

Dutch speakers take referent predictability into account, irrespective of addressee presence

Jorrig Vogels, Center for Language and Cognition (CLCG), University of Groningen, NL, jorrigvogels@gmail.com

Language comprehension involves continuously making predictions about what will be mentioned next. If speakers take these predictions into account, one would expect that they try to be extra clear (e.g., by saying “the girl with the big earrings”) when they are going to say something less predictable. Conversely, speakers do not need to be as clear when the listener already expects the thing that they are about to mention, and can therefore suffice with a pronoun such as *she*. Previous research testing this hypothesis has found mixed results, with some studies finding that the referent’s predictability in discourse affects pronoun use and others finding that it does not. One explanation might be that speakers are more likely to take predictability into account when there is a co-present addressee who is predicting the next referent. To test this possibility, I conducted a language production experiment in which participants produced spoken continuations of narrative fragments. The fragments were accompanied by pictures that made clear how the story continued. Half of the participants performed the task without anyone else being present, while the other half told the stories to another person, who had to pick out the correct picture. Referent predictability was varied by manipulating the coherence relation in the narrative context. In addition, I calculated a *surprisal* score for each character in each narrative, as a more direct measure of its predictability. The results showed that with higher predictability, speakers were indeed more likely to use a pronoun than a definite NP to refer to the target character in their continuations. However, it did not matter whether the speaker was telling the stories to a co-present addressee or not. The results are discussed in light of accounts that distinguish between taking the perspective of a specific and that of a hypothetical listener.



1. Introduction

A common view in psycholinguistics is that language use is a joint activity, in which speakers and listeners coordinate with each other to match the listener's interpretation to the speaker's intention (also called *audience design*; Clark, 1996; Clark & Murphy, 1982). This view suggests that both speakers and listeners regularly take each other's perspective to make sure they are still on the same page. Although there is ample experimental evidence supporting this view (e.g., Brennan et al., 2010; Brown-Schmidt & Hanna, 2011; Galati et al., 2013; Hanna & Tanenhaus, 2004; Heller et al., 2008), other experimental research suggests that there are limits to when and how perspective taking takes place (e.g., Caffarra et al., 2020; Epley et al., 2004; Fukumura, 2015; Fukumura & van Gompel, 2012; Keysar et al., 2000; Kronmüller & Guerra, 2020; Wu & Keysar, 2007). Perspective taking may be a cognitively effortful process, which is dropped under time pressure or cognitive load, with egocentric language processing seeping through (Horton & Keysar, 1996; Kronmüller & Barr, 2007; Wardlow, 2013; see also Hendriks et al., 2014; Van Rij et al., 2010). In this study, I address the question to what degree expectations from the listener's perspective about what will be mentioned next inform the speaker's choice of a certain type of referring expression.

There is accumulating evidence that in comprehending a sentence, listeners actively anticipate upcoming material based on the (linguistic as well as non-linguistic) context (e.g., Altmann & Kamide, 1999; Bosker et al., 2014; Grüter & Rohde, 2021; Hall et al., 2018; Kuperberg & Jaeger, 2016; Kutas et al., 2011; Otten & Van Berkum, 2009; Pickering & Garrod, 2007; see Pickering & Gambi, 2018, for a review). If speakers take predictions about what will be mentioned next into account, one would expect that they try to be extra clear when they are going to say something less predictable (cf. Arnold, 2008). Conversely, speakers do not need to be as clear when the listener already knows what they are about to say. Before continuing with the present study, I will first review some of the evidence for the role of predictability in language production.

2. The role of predictability in language production

Since speakers will most likely want their utterance to contribute new information, they will not merely choose those words that are most predictable at each point in their utterance (Rohde et al., 2021). However, certain linguistic choices speakers make, especially below the message level, could be influenced by predictability. On the lexical level, for example, Lieberman (1963) found that the word *nine* is pronounced with a shorter duration when it is highly predictable in the context, such as in the sentence *A stitch in time saves nine*, than when it is less predictable, as in *The next number is nine* (see also Hall et al., 2018; Watson et al., 2008). In addition, shortened word forms (such as *math* for *mathematics*) are used more often in a predictive context (Mahowald et al., 2013; Zarcone & Demberg, 2021). Similarly, when pointing out an object, people often mention its color, but less so when the object's color is predictable (e.g., Rubio-Fernández, 2016;

Westerbeek et al., 2015). On the syntactic level, the *that*-complementizer in English has been found to be more frequently omitted when a complement clause is highly likely to follow, which may be analyzed as an effect of syntactic predictability (Jaeger, 2010).

More generally, several researchers have proposed that whenever the language permits a choice between a more prominent and a more reduced linguistic form to express the same content, speakers tend to use the more prominent form for less predictable information, and the more reduced form for more predictable information (e.g., Aylett & Turk, 2004; Levy & Jaeger, 2007). This division ensures more efficient communication, balancing speaker effort on the one hand and informativeness on the other (see also Frank & Goodman, 2012; Piantadosi et al., 2012; Winters et al., 2018).

In the same vein, some researchers have proposed that whenever there are multiple ways to pick out a referent in a discourse, speakers choose a more reduced referring expression, such as a pronoun, when the mention of this referent is highly predictable, and a more elaborate referring expression when the referent is not predictable (e.g., Arnold, 2008, 2010; Tily & Piantadosi, 2009). However, this idea is heavily debated, because whereas some psycholinguistic studies found evidence for it (Arnold, 2001; Rosa & Arnold, 2017), other studies did not (e.g., Fukumura & Van Gompel, 2010; Hoek et al., 2021; Rohde & Kehler, 2014; Stevenson et al., 1994). For example, Rosa and Arnold (2017) found that referents with a goal thematic role (e.g., *the maid* in *The chef handed a cookbook to the maid*), which are generally more likely to be mentioned next, were also more likely to be pronominalized. Fukumura and Van Gompel (2010), however, did not find a similar tendency with next-mention biases arising from implicit causality in stimulus-experiencer verbs: a subject with a stimulus role (e.g., *Gary* in *Gary scared Anna after the long discussion ended in a row, because...*) was more likely to be mentioned next but not more likely to be pronominalized compared to a subject with an experiencer role (*Gary feared Anna...*). Similar mixed results have been found in corpus studies, where some found effects of predictability on pronoun use (Aina et al., 2021; Tily & Piantadosi, 2009), while others did not (Modi et al., 2017). Other researchers have proposed that predictability only affects acoustic reduction, and not the choice of a more reduced lexical item (Kaiser et al., 2011). That is, more predictable referents receive shorter or otherwise less prominent pronunciations, whereas the choice between different lexical forms, such as pronouns vs. definite NPs, is unaffected. This view would entail that predictability only affects referential choices as far as they are articulatory in nature, such as stressed versus unstressed pronouns.

To investigate the divergent findings in the experimental literature, Weatherford and Arnold (2021) revisited the implicit causality biases used in earlier studies, making several methodological changes that might make it easier to detect effects of these biases on referring expression choice. Firstly, they contextualized the usually one-sentence continuation fragments of previous studies, by adding an introductory context sentence as well as pictures of the characters in the story.

They argued that a more contextualized continuation task may encourage the creation of a richer discourse representation, which may in turn facilitate the activation of semantic next-mention biases. Support for this view was recently provided by Demberg et al. (2023), who found a predictability effect on pronoun use, but only in the context of a story. Secondly, Weatherford and Arnold controlled which event participants would be talking about in their continuations, by instructing them to use one of two textually presented facts about the story characters. In contrast to typical story continuation experiments, in which participants plan their continuations on the fly as they read the story fragments, speakers thus knew beforehand what they were going to say. This matches the time course of natural language production more closely, since speakers are likely to have a plan of the message they want to convey before they start speaking.

Using these methodological adjustments, Weatherford and Arnold (2021) were able to show an effect of implicit causality biases on pronoun production, with more pronouns produced for the referent that formed the implicit cause than for the non-implicated referent. These findings would be in line with the ubiquity of predictability effects in acoustic, lexical and syntactic variation. The authors argue that previous studies did not find this effect, because they used unnatural, decontextualized fragment completion tasks (see also Demberg et al., 2023). However, it remains unclear what underlies this implicit causality effect on pronoun use. Several possible accounts have been discussed in the literature, which I will briefly present here.

First of all, in most studies investigating the effect of predictability on pronoun use, the assumed predictability of referents is based on the fact that certain thematic roles have a higher chance of being re-mentioned given a particular coherence relation. For example, a referent in the stimulus role of a stimulus-experiencer verb, such as *Amanda* in *Amanda amazed Brittany*, is more likely to be mentioned in the next sentence if that sentence provides an explanation for the event in the first sentence (Rohde & Kehler, 2014). That is, in the second sentence one would expect to hear about what Amanda had done to amaze Brittany. Similarly, the goal referent of a source-goal verb, such as *the maid* in *The chef handed a cookbook to the maid*, is more likely to be mentioned next when the following sentence talks about a consecutive event (Kehler et al., 2008; Rosa & Arnold, 2017; Stevenson et al., 1994). At the same time, there is research suggesting that the referent's thematic role may have a separate effect on referential choices, independently of the referent's predictability (Fukumura & Van Gompel, 2010; Kaiser et al., 2011; Vogels, 2019). For example, while in Vogels (2019) a stronger next-mention bias was associated with more use of (reduced) personal pronouns, referents in goal roles were more likely to be referred to with the full form of the Dutch personal pronoun or with a demonstrative pronoun (see also Medina Fetterman et al., 2022, for a similar result on overt vs. null forms in Spanish). Therefore, it remains somewhat unclear to what degree implicit causality or implicit consequentiality effects based on thematic roles are related to predictability.

Second, it is possible that referential choices are influenced by the predictability of the event, rather than (or in addition to) the predictability of the referent itself (cf. Rosa & Arnold, 2017). After all, listeners are not merely predicting who or what will be referred to next, but these predictions arise from the listener's expectation of the message the speaker will want to convey (e.g., Grüter & Rohde, 2021; Guan & Arnold, 2021; Hartshorne et al., 2015; Kehler et al., 2008; Rohde et al., 2021). For example, when hearing the utterance in (1), a listener might expect that the speaker will say something about what the housewife is going to do with the broom. If that expectation is fulfilled (as in (1a)), the speaker could suffice with a pronoun, since the listener will likely resolve it to the housewife. However, if the speaker actually continues with (1b), they might want to use a repeated noun phrase to refer to the housewife, since although the referent was expected, the ensuing event was not.

- (1) The saleswoman sold the housewife a broom. Next ...
- a. ... she went home to sweep the driveway.
 - b. ... she/the housewife badly needed to go to the bathroom.

Thus, even though the mention of a particular referent may be expected based on the most likely continuation, if that referent is actually involved in an event that is less expected, speakers might be prompted to signal the unexpectedness by choosing a more informative expression to mention the referent.

Third, an important difference between referring expression reduction and other types of reduction is that the latter are about the properties of the linguistic forms themselves, such as the predictability of the next word or the following syntactic element. A more predictable lexical item given the context may facilitate the production process, leading to reduction in speech. For instance, the production of the word *nine* will be faster, and hence the word will have a shorter duration, when it is in a facilitating context (Arnold et al., 2012; Bard et al., 2000; Kahn & Arnold, 2012, 2015). The choice between a definite noun phrase and a pronoun, however, is not about the predictability of the referring expression itself, but of the entity the expression refers to (Arnold, 2008; Demberg et al., 2023). From the speaker's perspective, it is not clear why a more predictable concept would result in the selection of a different lexical form. After all, what the speaker will refer to is part of the message plan, and hence in a sense is already predictable for the speaker (Arnold & Zerkle, 2019). Indeed, recent studies have not found evidence for a correlation between ease of production and pronoun use (Rosa & Arnold, 2017; Zerkle & Arnold, 2016, 2019).

Instead of the speaker's own production system, predictability effects may arise from the speaker taking into account the predictability of discourse entities from the listener's perspective (cf. Arnold, 2010; Demberg et al., 2023; Orita et al., 2015; Weatherford & Arnold, 2021). In this case, it is easier to see how predictability at the conceptual or discourse level may affect

production. That is, while from the speaker's perspective the entity to be mentioned next will be highly predictable to the extent that it was part of the message plan, the predictability of that entity may vary as seen from the listener's perspective. If the speaker takes this predictability for the listener into account in her production process, this may affect her choice of referring expression. For example, if the entity is not predictable for the listener, using a semantically and acoustically reduced expression such as a pronoun may lead to ambiguity or high processing load for the listener. If the speaker is sensitive to this, she will resort to using a more elaborate expression. If, on the other hand, the referent is considered highly predictable for the listener, the speaker can suffice with using a reduced expression.

Many theories and models of reference production assume that referential choices are in some way oriented towards the addressee (Ariel, 1990; Gundel et al., 1993; Orita et al., 2015; Rubio-Fernandez, 2019; see Arnold & Zerkle, 2019, for an overview). There is also ample empirical evidence that speakers adjust at least some of their referential choices to accommodate their addressees (e.g., Arts et al., 2011; Brennan & Clark, 1996; Galati & Brennan, 2010; Kantola & van Gompel, 2016; Koolen et al., 2011; Rosa et al., 2015; Rosa & Arnold, 2017; Tal et al., 2023; Van Der Wege, 2009). For example, Kantola and Van Gompel (2016) found that speakers produced fewer pronouns when a competitor referent was visually present, but only when they were talking to an actual addressee. This was taken as evidence that visual salience effects on the choice of referring expression are due to audience design. If, likewise, predictability effects on referring expression choice arise partly or entirely from the speaker's belief about how the addressee predicts that the discourse will continue, the presence or absence of an actual addressee might be important for detecting these effects. Most psycholinguistic studies that did not find predictability effects on the choice of referring expression involved experiments in which participants produced continuations of story fragments that were not intended for an addressee (e.g., Fukumura & Van Gompel, 2010; Kaiser et al., 2011; Rohde & Kehler, 2014), so this could be another factor in explaining the mixed findings in the literature.

3. The present study

In the present study, I investigate the effect of referent predictability on the choice of referring expression type in Dutch. More specifically, I ask whether predictability affects the choice between pronouns on the one hand, and descriptive NPs on the other, as well as between full and reduced pronouns. In the Dutch pronominal paradigm, various forms have both a full and a reduced variant, such as *jou* vs. *je* ('you' – object form), *mijn* vs. *m'n* ('my'), and *hij* vs. *ie* ('he'). This paradigm contains gaps (there is no reduced form of *ons* 'us', 'our'), and some forms are not generally used in written language (*m'n*, *ie*) and/or behave as a clitic that can only occur post-verbally (*ie*). However, the full and the reduced form of the feminine 3rd person subject pronoun (*zij* 'she.FULL'; *ze* 'she.RED') are both used in spoken as well as written Dutch (Kaiser, 2011;

Kaiser & Trueswell, 2004). Kaiser (2011; see also Kaiser & Trueswell, 2004) has shown that the choice between *zij* and *ze* is partly driven by contrastiveness, with the full form often marking a contrast with another referent. However, as Kaiser notes, the full form does not seem inherently related to contrast, leaving open a potential role for other factors, such as predictability. Thus, if predictability affects the choice between acoustically more and less pronounced variants of a referring expression, one would predict more reduced rather than full pronouns for predictable referents. Alternatively, if predictability effects are driven by addressee-oriented processes, they might not affect fast automatic processes such as acoustic reduction (Bard et al., 2000).¹ If predictability affects the choice between different lexical forms, one would predict more pronouns in general for predictable referents, as opposed to definite NPs.

Furthermore, I ask whether speakers' use of predictability as a cue for referring expression choice is dependent on the presence of an addressee. If speakers use predictability as a cue for referential choices to aid their specific addressee, I predict that speakers are more likely to use referent predictability in their choice of referring expression when an addressee is present than when no addressee is present, especially when it is clear to the speaker that the addressee needs to understand the message correctly. That is, they should use more reduced expressions for predictable referents, and more informative or marked expressions for less predictable referents. Alternatively, speakers may generally increase the informativeness or markedness of their referring expressions in the presence of an actual addressee, irrespective of the referent's predictability.

To test these predictions, I designed a referential communication task, in which an addressee is either present or absent. The speaker refers back to a character in a narrative fragment, while recounting a depicted event. The character varies in its likelihood of being mentioned next, based on established next-mention biases. To alleviate the concern that effects of next-mention biases based on thematic roles may conflate predictability with other properties of those thematic roles, I aimed to elicit next-mention biases that were independent of thematic roles. This was done by taking semantically object-biased critical sentence fragments and changing the sentence-initial adverb to *eerst* 'first'. This was expected to create a parallel coherence relation with a subsequent continuation sentence starting with 'next', and hence to induce a bias to start the continuation with the same subject referent (Kehler et al., 2008), counter to the semantic bias of the verb. In Vogels (2019), this manipulation was shown to transform a strong object bias into a strong subject bias. To replicate this result, I conducted a web-based pretest preceding the main experiment (Section 4). Since next-mention biases are never absolute (i.e., it is still possible to refer to the object referent in a context that strongly biases towards the subject), I also calculated the *surprisal* of each referent in the pretest as a more direct measure of predictability.

¹ I thank an anonymous reviewer for pointing me to this alternative view.

The surprisal of a referent is defined as the negative log probability of referring to that referent, where higher values correspond to a lower referent predictability. Thus, as surprisal goes up, the likelihood of a pronoun is predicted to decrease (see also Tily & Piantadosi, 2009).

Furthermore, a second pretest collected ratings of the expectedness of the event following the narrative fragment, in order to check whether referential choice in the main experiment may be better explained by event expectancy than by referent predictability (Section 5).

4. Pretest 1: Next-mention biases

Pretest 1 aimed to replicate the next-mention biases after sentences starting with or without the adverb *eerst* ‘first’ found in Vogels (2019).

4.1 Methods

4.1.1 Participants

Twenty Dutch-speaking participants (mean age 34.4 years; range 18–62) were recruited via social media. They were entered in a raffle for a €20 book coupon.

4.1.2 Materials

Twenty-four stories were created, consisting of a photograph showing two female LEGO® minifigures, accompanied by two context sentences. The first context sentence introduced the two characters, and the second context sentence described a transitive action performed by one of the characters. The verbs describing the actions were either Agent-Patient or Source-Goal verbs, and all had a next-mention bias towards the second NP (NP2) when followed by a temporal or a consequence coherence relation, as established by earlier studies (Commandeur, 2010; Koornneef & Sanders, 2013).² To ensure a temporal or consequence coherence relation, the sentence to be completed by the participant always started with the connective *vervolgens* ‘next’. In the critical condition, the second context sentence started with the adverb *eerst* ‘first’ (henceforth, the *‘first’* condition), which was expected to change the coherence relation to a parallel coherence relation. Therefore, I expected a preference for the next sentence to start with the same referent as the current sentence, and hence the next-mention bias should shift towards the first NP (NP1). In the control condition (henceforth, the *neutral* condition), the second context sentence started with either a temporal adverb (e.g., *meteen* ‘immediately’), or a manner adverb (e.g., *geduldig* ‘patiently’), which were not expected to cause a shift in the next-mention bias, with NP2 remaining more likely to be mentioned next. An example of the two conditions is given in (2). A screenshot of a trial is given in **Figure 1**.

² For some verbs, no Dutch data were available, and I took Dutch translations of English NP2-verbs (from Cheng, 2016, and Rosa & Arnold, 2017).

(2) *Introductory sentence*

In de bouwmarkt werd de huisvrouw geholpen door de
 in the hardware.store became the housewife helped by the
 verkoopster.
 saleswoman

‘At the hardware store, the housewife was helped by the saleswoman.’

a. *‘First’ condition*

Eerst verkocht de verkoopster de huisvrouw een bezem.
 first sold the saleswoman the housewife a broom
 Vervolgens ...

next

‘First, the saleswoman sold the housewife a broom. Next ...’

b. *Neutral condition*

Even later verkocht de verkoopster de huisvrouw een bezem.
 a.while later sold the saleswoman the housewife a broom
 Vervolgens ...

next

‘Some time later, the saleswoman sold the housewife a broom. Next ...’



Figure 1: Screenshot of a trial of Pretest 1. English translation of the context sentences: ‘At the hardware store, the housewife was helped by the saleswoman. First, the saleswoman sold the housewife a broom. Next ...’.

Additionally, 24 filler items were created, which were similar to the experimental items, except that these also contained stories with two male characters, one female and one male character,

just one male character, or animal characters. The context sentences also varied in structure, and none of the adverbials from the experimental items appeared in the fillers. A range of different connectives was used for the fillers.

4.1.3 Procedure

A web-based story completion experiment was conducted via the online platform Qualtrics (www.qualtrics.com). For each story, participants were shown the picture with the accompanying context sentences, and were asked to complete the third sentence with the first thing that came to mind, by typing it in the input bar below the sentences. The items were shown in a semi-random order, with no more than two experimental items succeeding each other, and in which both the first and last item were always fillers. At regular intervals, an unrelated cute-animal picture was shown, where participants could take a short break before continuing. The experiment took about 20 minutes to complete.

4.1.4 Coding and analysis

For each trial, I coded whether the subject of the continuation (the first NP after the finite verb) referred to NP1, NP2, to both, or to something else. I excluded 18 cases in which reference was unclear, 8 cases that were not complete sentences, and 4 non-referential cases (mostly dummy subjects such as in “next, *it* seems that...”), leaving 450 references for analysis. All data were also coded independently by a trained research assistant. Inter-annotator agreement for referent was very high (Cohen’s κ : .92).

A logit mixed effects analysis was conducted on the log odds of a reference to NP1, with Adverb (neutral, ‘first’) as fixed factor. This factor was centered to reduce collinearity. I started with a model containing the full random-effects structure, and then did stepwise removal of random correlations and random slopes with the lowest variance until I arrived at the optimal model justified by the data (see Bates et al., 2015), testing for significance at each step using a Likelihood Ratio test (with α set to .20; Matuschek et al., 2017). The final logit mixed model included random intercepts for participant and item, as well as a by-participant random slope for Adverb.

4.2 Results and discussion

The results of Pretest 1 showed that the next-mention manipulation was effective: There was a significant main effect of Adverb (see **Table 1**), with more references to NP1 when the second context sentence started with *eerst* ‘first’ than with a neutral (temporal or manner) adverb. With neutral adverbs there was a clear preference for references to NP2 (58.3%, as opposed to 20.0% for NP1), in line with the default NP2 next-mention bias of the verbs. With the adverb *eerst* the preference for NP1 (49.5%) was larger on average than for NP2 (39.5%; see **Figure 2**). Thus, when the context sentence contained the adverb *eerst* ‘first’, the NP2 bias decreased.

Table 1: Logit mixed model output for the effect of Adverb on the log odds of a reference to NP1 in Pretest 1. Model specification: $NP1 \sim \text{Adverb} + (1 + \text{Adverb} || \text{Participant}) + (1 | \text{Item})$.

Random effects	Var.	SD			
Participant: intercept	1.06	1.03			
Participant: adverb	0.86	0.93			
Item: intercept	0.45	0.67			
Fixed effects	β	SE	z	p	95% CI
Intercept	-0.93	0.30	-3.07	<.01	[-1.52, -0.34]
Adverb	1.67	0.34	4.90	<.001	[1.002, 2.339]

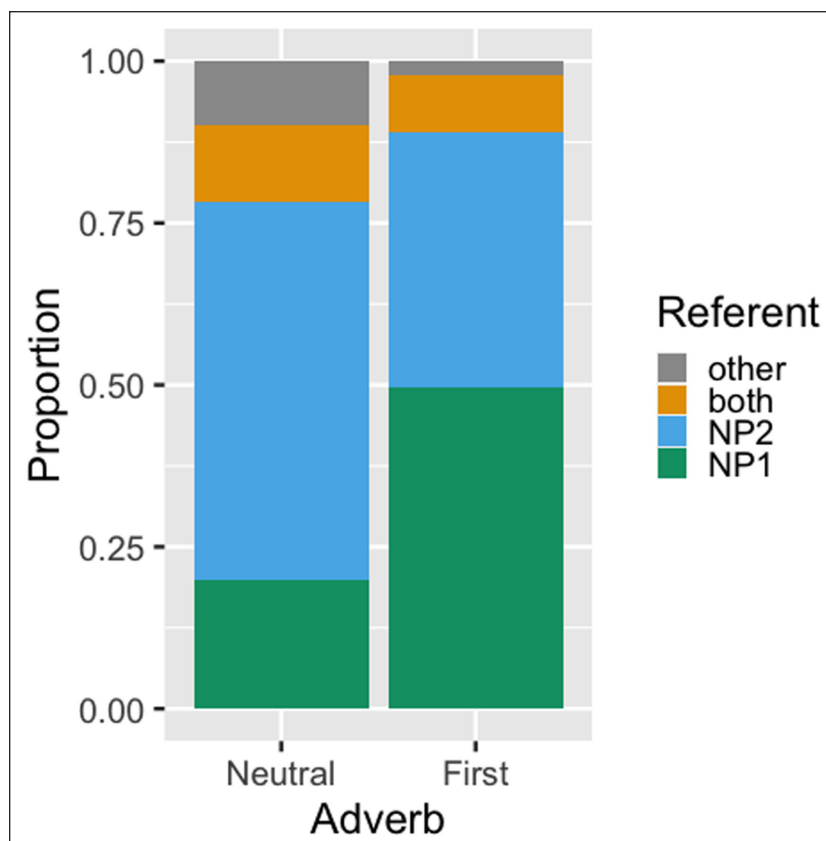


Figure 2: Effects of Adverb on the proportion of NP1, NP2, both, and other references in Pretest 1.

These results thus confirm that the experimental materials evoke preferences of varying strength to continue the narrative contexts with either the subject or the non-subject referent. These preferences will be used as a measure of referent predictability in the main experiment, where more pronouns are expected when the referent is in congruence with the bias induced by the adverb. That is, I predict pronouns to be more likely in the *first* than in the *neutral* condition when referring to NP1 (and more likely in the *neutral* than in the *first* condition when referring to NP2).

A drawback of using overall next-mention preferences as a measure of predictability, as with manipulations based on thematic role biases, is that they are averages over multiple items and multiple participants, both of which show individual variation. On the item level, most items showed a difference between the adverb conditions in the predicted direction, with more NP1 references after ‘first’ than after neutral adverbs. However, the items varied in the size of this difference as well as in the size of the original NP2-bias. Taking this variation into account allows for a more fine-grained and more direct measure of referent predictability. Therefore, I also calculated the by-item next-mention biases in Pretest 1. That is, for each item in each adverb condition, I calculated the proportion of references to NP1 and to NP2 out of all references, as an estimation of each referent’s probability of mention. Based on these probabilities, I then calculated the *surprisal* of each referent, defined as the negative log probability of referring to that referent. Surprisal is a measure from information theory corresponding to the amount of information in a signal, where higher values signify more information, and 0 means no information at all (Levy, 2008; Shannon, 1948). It has been used as a measure of predictability in language and has been shown to correlate with processing cost in comprehension (e.g., Demberg & Keller, 2008; Smith & Levy, 2013). Here, it will complement the use of the two adverb conditions as an estimate of referent predictability.³ The full list of by-item proportions and surprisal values is given in Table A.1 in the Appendix.

5. Pretest 2: Event expectancy

Pretest 2 tested the possibility that the next-mention biases found in Pretest 1 do not reflect the predictability of referents per se, but rather the predictability of the ensuing event, which in turn makes a certain referent more likely to be mentioned. If that is the case, then there should be a strong correlation between event predictability and next mention.

³ Because of the relatively small sample size of Pretest 1, the variance around the proportions is substantial. The surprisal analysis should therefore be seen as complementary to the analysis of the adverb conditions.

5.1 Methods

5.1.1 Participants

Twenty-two new Dutch-speaking participants⁴ (mean age 43.3 years; range 19–69) were recruited via social media. They were entered in a raffle for a €20 book coupon.

5.1.2 Materials

The same items as in Pretest 1 were used, except that each photo was accompanied by a second photo depicting an intransitive action (e.g., walking away) performed by one of the characters from the first photo (see **Figure 3**). Which referent performed the action in the second photo (the target referent), either the one mentioned in NP1 or the one mentioned in NP2, was added as a condition. I took the adverb conditions as well as the surprisal values from Pretest 1 as measures of the likelihood of mention of the target referent in each story. If the event in the second photo is judged as less expected when it involves a less likely/more surprising referent, one would expect a main effect of surprisal, with lower expectancy scores for continuations involving referents with a higher surprisal, and an interaction between Referent and Adverb, with lower expectancy scores for NP1 referents after a neutral adverb and NP2 referents after ‘first’ than for NP1 referents after ‘first’ and NP2 referents after a neutral adverb.

5.1.3 Procedure

Pretest 2 was a rating study, also using Qualtrics, in which participants rated the expectedness of events on a 5-point Likert scale. Participants were shown two adjacent photos, of which the second photo depicted what happened next after the action shown in the first photo (see **Figure 3**). Participants were asked to rate how (un)expected they found the event in the second photo, given the action in the first photo, from 1 (very unexpected) to 5 (very expected). The items were shown in a semi-random order, with no more than two experimental items succeeding each other, and in which both the first and last item were always fillers. Again, an unrelated cute-animal picture appeared at regular intervals to signal a short break. The experiment took about 20 minutes to complete.

5.1.4 Analysis

The ratings were analyzed using two ordinal mixed effects analyses, using the *clmm2* package in R (Christensen, 2015). One analysis included Adverb (neutral, ‘first’) and Referent (NP1, NP2) and their interaction as fixed factors, random intercepts for participant and item, and by-item random slopes for Adverb, Referent, and their interaction. The other analysis replaced the adverb

⁴ One additional participant was removed for having already participated in Pretest 1.



In de bouwmarkt werd de huisvrouw geholpen door de verkoopster.
Eerst verkocht de verkoopster de huisvrouw een bezem. Vervolgens ...

Hoe verwacht vind je de gebeurtenis in het tweede plaatje?

	Heel onverwacht	Een beetje onverwacht	Niet verwacht, maar ook niet onverwacht	Een beetje verwacht	Heel verwacht
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3: Example trial from Pretest 2. English translation of the context sentences: ‘At the hardware store, the housewife was helped by the saleswoman. First, the saleswoman sold the housewife a broom. Next ...’; English translation of the rating question: ‘How expected do you find the event in the second picture? Very unexpected; A little unexpected; Not expected, but not unexpected either; A little expected; Very expected’.

condition with surprisal as an alternative predictability measure, and included random intercepts for participant and item, and a by-item random slope for Referent. Fixed factors were centered to reduce collinearity.

5.2 Results and discussion

Neither the model analyzing the interaction of Adverb and Referent on event expectancy ratings nor the model taking Surprisal as the independent variable showed significant effects of a referent's likelihood of next mention on event expectancy (see **Tables 2** and **3**, respectively). By-item surprisal and the by-item average event expectancy ratings were not correlated either (Pearson's $r = -.022$, $p = .84$; see **Figure 4**). Thus, I found no evidence that in the present materials, events involving a target referent that is less likely to be mentioned given the context were perceived as less expected. That means there is no reason to assume that effects of referent predictability are actually driven by event expectancy. The ratings are added to the by-item results of Pretest 1 in Table A.1 in the Appendix.

Table 2: Ordinal mixed model output for the effect of Adverb, Referent and their interaction on event expectancy ratings. Model specification: Rating \sim Adverb * Referent + (1 | Participant) + (1 + Adverb * Referent || Item).

Random effects	Var.	SD			
Participant: intercept	0.45	0.67			
Item: intercept	1.27	1.13			
Item: adverb	0.14	0.38			
Item: referent	2.63	1.62			
Item: adverb * referent	1.46	1.21			
Fixed effects	β	SE	z	p	95% CI
Adverb	0.07	0.18	0.39	.70	[-0.289, 0.432]
Referent	0.00	0.37	0.01	1.00	[-0.725, 0.729]
Adverb * Referent	-0.36	0.42	-0.86	.39	[-1.173, 0.475]

Table 3: Ordinal mixed model output for the effect of Surprisal, Referent and their interaction on event expectancy ratings. Model specification: Rating \sim Surprisal * Referent + (1 | Participant) + (1 + Referent || Item).

Random effects	Var.	SD			
Participant: intercept	0.41	0.64			
Item: intercept	1.18	1.09			
Item: referent	2.43	1.56			

(Contd.)

Fixed effects	β	SE	z	p	95% CI
Surprisal	0.04	0.20	0.19	.85	[-0.356, 0.430]
Referent	-0.02	0.37	-0.04	.97	[-0.746, 0.713]
Surprisal * Referent	-0.38	0.41	0.94	.35	[-0.415, 1.177]

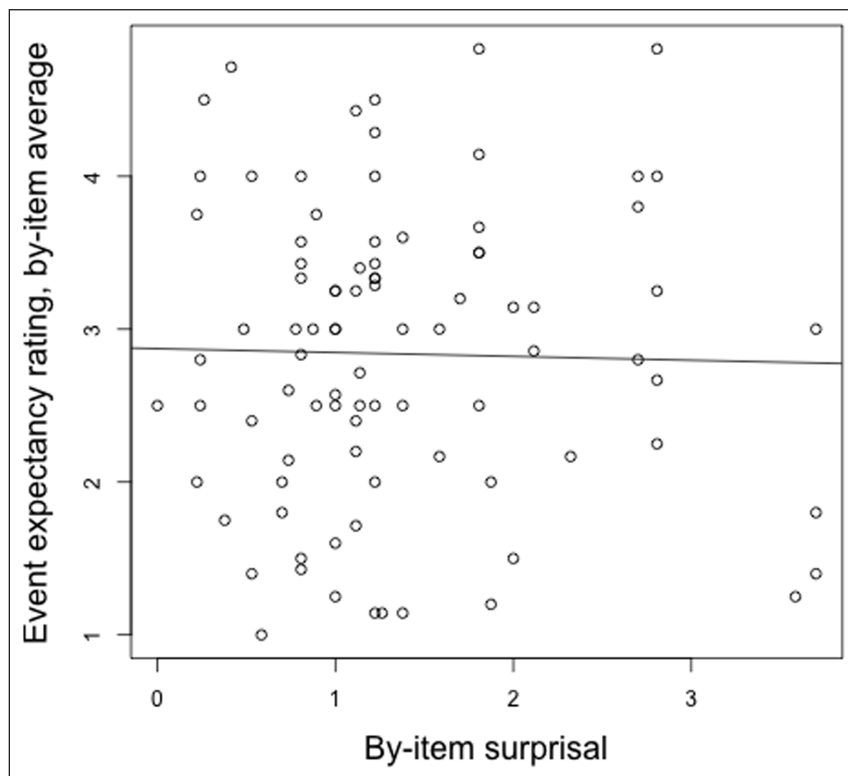


Figure 4: The relationship between by-item surprisal (from Pretest 1) and by-item average event expectancy ratings (from Pretest 2). The regression line results from a simple linear model.

6. Main experiment

Given the effect of the adverb manipulation on the likelihood that a referent will be mentioned again (Pretest 1), independent of how likely the ensuing situation is (Pretest 2), the next question is how this variation in likelihood of next mention affects the choice of referring expression. If speakers take into account that referents likely to be re-mentioned do not need an explicit reference with a definite NP, one would expect a higher rate of pronouns for subject referents after ‘first’ and for referents with a lower surprisal value. In addition, if the presence of an addressee is

an important factor in whether speakers take into account listeners' expectations, the effect of the adverb condition and/or surprisal should be larger when an actual addressee is present. Finally, if likelihood of mention affects acoustic reduction, one would expect more reduced compared to full pronouns for subject referents after 'first' and for referents with a lower surprisal value.

6.1 Methods

6.1.1 Participants

Eighty Dutch-speaking participants (mean age 23.3; age range 18–43) were recruited for this experiment, mainly from the University of Groningen student population. They received monetary compensation for their participation.⁵ None of them had participated in either of the pretests. All participants were native speakers of Dutch, although four participants reported having grown up bilingually.⁶ In the Addressee-present condition, two native Dutch research assistants who had not been informed of the purpose of the experiment took turns in serving as the addressee.

6.1.2 Materials

The same twenty-four short stories from Pretest 2 were used. Each story consisted of three photographs of LEGO® minifigures: an introductory picture, and two adjacent pictures depicting the actions in the story. **Figure 5** shows an example item. The introductory picture showed the two characters appearing in the story, idly facing the observer, with their labels printed above them. Characters were always two females, such that pronouns would be ambiguous, and to allow variation between the full and reduced form of the feminine pronoun. The first story picture showed the same characters in the same relative position as in the introductory picture. Here, they were depicted in a situation described by two context sentences appearing below the picture. The second picture was combined with the onset of a third sentence, which was always *vervolgens* 'next' (see **Figure 5**). The left-right position of the characters was counterbalanced between items.

The stories contained the same adverb manipulation as in the pretests, with the second context sentence starting either with the adverb *eerst* 'first' or with a temporal or manner adverb. In addition, as in Pretest 2, the second story picture showed either the character mentioned by NP1 or the character mentioned by NP2 performing some intransitive action (e.g., walking away). This resulted in four variants of each item, which are presented with an example in (3) below. The referent in italics has a higher expectancy to be mentioned next compared to the

⁵ Initially, compensation was set at €4. When 24 participants had been tested, teaching went mostly online due to the COVID-19 pandemic, and few students were present at the university. To boost willingness to participate, I increased compensation to €10.

⁶ Removing these four participants from the data gave very similar results.

variants with the other adverb, as established by the next-mention biases from Pretest 1. The target referent is given in bold in the part between square brackets describing the action in the second picture. Thus, in conditions a and b in (3), the target referent is congruent with the more expected referent, while in conditions a' and b' the target referent is incongruent with the more expected referent.

- (3) *Op de feestdag protesteerde de activiste tegen de koningin.*
 ‘On the holiday, the activist_{FEM} protested against the queen.’
- a. *Geïrriteerd bespote de koningin de activiste.* *Vervolgens ...*
 ‘Irritated, the queen mocked the activist_{FEM}.’ *Next, ...’ [activist walks away]*
- a'. ‘Irritated, the queen mocked the activist_{FEM}.’ *Next, ...’ [queen walks away]*
- b. *Eerst bespote de koningin de activiste.* *Vervolgens ...*
 ‘First, the queen mocked the activist_{FEM}.’ *Next, ...’ [queen walks away]*
- b'. ‘First, the queen mocked the activist_{FEM}.’ *Next, ...’ [activist walks away]*

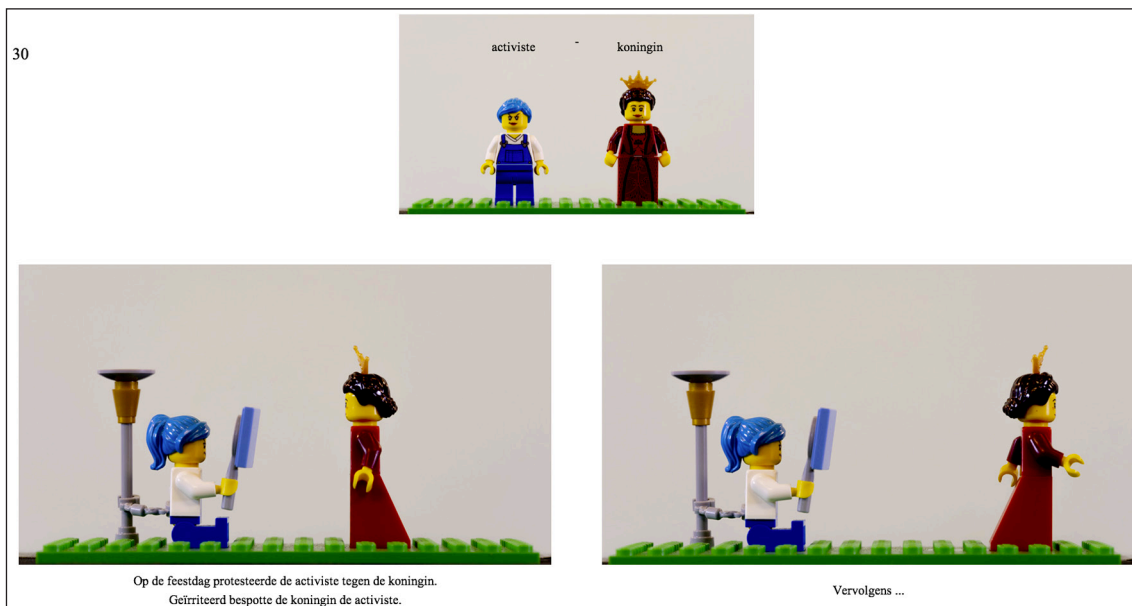


Figure 5: Example trial from the main experiment. English translation of the character labels: ‘activist_{FEM} – queen’. English translation of the context sentences: ‘On the holiday, the activist protested against the queen. Irritated, the queen mocked the activist. Next ...’.

Besides the 24 filler items, which were the same as in the pretests, there was one practice item, featuring two male characters. In the Addressee-present condition, one of the fillers was used as an additional practice item.

6.1.3 Design

The experiment was created using Experiment Builder⁷ (SR Research, Ottawa, Canada). Addressee presence was manipulated between participants. Crossing the within-participant and within-item variables Adverb (neutral, 'first') and Referent (NP1, NP2) resulted in a 2 × 2 item design. The 24 experimental items were distributed over four lists, such that each list contained only one version of each item, but each participant saw each condition equally often. Items were presented in a pseudo-random order, intermixed with the fillers such that no more than two experimental items appeared consecutively.

6.1.4 Procedure

Participants were assigned to either the Addressee-present or the Addressee-absent condition randomly, or based on the availability of the confederate addressee. In the Addressee-absent condition, no potential addressee was present at any point during the experiment. Participants first signed an informed consent form,⁸ and were then seated behind a computer screen in the lab. A Devine M-MIC USB BK microphone was mounted on the table next to the screen. Participants received both written and spoken instructions. They were told that they would be telling short stories about LEGO® minifigures, and that each story would consist of two pictures, plus an introductory picture to help them remember the characters. Participants were asked to first briefly examine the pictures and the accompanying sentences until they understood the actions depicted, and then start reading aloud the sentences given, finishing the story with one or two sentences describing what happened in the second story picture. Participants were instructed to think briefly about how they would complete the story before saying it out loud, but to not ponder too long and go with the first thing that came to mind. No mention was made of an addressee and nothing was said about who the stories should be directed to, only that they would be recorded for research purposes. Next, participants performed the practice trial together with the experimenter, after which there was room for questions. As soon as everything was clear to the participant, the experiment was started. The experimenter left the room at this point, so that there was no one in the room who could be seen as a potential addressee.

In the Addressee-present condition, a confederate addressee was already sitting at the far end of the table behind a laptop when the participant entered the lab. To minimize potential adverse effects of the use of confederates (see Kuhlen & Brennan, 2013), the addressee was

⁷ While Experiment Builder is normally used for constructing eye-tracking experiments, no eye-tracking data was collected with this experiment. However, Experiment Builder was chosen to allow for the possibility of future eye-tracking experiments using similar materials and setup.

⁸ During the COVID-19 pandemic, participants first applied for the experiment online, giving consent and declaring that they were not showing any COVID-19 symptoms.

oblivious to the purpose of the experiment and had a real task (although they inevitably saw the same version of the task multiple times). In addition, participants were not misled, the addressee being truthfully introduced to them as a lab assistant without knowledge of the experiment's purpose. The participant and the addressee were seated opposite from each other, with the computer monitor and microphone shifted a bit to the side so that they could see each other. In line with the COVID-19 measures, there was a transparent screen between them, and moving around the lab was kept to a minimum. Participants were given the same instructions about how to complete the stories as in the Addressee-absent condition, but with the following additions: First, participants were instructed that they had to tell the stories to the addressee, who then had to pick out the correct final picture of the story on the laptop screen. Second, it was made clear that the addressee saw the pictures "from the other side", as if the scenes had been there on the table between the two of them. This was illustrated with actual LEGO® scenes in two practice rounds. Third, participant and addressee also switched roles once in both practice rounds, so that the addressee told the story to the participant, who saw the addressee's task on their own computer screen. In the second practice round, it was pointed out that stories might become less clear for the addressee if one did not take into account their visual perspective. This was all done to make participants aware of the other person's perspective. Although the focus was on their visual perspective rather than the discourse, and hence not related to the choice of referring expression with respect to predictability, the rationale was that making speakers aware of potential differences in perspective would also make them more sensitive to factors dependent on perspective taking, such as predictability.

The addressee in the Addressee-present condition saw the same pictures for each story as the participant, but indeed photographed from the other side. In addition, a choice between two final pictures was given in each story. The addressee was instructed, in front of the participant, to listen carefully to the participant's stories, pick out the correct picture, and mark their answer on a response form. Next, they would signal to the participant that they were ready for the next story. They could also ask for clarifications, if necessary. In reality, however, the addressee had been instructed beforehand to give only very minimal feedback, except on five particular items (all fillers) where they were to express uncertainty by hesitating a bit. This was done to keep the participant attentive to the addressee over the course of the experiment, while at the same time keeping addressee behavior as constant as possible across participants.

In both the Addressee-present and the Addressee-absent conditions, participants cycled through the trials at their own pace by pressing the space bar. In each trial, they first saw the introductory picture in the center of the screen, so they could familiarize themselves with the story characters. With a press on the space bar, the two story pictures and the accompanying sentences appeared, while the introductory picture remained in view in smaller size at the top of the screen (see **Figure 5**). All speech uttered during a trial was recorded. There was no time limit for the participants' stories, but recording stopped after 60 seconds and any speech

occurring after that was not analyzed. A single experiment run took about 30 minutes. After the experiment, participants filled in a short questionnaire and came out of the lab to be debriefed. When asked, three participants noted that they found it somewhat strange that the addressee hesitated in just these five cases. One of them additionally guessed correctly that the experiment was about predictability, as did another participant. One further participant thought it was clear the addressee had done the task before. Removing these five participants from the data did not change the results in any meaningful way.

6.1.5 Coding and analysis

I removed 62 trials that did not start with the pre-given connective *vervolgens* ‘next’ and/or did not have a finite verb in the second position in the sentence, which is the standard position for a finite verb in main clauses in Dutch. In addition, I excluded 35 trials with a wrong-gender pronoun, 31 with a plural subject, 23 self-corrections, 20 non-responses, 12 trials in which reference was unclear, 8 where the referent was something other than NP1 or NP2, and 1 trial with a technical failure. In total, 192 trials (10%) were removed.

In addition, there were 196 trials in which the first reference was not to the target character (i.e., the referent that performed an action in the second picture). For the analysis testing the effect of the adverb conditions, I excluded those cases as well, since this analysis builds on the premise that references to the subject after the adverb ‘first’ are more predictable. The final data set included 1214 full NPs and 318 pronouns, of which 259 were reduced pronouns. Three possessive NPs and 11 bare nouns (mostly kinship terms, such as ‘(grand)mother’) were counted as full NPs; 1 demonstrative pronoun was counted as a full pronoun, and 2 null forms were counted as reduced pronouns. Fifty-five percent of the data was also coded by a trained research assistant. Inter-annotator agreement for referring expression type was very high (Cohen’s κ : .92).

For the analysis with the by-item surprisal values from Pretest 1 as a measure of referent predictability, I kept the non-target references as well-formed responses, and updated the Referent variable to reflect which referent participants actually referred to (as compared to what the intended referent was). Note that in the large majority of cases, participants did refer to the target character. For this analysis, 111 cases in which by-item surprisal was calculated to be infinite (a likelihood of mention of 0) were further removed. Here, the data consisted of 1284 full NPs (of which 3 were possessive NPs and 9 were bare nouns) and 333 pronouns (of which 272 were reduced pronouns, including 2 null forms). Two demonstrative pronouns were counted as full pronouns.

For each of the two analyses, I had originally planned to run two mixed effects models, one on the log odds of a pronoun (full and reduced combined) vs. a full NP, and one on the log odds of a reduced pronoun (including null forms) out of all pronouns. However, because full pronouns turned out to be relatively infrequent, I dropped the second model. In the adverb analysis, Adverb (neutral, ‘first’), Referent (NP1, NP2), Addressee presence (absent, present) and the interactions between them were entered as fixed effects. All predictors were centered to reduce collinearity.

In the surprisal analysis, Surprisal, Addressee presence, and the interaction between the two were entered as fixed effects and centered. As the grammatical function of the referent probably has the largest impact on referring expression choice, I also added Referent (NP1/Subject, NP2/Object) as a main effect, as well as its interactions with Surprisal and Addressee. Finally, to test for the possibility that referring expression choice is influenced by event expectancy, I further ran a model on the full data set (including non-target references and items with infinite surprisal) with Event expectancy rating, Addressee presence, Referent, and all two-way interactions as independent predictors.

As before, I simplified each of the models by sequentially removing first the random correlations and then random slopes with the lowest variances (see Bates et al., 2015). Each simplification step was tested for significance using a Likelihood Ratio Test. The specifications of the final models are reported in the captions of **Tables 4–6** below. For interactions between factors, simple effects were determined by recoding the factor of interest as nested within the interacting factor. When there was an interaction between a factor and a covariate, the factor was recoded with treatment contrasts to test the covariate's simple effect.

6.2 Results

6.2.1 Adverb manipulation

Figure 6 shows the proportions of the three referring expression types (full NP, full pronoun, reduced pronoun) for each combination of predictors (Adverb, Referent, Addressee presence). Analyzing the effect of the predictors on the use of pronouns (full and reduced) vs. full NPs showed significant effects of both Adverb and Referent, with more pronouns when the previous sentence started with *erst* 'first' (26.0%) than with a neutral adverb (15.7%), and more pronouns for references to the subject (33.6%) than to the object (9.3%; see **Table 4**). Crucially, there was also a significant interaction between the two predictors, suggesting that the adverb *erst* led to more pronouns in references to the subject. Testing for the simple effects of Adverb indicated that there were indeed significantly more pronouns after *erst* than after a neutral adverb for references to the subject ($\beta = 1.13$, $SE = 0.22$, $z = 5.14$, $p < .001$), whereas there was no effect of Adverb for references to the object ($\beta = 0.36$, $SE = 0.29$, $z = 1.23$, $p = .22$). Although pronoun use was slightly decreased in the presence of an addressee (18.9%) as compared to the absence of an addressee (23.0%), there was no significant main effect of Addressee presence, and no interactions with Adverb or Referent. However, because I predicted an effect of predictability especially in the Addressee-present condition, I also tested for the presence of a simple interaction effect between Adverb and Referent in both the Addressee-present and the Addressee-absent conditions, using a nested model. Here, the interaction between Adverb and Referent did not reach significance in either the Addressee-absent ($\beta = -0.77$, $SE = 0.49$, $z = -1.58$, $p = .11$) or the Addressee-present ($\beta = -0.77$, $SE = 0.47$, $z = -1.65$, $p = .10$) condition.

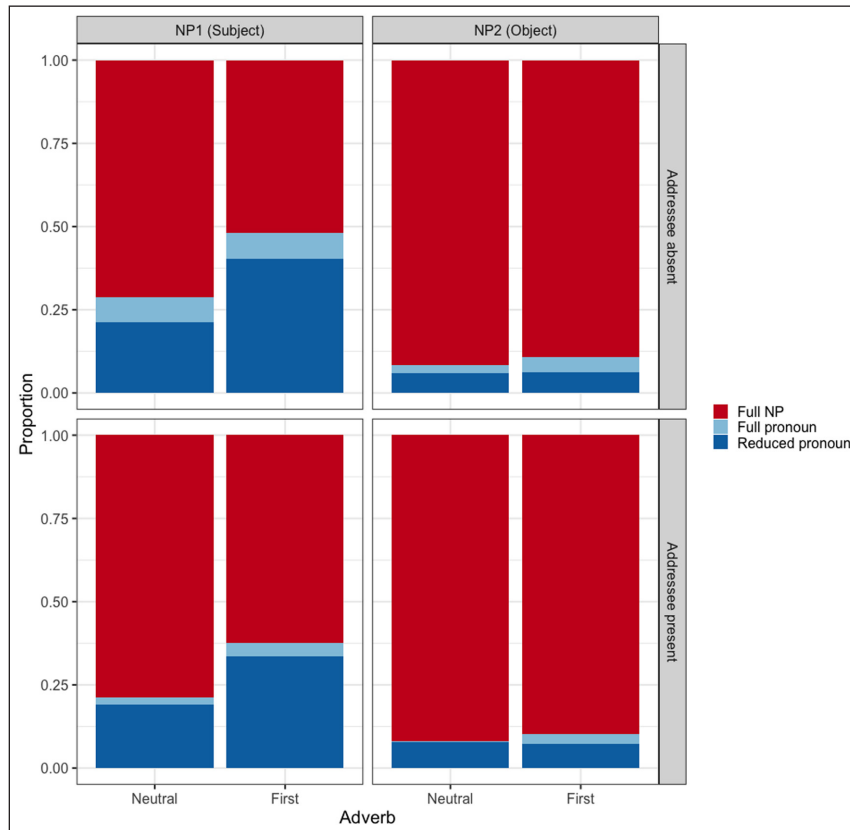


Figure 6: Proportion of full NPs, full pronouns and reduced pronouns by Adverb, Referent and Addressee presence.

Table 4: Logit mixed model output for the effect of Adverb, Referent, Addressee and their interactions on the log odds of a pronoun out of all referring expressions. Model specification: $\text{Pronoun} \sim \text{Adverb} * \text{Referent} * \text{Addressee} + (1 + \text{Adverb} || \text{Participant}) + (1 + \text{Referent} || \text{Item})$.⁹

Random effects	Var.	SD			
Participant: intercept	1.44	1.20			
Participant: adverb	0.66	0.81			
Item: intercept	0.11	0.33			
Item: referent	1.48	1.22			

(Contd.)

⁹ This model gave a convergence warning, but checking all available optimizers showed very similar results (see <https://rdrr.io/cran/lme4/man/convergence.html>), so I consider this warning a false positive.

Fixed effects	β	SE	z	p	95% CI
Intercept	-2.03	0.19	-10.69	<.001	[-2.399, -1.655]
Adverb	0.73	0.20	3.65	<.001	[0.336, 1.116]
Referent	-1.98	0.31	-6.40	<.001	[-2.581, -1.370]
Addressee	-0.29	0.33	-0.89	.37	[-0.926, 0.349]
Adverb*Referent	-0.77	0.34	-2.28	.02	[-1.430, -0.109]
Adverb*Addressee	-0.18	0.40	-0.45	.65	[-0.954, 0.598]
Referent*Addressee	0.47	0.34	1.37	.17	[-0.201, 1.138]
Adverb*Referent*Addressee	-0.00	0.67	-0.01	1.00	[-1.324, 1.317]

6.2.2 By-item surprisal

Figure 7 shows the proportion of pronouns plotted against the by-item surprisal of NP1 (subject) and NP2 (object) referents. The left-hand graph shows the pronoun production results for participants that were not talking to an addressee, and the right-hand graph shows pronoun production in the presence of an addressee. The model (see **Table 5**) showed a significant effect of Referent, with more pronouns for NP1 (the subject) than for NP2 (the object). There was also a significant main effect of Surprisal, with a lower rate of pronoun use for more surprising referents. Judging by **Figure 7**, this effect of surprisal seems to be driven mainly by references to NP1, which is supported by a significant interaction between Surprisal and Referent. Testing for the simple effects of Surprisal indeed indicated that Surprisal significantly affected references to NP1 (the subject; $\beta = -0.58$, $SE = 0.17$, $z = -3.52$, $p < .001$), but not references to NP2 (the object; $\beta = -0.12$, $SE = 0.20$, $z = -0.58$, $p = .56$). In line with the adverb analysis, the main effect of Addressee presence did not reach significance, and the interactions between Addressee presence and Referent, as well as between Surprisal and Addressee presence, were not significant. However, since I predicted predictability to affect referring expression choice primarily in the presence of an addressee, I also tested for the simple effects of Surprisal in the Addressee-absent and the Addressee-present conditions. This showed that the effect of Surprisal was significant in the Addressee-present condition ($\beta = -0.38$, $SE = 0.18$, $z = -2.06$, $p = .04$), but did not reach significance in the Addressee-absent condition ($\beta = -0.25$, $SE = 0.18$, $z = -1.41$, $p = .16$). While this result is in line with the prediction, the lack of a surprisal effect in the Addressee-absent condition may be due to just two items with a relatively high rate of pronoun use despite a relatively high surprisal, for references to the object (see **Figure 7**). Without these two items, the simple effect of Surprisal is significant in the Addressee-absent condition ($\beta = -0.44$, $SE = 0.20$, $z = -2.22$, $p = .03$) and does not reach significance in the Addressee-present condition ($\beta = -0.37$, $SE = 0.20$, $z = -1.89$, $p = .06$).

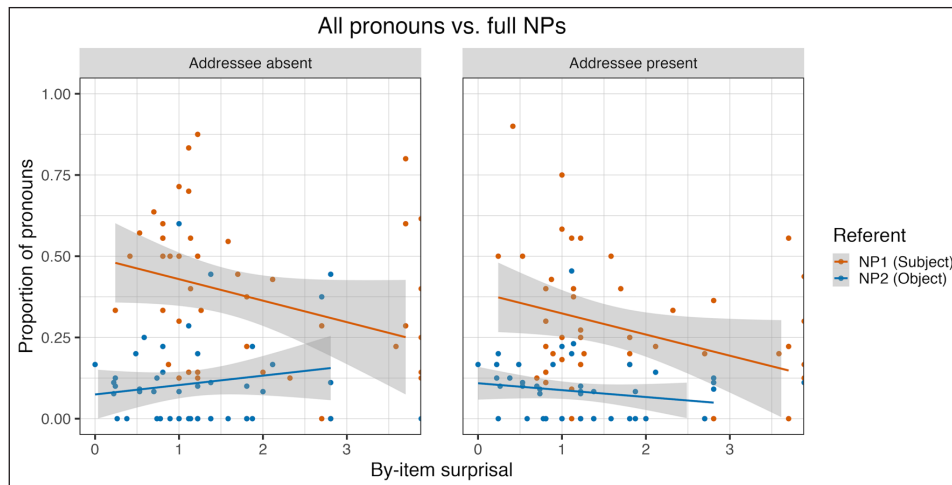


Figure 7: Proportion of pronouns by by-item surprisal (in bits), Referent and Addressee presence. Each dot represents one item in one of its four versions. The semi-circles on the right edge of each graph represent items with estimated infinite surprisal, i.e., items in which the target referent was never mentioned in Pretest 1 (not included in the statistical analysis). The shaded area around regression lines represents the 95% confidence interval based on a generalized linear regression model.¹⁰

Table 5: Logit mixed model output for the effect of Surprisal, Addressee, Referent and their interactions on the log odds of a pronoun out of all referring expressions. Model specification: $\text{Pronoun} \sim \text{Surprisal} * \text{Addressee} + \text{Surprisal} * \text{Referent} + \text{Addressee} * \text{Referent} + (1 + \text{Surprisal} || \text{Participant}) + (1 + \text{Surprisal} + \text{Addressee} + \text{Referent} || \text{Item})$.¹¹

Random effects	Var.	SD			
Participant: intercept	1.52	1.23			
Participant: surprisal	0.19	0.44			
Item: intercept	0.07	0.26			
Item: surprisal	0.09	0.31			
Item: addressee	0.18	0.43			
Item: referent	1.00	1.00			

(Contd.)

¹⁰ The data in **Figure 7** might suggest a U-shaped rather than a linear relation between surprisal and pronoun use. This would be in line with the idea that high surprisal results in high cognitive load, which in turn increases the use of pronouns as more production-efficient referring expressions (see Hendriks et al., 2014; Vogels et al., 2015; Zarcone et al., 2016). Since the size of the current data set does not allow the prediction of non-linear relationships, I leave this question to future research.

¹¹ Again, the model gave a convergence warning, but checking all available optimizers showed similar results.

Fixed effects	β	SE	z	p	95% CI
Intercept	-2.02	0.19	-10.68	<.001	[-2.399, -1.655]
Surprisal	-0.32	0.15	-2.11	.04	[0.336, 1.116]
Addressee	-0.36	0.34	-1.06	.29	[-0.926, 0.349]
Referent	-2.22	0.28	-7.87	<.001	[-2.581, -1.370]
Surprisal*Addressee	-0.12	0.20	-0.61	.54	[-0.954, 0.598]
Surprisal*Referent	0.47	0.22	2.11	.03	[-1.430, -0.109]
Addressee*Referent	0.46	0.34	1.37	.17	[-0.201, 1.138]

6.2.3 Event expectancy

In contrast to the adverb conditions and surprisal, using the event expectancy ratings from Pretest 2 as a measure of predictability did not show an effect on the choice of referring expression (see **Table 6**). There were no interactions with either Addressee presence or Referent. The only significant predictor was Referent, with more pronouns for references to NP1 (the subject) than NP2 (the object).

Table 6: Logit mixed model output for the effect of Event expectancy, Addressee, Referent and their interactions on the log odds of a pronoun out of all referring expressions. Model specification: Pronoun \sim Event expectancy * Addressee + Event expectancy * Referent + Addressee * Referent + (1 | Participant) + (1 + Event expectancy + Referent || Item).

Random effects	Var.	SD			
Participant: intercept	1.30	1.14			
Item: intercept	0.11	0.33			
Item: event expectancy	0.23	0.48			
Item: referent	0.30	0.55			
Fixed effects	β	SE	z	p	95% CI
Intercept	-1.92	0.18	-10.49	<.001	[-2.277, -1.560]
Event expectancy	0.01	0.15	0.04	.97	[-0.286, 0.297]
Addressee	-0.35	0.31	-1.16	.25	[-0.952, 0.245]

(Contd.)

Fixed effects	β	SE	z	p	95% CI
Referent	-2.09	0.21	-9.74	<.001	[-2.510, -1.669]
Event expectancy*Addressee	0.15	0.14	1.02	.31	[-0.135, 0.430]
Event expectancy*Referent	-0.16	0.20	-0.81	.42	[-0.558, 0.232]
Addressee*Referent	0.47	0.31	1.52	.13	[-0.135, 1.079]

6.3 Discussion

Given the results of Pretest 1, in which participants were more likely to continue a fragment with the subject of the previous sentence when that sentence started with the adverb *eerst* ‘first’ than with another temporal or manner adverb, I predicted that references to the subject would be more likely to be pronominalized if the previous sentence started with *eerst* than with another type of adverb. In addition, I calculated the surprisal of each character in each narrative, based on participants’ continuations of the narrative fragments in Pretest 1, as a more direct measure of its predictability. Here, I predicted that as surprisal increases, speakers would refer more clearly to the target character in their continuations, for example by repeating the full NP instead of using a pronoun (e.g., ‘Next, *the activist* walked away’). The results indeed show the predicted patterns: there were more pronouns referring to the subject when the previous sentence started with *eerst* ‘first’ than with another adverb. In addition, the rate of pronoun use decreased significantly as surprisal increased. This latter effect was also restricted to references to the subject, possibly due to the fact that pronoun rate was already generally low for NP2 (object) referents (floor effect). Alternatively, the surprisal effect may have been reduced for object referents because there were fewer high-surprisal object than subject referents, perhaps combined with the fact that the Addressee-absent condition contained two items with object references that seemed to deviate from the predicted trend. Finally, event expectancy did not affect referring expression choice, confirming that referent predictability is the driving factor.

I also predicted that if addressee presence is an important factor in whether speakers take listeners’ expectations into account, the effect of predictability on referring expression choice should be more pronounced in the presence of an actual addressee. This latter prediction was not borne out: while speakers’ pronominalization rate was somewhat lower (although not reliably) when an addressee was present, the effect of the adverb ‘first’ was not more pronounced in the Addressee-present condition. In addition, while the effect of surprisal seemed to be restricted to the Addressee-present condition, this effect appeared to hinge only on the presence of the two deviating items. Crucially, then, predictability appears to affect the choice of referring expression in Dutch, irrespective of addressee presence. This result is not in line with some earlier findings for English (e.g., Fukumura & Van Gompel, 2010; Rohde & Kehler, 2014), but is supported by more recent findings for both English (Demberg et al., 2023; Rosa & Arnold, 2017; Weatherford & Arnold, 2021) and German (Bott et al., 2018; Bott & Solstad, 2022).

7. General discussion

In the present study, I investigated whether Dutch speakers take into account a referent's predictability to choose a particular referring expression type. I also investigated whether taking into account referent predictability is related to the presence of an actual addressee. I operationalized predictability in two ways: First, I varied the next-mention biases in a story continuation experiment, such that either the subject or the object of the critical context sentence was more likely to be mentioned in the continuation. This was done by manipulating the adverb at the start of the critical context sentence: with a temporal or manner adverb, the next-mention bias was expected to follow the implicit consequentiality bias of the verb in the context sentence, which was always towards the object. With the adverb *eerst* 'first', the next-mention bias was expected to shift towards the subject, due to a parallel coherence relation between the context sentence and the continuation. This manipulation was found to produce the expected results in the written experiment reported in Vogels (2019) and in Pretest 1 of the present study. Second, based on the next-mention proportions of Pretest 1, I calculated the *surprisal* of each referent – the negative log probability of mentioning the referent as the subject of the story fragment continuation. The results of the main, spoken experiment showed that speakers indeed produced a higher rate of pronouns for the subject referent after *eerst* than after another type of adverb, and a lower rate of pronouns when the referent was more surprising. I found no evidence that this effect was more pronounced when an actual addressee was present who had to find the correct picture based on the speaker's story continuation, than when no addressee was present. This suggests that the presence of an actual addressee is not the driving factor behind the occurrence of the predictability effect on pronoun use.

Crucially, however, the fact that an effect of predictability on pronoun use was found in the first place is noteworthy in itself, given that several story continuation experiments have failed to find a relationship between predictability and rate of pronoun use (Fukumura & Van Gompel, 2010; Hoek et al., 2021; Rohde & Kehler, 2014; Stevenson et al., 1994). The present study differed in several respects from these previous studies. First, while in previous work predictability has often been operationalized as the next-mention biases that are assumed to exist for certain types of verbs, I elicited such biases independently from a separate experiment, and also used those biases to calculate the surprisal of each referent in each context as a more direct measure of predictability (following Tily & Piantadosi, 2009). Variation in the degree of predictability was created by manipulating the default assumed biases, which had the additional advantage of making the biases independent of the referents' thematic roles.

Second, whereas continuation experiments often present participants with single-sentence written fragments, the stories in the current study were all accompanied by pictures. This provides more control over which referent speakers will mention, allowing the researcher to avoid stories always continuing with the most predictable character. On top of that, the stories had an introductory sentence preceding the critical fragment, which set the scene for the story. Together

with the pictures, this additional contextualization may constrain speakers' conceptualization of the story, since speakers do not need to fall back on their own imagination as much to make sense of it. As Rosa and Arnold (2017) argue, contextual constraints may make it easier for speakers to create a mental model of a story, which may facilitate the role of semantic factors in language production. Recent support for this view comes from a large-scale web-based story fragment continuation study, in which predictability effects on pronoun use were only found if the story fragments were presented in a larger story context (Demberg et al., 2023). In addition, although story continuation tasks are somewhat unnatural by definition, providing participants with a conception of how the story continues more closely matches a natural story-telling situation, in which speakers know beforehand what they are going to tell, instead of having to invent that on the fly (Weatherford & Arnold, 2021).

Finally, most previous studies on the effect of predictability on the choice of referring expression have been done on English (but see Bott et al., 2018, for research on German, Kravtchenko, 2014, for Russian, and Medina Fetterman et al., 2022, for Spanish). To my knowledge, the present study is the first to investigate this effect in Dutch. The Dutch pronoun paradigm differs from the English one in that it is characterized by a distinction between full and reduced forms. As such, it forms in principle a good testing ground for the hypothesis that predictability only affects referential choices as far as they are acoustic or prosodic in nature (e.g., Kaiser et al., 2011). Although the present study cannot provide a conclusive answer, due to the low frequency of full pronouns, the finding that predictability affected the choice between a (reduced) pronoun and a full NP is inconsistent with the acoustic reduction hypothesis. Rather, full pronouns may pragmatically implicate that although the referent is salient, it is not the one that was most expected (Vogels, 2019; cf. also findings for null vs. overt forms in pro-drop languages; Kravtchenko, 2014; Medina Fetterman et al., 2022). Whether the choice between a full and a reduced pronoun in Dutch is part of a fast automatic process (Bard et al., 2000) or of a more effortful perspective-taking process remains to be seen.

Any combination of the three factors mentioned above (operationalization of predictability, contextual constraints, and language-specific properties) may have contributed to the presence of a predictability effect in the current study, where it was absent in other studies. However, these factors do not explain *why* predictability should have an effect on the choice of referring expression. According to the strong version of the Bayesian model of reference production and interpretation proposed by Kehler and colleagues (Kehler et al., 2008; see also Hoek et al., 2021; Kehler & Rohde, 2013; Rohde & Kehler, 2014), the likelihood that a referent will be mentioned should only affect reference interpretation and not production: For a listener, the probability that a given pronoun refers to a particular referent is determined by the probability that the speaker would have used a pronoun to refer to that referent as well as the probability that this referent is mentioned at all. In this way, the Bayesian model captures the idea that listeners take into account the speaker's perspective when interpreting referring expressions. However, the choice

of referring expression type is assumed to be determined instead by grammatical or information structural characteristics of the discourse, such as subjecthood or topicality (Rohde & Kehler, 2014). This implies that speakers are not taking into account expectations from the listener's perspective in their choice to produce a pronoun.

Still, there is ample evidence that speakers, at least in certain situations, adjust their referential choices to properties of the listener's perspective (e.g., Ahn & Brown-Schmidt, 2020; Brennan & Hanna, 2009; Galati & Brennan, 2010; Hawkins et al., 2021; Hendriks et al., 2014; Kuhlen & Brennan, 2010; Loy et al., 2020; Tal et al., 2023; Vogels et al., 2020). In this paper, I have assumed that taking into account the predictability of referents is related to speakers taking the perspective of their addressee (cf. Demberg et al., 2023; Orita et al., 2015). If speakers only calculate the listener's perspective in some cases, for instance when there are sufficient processing resources available (Hawkins et al., 2021; Hendriks, 2016; Horton & Keysar, 1996) and/or the risk of miscommunication is high (Mustajoki, 2012), this might explain some of the divergent results in predictability effects on pronoun use. Since previous story completion experiments that do not show effects of predictability typically do not include an addressee, but just require participants to write (or speak) a plausible continuation to each sentence fragment (e.g., Fukumura & Van Gompel, 2010; Hoek et al., 2021; Kaiser et al., 2011; Rohde & Kehler, 2014), I hypothesized that including a co-present addressee in the study design may enhance the role of predictability. This would be in line with research showing that the presence or absence of an actual addressee matters in the speaker's choice of referring expression (e.g., Kantola & van Gompel, 2016; Van Der Wege, 2009).

The finding in the present study that the presence of an actual addressee did not interact with the predictability effect seems to be at odds with the perspective-taking hypothesis. One possible explanation for the lack of an addressee effect is that perspective taking does not need to entail the creation of an explicit representation of a specific interlocutor's mental state; the listener's perspective may also be more generically integrated in the speaker's linguistic experience (both as a speaker and as a listener) over time (Dell & Brown, 1991; Grigoroglou & Papafragou, 2019). Under this broader definition of perspective taking, speakers only need to establish, based on their linguistic experience, whether their own preferred form (e.g., a pronoun) is also the form that a generic or hypothetical listener would interpret as referring to the intended referent (Hendriks, 2016). This explanation is in line with rational models of reference, in which referential choices are determined by a tradeoff between speaker effort and informativity, of which the latter is based on the speaker's estimate of listeners' beliefs (e.g., Degen et al., 2020; Demberg et al., 2023; Orita et al., 2015). Crucially, this listener as modelled in the speaker's mind is typically a *literal listener*, that is, a simplified model of a listener that does not, in turn, also take the speaker's perspective into account. This simple listener model may then be complemented with beliefs about the actual interlocutor's mental state (Hawkins et al., 2021). Hence, while predictability effects in referential choices may be driven by a perspective-taking

process, it is not necessarily the actual addressee's perspective that is calculated on the spot, but may rather be a more abstract, hypothetical addressee that is represented in the speaker's mind through repeated experience of being a listener herself.

An alternative explanation for the lack of an addressee effect may be that predictability effects on referential choices are due to production facilitation after all. That is, predictable referents may be easier to retrieve from memory, which in turn results in shorter forms to refer to predictable referents. So far, empirical studies have only shown production facilitation effects of predictable referents at the acoustic level (e.g., duration; Arnold et al., 2012; Kahn & Arnold, 2012, 2015; Zerkle et al., 2017), but not on the selection of a particular form of expression (Rosa & Arnold, 2017; Zerkle & Arnold, 2016). Recently, however, Karimi (2022) has argued that a more accurate measure for predictability as a factor in speaker effort may be *referential entropy*, which takes into account the probability distribution over all potential referents in the context instead of just the likelihood of mentioning the target referent (see also Modi et al., 2017; Pickering & Gambi, 2018; Tily & Piantadosi, 2009). Referential entropy captures the uncertainty about which referent will be mentioned: when there are multiple equiprobable referents, entropy is high, whereas entropy is low when there is a single high-probability referent (Hale, 2006). Karimi argues that in language production, referential entropy reflects the distribution of the speaker's attentional resources over multiple possible referents, with higher entropy resulting in a lower activation of each referent's mental representation and, in turn, a decrease in pronoun use. This account is parallel to speaker-internal accounts of accessibility effects on pronoun use: According to these accounts, as the discourse contains more potential referents, the speaker's attentional resources will be distributed over more mental representations, leading to lower activation of each referent's representation. In consequence, more effort is needed to retrieve a referent's representation from memory, resulting in more elaborate referring expressions (e.g., Arnold & Griffin, 2007; Fukumura & van Gompel, 2012). Given the potential relation between referential entropy and activation of mental representations of referents in the speaker's discourse model, it is not clear that speaker-internal explanations for predictability effects can be fully excluded yet.

In sum, the challenge for future studies on the role of predictability in referential choices is to find experimental paradigms that can dissociate a specific- from a generic-listener perspective, as well as listener- from speaker-oriented production processes. The crucial finding to take home from the present study is, however, that irrespective of the presence of an actual addressee, referent predictability affects the rate of pronoun use in Dutch. This result supports earlier findings for German and English showing effects of next-mention biases associated with certain thematic roles (e.g., Bott & Solstad, 2022; Demberg et al., 2023; Rosa & Arnold, 2017; Weatherford & Arnold, 2021). The present study thereby contributes to the debate on when predictability affects the choice of referring expression by providing evidence that addressee presence is not the determining factor.

Appendix

Table A.1. By-item proportions and surprisal values for NP1 and NP2 referents in each adverb condition, calculated from Pretest 1, as well as event expectancy ratings taken from Pretest 2.

Item	Referent	Adverb	Proportion	Surprisal	Event expectancy rating
1	NP1	first	.57	0.81	3.43
1	NP2	first	.29	1.81	3.67
1	NP1	neutral	.31	1.70	3.20
1	NP2	neutral	.38	1.38	2.50
2	NP1	first	.57	0.81	4.00
2	NP2	first	.14	2.81	3.25
2	NP1	neutral	.23	2.12	3.14
2	NP2	neutral	.69	0.53	4.00
3	NP1	first	.45	1.14	2.50
3	NP2	first	.27	1.87	1.20
3	NP1	neutral	.20	2.32	2.17
3	NP2	neutral	.60	0.74	2.14
4	NP1	first	.69	0.53	2.40
4	NP2	first	.23	2.12	2.86
4	NP1	neutral	.29	1.81	3.50
4	NP2	neutral	.71	0.49	3.00
5	NP1	first	.57	0.81	1.43
5	NP2	first	.29	1.81	3.50
5	NP1	neutral	.08	3.70	1.40
5	NP2	neutral	.54	0.89	3.75
6	NP1	first	.14	2.81	2.67
6	NP2	first	.86	0.22	3.75

(Contd.)

Item	Referent	Adverb	Proportion	Surprisal	Event expectancy rating
6	NP1	neutral	.08	3.70	3.00
6	NP2	neutral	.69	0.53	1.40
7	NP1	first	.54	0.89	2.50
7	NP2	first	.38	1.38	3.00
7	NP1	neutral	.57	0.81	2.83
7	NP2	neutral	.43	1.22	4.29
8	NP1	first	.50	1.00	3.00
8	NP2	first	.50	1.00	3.00
8	NP1	neutral	.00	Inf	3.50
8	NP2	neutral	.86	0.22	2.00
9	NP1	first	.50	1.00	2.57
9	NP2	first	.50	1.00	2.50
9	NP1	neutral	.08	3.70	1.80
9	NP2	neutral	.85	0.24	2.50
10	NP1	first	.00	Inf	3.33
10	NP2	first	1.00	0.00	2.50
10	NP1	neutral	.00	Inf	2.57
10	NP2	neutral	.60	0.74	2.60
11	NP1	first	.85	0.24	4.00
11	NP2	first	.15	2.70	4.00
11	NP1	neutral	.29	1.81	4.83
11	NP2	neutral	.57	0.81	3.57
12	NP1	first	.46	1.12	2.40
12	NP2	first	.46	1.12	4.43
12	NP1	neutral	.43	1.22	2.00

(Contd.)

Item	Referent	Adverb	Proportion	Surprisal	Event expectancy rating
12	NP2	neutral	.14	2.81	4.83
13	NP1	first	.75	0.42	4.71
13	NP2	first	.25	2.00	1.50
13	NP1	neutral	.15	2.70	3.80
13	NP2	neutral	.77	0.38	1.75
14	NP1	first	.43	1.22	4.00
14	NP2	first	.43	1.22	4.50
14	NP1	neutral	.25	2.00	3.14
14	NP2	neutral	.58	0.78	3.00
15	NP1	first	.46	1.12	3.25
15	NP2	first	.38	1.38	3.60
15	NP1	neutral	.43	1.22	3.33
15	NP2	neutral	.43	1.22	3.43
16	NP1	first	.62	0.70	1.80
16	NP2	first	.38	1.38	1.14
16	NP1	neutral	.50	1.00	1.25
16	NP2	neutral	.33	1.58	2.17
17	NP1	first	.43	1.22	3.29
17	NP2	first	.43	1.22	3.33
17	NP1	neutral	.00	Inf	4.20
17	NP2	neutral	.83	0.26	4.50
18	NP1	first	.57	0.81	1.50
18	NP2	first	.14	2.81	2.25
18	NP1	neutral	.42	1.26	1.14
18	NP2	neutral	.50	1.00	1.60

(Contd.)

Item	Referent	Adverb	Proportion	Surprisal	Event expectancy rating
19	NP1	first	.08	3.58	1.25
19	NP2	first	.67	0.58	1.00
19	NP1	neutral	.00	Inf	1.67
19	NP2	neutral	.43	1.22	1.14
20	NP1	first	.46	1.12	2.20
20	NP2	first	.46	1.12	1.71
20	NP1	neutral	.50	1.00	3.25
20	NP2	neutral	.00	Inf	2.83
21	NP1	first	.43	1.22	3.57
21	NP2	first	.57	0.81	3.33
21	NP1	neutral	.15	2.70	2.80
21	NP2	neutral	.62	0.70	2.00
22	NP1	first	.33	1.58	3.00
22	NP2	first	.50	1.00	3.25
22	NP1	neutral	.00	Inf	1.14
22	NP2	neutral	.85	0.24	2.80
23	NP1	first	.55	0.87	3.00
23	NP2	first	.27	1.87	2.00
23	NP1	neutral	.29	1.81	2.50
23	NP2	neutral	.29	1.81	4.14
24	NP1	first	.45	1.14	3.40
24	NP2	first	.45	1.14	2.71
24	NP1	neutral	.14	2.81	4.00
24	NP2	neutral	.43	1.22	2.50

Abbreviations

FEM feminine

FULL full pronoun

RED reduced pronoun

Data accessibility statement

The materials, raw data and analysis scripts for the experiments in this paper can be accessed from the following OSF repository: <https://doi.org/10.17605/OSF.IO/CMQKA>

Ethics and consent

The research was approved by the faculty's ethical committee (CETO: 64503791), and all participants gave informed consent for the use of their data.

Acknowledgements

This research was supported by the Netherlands Organization for Scientific Research (NWO) under grant 275-89-0360. I thank the members of the Semantics and Cognition group at the University of Groningen and two anonymous reviewers for their helpful comments. I am also grateful to Maeike Slikkerveer, Paula Sportel and Emke Sijtsma for their help in conducting the experiments and coding the data.

Competing interests

The author has no competing interests to declare.

Authors' contributions

JV: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, supervision, visualization, writing – original draft, writing – review and editing.

References

Ahn, S., & Brown-Schmidt, S. (2020). Retrieval processes and audience design. *Journal of Memory and Language*, 115, 104149. DOI: <https://doi.org/10.1016/j.jml.2020.104149>

Aina, L., Liao, X., Boleda, G., & Westera, M. (2021). Does referent predictability affect the choice of referential form? A computational approach using masked coreference resolution. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pp. 454–469, Online. Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/2021.conll-1.36>

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. DOI: [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge.
- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, 31(2), 137–162. DOI: https://doi.org/10.1207/S15326950DP3102_02
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527. DOI: <https://doi.org/10.1080/01690960801920099>
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4), 187–203. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00193.x>
- Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56(4), 521–536. DOI: <https://doi.org/10.1016/j.jml.2006.09.007>
- Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, 19(3), 505–512. DOI: <https://doi.org/10.3758/s13423-012-0233-y>
- Arnold, J. E., & Zerkle, S. A. (2019). Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience*, 34(9), 1152–1175. DOI: <https://doi.org/10.1080/23273798.2019.1636103>
- Arts, A., Maes, A., Noordman, L. G. M., & Jansen, C. (2011). Overspecification in written instruction. *Linguistics*, 49(3), 555–574. DOI: <https://doi.org/10.1515/ling.2011.017>
- Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56. DOI: <https://doi.org/10.1177/00238309040470010201>
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1–22. DOI: <https://doi.org/10.1006/jmla.1999.2667>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv Preprint arXiv:1506.04967*. DOI: <https://doi.org/10.48550/arXiv.1506.04967>
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native ‘um’s elicit prediction of low-frequency referents, but non-native ‘um’s do not. *Journal of Memory and Language*, 75, 104–116. DOI: <https://doi.org/10.1016/j.jml.2014.05.004>
- Bott, O., & Solstad, T. (2022). *The production of referring expressions is influenced by next-mention likelihood*. HSP 2022.
- Bott, O., Solstad, T., & Pryslopska, A. (2018). *Implicit causality affects the choice of anaphoric form*. AMLaP 2018.

- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482. DOI: <https://doi.org/10.1037/0278-7393.22.6.1482>
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Chapter 8 – Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 53, pp. 301–344). Academic Press. DOI: [https://doi.org/10.1016/S0079-7421\(10\)53008-1](https://doi.org/10.1016/S0079-7421(10)53008-1)
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2), 274–291. DOI: <https://doi.org/10.1111/j.1756-8765.2009.01019.x>
- Brown-Schmidt, S., & Hanna, J. (2011). Talking in another person’s shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, 2(1), 11–33. DOI: <https://doi.org/10.5087/dad.2011.102>
- Caffarra, S., Wolpert, M., Scarinci, D., & Mancini, S. (2020). Who are you talking to? The role of addressee identity in utterance comprehension. *Psychophysiology*, 57(4), e13527. DOI: <https://doi.org/10.1111/psyp.13527>
- Cheng, W. (2016). *Implicit causality and consequentiality in native and non-native coreference processing*. <http://scholarcommons.sc.edu/etd/3830>
- Christensen, R. H. B. (2015). *A tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal package*. <https://CRAN.R-project.org/package=ordinal>
- Clark, H. H. (1996). *Using language*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511620539>
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Le Ny & W. Kintsch (Eds.), *Advances in psychology* (Vol. 9, pp. 287–299). North-Holland. DOI: [https://doi.org/10.1016/S0166-4115\(09\)60059-5](https://doi.org/10.1016/S0166-4115(09)60059-5)
- Commandeur, E. (2010). *Implicit causality and implicit consequentiality in language comprehension*. https://pure.uvt.nl/portal/files/1240111/Proefschrift_Edwin_Commandeur_300610.pdf
- Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to “overinformative” referring expressions. *Psychological Review*, 127(4), 591–621. DOI: <https://doi.org/10.1037/rev0000186>
- Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the “model of the listener”. In D. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Lawrence Erlbaum Associates Inc.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210. DOI: <https://doi.org/10.1016/j.cognition.2008.07.008>
- Demberg, V., Kravtchenko, E., & Loy, J. E. (2023). A systematic evaluation of factors affecting referring expression choice in passage completion tasks. *Journal of Memory and Language*, 130, 104413. DOI: <https://doi.org/10.1016/j.jml.2023.104413>
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327–339. DOI: <https://doi.org/10.1037/0022-3514.87.3.327>

- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998–998. DOI: <https://doi.org/10.1126/science.1218633>
- Fukumura, K. (2015). Interface of linguistic and visual information during audience design. *Cognitive Science*, 39(6), 1419–1433. DOI: <https://doi.org/10.1111/cogs.12207>
- Fukumura, K., & Van Gompel, R. P. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, 62(1), 52–66. DOI: <https://doi.org/10.1016/j.jml.2009.09.001>
- Fukumura, K., & van Gompel, R. P. (2012). Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model? *Cognitive Science*, 36(7), 1289–1311. DOI: <https://doi.org/10.1111/j.1551-6709.2012.01255.x>
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62(1), 35–51. DOI: <https://doi.org/10.1016/j.jml.2009.09.002>
- Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, 68(2), 140–159. DOI: <https://doi.org/10.1016/j.jml.2012.10.001>
- Grigoroglou, M., & Papafragou, A. (2019). Children's (and adults') production adjustments to generic and particular listener needs. *Cognitive Science*, 43(10), e12790. DOI: <https://doi.org/10.1111/cogs.12790>
- Grüter, T., & Rohde, H. (2021). Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics*, 51–75. DOI: <https://doi.org/10.1017/S0142716420000582>
- Guan, S., & Arnold, J. E. (2021). The predictability of implicit causes: Testing frequency and topicality explanations. *Discourse Processes*, 58(10), 943–969. DOI: <https://doi.org/10.1080/0163853X.2021.1974690>
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307. DOI: <https://doi.org/10.2307/416535>
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4), 643–672. DOI: https://doi.org/10.1207/s15516709cog0000_64
- Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*, 4(s2). DOI: <https://doi.org/10.1515/lingvan-2017-0027>
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28(1), 105–115. DOI: https://doi.org/10.1207/s15516709cog2801_5
- Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience*, 30(6), 716–734. DOI: <https://doi.org/10.1080/23273798.2015.1008524>

- Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, 45(3), e12926. DOI: <https://doi.org/10.1111/cogs.12926>
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108(3), 831–836. DOI: <https://doi.org/10.1016/j.cognition.2008.04.008>
- Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology*, 7. DOI: <https://doi.org/10.3389/fpsyg.2016.00506>
- Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, 29(4), 391–407. DOI: <https://doi.org/10.1080/01690965.2013.766356>
- Hoek, J., Kehler, A., & Rohde, H. (2021). Pronominalization and expectations for re-mention: Modeling coreference in contexts with three referents. *Frontiers in Communication*, 6, 192. DOI: <https://doi.org/10.3389/fcomm.2021.674126>
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117. DOI: [https://doi.org/10.1016/0010-0277\(96\)81418-1](https://doi.org/10.1016/0010-0277(96)81418-1)
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. DOI: <https://doi.org/10.1016/j.cogpsych.2010.02.002>
- Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, 67(3), 311–325. DOI: <https://doi.org/10.1016/j.jml.2012.07.002>
- Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, 30(1–2), 103–119. DOI: <https://doi.org/10.1080/01690965.2013.848989>
- 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>
- Kaiser, E., Li, D. C.-H., & Holsinger, E. (2011). Exploring the lexical and acoustic consequences of referential predictability. In I. Hendrickx, S. Lalitha Devi, A. Branco & R. Mitkov (Eds.), *Anaphora processing and applications* (Vol. 7099, pp. 171–183). Springer. DOI: https://doi.org/10.1007/978-3-642-25917-3_15
- Kaiser, E., & Trueswell, J. C. (2004). The referential properties of Dutch pronouns and demonstratives: Is salience enough? *Proceedings of Sinn und Bedeutung*, 8, 137–150. DOI: <https://doi.org/10.18148/sub/2004.v8i0.754>
- Kantola, L., & van Gompel, R. P. (2016). Is anaphoric reference cooperative? *The Quarterly Journal of Experimental Psychology*, 69(6), 1109–1128. DOI: <https://doi.org/10.1080/17470218.2015.1070184>
- Karimi, H. (2022). Greater entropy leads to more explicit referential forms during language production. *Cognition*, 225, 105093. DOI: <https://doi.org/10.1016/j.cognition.2022.105093>
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1–44. DOI: <https://doi.org/10.1093/jos/ffm018>

- 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>
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38. DOI: <https://doi.org/10.1111/1467-9280.00211>
- Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13), 3231–3250. DOI: <https://doi.org/10.1016/j.pragma.2011.06.008>
- Koornneef, A. W., & Sanders, T. J. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, 28(8), 1169–1206. DOI: <https://doi.org/10.1080/01690965.2012.699076>
- Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in Russian. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (Vol. 36, pp. 785–790). <https://cogsci.mindmodeling.org/2014/>
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56(3), 436–455. DOI: <https://doi.org/10.1016/j.jml.2006.05.002>
- Kronmüller, E., & Guerra, E. (2020). Processing speaker-specific information in two stages during the interpretation of referential precedents. *Frontiers in Psychology*, 11. DOI: <https://doi.org/10.3389/fpsyg.2020.552368>
- Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, 47(7), 567–587. DOI: <https://doi.org/10.1080/01638530903441339>
- Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20(1), 54–72. DOI: <https://doi.org/10.3758/s13423-012-0341-8>
- 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>
- 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>
- Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. DOI: <https://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, 849–856. DOI: <https://doi.org/10.7551/mitpress/7503.003.0111>

- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6(3), 172–187. DOI: <https://doi.org/10.1177/002383096300600306>
- Loy, J. E., Bloomfield, S. J., & Smith, K. (2020). Effects of priming and audience design on the explicitness of referring expressions: Evidence from a confederate priming paradigm. *Discourse Processes*, 57(9), 808–821. DOI: <https://doi.org/10.1080/0163853X.2020.1802192>
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. DOI: <https://doi.org/10.1016/j.cognition.2012.09.010>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. DOI: <https://doi.org/10.1016/j.jml.2017.01.001>
- Medina Fetterman, A. M., Vazquez, N. N., & Arnold, J. E. (2022). The effects of semantic role predictability on the production of overt pronouns in Spanish. *Journal of Psycholinguistic Research*, 51(1), 169–194. DOI: <https://doi.org/10.1007/s10936-021-09832-w>
- Modi, A., Titov, I., Demberg, V., Sayeed, A., & Pinkal, M. (2017). Modeling semantic expectation: Using script knowledge for referent prediction. *Transactions of the Association for Computational Linguistics*, 5, 31–44. DOI: https://doi.org/10.1162/tacl_a_00044
- Mustajoki, A. (2012). A speaker-oriented multidimensional approach to risks and causes of miscommunication. *Language and Dialogue*, 2(2), 216–243. DOI: <https://doi.org/10.1075/ld.2.2.03mus>
- Orita, N., Vornov, E., Feldman, N., & Daumé III, H. (2015). Why discourse affects speakers' choice of referring expressions. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1639–1649. DOI: <https://doi.org/10.3115/v1/P15-1158>
- Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, 1291, 92–101. DOI: <https://doi.org/10.1016/j.brainres.2009.07.042>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291. DOI: <https://doi.org/10.1016/j.cognition.2011.10.004>
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002–1044. DOI: <https://doi.org/10.1037/bul0000158>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. DOI: <https://doi.org/10.1016/j.tics.2006.12.002>
- Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, 209, 104491. DOI: <https://doi.org/10.1016/j.cognition.2020.104491>
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912–927. DOI: <https://doi.org/10.1080/01690965.2013.854918>

- Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, 94, 43–60. DOI: <https://doi.org/10.1016/j.jml.2016.07.007>
- Rosa, E. C., Finch, K. H., Bergeson, M., & Arnold, J. E. (2015). The effects of addressee attention on prosodic prominence. *Language, Cognition and Neuroscience*, 30(1–2), 48–56. DOI: <https://doi.org/10.1080/01690965.2013.772213>
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7. DOI: <https://doi.org/10.3389/fpsyg.2016.00153>
- Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science*, 43(11), e12797. DOI: <https://doi.org/10.1111/cogs.12797>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656. DOI: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. DOI: <https://doi.org/10.1016/j.cognition.2013.02.013>
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9(4), 519–548. DOI: <https://doi.org/10.1080/01690969408402130>
- Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant referents with language learners: Evidence for communicatively-efficient referential choice. *Journal of Memory and Language*, 128, 104378. PsyArXiv. DOI: <https://doi.org/10.31234/osf.io/cw2be>
- Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference*.
- Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language*, 60(4), 448–463. DOI: <https://doi.org/10.1016/j.jml.2008.12.003>
- Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, 37(3), 731–766. DOI: <https://doi.org/10.1017/S0305000909990560>
- Vogels, J. (2019). Both thematic role and next-mention biases affect pronoun use in Dutch. In A. K. Goel, C. M. Seifert & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 3029–3035). Cognitive Science Society.
- Vogels, J., Howcroft, D. M., Tourtouri, E., & Demberg, V. (2020). How speakers adapt object descriptions to listeners under load. *Language, Cognition and Neuroscience*, 35(1), 78–92. DOI: <https://doi.org/10.1080/23273798.2019.1648839>
- Vogels, J., Krahmer, E., & Maes, A. (2015). How cognitive load influences speakers' choice of referring expressions. *Cognitive Science*, 39(6), 1396–1418. DOI: <https://doi.org/10.1111/cogs.12205>

- Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, 20(4), 766–772. DOI: <https://doi.org/10.3758/s13423-013-0396-1>
- Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106(3), 1548–1557. DOI: <https://doi.org/10.1016/j.cognition.2007.06.009>
- Weatherford, K. C., & Arnold, J. E. (2021). Semantic predictability of implicit causality can affect referential form choice. *Cognition*, 214, 104759. DOI: <https://doi.org/10.1016/j.cognition.2021.104759>
- Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology*, 6, 935. DOI: <https://doi.org/10.3389/fpsyg.2015.00935>
- Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, 176, 15–30. DOI: <https://doi.org/10.1016/j.cognition.2018.03.002>
- Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, 18(7), 600–606. DOI: <https://doi.org/10.1111/j.1467-9280.2007.01946.x>
- Zarcone, A., & Demberg, V. (2021). A bathtub by any other name: The reduction of German compounds in predictive contexts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43.
- Zarcone, A., Van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and attention in surprisal-based accounts of language processing. *Frontiers in Psychology*, 7. DOI: <https://doi.org/10.3389/fpsyg.2016.00844>
- Zerkle, S. A., & Arnold, J. E. (2016). Discourse attention during utterance planning affects referential form choice. *Linguistics Vanguard*, 2(s1). DOI: <https://doi.org/10.1515/lingvan-2016-0067>
- Zerkle, S. A., & Arnold, J. E. (2019). Does pre-planning explain why predictability affects reference production? *Dialogue & Discourse*, 10(2), 34–55. DOI: <https://doi.org/10.5087/dad.2019.202>
- Zerkle, S. A., Rosa, E. C., & Arnold, J. E. (2017). Thematic role predictability and planning affect word duration. *Laboratory Phonology*, 8(1). DOI: <https://doi.org/10.5334/labphon.98>

