

training sentences. The participant was presented with a set of two asterisks at the beginning of each sentence in order to orient themselves and direct their attention to the appropriate point on the screen.

3 Results

3.1 Outlier analysis

The analysis was completed in R (R Core Team, 2021). The results were calculated using data from 55 of the 56 original participants. The overall data showed a high level of task compliance; however, one participant (P21) showed significantly more errors than the others, with only 86.2% (462/536) correct. This is illustrated in an association plot in Figure 2 that was created using the *vcd* package in R (Meyer et al., 2006; Zeileis et al., 2007; Meyer et al., 2021). Therefore, we chose to exclude all data recorded for this participant from the data analysis, since this participant may have solved the task differently.

Our data contained a few values for RT that were significantly longer than the average. To prevent extreme outliers from driving effects in the data, we set upper and lower RT boundaries based on the findings of a comparative study on RT outlier

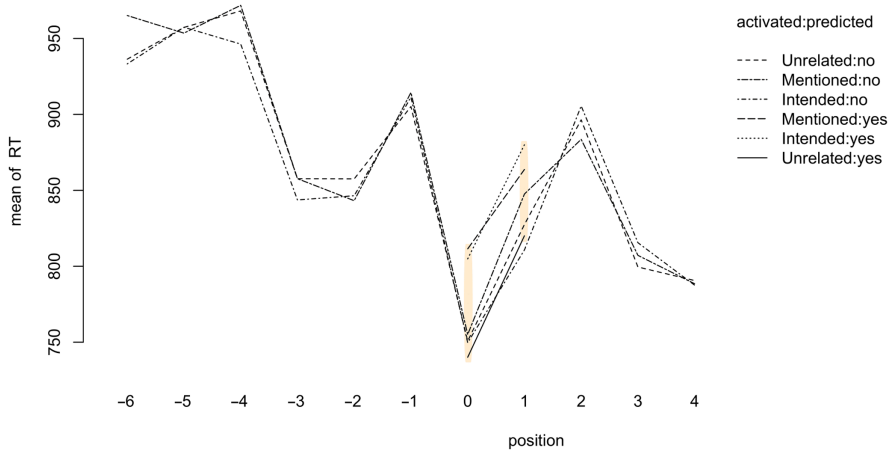


Figure 3. Interaction Plot showing the mean RT for the levels of *activated* and *predicted* across *position*. The shaded areas mark the difference between the levels of activated in the test positions.

exclusion methods. Berger & Kiefer (2021) found that methods based on standard deviations (SDs) and z-scores showed small (absolute) biases, few Type-I errors, and excluded only small proportions of RTs. Therefore, the upper boundary was set at 1928ms, which was the mean of RT plus 2 SDs. The lower boundary could not be a negative value, so it could not be set based on SDs.

We still needed to account for the time it took for a participant to encode and react, so the lower boundary was set at 200ms (Ashby & Townsend, 1980; Luce, 1991; Whelan, 2008). These boundaries kept 96% of the data and about 139 out of 144 data points per participant.

3.2 Overall effects for all sentence types and test positions

Figure 3 provides an illustration of the changes in mean reaction time across sentence position. Figure 4, created using the *ggplot2* (Wickham, 2016), *dplyr* (Wickham et al., 2021), and *tidyr* (Wickham, 2021) packages in R, focuses on the mean reaction time from two words before the infinitive marker/complementizer up until 2 words after. For both Figure 3 and 4, the x-axis represents sentential positions, in words, relative to the infinitive marker or the complementizer (position 0). For test position zero, we expect interference at position 0, and for test position one, we expect it at position 1. Interference can be observed as a spike in slower reaction times.

The y-axis shows the mean reaction time. The overall processing of the sentences shows that potential antecedents (Mentioned and Intended) stand out, creating a visible interference effect (Figure 3) that we will analyze in more detail. The typical sentence processing signatures are present with higher reaction times for the first (unpredictable) item of a sentence or clause (position -6 or -5, depending on sentence length). The contrast between unrelated and potential antecedents is highlighted by a shading in the graph. The two lines that stand out at position 0

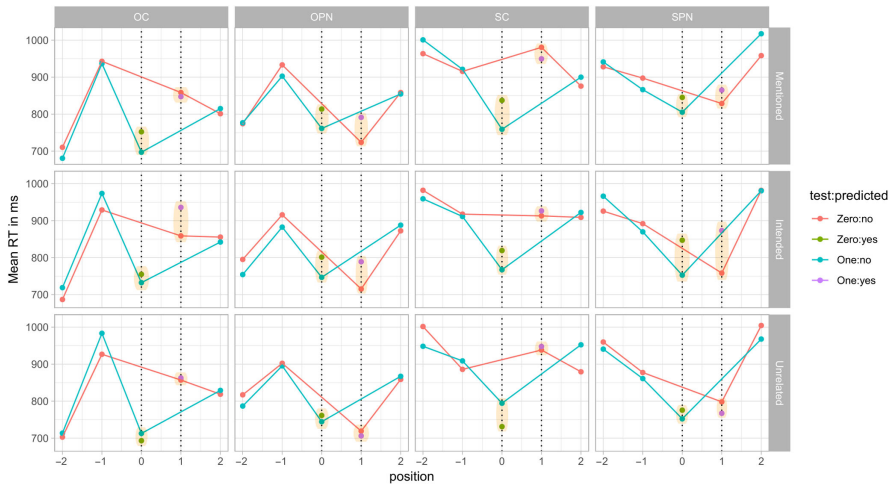


Figure 4. The mean RT for each sentence type for each level of *activated*. A vertical dotted line marks the test positions. The position being tested and whether or not an effect is predicted is displayed in the legend. Interference effects are noted as a positive difference between unrelated and the other two levels of *activated*, as indicated by the shaded regions.

and 1 are the lines for potential antecedents (mentioned or intended). This provides additional evidence that the experimental design is capable of capturing the intended interference effects at our test positions.

Figure 4 gives the mean RT of the levels of *activated* and the levels of *type* across sentence positions, after honing in on the immediate context. It shows interference effects as peaks at either position 1 (see x-axis) if the test position is one or position 0 if the test position is zero. No effects were predicted at any positions other than 0 and 1. Shaded ovals are used to highlight the differences between when an effect was predicted or not in one of the test positions. The most important differences lie in the comparison of mentioned and intended to the unrelated baseline condition. The mean reaction time for SC and OC when the test position was zero is much larger than that for unrelated. The same trend occurs for SPN and OPN but in test position one.

To get an idea of which factors and conditions should be looked at more in depth, we ran an Analysis of Variance (ANOVA) on a linear mixed effects model. This was done using the *lmerTest* package in R (Kuznetsova et al., 2017).

The following factors were included in our initial model:

- **Predicted** (yes, no)
 - Presentations in which the position 0 or 1 was indeed the *test* position and an interference effect was predicted. Both positions occur in all sentence presentations, but they cannot both be the test position in the same presentation.
- **Type** (Object Control, Subject Control, Object Pronoun, and Subject Pronoun)
- **Activated** (intended, mentioned, unrelated)

- **Position** (0/zero, 1/one)
- **Order**
 - This represents the unique order of presentation for all the items for each participant.
- **Language** (Bokmål, Nynorsk, NA)
- **Gender** (Man, Woman, NA)
- **Education** (High School, Bachelors, Masters, NA)
- **Age** (18-25, 26-35, 36-45, 46+)

Participant and *item* (each individual word) were included as random effects with a random intercept. The items were encoded so that each item was unique per condition, in total 144 items. Recall that the participant is supposed to choose the non-antecedent, so the item is the (unique) combination of the two choices in that situation. Therefore, there was no possibility to add slopes, as we do not have repetitions for these items. Gender, Education, and Language did not show any significant effects. Thus we removed them from our model. The complete ANOVA table (Table 11) and summary for the model (Table 10) are shown in Appendix D. The mixed effects model formula used to obtain the variance for the ANOVA is **lmer(RT predicted * type * activated * position + order + age + (1 | participant) + (1 | item))**. The factors *age* and *order* are used to regress out the effect of age and presentation order on the reaction times. The main factors are: *predicted* (yes, no), *type* (OC, SC, OPN, SPN), *activated* (mentioned, intended, unrelated), *position* (0/zero, 1/one). For *activated*, it is possible to think of mentioned and intended as potential antecedents and unrelated as a non-antecedent.

Table 11 shows significant effects for *predicted*, *type*, *activated*, *position*, *order*, and *age*. There are significant interaction effects between *predicted* and *activated*, and between *type* and *position*. Since the ANOVA shows a highly significant effect ($p = 0.0008$) of *predicted*, we concluded that our design is effective in provoking effects in the selected test positions. We then focused on the *yes* level of *predicted*.

Our final model was made using this formula: **lmer(RT type * activated * position + order + age + (1 | participant) + (1 | item))**. The significant factors remain the same as in our previous model; however, removing *predicted* as a factor allows the model to better calculate the effect of *activated* since there is no longer an interaction effect involved. These effects will be discussed in detail in the following subsections. Estimated Marginal Means (EMMs) from the *emmeans* package in R (Lenth, 2021) will be used as post-hoc results for the factors with more than two levels. This post-hoc function uses the variance estimated by the mixed effects model.

3.3 Activated

The significant effect for *activated* shows an observed difference between the unrelated, mentioned, and intended levels ($p = 0.0009$). Table 5 includes the EMMs for our post-hoc investigation within the factor *activated*. The unrelated level is 62.89ms faster than the intended ($p = 0.0054$) and 53.34ms faster than the mentioned ($p = 0.0007$). These differences do not take into account other experimental factors though.

Table 4. Type III ANOVA table for the final model. Degrees of freedom within conditions (df_W) are estimated using Satterthwaite's method

Factor	SS	MSS	df_B	df_W	F	Pr(> F)	
type	1764314	588105	3	61.4	8.7430	0.0000	***
activated	1048188	524094	2	60.5	7.7914	0.0010	***
position	1700729	1700729	1	60.5	25.2838	0.0000	***
order	9902666	9902666	1	3696.0	147.2176	0.0000	***
age	664191	221397	3	50.8	3.2914	0.0279	*
type:activated	198014	33002	6	61.4	0.4906	0.8129	
type:position	2262501	754167	3	61.4	11.2118	0.0000	***
activated:position	28373	14186	2	60.5	0.2109	0.8104	
type:activated:position	327206	54534	6	61.4	0.8107	0.5656	

Table 5. Pairwise model comparisons for the levels of *activated*

contrast	estimate	SE	df	asympt.LCL	asympt.UCL	z.ratio	p.value
Mentioned - Intended	-9.57	17.2	Inf	-49.9	30.7	-0.557	0.8431
Mentioned - Unrelated	53.31	17.2	Inf	13.1	93.5	3.105	0.0054 **
Intended - Unrelated	62.89	17.2	Inf	22.6	103.1	3.661	0.0007 ***

3.4 Type

There is a significant difference between the sentence types ($p = 0.0000$). Both types assigning the object as the correct antecedent are faster than their subject-assigning counterparts. EMMs were calculated as a poc-hoc test to determine where the significant differences lie within the levels of the factor *type*. Table 6 presents the results. There is a significant difference between OC and SC ($p = 0.0253$), OPN and SC ($p < 0.0001$), and OPN and SPN ($p = 0.0190$). The main contrast is between types that assign the subject as the antecedent and those that assign object as the antecedent. The interaction between *type* and *position* will affect these relationships as well.

3.5 Position

A significant effect of position ($p = 0.0000$) was detected. Participants made quicker decisions in position zero (826.34ms) in comparison to position one (938.87ms). This is likely because starting a new clause causes an increase in cognitive load at the first content word. Since *position* has a significant interaction effect with *type*, this relationship between position zero and one may be related to a larger interference effect.

Table 6. Pairwise model comparisons between types of sentences

contrast	estimate	SE	df	asympt.LCL	asympt.UCL	z.ratio	p.value
OC - OPN	36.4	18.9	Inf	-12.1	84.82	1.928	0.2163
OC - SC	-58.2	20.7	Inf	-111.4	-5.06	-2.813	0.0253 *
OC - SPN	-18.7	20.7	Inf	-72.0	34.57	-0.902	0.8039
OPN - SC	-94.6	18.9	Inf	-143.1	-46.08	-5.008	<0.0001 ***
OPN - SPN	-55.1	18.9	Inf	-103.7	-6.44	-2.909	0.0190 *
SC - SPN	39.5	20.8	Inf	-13.8	92.90	1.904	0.2261

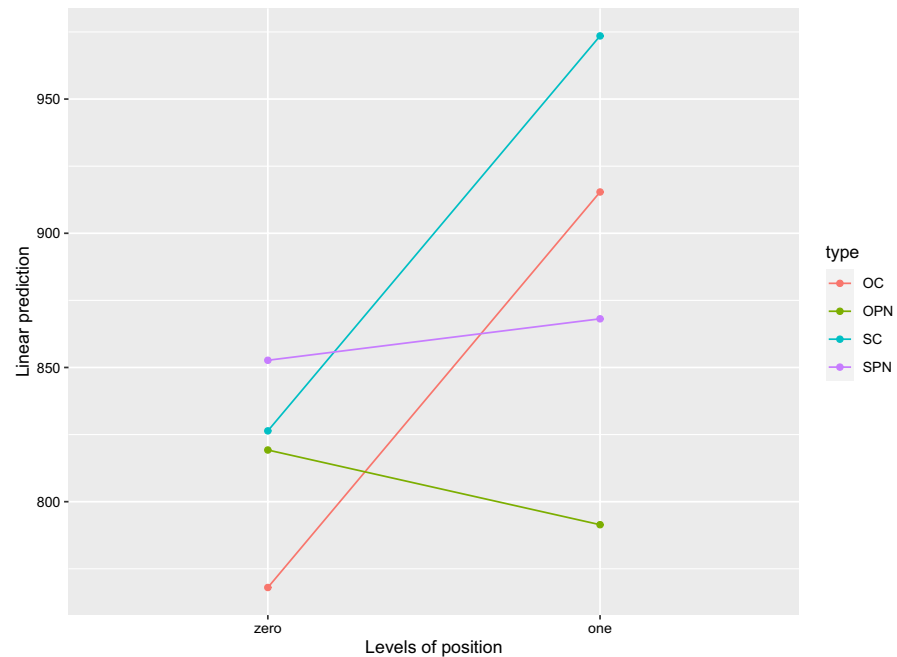


Figure 5. The model prediction of difference in estimated marginal means of RT per sentence type and position. Position one for SC and OC involves competition with a verb.

3.6 Interaction of type and position

The factors of *type* and *position* are significant main effects. There is also a significant interaction between the two ($p = 0.0000$, Figure 5). Table 7 lists the results of the post-hoc EMMs for the differences between all combinations of *type* and *position*. The important comparisons to consider are those between the levels of *type* for the same level of *position* and those between the levels of *position* for the same level of *type*.

In position one, OC is 123.95ms slower than its pronoun counterpart ($p < 0.0001$) and SC is 105.40ms slower than its pronoun counterpart ($p = 0.0080$). There are no significant differences between any of the sentence types at position

Table 7. Pairwise model comparisons of *type* and *position*. The interesting comparisons are highlighted

contrast	estimate	SE	df	asympt.LCL	asympt.UCL	z.ratio	p.value	
OC one - OPN one	123.95	26.7	Inf	43.123	204.78	4.648	0.0001	***
OC one - SC one	-58.15	29.3	Inf	-147.004	30.71	-1.983	0.4931	
OC one - SPN one	47.25	29.3	Inf	-41.651	136.15	1.611	0.7440	
OC one - OC zero	147.34	29.2	Inf	58.717	235.97	5.039	<0.0001	***
OC one - OPN zero	96.12	26.7	Inf	15.098	177.14	3.596	0.0078	**
OC one - SC zero	89.00	29.3	Inf	0.208	177.78	3.038	0.0490	*
OC one - SPN zero	62.69	29.4	Inf	-26.329	151.71	2.134	0.3925	
OPN one - SC one	-182.10	26.7	Inf	-263.013	-101.18	-6.821	<0.0001	***
OPN one - SPN one	-76.70	26.7	Inf	-157.671	4.27	-2.871	0.0785	.
OPN one - OC zero	23.39	26.6	Inf	-57.269	104.06	0.879	0.9880	
OPN one - OPN zero	-27.83	23.8	Inf	-100.056	44.40	-1.168	0.9410	
OPN one - SC zero	-34.95	26.7	Inf	-115.800	45.89	-1.310	0.8953	
OPN one - SPN zero	-61.26	26.8	Inf	-142.361	19.85	-2.289	0.2994	
SC one - SPN one	105.40	29.4	Inf	16.414	194.38	3.590	0.0080	**
SC one - OC zero	205.49	29.3	Inf	116.790	294.19	7.022	<0.0001	***
SC one - OPN zero	154.27	26.8	Inf	73.158	235.38	5.765	<0.0001	***
SC one - SC zero	147.14	29.3	Inf	58.272	236.02	5.018	<0.0001	***
SC one - SPN zero	120.84	29.4	Inf	31.734	209.95	4.110	0.0010	**
SPN one - OC zero	100.10	29.3	Inf	11.344	188.85	3.418	0.0146	*
SPN one - OPN zero	48.87	26.8	Inf	-32.287	130.03	1.825	0.6028	
SPN one - SC zero	41.75	29.3	Inf	-47.168	130.66	1.423	0.8469	
SPN one - SPN zero	15.45	29.4	Inf	-73.706	104.60	0.525	0.9995	
OC zero - OPN zero	-51.22	26.7	Inf	-132.090	29.64	-1.920	0.5371	
OC zero - SC zero	-58.35	29.2	Inf	-146.998	30.30	-1.995	0.4853	
OC zero - SPN zero	-84.65	29.3	Inf	-173.526	4.23	-2.887	0.0752	.
OPN zero - SC zero	-7.12	26.7	Inf	-88.156	73.91	-0.266	1.0000	
OPN zero - SPN zero	-33.43	26.8	Inf	-114.722	47.87	-1.246	0.9180	
SC zero - SPN zero	-26.30	29.4	Inf	-115.345	62.74	-0.895	0.9866	

zero. SC is 147.14ms slower at position one than it is at position zero ($p < 0.0001$). Similarly, OC is 147.34ms slower at position one than zero ($p < 0.0001$). SPN and OPN show no significant difference in reaction time between test positions.

Figure 5 illustrates the interaction between *type* and *position*. The two lines for SPN and OPN show no significant differences between the levels of position. Both SC and OC are represented with lines that have comparable slopes, which is confirmed by the EMMs.

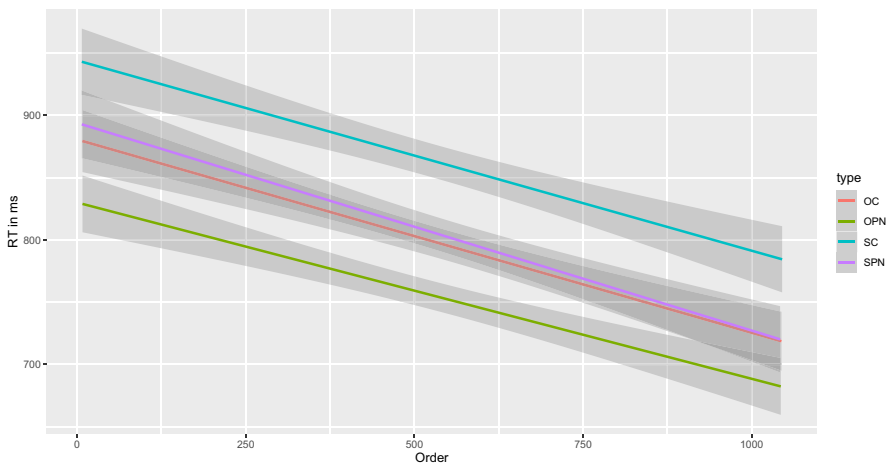


Figure 6. The mean RT for the learning effect for each sentence type over the time course of the experiment. Ellipses of uncertainty are included as shaded areas in the graph.

3.7 Order: The learning effect

Order of presentation showed a significant effect in our model ($p = 0.0000$). With each sentence presentation, the participant's decision time was on average 0.17ms faster. This seems like a small effect, but it adds up and is highly significant. The learning effect due to order of presentation is plotted in Figure 6.

The graph shows a similar negative slope for all sentence types as the experiment progresses; it is just the starting point that is different. Including the presentation order as a factor in the model allows us to regress out the unavoidable learning effect that takes place when a participant is presented with the same, and very similar, stimulus in various conditions. Note that the order of presentation was randomized so that each test item is presented in a different order for each participant.

The participants become more confident in their decisions, resulting in faster reaction times. Regressing out such “learning effects” makes it possible to better account for variance in the main contrasts. The repeated measures design makes it possible to collect data for all conditions from each participant, permitting us to better account for variance due to individual participants. Individual differences related to our conditions are thus accounted for by repeated measures and explicitly regressing out the expected learning effect. Detrending is often necessary in longer experiments, and randomization of items is not enough since the participants change their behavior as they gain experience.

3.8 Age

Age significantly contributed to the RT of participants ($p = 0.0279$). The group of participants who were 46 years of age or older were 187.75ms slower than the youngest age group (18-25) and 189.6ms slower than the 26-35 age group. These significant differences ($p = 0.0111$ and $p = 0.0274$) are apparent in the EMMs in

Table 8. Pairwise model comparisons between age groups

contrast	estimate	SE	df	asympt.LCL	asympt.UCL	z.ratio	p.value
(18-25) - (26-35)	1.85	42.4	Inf	-107	110.9	0.044	1.0000
(18-25) - (36-45)	12.41	129.3	Inf	-320	344.7	0.096	0.9997
(18-25) - (46+)	-187.75	61.0	Inf	-344	-31.2	-3.080	0.0111 *
(26-35) - (36-45)	10.56	132.8	Inf	-331	351.8	0.080	0.9998
(26-35) - (46+)	-189.60	68.1	Inf	-364	-14.7	-2.785	0.0274 *
(36-45) - (46+)	-200.17	139.9	Inf	-559	159.2	-1.431	0.4798

Table 8 as well. We chose to regress out this effect of age rather than remove the few older participants.

3.9 Control

Since no difference between mentioned and intended antecedents was found in the EMMs of our overall model, we chose to combine the mentioned and intended levels of *activated* to take a closer look at the levels of Control. The data set was subsetting such that only SC and OC remained for sentence types. We ran a linear mixed effects model with the same formula as our main model. Like in the results from the main model, the ANOVA reported a significant effect of *type*, *order*, and *position*. Activated trends towards significance ($p = 0.0502$). Both types of Control presented with slower reaction times in position one in comparison to position zero ($p \leq 0.0001$), which is an additional effect after the main effects are accounted for. From the EMMs, we found a trend towards significance for *activated*, but this was only evident for the difference between the potential antecedents and the unrelated NP in position zero ($p = 0.0728$).

3.10 Pronoun

To investigate the levels of the pronoun constructions, we combined the mentioned and intended levels of *activated* and subsetting the data set such that only SPN and OPN remained for sentence types. The same formula as our main model was used to create a new linear mixed effects model. The ANOVA results confirmed the overall significant effect for *type*, *activated*, *order*, and *age*. Unlike the Control conditions, the Pronoun conditions show no effect of *position* and a significant effect of *activated* ($p = 0.0007$). There is a significant difference between potential antecedents and the unrelated NP in position one only (0.0104). We only see a significant difference between levels of *activated* for SPN ($p = 0.0458$). This difference in *activated* trends towards significance for OPN ($p = 0.0726$).

4 Discussion

A significant main effect of *position* was detected. Position zero marks the start of a new clause, but the grammatically correct option is a high-frequency function word.

Therefore, the large, general difference in RT across positions is likely due to the grammatically correct option in position one being a less frequent content word (Schmauder et al., 2000).

There is a significant main effect of *predicted*, which tells us the predicted test positions were effective. There is a significant main effect of *activated*, both potential antecedents are significantly slower compared to the unrelated NP. There is also a main effect of sentence type, which tells us (after looking at specific effects) that sentences involving subject antecedents (SC and to some extent SPN) are slower, indicating greater interference than sentences involving object antecedents. Finally, we found a significant main effect of *activated*, but we detected *no difference* between the two levels of potential antecedents (mentioned and intended). This indicates that both are activated to some extent, which is in line with previous research but in conflict with accounts that hypothesize immediate reactivation of only the correct antecedent of a referent-dependent item. The results are consonant with a dynamic model of language processing, where antecedents are activated based on probabilities estimated by the available context. Antecedents may not always be activated and can potentially always be activated to some degree.

There is one significant interaction effect in the final model, between *type* and *position*. This effect tells us that Control sentences and their pronoun counterparts have different effects related to *position*, so *PRO* and pronouns may use different processing strategies. When we restricted the data set to only data points obtained when an effect was predicted (and tested for), we found support for a difference between the sentence types.

4.1 Control sentences

For Control sentences, we found an additional slowdown in position one relative to position zero for both Object and Subject Control, but there was only a trend towards significance for a main effect of *activated* in position zero. Therefore, the difference between the potential antecedents (intended and mentioned together) and the unrelated NP is a rather weak effect. One possible explanation (or complication) is that the unrelated were similar animals that might be partly activated due to their semantic similarity to the potential antecedents (i.e., they are all names of exotic animals). A stronger contrast might provide a clearer difference between potential antecedents and non-antecedents. This would suggest a (lexical) semantic, or spreading activation, account.

The pattern of interference effects recorded for Control sentences can be interpreted as supporting Nicol & Swinney (1989)'s results. Nicol and Swinney reported processing effects for all potential antecedents in Control conditions. We found a trend towards interference effects for both potential antecedents in competition with the infinitive marker. These results conflict with the Trace Reactivation Hypothesis which presumes *immediate* reactivation of *only* the correct antecedent of *PRO*. This suggests that the parser lacks the complete information needed to assign the correct antecedent at the time that *PRO* is processed. The reactivation of two potential antecedents allows both to be stored in short-term memory for later use. This is also observed for the pronoun counterparts, as both mentioned and intended are associated with interference effects relative to unrelated.

Our results suggest that the antecedent selection process has already *begun* at the infinitive marker. We cannot claim to know when this selection has been finalized, but we can predict the start of this process by combining our current results with those from previous research on Control in Norwegian. Using a priming paradigm, we observed similar activation patterns for potential PRO antecedents (Larsen & Johansson, 2020). Both potential antecedents of PRO showed a facilitation effect directly after the presentation of the infinitive marker. There was no facilitation directly before the infinitive marker. Since no processing effects have been found before the infinitive marker or at subordinate clause verb, we can assume that antecedent selection begins following the processing of the infinitive marker. This suggests that the Norwegian infinitive marker allows for early activation.

The difference that was recorded between the potential antecedents and the unrelated NP was not as strong as it was for the pronoun condition. It is possible that PRO is less susceptible to interference effects in comparison to pronouns. Another reason for this could be that the syntactic and semantic constraints on PRO are not as absolute as those for pronouns. Pronouns can be matched for grammatical gender and other such features, unlike PRO. That would suggest discourse and sentence context play a greater role in PRO antecedent assignment.

4.2 Pronoun sentences

The pronoun sentences did not exhibit a significant main effect of *position*, but we saw a difference between potential antecedents and unrelated NPs specifically in position one, in competition with the pronoun itself. This is in line with results recorded for English pronouns: all structurally appropriate antecedents were reactivated simultaneously in a similar manner at the position of the pronoun (Corbett & Chang, 1983; Nicol, 1988; Nicol & Swinney, 1989; MacDonald & MacWhinney, 1990; Nicol & Swinney, 2002, a.o.). Our results imply that a similar reactivation pattern takes place in Norwegian.

The pronoun sentences show a similar pattern of antecedent reactivation to the Control sentences in that all potential antecedents are reactivated; however, we see that these effects occur in two different sentence positions. This reinforces previous research on English that found that PRO and pronouns undergo antecedent resolution at different points during sentence processing (Nicol & Swinney, 1989). One possible explanation is that the parser has enough information to predict the occurrence of PRO after encountering the infinitive marker in Norwegian but that it does not have enough information to predict the occurrence of a pronoun after encountering the complementizer *at*. This could also be a potential explanation for why pronoun antecedent reactivation patterns similarly to English, but PRO antecedent reactivation does not, as English does not have a distinctly unique infinitive marker.

4.3 Summary

The results show that our experimental design functioned as expected based on previous G-maze research and was capable of detecting interference effects in the selected test positions. Sentence types that assign the matrix clause subject as the antecedent differ significantly from their object-assigning counterparts. Test

position zero is always faster than position one, and there is a significant interaction effect between position and type. The possibility to control variance from presentation effects (learning effects) is demonstrated by using the order of individual word presentation as a model factor rather than trial number (cf. Baayen & Milin, 2010).

Considering the detected differences, our results support the presence of an interference effect for Control and pronoun conditions. We were able to detect these effects using written stimuli, disproving the suggestion that effects would only be found for auditory stimuli (Nicol & Swinney, 2002). Both the Control and pronoun sentences recorded no significant difference between the intended antecedent and the mentioned NP, so antecedent resolution is not complete at the test positions in this study, which opposes the predictions of the Trace Reactivation Hypothesis. We find no indication of a preference for the most recent antecedent as has been found previously for Japanese and nominal Control in English (Witzel & Witzel, 2016; Kwon & Sturt, 2016). These results suggest that coreference processing for Control and pronouns in Norwegian patterns similarly to English in that all potential antecedents show signs of processing differences compared to NPs that are not potential antecedents. If this similarity to English also extends to NP traces in that traces cause immediate reactivation of only their antecedent, then we can exclude a trace account of PRO.

In addition to a difference between Control and Pronoun sentences, our results reflect the results from our previous study that used a priming paradigm to investigate Control and Raising (Larsen & Johansson, 2020). The priming study revealed a priming effect for both potential PRO antecedents following the presentation of the infinitive marker in Subject Control sentences. The current study replicated a trend of this effect directly succeeding the presentation of and in competition with the infinitive marker. We found the predicted interference effect at the predicted position. We also detected a trend towards significance for *activated* for Object Control, which did not reach significance in the priming study.

5. Conclusion

Throughout this article we have discussed coreference processing in relation to Control sentences as well as their pronoun counterparts. The results demonstrate that antecedent reactivation effects can be observed using a G-maze design, similar to earlier (word-to-picture) priming and L-maze studies. We see a rise in reaction times at all test positions where interference was predicted, which supports the validity of the study.

Due to the prediction of interference rather than facilitation, G-maze studies complement the more commonly used priming studies. The combination of techniques demonstrates that we can predict the direction of reaction times by predicting either facilitation (priming) or interference (G-maze). G-maze can be used to test previous findings as an alternative to repeating a priming study.

Our results confirm the findings of our previous study, which found priming effects for both potential antecedents of PRO following the infinitive marker in Subject Control sentences in Norwegian. This study used a method that tested for interference, and thus slower reaction times rather than facilitation. Both sets

of results point to PRO antecedent processing beginning directly following the appearance of the infinitive marker. This data might support the infinitive marker being positioned in C as a complementizer in Norwegian, as opposed to T like in other languages with similar Control surface structures, such as English. Our findings indicate that Norwegian may trigger antecedent activation earlier at the infinitive marker. This activated tense phrase (TP) could be marked for tense and an active PRO before seeing the subordinate clause verb. This predicts that the effects we associate with PRO will occur later in languages with more ambiguous markers for the start of an infinitive clause, such as Swedish. The Swedish infinitive marker (“att”) and complementizer (“att”) share a common written form.

In a similar manner to our priming study, which exposed a difference between the processing of PRO and NP traces in Raising constructions in Norwegian, the current G-maze study recorded a difference between the processing of PRO and pronouns in complement clauses in Norwegian. Though the sentence structures of these three constructions are extremely similar to the untrained eye, our experimental work has detected differences in priming and inhibitory effects during sentence processing. That there is a difference suggests that speakers process the constructions differently despite the superficial similarity up to the test position. This argues that there is an underlying difference, such as the early commitment to a non-finite clause by means of a clear infinitive marker.

This article has also introduced some improvements to the analysis of reaction time experiments, which may explain earlier difficulties in detecting effects of coreference processing. We found a way to regress out the effect of increasingly faster responses from participants that get increasingly confident and familiar with the task. We extended this to account for the effect of slower responses in our older participants, and thus were able to keep more participants and a slightly more representative sample of the population. These improvements led to an increased power to detect small differences since more of the variance could be accounted for.

Our impression is that the length of the study was appropriate to avoid fatigue, especially since each decision was fairly easy to make, which is evidenced by a low average error rate. The randomized order of the trials makes it unlikely that a training effect can explain any association between our controlled factors, but rather the participants get increasingly confident in their decisions resulting in faster reaction times. This effect was handled by the statistical model.

It is interesting that neither the present nor the previous study (Larsen & Johansson, 2020) observed a significant difference between the mentioned NP and the intended antecedent for PRO. This suggests that the binding and eventual assignment, or selection, of the antecedent occurs at a *later* stage in sentence processing. In the Control constructions, the Control information supposedly provided by the matrix clause verb is not enough to assign *immediate* coreference upon encountering PRO. When is this information available, where does it come from, and can we exclude an trace account of PRO?

Future research may expand the range of languages for which coreference processing has been investigated and its relation to how clearly non-finite clauses are marked lexically and structurally. Swedish, which is similar to Norwegian, is one interesting case with an ambiguous infinitive/complementizer marker (“att”). Comparing data on the processing of referent-dependent items across languages

may provide insights into the role of syntactic, semantic, and pragmatic dependencies and restrictions in coreference resolution, and at what point coreference is resolved for different types of sentence constructions. Since the data shows that both potential antecedents (mentioned and intended) are activated, there is a chance that contextual information will affect the selection of antecedents and affect the timing of selection. Branching out to other methods such as eye-tracking and the use of visual world paradigms could provide additional insight into the time course of coreference processing. In controlled experiments contextual information could be provided that predicts only one potential antecedent at the relevant positions. Will both antecedents still be activated, with selection occurring at a later stage?

Acknowledgments. We are grateful for the feedback and suggestions from three anonymous reviewers that greatly improved the clarity and argumentation of our article. We would also like to thank Carl Börstell from the University of Bergen and Christine Howes from the University of Gothenburg for feedback on earlier drafts. Finally, we greatly appreciate the enthusiasm for our research that Marit Julien from Lund University showed prior to us writing this article.

References

- Aoshima, Sachiko, Yoshida, Masaya & Phillips, Colin. 2009. Incremental Processing of Coreference and Binding in Japanese. *Syntax*, 12(2), 93–134, doi: [10.1111/j.1467-9612.2009.00123.x](https://doi.org/10.1111/j.1467-9612.2009.00123.x).
- Arnold, Jennifer E. 1998. Reference form and discourse patterns, *Doctoral dissertation*, Stanford University, California.
- Arnold, Jennifer E., Eisenband, Janet G., Brown-Schmidt, Sarah, & Trueswell, John C. 2000. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76(1), B13–B26, doi: [10.1016/S0010-0277\(00\)00073-1](https://doi.org/10.1016/S0010-0277(00)00073-1).
- Ashby, F. Gregory, & Townsend, James T. 1980. Decomposing the reaction time distribution: pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, 21, 93–123, doi: [10.1016/0022-2496\(80\)90001-2](https://doi.org/10.1016/0022-2496(80)90001-2).
- Åfarli, Tor A. & Eide, Kristin M. 2003. *Norsk generativ syntaks*, Oslo: Novus forlag.
- Baayen, R. Harald & Milin, Petar. 2010. Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28, doi: [10.21500/20112084.807](https://doi.org/10.21500/20112084.807).
- Berger, Alexander & Kiefer, Markus. 2021. Comparison of different response time outlier exclusion methods: A Simulation study. *Frontiers in Psychology*, 12, 1–13, doi: [10.3389/fpsyg.2021.675558](https://doi.org/10.3389/fpsyg.2021.675558).
- Bickel, Balthasar, Witzlack-Makarevich, Alena, Choudhary, Kamal K., Schlesewsky, Matthias & Bornkessel-Schlesewsky, Ina. 2015a. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PloS One*, 10(8), 1–22, doi: [10.1371/journal.pone.0132819](https://doi.org/10.1371/journal.pone.0132819).
- Bickel, Balthasar, Witzlack-Makarevich, Alena & Zakharko, Taras. 2015b. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej L. Malchukov & Marc Richards (eds.), *Scales and Hierarchies: A Cross-Disciplinary Perspective*, 277(1), 7–44. Berlin: De Gruyter Mouton.
- Bisang, Waite. 2008. Finite vs. non finite languages. In Martin Haspelmath, et al.(eds.), *Language Typology and Language Universals 2.Teilband*, 20(2), 1400–1413. Berlin, Boston: De Gruyter. (Original work published 2001), doi: [10.1515/9783110194265-037](https://doi.org/10.1515/9783110194265-037).
- Blanchard, Harry. 1987. The effects of pronoun processing on information utilization during fixations in reading. *Bulletin of the Psychonomic Society*, 25, 171–174, doi: [10.3758/BF03330319](https://doi.org/10.3758/BF03330319).
- Bornkessel-Schlesewsky, Ina & Schlesewsky, Matthias. 2009. The Role of Prominence Information in the Real-Time Comprehension of Transitive Constructions: A Cross-Linguistic Approach. *Language and Linguistics Compass*, 3(1), 19–58, doi: [10.1111/j.1749-818X.2008.00099.x](https://doi.org/10.1111/j.1749-818X.2008.00099.x).
- Carnie, Andrew. 2013. Syntax: A generative introduction. *Introducing Linguistics*, 4. Chichester: Wiley-Blackwell.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

- Chomsky, Noam & Lasnik, Howard. 1993. The theory of principles and parameters. *Handbooks of Linguistics and Communication Science*, 9(1), 505–569.
- Clark, Herbert H. & Sengul, C.J. 1979. In search of referents for nouns and pronouns. *Memory & Cognition*, 7(1), 35–41. Stanford: Psychonomic Society, Inc.
- Clifton, Charles Jr, Kennison, Shelia, & Albrecht, Jason E. 1997. Reading the words her, his, and him: implications for parsing principles based on frequency and structure. *Journal of Memory and Language*, 36, 276–292.
- Corbett, Albert T. & Chang, Frederick R. 1983. Pronoun disambiguation: Accessing potential antecedents. *Memory and Cognition*, 11, 283–294.
- Cowles, H. Wind, Walenski, Matthew, & Kluender, Robert. 2007. Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi*, 26, 3–18, doi: [10.1007/s11245-006-9004-6](https://doi.org/10.1007/s11245-006-9004-6)
- Dahl, Østen & Fraurud, Kari. 1996. Animacy in Grammar and Discourse. In Tore Frøtheim & Jeanette K. Gundel (eds.), *Reference and Referent Accessibility*, 38, 47–64. Amsterdam: John Benjamins.
- Delgado, João, Raposo, Ana, & Santos, Ana Lúcia. 2021. Assessing Intervention Effects in Sentence Processing: Object Relatives vs. Subject Control. *Frontiers in Psychology*, 12, 1–19, doi: [10.3389/fpsyg.2021.610909](https://doi.org/10.3389/fpsyg.2021.610909).
- Dillon, Brian, Mishler, Alan, Sloggett, Shayne & Phillips, Colin. 2013. Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103, doi: [10.1016/j.jml.2013.04.003](https://doi.org/10.1016/j.jml.2013.04.003).
- Ehrlich, Kate. 1980. Comprehension of Pronouns. *Quarterly Journal of Experimental Psychology*, 32(2), 247–255, doi: [10.1080/14640748008401161](https://doi.org/10.1080/14640748008401161).
- Ehrlich, Kate & Rayner, Keith. 1983. Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, 22(1), 75–87, doi: [10.1016/S0022-5371\(83\)80007-3](https://doi.org/10.1016/S0022-5371(83)80007-3).
- Faarlund, Jan Terje, Svein, Lie, & Vannebo, Kjell Ivar. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Faarlund, Jan Terje. 2007. Parameterization and change in non-finite complementation. *Diachronica*, 24(1), 57–80.
- Faarlund, Jan Terje. 2015. The Norwegian Infinitive Marker. *Working Papers in Scandinavian Syntax*, 95, 1–10.
- Forster, Kenneth I., Guerrero, Christine, & Elliot, Lisa. 2009. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41(1), 163–171.
- Freedman, Sandra E. & Forster, Kenneth I. 1985. The psychological status of overgenerated sentences. *Cognition*, 19, 101–131.
- Garrod, Simon & Sanford, Anthony. 1977. Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 77–90, doi: [10.1016/S0022-5371\(77\)80009-1](https://doi.org/10.1016/S0022-5371(77)80009-1).
- Green, Jeffrey J., McCourt, Michael, Lau, Ellen, & Williams, Alexander. 2020. Processing adjunct control: Evidence on the use of structural information and prediction in reference resolution. *Glossa*, 5(1), 1–33, doi: [10.5334/gjgl.1133](https://doi.org/10.5334/gjgl.1133).
- Grosz, Barbara J., Joshi, Aravind, & Weinstein, Scott. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.
- Hestvik, Arild, Nordby, Helge, & Karlsen, Geir. 2005. Antecedent reactivation by surface and deep anaphora in Norwegian. *Scandinavian Journal of Psychology*, 46(3), 229–238.
- Hestvik, Arild, Schwartz, Richard G., & Tornøya, Lydia. 2010. Relative Clause Gap-Filling in Children with Specific Language Impairment. *Journal of Psycholinguistic Research*, 39(5), 443–456.
- Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science*, 3, 67–90.
- Hobbs, Jerry R. 1990. *Literature and cognition*. Stanford, CA: CSLI. Lecture Notes 21.
- Holmes, Virginia M. 1973. Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior*, 12, 285–293.
- Hornstein, Norbert. 1999. Movement and Control. *Linguistic Inquiry*, 30(1), 69–96.
- Hornstein, Norbert & Polinsky, Maria. 2010. Control as Movement. *Movement Theory of Control*. In Norbert Hornstein, and Maria Polinsky (eds.), 1–44, John Benjamins.
- Janke, Vikki. 2003. A PRO-less theory of control. *UCLWPL*, 15, 213–242.

- Janke, Vikki. 2007. Control without PRO. *Doctoral dissertation*. University College London.
- Janke, Vikki. 2008. Control without a Subject. *Lingua*, 118(1), 82–118, doi: [10.1016/j.lingua.2007.05.002](https://doi.org/10.1016/j.lingua.2007.05.002).
- Järvisikivi, Juhani, van Gompel, Roger P.G., & Hyönä, Jukka. 2017. The Interplay of Implicit Causality, Structural Heuristics, and Anaphor Type in Ambiguous Pronoun Resolution. *Journal of Psycholinguistic Research*, 46(3), 525–550, doi: [10.1007/s10936-016-9451-1](https://doi.org/10.1007/s10936-016-9451-1)
- Johshi, Aravind & Prasad, Rashmi. 2006. Anaphora Resolution: Centering Theory Approach. *Encyclopedia of Language & Linguistics*, 1, 223–230, doi: [10.1016/B0-08-044854-2/04366-2](https://doi.org/10.1016/B0-08-044854-2/04366-2).
- Kaan, Edith & Stowe, Laurie. 2002. Storage and Computation in Sentence Processing. *Storage and Computation in the Language Faculty*, 30, 257–295, doi: [10.1007/978-94-010-0355-1_9](https://doi.org/10.1007/978-94-010-0355-1_9).
- Karimi, Hossein & Ferreira, Fernanda. 2016. Processing Control Information in a Nominal Control Construction: An Eye-Tracking Study. *Journal of Psycholinguistic Research*, 45, 779–793.
- Kehler, Andrew & Rohde, Hannah. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39(1-2), 1–37, doi: [10.1515/tl-2013-0001](https://doi.org/10.1515/tl-2013-0001).
- King, Jonathan & Just, Marcel Adam. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26, doi: [10.18637/jss.v082.i13](https://doi.org/10.18637/jss.v082.i13).
- Kwon, Nayoung & Sturt, Patrick. 2016. Processing Control Information in a Nominal Control Construction: An Eye-Tracking Study. *Journal of Psycholinguistic Research*, 45, 779–793, doi: [10.1007/s10936-015-9374-2](https://doi.org/10.1007/s10936-015-9374-2).
- Larsen, Tori & Johansson, Christer. 2020. Gone with a Trace?: Reactivation at PRO Positions. *Typical and impaired processing in Morphosyntax*, 64, 91–116, John Benjamins.
- Larsen, Espen & Johansson, Christer. 2007. Animacy and word order effects in language processing. *1st proc. of the Swedish Association of Language and Cognition*.
- Larsen, Espen & Johansson, Christer. 2009. Animacy and Canonical Word Order. In Jordan Zlatev, Mats André, Marlene Johansson Falck & Carita Lundmark (eds.), *Studies in Language and Cognition*, 287–297, Cambridge Scholars Publishing.
- Lenth, Russell V. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.0. <https://CRAN.R-project.org/package=emmeans>.
- Luce, R. Duncan. 1991. *Response Times: Their Role in Inferring Elementary Mental Organization*. New York, NY: Oxford University Press, doi: [10.1093/acprof:oso/9780195070019.001.0001](https://doi.org/10.1093/acprof:oso/9780195070019.001.0001).
- MacDonald, Maryellen C. & MacWhinney, Brian. 1990. Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29, 469–492.
- Meyer, David, Zeileis, Achim, & Hornik, Kurt. 2006. The Strucplot Framework: Visualizing Multi-Way Contingency Tables with vcd. *Journal of Statistical Software*, 17(3), 1–48, doi: [10.18637/jss.v017.i03](https://doi.org/10.18637/jss.v017.i03).
- Meyer, David, Zeileis, Achim, & Hornik, Kurt. 2021. vcd: Visualizing Categorical Data. R package version 1.4-9.
- Neeleman, Ad & van de Koot, Hans. 2002. The configurational matrix. *Linguistic Inquiry*, 33, 529–574.
- Nicol, Janet. 1988. Coreference processing during sentence comprehension. *Unpublished doctoral dissertation*, MIT.
- Nicol, Janet & Swinney, David. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1), 5–19.
- Nicol, Janet & Swinney, David. 2002. The psycholinguistics of anaphora. In Barss A. (eds), *Anaphora: a reference guide*, 5, 72–104. Blackwell Publishing, Cambridge.
- Parker, Dan, Lago, Sol & Phillips, Colin. 2015. Interference in the processing of adjunct control. *Frontiers in Psychology*, 6, 1–13, doi: [10.3389/fpsyg.2015.01346](https://doi.org/10.3389/fpsyg.2015.01346).
- Primus, Beatrice. 2006. Hierarchy mismatches and the dimensions of role semantics. In Ina Bornkessel, Matthias Schlesewsky, Bernard Comrie & Angela D. Friederici (eds.), *Semantic role universals and argument linking: Theoretical, typological and psycholinguistic perspectives*, 165(1), 53–88. Berlin: Mouton de Gruyter, doi: [10.1515/9783110219272.53](https://doi.org/10.1515/9783110219272.53).
- R Core Team. 2021. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>.
- Reed, Lisa A. 2014. Strengthening the PRO hypothesis. *Studies in Generative Grammar*, 110, Berlin: De Gruyter Mouton, doi: [10.1515/9781614510413](https://doi.org/10.1515/9781614510413).

- Schmauder, A. René, Morris, Robin K., & Poynor, David V. 2000. Lexical processing and text integration of function and content words: Evidence from priming and eye fixations. *Memory & Cognition*, **28**(7), 1098–1108, doi: [10.3758/BF03211811](https://doi.org/10.3758/BF03211811).
- Sekerina, Irina A., Laurinavichyute, Anna K., & Dragoy, Olga. 2019. What eye movements can and cannot tell us about Wh-movement and Scrambling. *Grammatical Approaches to Language Processing*, **48**, 147–165, Springer.
- Sigursson, Halldór Ármann. 2008. The case of PRO, *Natural Language & Linguistic Theory*, **26**(2), 403–450.
- Stewart, Andrew J., Holler, Judith, & Kidd, Evan. 2007. Shallow processing of ambiguous pronouns: Evidence for delay. *The Quarterly Journal of Experimental Psychology*, **60**(12), 1680–1696, Psychology Press.
- Stewart, Andrew J., Pickering, Martin J., & Sanford, Anthony J. 2000. The Time Course of the Influence of Implicit Causality Information: Focusing versus Integration Accounts. *Journal of Memory and Language*, **42**(3), 423–443, doi: [10.1006/jmla.1999.2691](https://doi.org/10.1006/jmla.1999.2691).
- Stowe, Laurie & Kaan, Edith. 2006. Developing an Experiment: Techniques and Design. *Methods in Psycholinguistics*, Unpublished Manuscript. <http://ufdc.ufl.edu/IR00005517/00001>.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, **48**, 542–562, doi: [10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3).
- Sturt, Patrick & Kwon, Nayoung. 2015. The processing of raising and nominal control: an eye-tracking study. *Frontiers in Psychology*, **6**(331), 542–562, doi: [10.3389/fpsyg.2015.00331](https://doi.org/10.3389/fpsyg.2015.00331).
- Swinney, David. 1979. Lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behaviour*, **18**, 645–660, doi: [10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4).
- Walenski, Matthew. 2002. Relating parsers and grammars: On the structure and realtime comprehension of English infinitival complements. *Unpublished doctoral dissertation*, UCSD, California.
- Wanner, Eric, & Maratsos, Michael. 1978. An atn approach in comprehension. In M. Halle, J. B. and Miller, G., editors, *Linguistic Theory and Psychological Reality*, Cambridge, MA: MIT Press.
- Weatherford, Kathryn C., & Arnold, Jennifer E. 2021. Semantic predictability of implicit causality can affect referential form choice. *Cognition*, **214**, 104759, doi: [10.1016/j.cognition.2021.104759](https://doi.org/10.1016/j.cognition.2021.104759).
- Whelan, Robert. 2008. Effective analysis of reaction time data. *The Psychological Record*, **58**, 475–482, doi: [10.1007/BF03395630](https://doi.org/10.1007/BF03395630).
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, Hadley. 2021. tidy: Tidy Messy Data. R package version 1.1.4. <https://CRAN.R-project.org/package=tidy>.
- Wickham, Hadley, François, Romain, Henry, Lionel & Müller, Kirill. 2021. dplyr: A Grammar of Data Manipulation. R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>.
- Witzel, Jeffrey & Witzel, Naoko. 2016. Incremental Sentence Processing in Japanese: A Maze Investigation into Scrambled and Control Sentences. *Journal of Psycholinguistic Research*, **45**, 475–505, doi: [10.1007/s10936-015-9356-4](https://doi.org/10.1007/s10936-015-9356-4).
- Xiang, Ming, Dillon, Brian, & Phillips, Colin. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, **108**(1), 40–55, doi: [10.1016/j.bandl.2008.10.002](https://doi.org/10.1016/j.bandl.2008.10.002).
- Zeileis, Achim, Meyer, David, & Hornik, Kurt. 2007. Residual-based Shadings for Visualizing (Conditional) Independence. *Journal of Computational and Graphical Statistics*, **16**(3), 507–525.

Appendix

A. Test sentence templates

Subject Control

1. Den snille apekatten tilbyr elefanten å bære posen hjem.
The kind monkey offers the elephant to carry the bag home.
2. Den grønne krokodillen skylder sjiraffen å lage lunsj.
The green crocodile owes the giraffe to make lunch.
3. Den store neshornet lover krokodillen å spise alle grønnsakene sine.
The big rhinoceros promises the crocodile to eat all his vegetables.

Subject Pronoun

4. Den raske leoparden tilbyr elefanten at han betaler for taxien.
The fast leopard offers the elephant that he pays for the taxi.
5. Den høye sjiraffen skylder apekatten at han må ut søppelet.
The tall giraffe owes the monkey that he takes out the trash.
6. Den grå elefanten lover neshornet at han drikker tre liter vann hver dag.
The gray elephant promises the rhinoceros that he drinks three liter of water each day.

Object Control

7. Elefanten anbefaler krokodillen som er søvnig å meditere før sengetid.
The elephant recommends the crocodile who is sleepy to meditate before bedtime.
8. Apekatten tillater leoparden som er sulten å stjele en moden banan.
The monkey allows the leopard who is hungry to steal a banana.
9. Leoparden lærer neshornet som er uerfaren å finne mat i skogen.
The leopard teaches the rhinoceros who is inexperienced to find food in the forest.

Object Pronoun

10. Krokodillen anbefaler apekatten som lukter dårlig at han dusjer hver dag.
The crocodile recommends the monkey who smells bad that he shower every day.
11. Sjiraffen tillater leoparden som er rar at han har på seg to forskjellige sko.
The giraffe allows the leopard who is unconventional that he wears two different shoes.
12. Neshornet lærer sjiraffen som er usikker at han tenker for mye på hvordan han ser ut.
The rhinoceros teaches the giraffe who is self-conscious that he thinks too much about how he looks.

B. Filler sentences

1. Apekatten som er forelsket tenker på krokodillen når han ser en fin blomst.
The monkey who is in love thinks about the crocodile when he sees a pretty flower.
2. Elefanten som er bekymret ringer leoparden fem ganger i uken bare for å høre hvordan han har det.
The elephant who is worried calls the leopard five times a week just to hear how he is doing.
3. Elefanten treffer den vennlige apekatten på den store og flate steinen hver annen dag.
The elephant meets the friendly monkey at the big and flat stone every other day.

4. Leoparden som alltid står opp tidlig vekker apekatten med en gøy sang om egg og bacon.
The leopard who always wakes up early wakes the monkey with a fun song about eggs and bacon.
5. Leoparden synger bursdagssangen til den sjenerte sjiraffen hvert år i desember.
The leopard sings the birthday song to the shy giraffe every year in December.
6. Neshornet overnatter hos apekatten på fredag og de skal spille fotball sammen.
The rhinoceros is sleeping over at the monkey's on Friday and they will play soccer together.
7. Neshornet leter etter elefanten, men han fant bare katten hans.
The rhinoceros is looking for the elephant, but he only found his cat.
8. Sjiraffen sender den kjekke krokodillen en pakke med godteri på Valentinsdagen.
The giraffe sends the handsome crocodile a pack of candy on Valentine's Day.
9. Sjiraffen dusjer mye tidligere på dagen enn elefanten hver dag i uken.
The giraffe showers much earlier in the day than the elephant each day of the week.
10. Sjiraffen hjelper det trette neshornet med hans mattelekser fordi han klarer det ikke selv.
The giraffe helps the tired rhinoceros with his math homework because he cannot do it himself.
11. Krokodillen som er fattig føler seg trist fordi leoparden har mer penger enn ham.
The crocodile who is poor feels sad because the leopard has much more money than him.
12. Krokodillen gifter seg med neshornet, og de vil være sammen for resten av livet.
The crocodile marries the rhinoceros, and they will be together for the rest of their lives.

C. Training sentences

1. Apekatten arbeider sammen med det grinete neshornet på onsdager.
The monkey works together with the grumpy rhinoceros on Wednesdays.
2. Apekatten som allerede har en kjæreste, vet at sjiraffen fortalte alle om hans forelskelse.
The monkey who already has a significant other knows that the giraffe told everyone about his crush.
3. Elefanten som er omtenksum stemmer for at sjiraffen blir kongen av jungelen ved neste valg.
The elephant who is thoughtful votes for the giraffe to become the king of the jungle in the next election.
4. Leoparden forklarer til den unge krokodillen alt om hvordan han klatrer i trær.
The leopard explains to the young crocodile everything about how he climbs in trees.
5. Neshornet barberer leoparden gratis fordi de har vært venner siden de var små.
The rhinoceros shaves the leopard free of charge because they have been friends since they were little.
6. Krokodillen eter middag så fort at elefanten som også er sulten er redd for at han ikke får noe mat.
The crocodile eats dinner so quickly that the elephant who is also hungry is scared that he will not get any food.

D. Tables

Table A1. Estimated Model characteristics for Table 10

Scaled residuals:				
Min	1Q	Median	3Q	Max
−2.9725	−0.6297	−0.2226	0.3760	5.1854
Random effects				
Groups	Name	Variance	Std. Dev.	
Item	(Intercept)	2532	50.31	
Participant	(Intercept)	13066	114.31	
Residual		59002	242.90	
Number of obs: 7645, groups: Item, 144; Participant, 55				
REML criterion at convergence:105539.2				

Table A2. Model summary for the analysis of all conditions. *Predicted* is included as a factor

Model Summary for Analysis of all Conditions						
Fixed Effects	Estimate	Std. Error	df	t-value	Pr(> t)	
(Intercept)	9.39.1e+02	4.007e+01	1.504e+02	23.438	<2e-16	***
predictedyes	−9.432e+00	4.926e+01	1.014e+02	−0.191	0.84853	
typeOPN	−1.499e+02	4.474e+01	1.227e+02	−3.351	0.00107	**
typeSC	6.853e+01	4.933e+01	1.020e+02	1.389	0.16777	
typeSPN	−7.275e+01	4.937e+01	1.024e+02	−1.473	0.14372	
activatedMentioned	−1.200e+01	4.928e+01	1.016e+02	−0.244	0.80809	
activatedIntended	−6.238e+00	4.923e+01	1.012e+02	−0.127	0.89942	
positionzero	−1.597e+02	4.924e+01	1.012e+02	−3.244	0.00160	**
order	−1.601e-01	9.342e-03	7.497e+03	−17.135	<2e-16	***
age26-35	6.433e+00	3.858e+01	5.092e+01	0.167	0.86824	
age36-45	−8.405e+00	1.176e+02	5.083e+01	−0.071	0.94333	
age46+	1.822e+02	5.536e+01	5.102e+01	3.290	0.00182	**
predictedyes:typeOPN	−6.146e+00	6.322e+01	1.223e+02	−0.097	0.92272	
predictedyes:typeSC	2.664e+01	6.975e+01	1.019e+02	0.382	0.70325	
predictedyes:typeSPN	−1.539e+01	6.974e+01	1.019e+02	−0.221	0.82582	
predictedyes: activatedMentioned	1.205e+01	6.973e+01	1.018e+02	0.173	0.86318	

(Continued)

Table A2. (Continued)

Model Summary for Analysis of all Conditions					
Fixed Effects	Estimate	Std. Error	df	t-value	Pr(> t)
predictedyes:activatedIntended	8.270e+01	6.962e+01	1.012e+02	1.188	0.23764
typeOPN:activatedMentioned	5.183e+00	6.318e+01	1.220e+02	0.082	0.93476
typeSC:activatedMentioned	5.079e+01	6.964e+01	1.013e+02	0.729	0.46751
typeSPN:activatedMentioned	4.090e+01	6.974e+01	1.019e+02	0.586	0.55887
typeOPN:activatedIntended	-2.710e+00	6.316e+01	1.219e+02	-0.043	0.96585
typeSC:activatedIntended	-2.357e+01	6.969e+01	1.016e+02	-0.338	0.73588
typeSPN:activatedIntended	-2.933e+01	6.970e+01	1.017e+02	-0.421	0.67484
predictedyes:positionzero	2.889e+00	6.958e+01	1.010e+02	0.042	0.96696
typeOPN:positionzero	1.813e+02	6.322e+01	1.223e+02	2.868	0.00487 **
typeSC:positionzero	1.913e+01	6.966e+01	1.014e+02	0.275	0.78421
typeSPN:positionzero	1.195e+02	6.971e+01	1.017e+02	1.715	0.08942 .
activatedMentioned: positionzero	-1.731e+00	6.957e+01	1.009e+02	-0.025	0.98020
activatedIntended:positionzero	2.446e+01	6.952e+01	1.006e+02	0.352	0.72572
predictedyes:typeOPN: activatedMentioned	8.094e+01	9.404e+01	1.100e+02	0.861	0.39127
predictedyes:typeSC: activatedMentioned	-5.982e+01	9.855e+01	1.016e+02	-0.607	0.54519
predictedyes:typeSPN: activatedMentioned	5.330e+01	9.870e+01	1.022e+02	0.540	0.59035
predictedyes:typeOPN: activatedIntended	1.492e+01	9.395e+01	1.096e+02	0.159	0.87411
predictedyes:typeSC: activatedIntended	-7.745e+01	9.866e+01	1.020e+02	-0.785	0.43428
predictedyes:typeSPN: activatedIntended	5.827e+01	9.860e+01	1.018e+02	0.591	0.55585
predictedyes:typeOPN: positionzero	3.628e+01	8.939e+01	1.222e+02	0.406	0.68556
predictedyes:typeSC: positionzero	-8.352e+01	9.853e+01	1.015e+02	-0.848	0.39858
predictedyes:typeSPN: positionzero	4.057e+01	9.855e+01	1.016e+02	0.412	0.68146
predictedyes: activatedMentioned:position- zero	4.784e+01	9.845e+01	1.012e+02	0.486	0.62807
predictedyes: activatedIntended:positionzero	-4.200e+01	9.835e+01	1.008e+02	-0.427	0.67028
typeOPN:activatedMentioned: positionzero	2.993e+01	8.938e+01	1.221e+02	0.335	0.73829

(Continued)

Table A2. (Continued)

Model Summary for Analysis of all Conditions					
Fixed Effects	Estimate	Std. Error	df	t-value	Pr(> t)
typeSC:activatedMentioned: positionzero	-7.423e+01	9.845e+01	1.012e+02	-0.754	0.45263
typeSPN:activatedMentioned: positionzero	2.387e+01	9.855e+01	1.016e+02	0.242	0.80907
typeOPN:activatedIntended: positionzero	-2.195e+00	8.932e+01	1.219e+02	-0.025	0.98043
typeSC:activatedIntended: positionzero	-2.616e+01	9.844e+01	1.011e+02	-0.266	0.79097
typeSPN:activatedIntended: positionzero	1.553e+01	9.853e+01	1.015e+02	0.158	0.87507
predictedyes:typeOPN: activatedMentioned: positionzero	-1.235e+02	1.330e+02	1.099e+02	-0.928	0.35523
predictedyes:typeSC: activatedMentioned: positionzero	1.414e+02	1.393e+02	1.014e+02	1.015	0.31265
predictedyes:typeSPN: activatedMentioned: positionzero	-9.699e+01	1.395e+02	1.019e+02	-0.695	0.48843
predictedyes:typeOPN: activatedIntended;positionzero	-3.179e+01	1.329e+02	1.098e+02	-0.239	0.81141
predictedyes:typeSC: activatedIntended;positionzero	1.528e+02	1.393e+02	1.015e+02	1.097	0.27535
predictedyes:typeSPN: activatedIntended;positionzero	-2.631e+01	1.394e+02	1.018e+02	-0.189	0.85072

Table A3. Type III ANOVA table for the final model. Degrees of freedom within conditions (df_w) are estimated using Satterthwaite's method

Factor	SS	MSS	df _B	df _w	F	Pr(> F)	
Predicted	700355	700355	1	123.4	11.8701	0.0008	***
Type	3524024	1174675	3	101.7	19.9092	0.0000	***
Activated	507652	253826	2	116.6	4.3020	0.0158	*
Position	3078287	3078287	1	101.7	52.1729	0.0000	***
Order	17323194	17323194	1	7496.8	293.6052	0.0000	***
Age	651193	217064	3	50.9	3.6790	0.0179	*
Predicted:Type	133914	44638	3	134.8	0.7566	0.5204	
Predicted:Activated	543240	271620	2	105.9	4.6036	0.0121	*

(Continued)

Table A3. (Continued)

Factor	SS	MSS	df _B	df _W	F	Pr(> F)
Type:Activated	302405	50401	6	121.5	0.8542	0.5308
Predicted:Position	3521	3521	1	123.4	0.0597	0.8074
Type:Position	3978128	1326043	3	101.7	22.4747	0.0000 ***
Activated:Position	16531	8265	2	116.5	0.1401	0.8694
Predicted:Type:Activated	65836	10973	6	106.2	0.1860	0.9802
Predicted:Type:Position	28557	9519	3	134.9	0.1613	0.9222
Predicted:Activated:Position	55398	27699	2	105.9	0.4695	0.6266
Type:Activated:Position	94037	15673	6	121.5	0.2656	0.9519
Predicted:Type:Activated:Position	304082	50680	6	106.2	0.8590	0.5277

Table A4. Estimated Model characteristics for Table 13

Scaled residuals:				
Min	1Q	Median	3Q	Max
−2.8175	−0.6396	−0.2317	0.4025	4.1473
Random effects				
Groups	Name	Variance	Std. Dev.	
Item	(Intercept)	2587	50.86	
Participant	(Intercept)	15350	123.90	
Residual		67266	259.36	
Number of obs: 3798, groups: Item, 90; Participant, 55				
REML criterion at convergence:52969.7				

Table A5. Model summary for the analysis of all conditions. *Predicted* is excluded as a factor

Model Summary for Analysis of all Conditions						
Fixed Effects	Estimate	Std. Error	df	t-value	Pr(> t)	
(Intercept)	826.33667	41.91831	91.18207	19.713	0.0000	***
typeOPN	53.62174	46.15438	67.02931	1.162	0.24944	
typeSC	89.32465	50.68605	54.84211	1.762	0.08359	.
typeSPN	93.58935	50.80230	55.33580	1.842	0.07080	.
activatedIntended	13.83951	50.55442	54.27635	0.274	0.78531	
activatedUnrelated	−44.97975	50.55804	54.29187	−0.890	0.37757	

(Continued)

Table A5. (Continued)

Model Summary for Analysis of all Conditions						
Fixed Effects	Estimate	Std. Error	df	t-value	Pr(> t)	
positionone	112.52955	50.74352	55.09440	2.218	0.03072	*
order	−0.17179	0.01416	3695.98954	−12.133	0.0000	***
age26-35	−1.85218	42.43896	50.75249	−0.044	0.96536	
age36-45	−12.41459	129.34028	50.55906	−0.096	0.92391	
age46+	187.75105	60.95007	51.02220	3.080	0.00333	**
typeOPN:activatedIntended	−14.41082	65.37451	67.43293	−0.220	0.82620	
typeSC:activatedIntended	−33.54997	71.66154	54.78286	−0.468	0.64152	
typeSPN:activatedIntended	−4.28009	71.90720	55.52677	−0.060	0.95275	
typeOPN:activatedUnrelated	7.21579	65.26558	66.99784	0.111	0.91230	
typeSC:activatedUnrelated	−59.38045	71.62585	54.67434	−0.829	0.41069	
typeSPN:activatedUnrelated	−22.53683	71.74130	55.02264	−0.314	0.75460	
typeOPN:positionone	−125.63484	65.39059	67.51873	−1.921	0.05892	.
typeSC:positionone	−4.74240	71.70474	54.91630	−0.066	0.94751	
typeSPN:positionone	−88.66239	71.95473	55.67791	−1.232	0.22305	
activatedIntended:positionone	60.90981	71.68875	54.86726	0.850	0.39922	
activatedUnrelated:positionone	43.53164	71.63781	54.71250	0.608	0.54593	
typeOPN:activatedIntended:positionone	−57.29539	92.47244	67.49467	−0.620	0.53761	
typeSC:activatedIntended:positionone	−56.26829	101.48970	55.09780	−0.554	0.58153	
typeSPN:activatedIntended:positionone	−59.73479	101.72778	55.60620	−0.587	0.55944	
typeOPN:activatedUnrelated:positionone	−91.31931	92.36044	67.17920	−0.989	0.32634	
typeSC:activatedUnrelated:positionone	69.89569	101.37371	54.84704	0.689	0.49342	
typeSPN:activatedUnrelated:positionone	−69.97147	101.54331	55.21046	−0.689	0.49366	