GLOSSA psycholinguistics

Giskes, A., & Kush, D. 2022. Abstract prediction of morphosyntactic features: Evidence from processing cataphors in Dutch. *Glossa Psycholinguistics* 1(1): X, pp. 1–30. DOI: https://doi.org/10.5070/G6011152

eScholarship University of California

Abstract prediction of morphosyntactic features: Evidence from processing cataphors in Dutch

Anna Giskes, Norwegian University of Science and Technology, NO, anna.giskes@ntnu.no

Dave Kush, University of Toronto, Norwegian University of Science and Technology, CA, dave.kush@utoronto.ca

When comprehenders predict a specific lexical noun in a highly constraining context, they also activate the grammatical features, such as gender, of that noun. Evidence for such *lexically mediated* prediction comes from ERP studies that show that comprehenders are surprised by adjectives and determiners that mismatch the features of a highly predictable noun. In this study, we investigated whether comprehenders can (i) predict an abstract noun phrase in an upcoming argument position (without pre-activating a specific lexical item) and (ii) assign morphosyntactic features to the head noun of that phrase. To do so we used the processing of Dutch cataphors as a test case. We tested whether seeing a cataphor in a preposed clause triggered a prediction of a feature-matching antecedent NP in main subject position. If comprehenders predicted a feature-matching subject, we reasoned that they should also expect an agreeing main verb, which comes before the subject because Dutch is a V2 language. A single-word prediction experiment showed that comprehenders expect a main verb matching the number of the cataphor. In a follow-up self-paced reading experiment, we found a number-mismatch effect if the V2 main verb did not agree with the cataphor. We take the results as evidence that comprehenders predicted a matching antecedent in subject position. We argue that the results are better explained as involving prediction of an abstract noun phrase marked for morphological features, rather than a specific lexical item.

1 Introduction

It is well established that comprehenders generate expectations about upcoming linguistic material (Federmeier, 2007; Van Berkum, 2009; Kutas et al., 2011; Kuperberg & Jaeger, 2016, Pickering & Gambi, 2018). Such expectations can include predictions of morphosyntactic features of words that have not yet been seen (Wicha et al., 2004; Van Berkum et al., 2005; Otten et al., 2007; Otten & Van Berkum, 2008; Szewczyk & Schriefers, 2013, Laszlo & Federmeier, 2009). Fine-grained morphosyntactic prediction has been demonstrated in studies that focused on lexical prediction. In highly constraining contexts, comprehenders are able to pre-activate specific nouns in upcoming positions. Pre-activating the lexical entry for a noun, in turn, activates its grammatical features, such as gender, which can generate morphosyntactic expectations, e.g., for agreement features on intervening words that enter into a relation with the predicted noun. We refer to the prediction of a specific noun (or a highly restricted set of nouns) and the collateral morphosyntactic commitments it engenders as lexically mediated predictions.

The main aim of this study is to investigate whether comprehenders are capable of making morphosyntactic predictions when it is not possible to pre-activate specific lexical nouns in an upcoming position. We ask whether comprehenders can (i) predict an abstract noun phrase in a yet-to-be-seen argument position, (ii) assign morphosyntactic features to the head noun of that phrase, and (iii) generate morphosyntactic expectations that follow as a consequence of their initial prediction. We investigate such non-lexical, abstract prediction by looking at the incremental processing of long-distance dependencies between cataphoric pronouns and their antecedents. Since cataphors, like *she* in (1), precede their antecedents, they provide an opportunity for predicting where in the sentence the antecedent might show up. Predicting an antecedent involves inserting an NP in an unseen position and marking the head noun with gender and number features that match the pronoun (**Figure 1**). We consider such prediction abstract because it can be done in the absence of a highly constraining context, without positing a specific lexical noun to head the NP. We present a completion study and a self-paced reading study that provide evidence that after seeing a cataphor, comprehenders can predict an abstract, subject NP headed by a feature-matching noun and that this prediction generates further expectations about morphosyntactic agreement relations.

(1) After she$_i$ signed the contract, Ingrid$_i$ celebrated with Magnus.

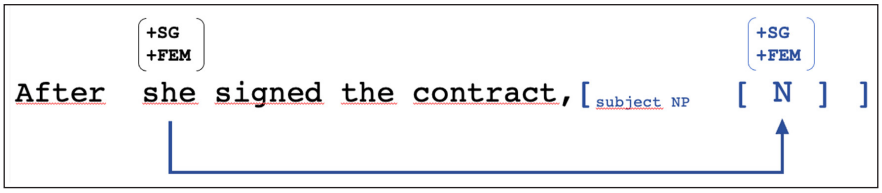

Figure 1: Abstract prediction of a noun with gender and number features.

Our research also informs a subsidiary research question related to the incremental processing of cataphora, namely, how far in advance comprehenders can predict an antecedent for a cataphor in a specific syntactic position. Prior work (Van Gompel & Liversedge, 2003; Kazanina et al., 2007; Drummer & Felser, 2018; Ackerman, 2015; Patterson & Felser, 2019; Xiang et al., 2009; Giskes & Kush, 2021) has shown that comprehenders eagerly and predictively associate cataphors with subject NPs, but there has been debate about the incremental time-course of this predictive association. Conservative models contend that comprehenders may wait to associate a cataphor with a subject NP until the head noun has been recognized (Van Gompel & Liversedge 2003; Brasoveanu & Dotlačil, 2020). Alternative models (e.g., Kush & Dillon 2021) propose that comprehenders can predictively link cataphors to an upcoming subject before the subject NP is reached. Our results suggest that comprehenders can predict that the subject is the antecedent to a cataphor at least before the head noun of the subject phrase is reached.

2 Background

In this section, we review previous work on the prediction of morphosyntactic features. In 2.1, we focus on experimental evidence for lexically mediated morphosyntactic prediction. In 2.2, we discuss previous work on the processing of cataphors, and the limitations of these studies for showing abstract prediction.

2.1 Lexically mediated morphosyntactic prediction

Evidence for lexically mediated morphosyntactic prediction comes from studies on highly predictable nouns in constraining contexts. ERP studies show that when comprehenders predict a specific noun in a highly constraining context, evidence of the prediction can be seen before that noun is encountered at preceding determiners or adjectives. In languages where such prenominal words agree with their nouns, a determiner or adjective that mismatches the predicted noun in morphosyntactic features results in increased processing cost (Wicha et al., 2004; Foucart et al., 2014; Martin et al, 2018; Szewczyk & Wodniecka, 2020; Ito et al., 2020; Fleur et al., 2020; Van Berkum et al., 2005; Ottenet al., 2007; Otten & Van Berkum, 2008 but cf. Kochari & Flecken, 2019). These prenominal effects provide compelling evidence for prediction because they cannot be attributed to the difficulty associated with integrating an unpredictable noun, a well-known challenge for interpreting effects localized on the target noun (e.g., Pickering & Gambi, 2018; Lau et al., 2013).

Otten & Van Berkum (2008) had participants read the Dutch equivalents of short stories like (2), which made a sentence-final noun predictable. They manipulated whether the sentences continued with adjectives that agreed in grammatical gender either with a high-cloze noun (the neuter gender *sword* in 2a) or a low-cloze, but contextually appropriate, noun (the common gender *lance* in 2b). The researchers reasoned that if comprehenders predicted the high-cloze

noun as the head of the NP, they should be surprised to encounter a prenominal adjective with agreement morphology that was inconsistent with that prediction.

(2) The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his …
   a. … big-[$\varnothing_{neuter}$] but rather old-[$\varnothing_{neuter}$] $sword_{neuter}$.
   b. … big-[e_{common}] but rather old-[e_{common}] $lance_{common}$.

The authors found a gender-mismatch effect on the adjective *big*: mismatching adjectives elicited a late negativity starting at roughly 900ms after adjective onset compared to matching adjectives, suggesting error detection or increased processing difficulty well before the critical noun. This effect suggests that comprehenders committed to an analysis where the NP was headed by the high-cloze noun, which enforced gender agreement with the preceding adjective.

In a similar design, Szewczyk & Schriefers (2013) reported evidence for morphosyntactic prediction in the absence of a single high-cloze noun prediction. They measured ERPs on prenominal adjectives in Polish, exploiting differential adjectival agreement with animate and inanimate nouns. They constructed stories that set up a strong expectation for an animate or inanimate noun as the direct object in the critical sentence. In manipulations like (3), the context creates a strong expectation for an inanimate object for the verb *clean*, but cloze norming showed that there was not a strong bias towards one specific lexical candidate. The researchers nevertheless observed an animacy-mismatch effect at the adjective preceding the noun (*entire* in 3).

(3) My mother decided that we should have a 'spring clean' in our house. She cleaned the living room and the kitchen, and my father's job was to clean the first floor. My job was to clean the entire[-$y_{inanimate}$] attic /the entire[-$ego_{animate}$] wolf which had not been used for ages.

Although the findings are consistent with participants having predicted a nominal head without lexical content, but marked for [+/–animate] features, the authors propose a model that employs lexically mediated prediction of multiple words. They argue that the stories constrained the number of plausible nouns to a relatively small set (e.g. 'parts of a house' in 3), which could be activated in parallel and therefore trigger agreement. Under this hypothesis, the findings do not reflect abstract prediction. If the model requires a small set of candidate nouns that is restricted by the semantic environment, we would predict that such prenominal effects would not be observed in less constraining contexts.

The results above show that comprehenders predictively build NPs headed by pre-activated nouns as (internal) arguments for verbs. In virtue of activating a specific lexical item (or small set of lexical items), the predicted nominal head bears features that control agreement with pre-nominal dependent elements. The experiments establish that these kinds of morphosyntactic

predictions are possible when lexically-mediated in highly constraining contexts.[1] They do not establish, however, whether comprehenders can generate similar morphosyntactic predictions in the absence of highly predictable lexical items. Can comprehenders predict an abstract noun phrase with agreement-controlling features in an upcoming position?

2.2 Abstract prediction: Processing cataphors

Encountering a pronoun like *he/she* in (4) in the absence of a prior discourse context triggers an expectation that an antecedent will come later in the sentence (Filik & Sanford, 2008). The earliest grammatical location for an antecedent in such sentences is immediately following the preposed adjunct phrase, usually corresponding to the main subject position.

(4) When [he/she] arrived at the party, the boy cruelly teased the girl.

Research shows that comprehenders actively posit coreference between a cataphor in a preposed adjunct and the main subject. Most evidence comes from gender-mismatch manipulations, where the cataphor either matches or mismatches the gender and/or number of the main subject. In sentences like (4), readers slow down at the noun *boy* when the cataphor is *she* as compared to *he* (Van Gompel & Liversedge, 2003; Kazanina et al., 2007; Yoshida et al., 2014 ; Drummer & Felser, 2018; Ackerman, 2015; Patterson & Felser, 2019; Xiang et al., 2009; Kush & Dillon, 2021; Giskes & Kush, 2021). These effects do not depend on elaborate or rich semantic contexts that could winnow down the set of plausible subject nouns to a small number, unlike prior prediction studies.

Active cataphor resolution suggests that readers do not wait until they have fully processed the head noun *boy* to posit coreference between the cataphor and the main subject in (4). On the one hand, the results are compatible with abstract prediction of a feature-matching noun as the subject. On the other hand, the results are compatible with no prediction taking place before the noun. For example, Van Gompel and Liversedge (2003) propose a mechanism that posits coreference between the cataphor and the noun only after the noun has been reached in the bottom-up input. That is, comprehenders wait to predict coreference until they encounter the noun *boy* in (4). In their model, mismatch effects occur because the bottom-up processing of gender and number features is delayed until after coreference has been established. The

[1] A reviewer asks whether results from visual world eye-tracking studies that show that comprehenders use the gender of a preceding determiner to anticipate reference to specific objects in a scene (e.g., Lew-Williams & Fernald 2007) provide evidence for abstract prediction of nominal features without lexical mediation. We consider such anticipatory looks to rely on lexically-mediated prediction. Since participants see the set of candidate referents before they make their prediction, their behavior in the task can be described as selecting from a contextually-constrained set of preactivated nouns.

mismatch effect is then triggered when the gender and number features are fully processed, and the impossibility of coreference becomes clear.

Subsequent research (Kazanina et al., 2007; Drummer & Felser, 2018; Ackerman, 2015; Patterson & Felser, 2019; Xiang et al., 2009; Kush & Dillon, 2021) has not been able to tease apart whether the active processing of cataphors employs abstract prediction of a subject noun or the more conservative strategy proposed by Van Gompel and Liversedge (2003) because they have looked for mismatch effects at the head noun of a subject phrase. Mismatch effects at the head noun are compatible with either option. Testing whether cataphors can trigger abstract predictions requires a way to test for prediction in a region that precedes the main subject noun. Such a design is not possible in English, but as we show below, it is possible in Dutch.

2.3 The present study

In the present study, we test whether cataphors in preposed adjuncts trigger abstract prediction of a feature-matching noun in main subject position in Dutch. To test for prediction in a region before the main subject, our design exploits the fact that Dutch is a language with V2 word order in main clauses (e.g., Zwart, 2011). The V2 property of Dutch guarantees that in sentences with preposed subordinate clauses, the finite main clause verb must precede the main clause subject. This is illustrated in (5) where the main verb *sleepte* ('dragged') comes before the main subject *de assistent* ('the assistant').

(5) a. Nadat hij de sleutels had ingeleverd, **sleepte** de assistent
After he the keys had.SG turned_in **dragged.SG** the asssistant
de koffers van de muzikanten naar de lobby.
the suitcases of the musicians to the lobby
'After he had turned in the keys, the assistant dragged the musicians' suitcases to the lobby.'

b. Nadat zij de sleutels hadden ingeleverd, **sleepte** de assistent de
After they the keys had.PL turned_in **dragged.SG** the asssistant the
koffers van de muzikanten naar de lobby.
suitcases of the musicians to the lobby
'After they had turned in the keys, the assistant dragged the musicians' suitcases to the lobby.'

Dutch verbs agree with their subjects in number, so when reading sentences like (5), the number features of the main subject are first revealed upon seeing the main verb. Thus, we can test whether comprehenders predict a feature-matching antecedent in subject position by assessing whether comprehenders are surprised to see a main verb that mismatches the cataphor in number (as in 5b). If participants predict a feature-matching antecedent in subject position, we reason that this should trigger number agreement on the V2 main verb. If they have made this prediction, we expect

number mismatch effects on V2 verbs that do not match the cataphor in number. On the other hand, if comprehenders wait until the head noun is recognized before they commit to coreference between the cataphor and an NP in subject position (as suggested by Van Gompel & Liversedge 2003), we do not expect number mismatch effects at the V2 verb (or in any pre-nominal position).

3 Experiment 1: Offline next-word completion task

Before testing the real-time processing of cataphors, we carried out an offline completion task to investigate whether and to what extent cataphors in preposed adjuncts trigger an expectation of a number-matching main verb in V2 position.

3.1 Materials

We created 18 sentence fragments that consisted of a temporal adjunct clause formatted as the start of a sentence. Following Dutch grammar, the first word to follow the adjunct clause must be the finite main clause verb in V2 position (Zwart, 2011). The subject NP must follow the V2 verb. Three different conditions were created as follows. In two conditions the preposed adjunct was finite and the subject of the clause was a cataphoric pronoun. In one condition the cataphor was singular (6a), in the other it was plural (6b). In the third condition the adjunct was non-finite and the subject of the adjunct was null (6c). For convenience we refer to this null subject as PRO (Chomsky, 1981). The PRO subject of a preposed non-finite clause is obligatorily co-referent with the main subject that would follow the V2 main verb (Control Theory; Chomsky, 1981). There is no subject-verb agreement in Dutch non-finite clauses, so there were no overt cues to the number of the main subject in the null-subject condition. Thus, singular and plural main verb continuations were equally acceptable and no expectations for subject number could be triggered by elements in the preposed adjunct. An example item set is presented in (6).

(6) a. **Singular cataphor**
Nadat hij de sleutels had ingeleverd, ___
After he the keys had.SG turned_in,
'After he turned the keys in, ___'

b. **Plural cataphor**
Nadat zij de sleutels hadden ingeleverd, ___
After they the keys had.PL turned_in,
'After they turned the keys in, ___'

c. **Null subject**
Na de sleutels te hebben ingeleverd, ___
After the keys INF have.INF turned_in,
'After turning the keys in, ___'

In Dutch, the 3rd person singular feminine pronoun and the 3rd person plural pronouns are homophonous in both their strong (stressed) and their weak (unstressed, clitic) nominative forms (strong form: *zij,* weak form: *ze*). The different readings are disambiguated via number agreement on the verb. To minimize ambiguity in our experiment we only used the masculine pronoun in our Singular cataphor items. Thus, the ambiguous pronoun was only used in Plural cataphor condition. In the plural condition the adjunct-internal finite plural auxiliary verb (*hadden*) disambiguated to the plural reading of the pronoun. The finite auxiliary verb can either precede or follow the past participle in Dutch, but in our items the auxiliary always came before the participle so that the finite auxiliary and the V2 main verb were not linearly adjacent.

We made one other choice in the experiment to avoid another potential ambiguity relating to pronominal form: As noted above, the plural pronoun has both a strong and weak form. We chose to use the strong form *zij* in the experiment, because the weak form *ze* may be more prone to be interpreted as a generic (impersonal) pronoun (e.g., *They always check the train for lost items at the final station*), which is always a weak pronoun in Dutch. A generic interpretation could diminish a drive to find an explicit antecedent (Filik et al., 2008; Sanford et al., 2008).

Experimental items were distributed over 3 lists according to a Latin Square design and pseudo-randomly mixed among 46 filler items. Completion items varied in structure and degree to which they constrained the category and identity of the next word.

3.2 Task

Participants read sentence fragments one at a time and were instructed to then fill in the next word in a plausible completion of the sentence. We chose to limit responses to a single word instead of allowing participants to finish the whole sentence in order to: (i) encourage continuations that reflected participants' initial expectations by minimizing the amount of time participants had to deliberate, and (ii) expedite the task.

As mentioned above, if participants chose to continue test sentences by starting the main clause, the V2 rule in Dutch determines that the only grammatical option for a next word was a finite verb. We were interested in checking whether participants chose singular main verbs in the Singular cataphor condition, and plural main verbs in the Plural cataphor condition. The Null subject condition served as a baseline, to see whether participants had a baseline preference for either plural or singular main verbs/subjects in the absence of a cataphor that provided number information.

3.3 Procedure

The experiment was implemented with the online survey tool Nettskjema from the University of Oslo and distributed online via a link. Participants filled out a short language background questionnaire to confirm that they were active native speakers of Dutch. The questionnaire was

followed by instructions and an example trial. The sentence fragments were displayed one by one in the center of the screen, together with a text box in which participants could write their response. Once they had moved on to the next trial, they could not go back to view or change their response. After the experiment participants answered a few debriefing questions intended to identify any participants that did not perform the task in good faith or who had non-native proficiency in Dutch.

3.4 Participants

55 self-reported native speakers of Dutch participated in the experiment. All participants were volunteers recruited through various social media platforms. One participant was excluded for providing ungrammatical responses on 12.5 percent of trials. 6 participants were excluded because they wrote full-sentence completions contrary to the instructions. Data from the remaining 48 participants (Mean age: 39.0) was included in the analysis. One further ungrammatical response was excluded.

3.5 Results

Verb responses were coded for number: singular or plural. Grammatical non-verb responses (e.g., further modification of the preposed adjunct) were coded as 'other'. **Figure 2** displays the proportion of responses for each condition.

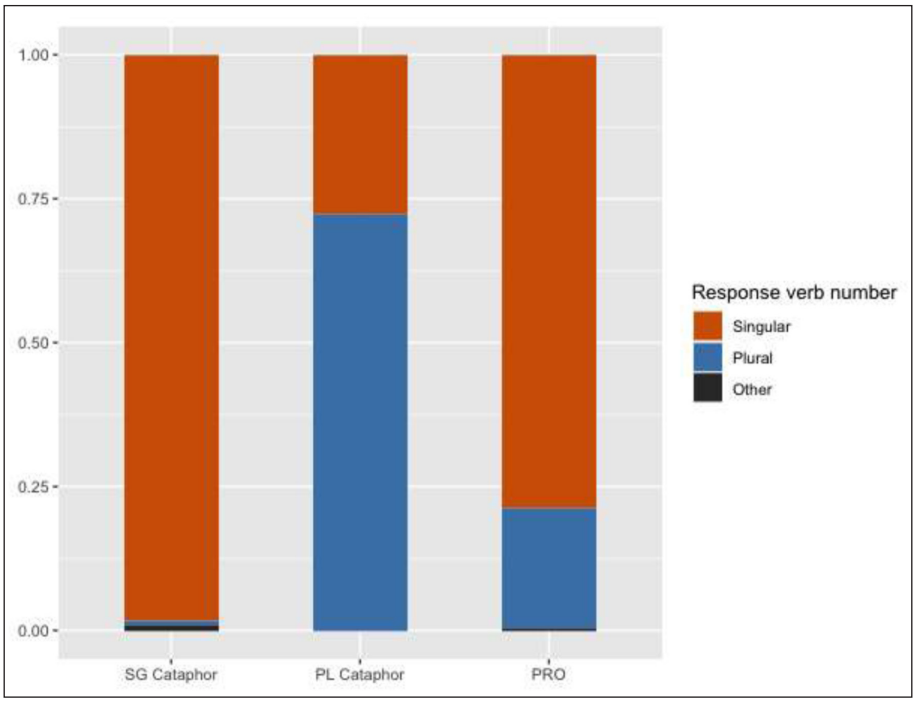

Figure 2: Proportion of responses by condition for the completion task in Experiment 1.

The preambles with cataphors showed a preference for responses in which the V2 main verbs matched the preceding cataphor in number. The preference for number-match was stronger with singular cataphors than plural cataphors. The preambles with singular cataphors elicited almost exclusively singular verbs (98.26 percent) versus 72.44 percent plural responses for the plural cataphors. For the null subject condition, we observed a preference for singular verbs (78.82 percent). The number of 'other' responses was negligible in all conditions (resp. 3, 0, and 1 response). A mixed effects logistic regression[2] confirmed that the response proportions were significantly different across conditions ($p < 0.001$ for all comparisons).

3.6 Discussion

The results show that comprehenders prefer main verbs that match the unresolved cataphor in number, consistent with the hypothesis that comprehenders anticipate a feature-matching antecedent for the cataphor in main subject position. The higher proportion of singular responses in the Null subject condition and the fact that roughly 27% of responses in the plural cataphor condition had singular verbs suggest a baseline preference for singular verbs.

A factor that might have contributed to the singular bias in the Null subject condition is the possibility that positing a singular referent is more economical. Entering a single referent into the discourse may be less complex than a plural referent, which may require more presuppositions than a singular referent (Patson, 2014). A parser that posits the simplest structure, using a least-effort strategy ('Minimal Everything', Fodor & Inoue, 1998) may prefer positing a singular referent, all else being equal. We should also note the possibility that a proportion of the singular completions may have been intended as first-person singular verbs, which would be consistent with participants predicting the first-person pronoun *ik* ('I') as the subject. We cannot rule out this possibility because, for most Dutch verbs, there is no distinctive morphology for person.

Regardless of what drives the number asymmetry, the fact that participants chose a main verb that matched a preceding cataphor in number at significantly greater-than-chance rates indicates that cataphors trigger an off-line expectation of a number-matching main verb, as predicted if they predict a feature-matching antecedent in main subject position. In the self-paced reading experiment below, we investigated whether we find online evidence for such a prediction in comprehension.

---

[2] Model: glmer(response ~ condition + (1 + condition | subject) + (1 | item), family = binomial). For the factor 'condition', the reference level was set as the Singular Cataphor condition. The model excluded the responses in the category 'other'.

4 Experiment 2: Self-paced reading

4.1 Materials and design

We constructed 24 sentences that began with a preposed temporal adjunct clause. A pronoun, the cataphor, was the subject of the preposed adjunct clause. The preposed clause was immediately followed by the main clause, which began with a tensed verb in V2 position. In a 2x2 design, we manipulated the factors CATAPHOR NUMBER (Singular/Plural) and NUMBER MATCH (Match/Mismatch), resulting in four conditions exemplified in (7) with slashes indicating how regions were divided.

(7) a. **Singular cataphor**
Nadat/ {hij}/ de sleutels/ had/ ingeleverd,/ {sleepte|sleepten}/ de
After {he} the keys had.SG turned_in dragged.SG|dragged.PL the
extreem/ chagrijnige/ {assistent|assistenten}/ de koffers/ van/ {de
extremely annoyed asssistant|assistants the suitcases of the
muzikanten| de muzikant}/ naar/ de lobby.
musicians the musician to the lobby
'After he had turned in the keys, the extremely annoyed assistant(s) dragged{SG/PL} the suitcases of the musician(s) to the lobby.'

b. **Plural cataphor**
Nadat/ {zij}/ de sleutels/ hadden/ ingeleverd,/ {sleepten|sleepte}/ de
After {they} the keys had.PL turned_in dragged.PL|dragged.SG the
extreem/ chagrijnige/ {assistenten|assistent}/ de koffers/ van/ {de
extremely annoyed asssistants|assistant the suitcases of the
muzikant| de muzikanten}/ naar/ de. lobby
musician the musicians to the lobby
'After they had turned in the keys, the extremely annoyed assistant(s) dragged{PL/SG} the suitcases of the musician(s) to the lobby.'

The adjunct clauses contained five regions so that comprehenders had enough time to predict a main subject antecedent before entering the main clause. The adjunct did not contain any other animate NPs to avoid introducing additional human referents. We selected main verbs that either could not take or did not require animate internal arguments to avoid the possibility that comprehenders would posit additional non-subject referents in the main clause, which could potentially serve as the antecedent. The main clause subject was a definite NP and consisted of three regions. The first region contained a determiner and an adverb (*de extreem* 'the extremely'). The second contained an adjective (*chagrijnige* 'grumpy') and the third was the head noun (*assistent/assistenten* 'assistant/s'). The two regions preceding the head noun did not contain any information about the number of the subject, so they could function as spillover regions for the main verb. The items contained another referent in a post-spillover region (*muzikant/*

muzikanten ‘musician/s’), to ensure that the cataphors were resolvable within the sentence, even in the Mismatch condition. As in the completion experiment, we only used the masculine pronoun (*hij*) in Singular conditions and the strong form of the plural pronoun (*zij*) in Plural conditions. Once again the finite plural auxiliary (*hadden*) disambiguated the pronoun *zij* to its plural interpretation.[3]

If participants do not posit an antecedent until the head noun is reached, we expect no effects at the V2 verb or spillover region. However, if participants predict a feature-matching antecedent NP in subject position, we expect a main effect of number match at or immediately following the V2 verb such that mismatch conditions should be read more slowly than match conditions. A CATAPHOR NUMBER × NUMBER MATCH interaction at the V2 verb would indicate that mismatch effects depended on cataphor number (e.g., if effects were more prominent with singular cataphors).

In addition to the 24 items that manipulated number-match, we constructed 12 independent test items in which we manipulated gender-match between a singular cataphor and a proper name in main subject position, as in (8). Slashes indicate region breaks. The critical subject region was followed by a spillover region consisting of an adverb.

(8) Nadat/ {zij|hij}/ de vliegtickets/ had/ besteld,/ schreef/ Diane/ meteen/
After {she|he} the plane_tickets had.SG booked wrote.SG Diane immediately
de datum/ van (Philips)/ aankomst/ op.
the date of Philip’s arrival up.
‘After she/he had booked the plane tickets, Diane immediately wrote down the date of (Philip’s) arrival.’

The gender items served as a control condition, so that we could establish that comprehenders were actively resolving the cataphors even if the number manipulation did not result in a mismatch effect. The gender items all had singular cataphors (6 masculine *hij*, 6 feminine *zij*) and singular main verbs/subjects. The cataphors always had a gender-matching referent in the main clause in both conditions, so that the cataphor was in principle always resolvable.

The items were distributed over 4 lists in a Latin Square design and added to a list of 55 fillers. 15 of the fillers had a structure similar to the test items. They started with a temporal adjunct clause that had a full DP as its subject, which was plausibly coreferent with a number-matching main clause subject pronoun (e.g., *When the lifeguards saw the dark clouds, they immediately tried to warn the reckless surfer*). The purpose of these fillers was to introduce variation in the sentences

[3] A reviewer asks whether number marking on the adjunct-internal auxiliary could have primed number-marking on the V2 verb. We cannot rule out this possibility, but we attempted to minimize any possible effects of immediate repetition priming in choosing where to place the auxiliary in the linear string. Dutch allows the auxiliary to either precede or follow the participle (e.g. *ingeleverd*) in such adjunct clauses. We chose to use the word order where the participle intervenes between the auxiliary and the V2 verb.

with temporal adjunct clauses, so that not all of them contained cataphors. Each sentence was followed by a yes/no comprehension question. For 8 of the number items, answering the question required interpreting the cataphor.

4.2 Procedure

The experiment was run on the online platform Ibex Farm (Drummond, 2013). Participants filled out a short questionnaire about their language background to confirm that they were native speakers of Dutch who spoke Dutch regularly. Trials started with a 1500 ms fixation cross. Upon pressing the spacebar, the sentence appeared chunk-by-chunk in the center of the screen.

After each sentence, a comprehension question appeared. Participants answered using the E ('yes') and I ('no') keys on their keyboard. Incorrect answers were followed by immediate feedback and a reminder to read carefully. For ambiguous questions, including some of the questions probing the interpretation of the cataphor, both 'yes' and 'no' were coded as correct responses, so that they were never followed by feedback. Following each 24th trial, a 'break'-screen appeared, encouraging participants to take a short break.

The experiment ended with a short debrief questionnaire to identify bots and non-native speakers. As part of this questionnaire, participants were asked to describe in a few sentences the most beautiful and the most ugly building on their street (Chmielewski & Kucker, 2020).

4.3 Participants

Prior to data collection we set an upper-bound of 160 eligible participants.[4] To reach this total we ran 179 participants on the Prolific Academic platform who were paid 4 GBP for their time. 2 of these participants were excluded because of reported technical difficulties, or because they appeared non-nativelike in the debrief questionnaire. 17 additional participants were excluded because they scored below 80% on comprehension question accuracy.

4.4 Analysis

The mean accuracy score of the 177 participants before exclusion was 85.7%. After exclusion, the mean accuracy score of the 160 remaining participants was 90.8%.

We excluded reaction times under 100 ms and over 3000 ms. This data trimming led to the exclusion of 321 data points (0.004% of the data set).

[4] The sample size is equal to the sample size of a pilot study with similar items. For this pilot, sample size was determined using a ROPE (Region of Practical Equivalence, see Kruschke, 2014 and Vasishth et al., 2018a). See Additional Materials for a description of the pilot study.

For the analysis of the reaction time data, we fitted Bayesian linear mixed effects regression models using the brms package (Buerkner, 2017) in the software environment R (R Core Team, 2013). The brms package can be used to fit Bayesian multilevel models using the programming language Stan (Gelman, Lee & Guo, 2015). We sum-coded the factors CATAPHOR NUMBER and NUMBER MATCH. We estimated main effects of CATAPHOR NUMBER, NUMBER MATCH, and their interaction. Pairwise comparisons were done using the emmeans package (Searle et al., 1980). For the verb region, we added scaled word length as a factor to account for the fact that the plural verbs were 1–3 characters longer than singular verbs. For the gender manipulation we sum-coded the levels Gender-Match and Gender-Mismatch as –0.5 and 0.5.

We fitted models with a 'maximal' random effect structure[5] with random slopes and intercepts for subject and item (Barr et al., 2013). We used a shifted log-normal link function to account for the non-normal distribution of RT data. For our fixed effects, we set weakly informative priors that did not make strong assumptions about the expected RT distributions. All fixed effects had normal priors with a mean value of 0, a variance of 1, and a variance of 10 for the intercept. For our random effects, we set a regularizing LKJ prior with $\eta = 2$. This is a prior that downweighs high values in the correlation matrices (Vasishth et al., 2018a; Vasishth et al., 2018b).

For each model, we ran 4 Monte Carlo Markov Chains (MCMC), each consisting of 6500 iterations (3250 warmup, 3250 sampling). For all models, we checked that the R-hat statistics were (very close to) 1.00, indicating that the chains converged. We also checked samples (n = 100) from the posterior against the data, assessing how well the posterior fit the data.

5 Results

5.1 Number manipulation

The results and analyses are plotted in **Figures 3–4**, and summarized in **Tables 1–2**. In the verb region, the model revealed an expected main effect of length and a main effect of cataphor number. The model revealed clear evidence that the verb region took longer to read for conditions with plural cataphors (Pr = 0.988). There was not strong evidence for a main effect of NUMBER MATCH (Pr = 0.764) or for a CATAPHOR NUMBER × NUMBER MATCH interaction (Pr = 0.739). In the first spillover region, we observed a main effect of match (Pr = 0.999). For both the sentences with singular and plural cataphors, the mismatch conditions yielded longer reading times. In the second spillover region, the model indicated a CATAPHOR NUMBER × NUMBER MATCH interaction. Pairwise comparisons using the emmeans package (Lenth, 2022) showed that the effect was driven by longer reading times for the Singular cataphor – Mismatch

[5] Models for the number manipulation were of the following form: RT ~ cataphorNumber * match + (1 + cataphorNumber * match | subject) + (1 + cataphorNumber * match | item). For the verb region, z-scaled length was added as a fixed effect, but length was omitted from the random effect structures.

condition compared to the Plural cataphor – Mismatch conditions (from Singular cataphor to Plural cataphor in the Mismatch condition: est. 0.042, 95% HDP 0.004 – 0.081; in the NUMBER MATCH conditions, est. –0.012, 95% HDP –0.052–0.025).

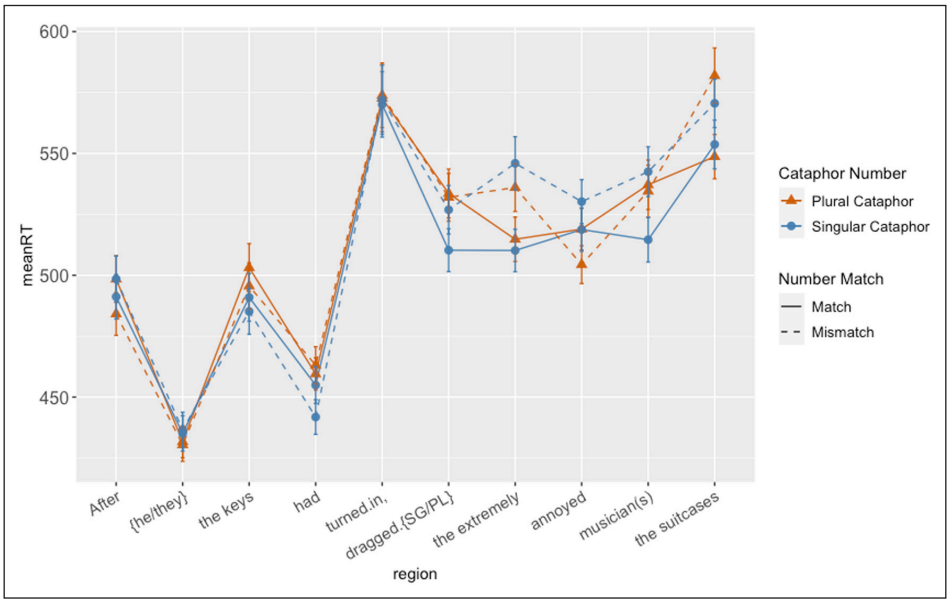

Figure 3: Average RTs per region for the critical number manipulation items in Experiment 2. Error bars represent standard error of the mean.

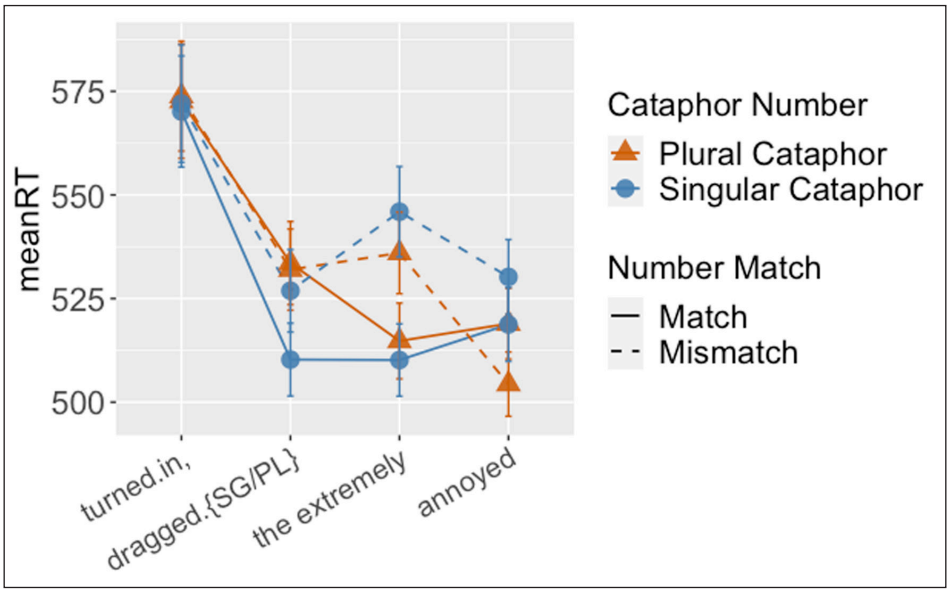

Figure 4: Closeup of the regions of interest for the critical number manipulation items Experiment 2. Error bars represent standard error of the mean.

Table 1: Average RTs (SE) in the regions of interest for the critical Number manipulation, Experiment 2.

	Verb	**Spillover 1**	**Spillover 2**
Singular – Match	510.28 (8.79)	510.17 (8.72)	518.75 (8.89)
Singular – Mismatch	526.88 (9.95)	545.94 (10.91)	530.22 (9.01)
Plural – Match	533.57 (10.03)	514.77 (9.13)	518.95 (8.54)
Plural – Mismatch	531.99 (9.80)	536.01 (9.83)	504.36 (7.76)

Table 2: Means and 95% CI of the posterior distributions for regions of interest in critical Number manipulation, Experiment 2.

	Estimate	**95% CI**	**Pr > 0**
Verb Region			
Cataphor Number	**0.03**	**0.00, 0.06**	**0.988**
Match	0.01	–0.02, 0.04	0.764
Cat. Number × Match	0.02	–0.05, 0.09	0.739
Length	**0.07**	**0.02 0.12**	**0.996**
Spillover Region 1			
Cataphor Number	0.00	–0.03, 0.03	0.528
Match	**0.05**	**0.02, 0.08**	**0.999**
Cat. Number × Match	-0.01	–0.08, 0.05	0.321
Spillover Region 2			
Cataphor Number	-0.01	–0.04, 0.01	0.135
Match	0.00	–0.03, 0.03	0.524
Cat. Number × Match	**–0.05**	**–0.11, 0.00**	**0.026**

5.2 Gender manipulation

The results from the gender manipulation are plotted in **Figure 5** and summarized in **Tables 3–4**. Results in the gender items were as expected. As shown in **Figures 5**, we observed a large gender-mismatch effect at the name region (43 ms, Pr = 0.999) and the spillover region (52ms, Pr > 0.999). The effect persisted past the analyzed spillover region, but we did not analyze any further regions.

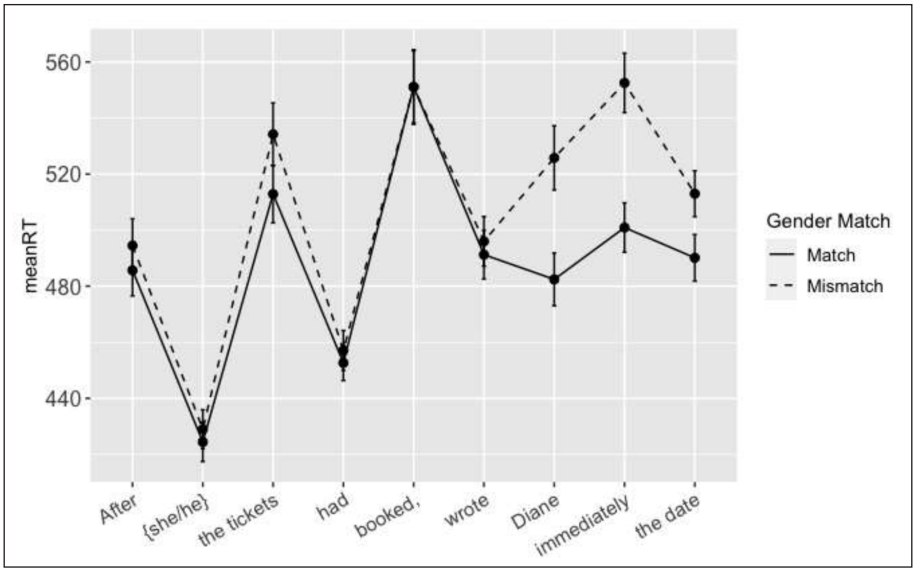

Figure 5: Average RTs per region for the control gender manipulation in Experiment 2. Error bars represent standard error of the mean.

Table 3: Average RTs (SE) for the regions of interest in control Gender manipulation, Experiment 2.

	Subject Noun	**Spillover 1**
Gender-Match	482.43 (9.40)	500.92 (8.82)
Gender-Mismatch	525.78 (11.45)	552.54 (10.61)

Table 4: Means and 95% CI of the posterior distributions for the regions of interest in control Gender manipulation, Experiment 2.

	Estimate	**95% CI**	**Pr > 0**
Verb Region			
Gender Match	0.07	0.03, 0.11	0.999
Spillover Region 1			
Gender Match	0.10	0.05, 0.16	> 0.999

5.3 Discussion

First, we found a strong gender-mismatch effect in the conditions that manipulated gender match between a cataphor and the main subject NP. The results replicate previous research and constitute strong evidence that comprehenders were actively attempting to complete the

cataphoric dependency with the subject noun when the preceding main verb matched the number of the cataphor.

The main novel finding of our self-paced reading experiment was a mismatch effect for the number manipulation. Participants slowed down in the region following the critical main verb when it mismatched the number features of the cataphor. We interpret this mismatch effect as evidence that comprehenders expected the antecedent of the cataphor to be the main subject and consequently predicted that the yet-to-be-seen subject noun bore features that matched the cataphor.

In the main V2 verb region, we found a main effect of CATAPHOR NUMBER: reading times were longer in conditions where the cataphor was plural than when it was singular, irrespective of the number-marking on the verb. This difference cannot be readily explained as an effect of prediction. It is possible that the effect at the verb is a spillover effect of processing the end of the preposed adjunct clause. There are several factors that could have increased processing difficulty for the adjuncts containing a plural cataphor. It is possible that the plural cataphors required more effort to process and integrate because they are compatible with – or ambiguous between – various different interpretations (e.g., Patson & Ferreira, 2009; Patson et al., 2014). The singular bias that we observed in the completion study might be caused by similar factors.

Another possibility is that the increased reading times are related to the temporary ambiguity of plural pronouns. As noted in the discussion of our materials, the form *zij* is ambiguous between a singular–feminine pronoun and a plural pronoun. Thus, when participants first read *zij* in the preposed adjunct, they could not be sure whether the pronoun was singular or plural. Disambiguation to the plural reading first came near the end of the adjunct clauses, at the tensed auxiliary (*had/hadden*), two regions before the main verb. If disambiguation was costly, or if participants initially interpreted *zij* as singular and were forced to reanalyze, it is possible that that difficulty spilled over onto the main verb region. We note that because the singular cataphors in the number manipulation were always masculine, the ambiguity did not arise in those conditions.

In the second spillover region, we observed an interaction indicating that this region was read faster in the Mismatch conditions when the verb was plural than when the verb was singular. We do not have a definitive explanation for this effect. We see two possible explanations. One explanation is that predictions of a matching antecedent triggered by plural cataphors are 'weaker' or less reliable than predictions triggered by singular cataphors. Under this interpretation, mismatch effects are expected to be less disruptive. This interpretation is potentially consistent with the results from Experiment 1, where plural cataphors elicited 27.6% singular verb responses. Our choice for strong pronouns may also have contributed to a lesser

degree of commitment in the conditions with plural cataphors. Although the exact conditions for using strong and weak pronouns in Dutch is debated (Kaiser, 2011), strong pronouns can be used in contexts of contrastive focus, which may have led to a weaker expectation for a coreferent subject.

An alternate interpretation is that predictions are equally strong, regardless of cataphor number, but that recovering from a mismatch requires less effort after a plural cataphor. Recovery from a violated expectation in the plural cataphor-mismatch condition requires participants to accommodate an NP that corresponds to a new singular referent. In the Singular cataphor-Mismatch condition, readers have to accommodate a new plural referent. Accommodating a singular referent may be easier in the context of a previous plural, which implies the existence of multiple referents, than accommodating a plural referent after a singular cataphor, which does not imply the existence of anyone else in the discourse.

6 General discussion

We investigated whether cataphors can trigger abstract prediction of features on yet-to-be-seen nouns. Previous work on prediction at a morphosyntactic level relied on highly constraining contexts to trigger a prediction of a specific lexical item (or a restricted set of lexical items) in object position of a transitive verb. We tested whether comprehenders can predict grammatical features of a noun in main subject position, driven by the pressure to complete a cataphoric dependency. Earlier studies showed that comprehenders actively attempt to establish coreference between a cataphor in a preposed clause and the subject of the subsequent main clause, but the work did not establish whether active completion reflects an abstract prediction that the antecedent was in main subject position.

We reasoned that if participants predicted that the antecedent of the cataphor was the main subject of the sentence, they would posit a corresponding noun in subject position. The lexical content of the noun could not be predicted, but morphological features on the noun such as number could be, as those features were supplied by the cataphor.

We looked for evidence of such abstract prediction in regions that preceded the subject noun by looking at Dutch sentences with V2 word order. In Dutch, preposed clauses are followed by the finite main verb in V2 position, which precedes the main subject. Because Dutch has subject-verb agreement for number, we assumed that predicting a feature-matching antecedent in subject position would entail matching number features on the V2 main verb.

In a completion study we asked participants to predict the next word following a preposed adjunct that contained a singular or plural cataphor. Consistent with the idea that they would predict a feature-matching antecedent in subject position, participants overwhelmingly produced V2 verbs that matched the preceding cataphor in number.

In a self-paced reading experiment we manipulated number-match between a cataphor and the finite main clause verb, which occurred in V2 position. We reasoned that if comprehenders predicted the antecedent in main subject position, they should be surprised to encounter a main verb whose number features were inconsistent with that prediction. In our study we observed the predicted number-mismatch effect at the spillover region immediately after the V2 verb. Because the spillover region still came before the subject noun, we took the effect as evidence for abstract prediction of number features on the unseen noun heading the subject NP.

6.1 Abstract vs. lexically mediated prediction

We take the effects above to indicate that comprehenders predicted that the antecedent of the cataphor was the main subject of the sentence and that making such a prediction involves projecting a noun in subject position. This prediction should be *abstract*: The comprehender would not pre-activate a specific noun (e.g., *boy*) or a highly constrained set of nouns, as argued for in previous studies. Instead the prediction would be for a 'dummy' N head that bears the morphological features matching the cataphor (in this case number at the very least).

One might argue that our data could, in principle, be explained by a model in which prediction involved pre-activating a specific noun, even in the absence of a highly constraining context. Comprehenders could have predicted, for example, specific lexical items corresponding to basic nouns such as *the man, the women, the person,* etc. in subject position. We deem this possibility unlikely for two reasons.

The first reason is conceptual. The preposed adjuncts that formed the context for our subject nouns described human referents engaging in everyday activities that can plausibly be executed by most human referents. Thus, a large number of potential nouns were compatible with our contexts. Predicting a single lexical item in such relatively unconstraining contexts would result in a highly inefficient processing mechanism. As illustrated by work on garden-path effects, there is a cost to reanalysis after incorrect predictions (e.g., Frazier & Rayner, 1982; Husband & Bovolenta, 2020). Under processing models that assume a limited working memory available for incremental processing (e.g., Frazier & Fodor, 1978; Lewis et al., 2006; Christiansen & Chater, 2016), a mechanism that commits to specific lexical predictions when such predictions are likely to be wrong, is suboptimal (Huettig, 2015).

A second, indirect, argument against predicting a single lexical noun in less constraining contexts comes from the Szewczyk & Schriefers (2013) study discussed above. As a reminder, in that study participants read stories that strongly biased towards a completion with either an animate or inanimate noun, but not a single noun in particular. The authors reasoned that if comprehenders were to predict a specific word in their experiment, the probability of each specific word being predicted should correspond to its cloze probability in the given context. If comprehenders made such a prediction, the size of the N400 associated with pre-nominal input

that was inconsistent with that specific word should correlate with its cloze probability. In a *post hoc* analysis, Szewczyk and Schriefers found no evidence for such results.[6]

Although we think that comprehenders do not predict the identity of the upcoming noun, one could argue that participants pre-activate number morphemes like the plural suffixes *-en* or *-s*. If these morphemes are considered separate lexical entries from the nouns to which they attach, then there could be a degree of lexical mediation at play. We do not think, however, that pre-activation of number-marking morphemes like the plural suffix undermines our claim that prediction of an abstract head noun is required. Any account that allows for the pre-activation of a suffix seems, to our mind, to presuppose the existence of a noun for the suffix to attach to. Moreover, if morphological agreement is controlled by nouns, the suffixes must attach to a noun in order to guarantee subject-verb agreement. First, we point out that the mere activation of a number morpheme alone is not expected to affect subject-verb agreement. As a final note, we point out that if our effects are mediated through pre-activation of number morphemes, pre-activation cannot be restricted to overt morphemes. Such an account will have to posit an abstract singular morpheme in addition to the overt plural markers *-en* and *-s*, otherwise we should only see mismatch effects with plural cataphors, contrary to fact.

6.2 Implications for the timing of abstract prediction

The mismatch effect in the number manipulation indicated that comprehenders predicted the subject number before reaching the head noun. This supports the hypothesis that cataphors can trigger a prediction for a noun in subject position and that the predicted noun bears morphosyntactic features that match the cataphor's. As we noted in the introduction, researchers interested in cataphor resolution have proposed different models that make different assumptions about the time-course and the extent of antecedent prediction during cataphor processing. Conservative models of active cataphor resolution (Van Gompel & Liversedge, 2003 and, to some extent, Brasoveanu & Dotlačil 2020) propose little to no role for prediction in the process of positing coreference between a cataphor and a main subject, holding that comprehenders wait until the head noun of the subject is recognized in the bottom-up input before they assume coreference. That is, the models hold that comprehenders do not predict feature-matching nouns. Our results are not consistent with these more conservative models. Comprehenders clearly do predict an antecedent in subject position before the head noun of the subject NP is ever encountered.

Our results show that prediction is made before the noun, but they do not determine decisively when the prediction occurs. The results are compatible with two different possibilities, which we name *pre-verbal* prediction and *verb-triggered* prediction.

[6] Szewczyk and Schriefers offer this argument as evidence against prediction of a specific word, but we think the argument works equally well against their model of set-based lexical prediction. To our mind, the results of the post-hoc analysis are most compatible with abstract prediction of simply animate or inanimate features.

The first option, *pre-verbal* prediction, posits that comprehenders project an abstract feature-matching antecedent in subject position before reaching the V2 verb. The prediction could come at any point in the preposed clause, potentially even at the moment the cataphor is encountered. Under this hypothesis, subject-verb agreement could be carried out, in the abstract. The mismatch effect would then reflect a violation of predicted verb agreement features.

The second option is *verb-triggered* prediction. Under this hypothesis, comprehenders would abstain from predicting a feature-matching antecedent until they encountered the V2 verb, at which point they would project the antecedent in main subject position. After projecting the antecedent, comprehenders could 'check' whether their projected subject was compatible with the number marking on the verb. The number mismatch effect would reflect, on this interpretation, realizing that there is a clash between the subject and the verb's number features. We note that for this option to work, prediction of the antecedent in subject position would have to happen immediately after the verb was recognized, but before number agreement was checked. To be concrete, the only way to produce a number mismatch effect is to predict a subject noun whose number features are incompatible with the overt number marking morphology on the verb. If comprehenders waited until after number morphology on the verb was checked to predict the subject, then there would be no reason to predict a mismatching subject NP. Thus, the verb-triggered prediction option commits to (i) a two-step model of verb processing that separates recognition of the verb from agreement-checking and (ii) the possibility that higher-level syntactic predictions can be ordered between these two steps.

We note that an analogous debate surrounds when the parser predictively posits (object) gaps during active filler-gap resolution. When incrementally processing sentences like (9), it has been shown that comprehenders actively predict that the gap corresponding to the filler *who* will be in object position at least by the time they encounter the verb *bring* (Stowe 1986, Traxler & Pickering, 1996, a.o.).

(9) Borghild wondered who Tor would bring …

The results are consistent with pre-verbal prediction: comprehenders may predict that the gap is the object of the upcoming verb before seeing *bring*. Alternatively, prediction could wait until the verb has been recognized. In a clever series of studies Omaki et al., (2015) attempted to tease the two possibilities apart by testing whether participants were surprised to encounter an intransitive verb (e.g., *chat*) during active gap-filling, which would be incompatible with a pre-verbal prediction of an object gap. The authors found that participants were surprised to see an intransitive verb, consistent with the predictions of the pre-verbal prediction account. However, they acknowledged that the results were also consistent with a variant of a verb-driven account where the parser projected an object gap as soon as a verb was recognized, but before the transitivity information of that verb was processed. Although the verb-driven account is in

principle compatible with the English filler-gap results, there is some evidence from filler-gap processing in verb-final languages that comprehenders commit to gap positions well before the verb is encountered (Aoshima et al., 2004). Insofar as expectation-driven prediction strategies are similar across different dependency types (Kazanina et al. 2007; Giskes & Kush 2021), the possibility of pre-verbal prediction in filler-gap processing may provide converging evidence for the possibility of pre-verbal prediction in active dependency resolution more generally.

6.3 Factors driving abstract prediction

As many have observed, encountering a cataphor seems to trigger a search for an antecedent in a later position. This search presumably reflects a pragmatic pressure to link pronouns to established entities in the discourse. We and other researchers have shown that comprehenders predict that that antecedent will be in subject position, even though the antecedent could, in principle, be found elsewhere in the sentence. Why should comprehenders predictively commit to subject position? There are various factors that conspire to make the main subject position a good candidate for an antecedent position. Subject NPs are syntactically required, so the position is guaranteed and can therefore be constructed in advance based on top-down knowledge. Furthermore, pronouns prefer antecedents that are topics or prominent (e.g., Givón, 1983; Ariel, 1990). As a canonical position for topics (e.g., Gordon & Hendrick, 1998; Grosz et al., 1995), subject position is a particularly felicitous place to expect an antecedent in. Subject position is also linearly one of the first grammatically licit antecedent positions, and it seems to be the most frequent antecedent position for cataphors in grammatical constructions like our test sentences (Crawley et al., 1990; Hobbs, 1978).

Our results are thus compatible with different theories of what drives prediction. They are in line with models that assume ‘least effort’ heuristics, such as a principle to resolve dependencies by positing the least structure possible (Frazier & Fodor, 1978; Fodor & Inoue, 1998), or a drive to complete dependencies as quickly as possible (e.g., Fodor, 1978; Stowe, 1986; Frazier, 1987; Frazier & Flores d’Arcais, 1989; Omaki et al., 2015). The results are also compatible with constraint-based models that propose prediction as a phenomenon sensitive to a broad range of (weighted) cues (e.g., Kazanina & Phillips, 2010; Keshev & Meltzer-Asscher, 2020; Federmeier, 2007). These models may propose that abstract prediction needs to be warranted by a combination of relevant cues that is sufficiently strong.

For the interpretation of non-cataphoric pronouns, it has been shown that comprehenders are indeed sensitive to factors such as coherence relations and information structure (Kehler & Rohde, 2019). But even if these factors, which generally favor a subject antecedent, are controlled for, comprehenders seem to have a general preference for subject antecedents (Kehler & Rohde, 2013). It is an issue for further research whether abstract prediction is confined to subject position, and whether it is sensitive to these factors.

7 Conclusion

In a self-paced reading experiment, we tested whether comprehenders can predict an abstract noun phrase in an upcoming argument position and assign morphosyntactic features to the nominal head in contexts where pre-activation of a specific lexical head is unlikely. We found evidence that cataphors can trigger prediction of a noun phrase specified for number features in main subject position: participants were surprised to encounter a V2 verb that mismatched the cataphor in number, suggesting that subject-verb agreement was controlled by the predicted abstract subject. Our findings suggest that morphosyntactic prediction does not need to be lexically mediated. In addition, results suggest that the prediction of the antecedent for the cataphor took place before readers encountered the head of the subject NP. This is inconsistent with processing models that view the prediction of features as a mechanism that is triggered by encountering a syntactic position bottom-up.

Data Accessibility Statement

The materials, data, and scripts for analysis are available online at https://osf.io/kcqt9/?view_only=962024c18a7c40698cb6d3253c4939a8. This repository also contains a description and analysis of a pilot study that was a self-paced reading experiment based on the materials from the completion study (Experiment 1). This pilot formed the basis for the experiment reported in the current paper as Experiment 2, which has an improved design.

Ethics and consent

All experiments reported in this study were conducted in compliance with the Declaration of Helsinki and was considered exempt from review by the Norwegian Centre for Research Data.

Acknowledgements

Special thanks to Brian Dillon for extended discussion and to the UMASS Linguistics community. Thanks as well to Anastasia Kobzeva, Charlotte Sant, and audiences at HSP 2022 and University of Toronto, Scarborough.

Competing interests

The authors have no competing interests to declare.

Author contributions

Anna Giskes: Conceptualization, Formal Analysis, Investigation, Writing – Original Draft preparation. **Dave Kush:** Conceptualization, Supervision, Writing – Reviewing & Editing

References

Ackerman, L. M. (2015). *Influences on parsing ambiguity* [Unpublished doctoral dissertation]. Northwestern University.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*(1), 23–54. DOI: https://doi.org/10.1016/j.jml.2004.03.001

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Brasoveanu, A., & Dotlačil, J. (2020). *Computational cognitive modeling and linguistic theory* (p. 294). Springer Nature. DOI: https://doi.org/10.1007/978-3-030-31846-8

Buerkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*(4), 464–473. DOI: https://doi.org/10.1177/1948550619875149

Chomsky, N. (1981). *Lectures on government and binding.* Foris, Dordrecht.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39,* E62. DOI: https://doi.org/10.1017/S0140525X1500031X

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19*(4), 245–264. DOI: https://doi.org/10.1007/BF01077259

Drummer, J. D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language, 101*, 97–113. DOI: https://doi.org/10.1016/j.jml.2018.04.001

Drummond, A. (2013). Ibex farm. *Online server: http://spellout. net/ibexfarm*. [now defunct]

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491–505. DOI: https://doi.org/10.1111/j.1469-8986.2007.00531.x

Filik, R., & Sanford, A. J. (2008). When is cataphoric reference recognised? *Cognition, 107*(3), 1112–1121. DOI: https://doi.org/10.1016/j.cognition.2007.11.001

Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 20*(7), 1315–1326. DOI: https://doi.org/10.1162/jocn.2008.20090

Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition, 204*, 104335. DOI: https://doi.org/10.1016/j.cognition.2020.104335

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9, 427–473.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In Fodor, J. & Ferreira, F. (Eds.). *Reanalysis in sentence processing* (pp. 101–141). Springer, Dordrecht. DOI: https://doi.org/10.1007/978-94-015-9070-9_4

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 14–61. DOI: https://doi.org/10.1037/a0036756

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory, 5*, 519–559. DOI: https://doi.org/10.1007/BF00138988

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory & Language, 28*, 331–344. DOI: https://doi.org/10.1016/0749-596X(89)90037-5

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. DOI: https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178–210. DOI: https://doi.org/10.1016/0010-0285(82)90008-1

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, *40*(5), 530–543. DOI: https://doi.org/10.3102/1076998615606113

Giskes, A., & Kush, D. (2021). Processing cataphors: Active antecedent search is persistent. *Memory & Cognition*, *49*, 1370–1386. DOI: https://doi.org/10.3758/s13421-021-01176-z

Givón, T. (1983). *Topic continuity in discourse*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tsl.3

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, *22*(4), 389–424. DOI: https://doi.org/10.1207/s15516709cog2204_1

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225. DOI: https://doi.org/10.21236/ADA324949

Hobbs, J. R. (1978). Resolving pronoun references. *Lingua*, *44*(4), 311–338. DOI: https://doi.org/10.1016/0024-3841(78)90006-2

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–35. DOI: https://doi.org/10.1016/j.brainres.2015.02.014

Husband, E. M., & Bovolenta, G. (2020). Prediction failure blocks the use of local semantic context. *Language, Cognition and Neuroscience*, *35*(3), 273–291. DOI: https://doi.org/10.1080/23273798.2019.1651881

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*, 107–291. DOI: https://doi.org/10.1016/j.neuropsychologia.2019.107291

Kaiser, E. (2011). Salience and contrast effects in reference resolution: The interpretation of Dutch pronouns and demonstratives. *Language and Cognitive Processes*, *26*(10), 1587–1624. DOI: https://doi.org/10.1080/01690965.2010.522915

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384–409. DOI: https://doi.org/10.1016/j.jml.2006.09.003

Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, *63*(2), 371–400. DOI: https://doi.org/10.1080/17470210902974120

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1/2), 1–37. DOI: https://doi.org/10.1515/tl-2013-0001

Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, *154*, 63–78. DOI: https://doi.org/10.1016/j.pragma.2018.04.006

Keshev, M., & Meltzer-Asscher, A. (2020). The effects of syntactic pressures and pragmatic considerations on predictive dependency formation. *Language, Cognition and Neuroscience*, *35*(2), 256–272. DOI: https://doi.org/10.1080/23273798.2019.1650946

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. DOI: https://doi.org/10.1080/23273798.2018.1524500

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. New York, NY: Academic Press. DOI: https://doi.org/10.1016/B978-0-12-405888-0.00008-8

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, cognition and neuroscience*, *31*(1), 32–59. DOI: https://doi.org/10.1080/23273798.2015.1102299

Kush, D., & Dillon, B. (2021). Principle B constrains the processing of cataphora: Evidence for syntactic and discourse predictions. *Journal of Memory and Language*, *120*, 104–254. DOI: https://doi.org/10.1016/j.jml.2021.104254

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195395518.003.0065

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. DOI: https://doi.org/10.1016/j.jml.2009.06.004

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. DOI: https://doi.org/10.1162/jocn_a_00328

Lenth, R. (2022). emmeans: estimated marginal means, aka least-squares means. R package version 1.4. 7. 2020.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. DOI: https://doi.org/10.1016/j.tics.2006.08.007

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. DOI: https://doi.org/10.1111/j.1467-9280.2007.01871.x

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1–9. DOI: https://doi.org/10.1038/s41598-018-19499-4

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*, 384. DOI: https://doi.org/10.3389/fpsyg.2015.00384

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*(1), 1–9. DOI: https://doi.org/10.1186/1471-2202-8-89

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming?. *Discourse Processes*, *45*(6), 464–496. DOI: https://doi.org/10.1080/01638530802356463

Patson, N. D. (2014). The processing of plural expressions. *Language and Linguistics Compass*, *8*(8), 319–329. DOI: https://doi.org/10.1111/lnc3.12085

Patson, N. D., & Ferreira, F. (2009). Conceptual plural information is used to guide early parsing decisions: Evidence from garden-path sentences with reciprocal verbs. *Journal of Memory and Language*, *60*(4), 464–486. DOI: https://doi.org/10.1016/j.jml.2009.02.003

Patson, N. D., George, G., & Warren, T. (2014). The conceptual representation of number. *Quarterly Journal of Experimental Psychology*, *67*(7), 1349–1365. DOI: https://doi.org/10.1080/17470218.2013.863372

Patterson, C., & Felser, C. (2019). Delayed application of binding condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research*, *48*(2), 453–475. DOI: https://doi.org/10.1007/s10936-018-9613-4

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002. DOI: https://doi.org/10.1037/bul0000158

R Core Team, (2013). R: A language and environment for statistical computing.

Sanford, A. J., Filik, R., Emmott, C., & Morrow, L. (2008). Short article: They're digging up the road again: The processing cost of institutional they. *Quarterly Journal of Experimental Psychology*, *61*(3), 372–380. DOI: https://doi.org/10.1080/17470210701634552

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means, *The American Statistician*, 34:4, 216–221. DOI: https://doi.org/10.1080/00031305.1980.10483031

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. DOI: https://doi.org/10.1080/01690968608407062

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314. DOI: https://doi.org/10.1016/j.jml.2012.12.002

Szewczyk, J. M., & Wodniecka, Z. (2020). The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1714. DOI: https://doi.org/10.1037/xlm0000835

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. DOI: https://doi.org/10.1006/jmla.1996.0025

Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Palgrave Macmillan.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443. DOI: https://doi.org/10.1037/0278-7393.31.3.443

Van Gompel, R. P., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 128. DOI: https://doi.org/10.1037/0278-7393.29.1.128

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018a). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. DOI: https://doi.org/10.1016/j.jml.2018.07.004

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018b). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–61. DOI: https://doi.org/10.1016/j.wocn.2018.07.008

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. DOI: https://doi.org/10.1162/0898929041920487

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55. DOI: https://doi.org/10.1016/j.bandl.2008.10.002

Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience*, *29*(7), 761–770. DOI: https://doi.org/10.1080/01690965.2013.788196

Zwart, J. W. (2011). *The syntax of Dutch*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511977763