Discourse attention during utterance planning affects referential form choice

Keywords (3-5):
discourse, utterance planning, referential form, individual variation, language production

Abstract
An unstudied source of linguistic variation is the use of discourse-appropriate language. Sometimes individuals use linguistic devices (anaphors, connectors) to connect utterances to the discourse context, and sometimes not. We asked how this variation is related to utterance planning, using eyetracking with a narrative production task. Participants saw picture pairs depicting two events. They heard a description of the first event (Context picture), then added to the story by describing the second event (Target picture). We found that one group of participants produced utterances that connected with the discourse context (Context-Users), using pronouns/zeros and connectors (and/then) as appropriate, while another group consistently used definite NP descriptions and virtually no connectors (Context-Ignorers). Eyetracking measures reflected utterance planning within a discourse context: all participants shifted their attention from the Context picture to the Target picture throughout a trial. We also observed group differences: Context-Users directed their attention in a more systematic way than Context-Ignorers. At trial onset, Context-Users looked more at the Context picture than Context-Ignorers. Right before speaking, they looked more at the Target picture than Context-Ignorers. The Context-Users also had shorter latency to begin speaking. This study provides a first step toward characterizing individual differences in terms of utterance planning.
1 Introduction

1.1 Variation in the production of contextually-appropriate language

Successful communication requires speakers and listeners to recognize the relationship between each utterance and the prior linguistic context. Speakers can use various linguistic devices to establish this connection, such as connector words, emphasis in prosody, and specific word orders. Yet strikingly, there is substantial variation across situations and speakers in the use of these devices. Sometimes speakers produce indicators of how an utterance relates to the prior discourse and current context, while in other cases an equivalent utterance may be produced as if it were in isolation. For example, imagine a discourse in which two speakers describe the second panel in Figure 1, within the same context.

(1) Discourse Context: The Duchess handed a painting to the Duke.
   Speaker 1: The Duke threw the painting into the closet.
   Speaker 2: and then he threw it in the closet.

Figure 1 here

Both speakers describe the same action, but Speaker 2 (a “Context-User”) uses connectors and pronouns to signal a link with the prior context, and Speaker 1 (a “Context-Ignorer”) does not. Descriptions or names can be produced without considering the relationship between the current event and the prior context, while pronouns and zeros are dependent on the prior discourse. That is, given the same task, some individuals produce utterances that signal their relation to the prior discourse, while others do not. This variation has received little attention in the literature.

This paper is a first step towards understanding this variation, by examining how it relates to measures of utterance planning. While we did not collect individual difference measures (e.g. working memory, print exposure), planning and other executive function processes are known to vary across individuals (e.g., Englehardt, Nigg, & Ferreira, 2013). We hypothesize that individuals differ in the utterance planning strategies they recruit for particular tasks, and in particular whether they pay attention to the discourse context.

We focus in particular on variation in referential form, e.g. the contrast between descriptions like The Duke, and reduced expressions like pronouns (e.g., he) or zeros\(^1\) (e.g., …and threw…). This choice is conditioned by the discourse context, where reduced referential

---

\(^1\) The construction we are coding as “zero” is not a zero pronoun per se, but typically is a VP coordination structure in our data. We group this structure with pronouns because it represents one way of describing the event without a repeated description.
Discourse Attention During Utterance Planning Affects Referential Form Choice

expressions tend to be used when the referent is given and accessible (Ariel, 1990, 1996; Arnold, 2001, 2008; Brennan, 1995; Chafe, 1976, 1995; Givon, 1983). Nevertheless, there is substantial variation in the way that different speakers utilize these devices. For example, Figure 2 illustrates the distribution of individuals’ overall reduced referential form production in three similar experiments.

Figure 2 here

Why do individuals vary in their sensitivity to the previous discourse? We tackle this novel question by asking how the usage of discourse devices, like referential form, is related to variation in utterance planning and production. Using an eyetracking task, we test the hypothesis that planning strategies influence the activation of the prior context and its relation to the current utterance, and that this influences the use of linguistic forms like pronouns.

Eyetracking has been used to study planning mechanisms, but usually without a discourse context. Evidence suggests that fixations to visual stimuli reflect the conceptual and linguistic planning occurring before language production. Trial-initial looks indicate rapid planning: for example, speakers direct their attention to the part of the picture that represents the first referent in their utterance (Bock et al., 2003). Speakers also tend to look at objects immediately before naming them (Griffin & Bock, 2000).

However, this tight coordination of gaze and speech might be more representative of the formulation of sentences out of context, where all information is in a to-be-described event that is new and unfocused. Supporting this hypothesis, Van der Meulen et al. (2001) found that speakers do not always look at the objects they name: speakers allocate less visual attention to given objects than to new ones, and less visual attention to objects they will refer to with a pronoun than with a full noun phrase. This suggests that discourse focus (given vs. new) and referential form (pronoun vs. name) modulate the amount of visual attention allocated to a referent. One drawback to this study is that participants were instructed to use a pronoun vs. name, instead of choosing their own reference form more naturally. Ganushchak et al. (2014) also manipulated discourse focus and found strong effects of top-down guidance from the message level and contextual facilitation of linguistic encoding: speakers deployed their gaze only to the character they needed to encode to answer a question. This suggests that speakers direct their visual attention to whichever part of the display they need to process with priority to produce a contextually fitting utterance.
Discourse Attention During Utterance Planning Affects Referential Form Choice

Our study extends this type of work to examine utterance planning within a rich discourse context. We use a task that presents visual stimuli for both a context and target utterance simultaneously, which allows speakers to begin planning during the prior utterance, similar to real-life situations (e.g., as soon as they see both pictures in Figure 1). One way to investigate planning within a context is by comparing relative preferences to fixate the Context vs. Target picture (Figure 1) throughout an entire trial.

1.2 Study goals and task overview

The goal of our study was to examine the relationship between the speaker’s attention during utterance planning, and their use of referential form to signal a connection with the prior context. We used a paradigm in which speakers viewed pairs of pictures (like Figure 1), and heard a sentence about the first picture. Their job was to describe the second picture. Thus, fixations on the Context (first) picture represented attention to the discourse context, and fixations on the Target (second) picture represented planning the target utterance.

Our paradigm was adapted from a publicly-available paradigm developed by Rosa & Arnold (under review), and modified for use with eyetracking. To allow for accurate timing, the task was modified from its original design by eliminating the live addressee. Instead, the context sentence was recorded and played through headphones on each trial. We also changed the names of three characters, such that all of our character names were role descriptions like “the duchess” and “the butler.”

Following Rosa & Arnold’s experimental design, our study manipulated two aspects of the discourse context that are expected to affect referential form choice: both the thematic roles (goal vs. source) and grammatical roles (subject vs. non-subject) of the characters in the discourse context. The context sentences described transfer events, e.g. *The chef handed a cookbook to the maid*. By reversing the order of mention of the goal and source, we manipulated the grammatical role position of each thematic role, e.g. *The maid took a cookbook from the chef*. If speakers take the context into account when choosing referential expressions, we expected them to use reduced expressions more often for reference to the subject of the context sentence than the non-subject, and more often for goal than source characters (Arnold, 2001; Rosa & Arnold, under review).

However, our results revealed an even stronger source of variation: individual differences in overall sensitivity to the discourse context. Our data revealed two clear groups that differed in their use of reduced references and therefore, in their sensitivity to the discourse context. Some participants varied their word choice in pragmatically appropriate ways, using a mix of pronouns and descriptions for reference. We called these participants the “Context Users.” Other
participants exhibited no variation at all on the critical trials. We called these participants the “Context Ignorers.” The focus of the analyses in this paper is on differences between these two groups of participants, as measured by both linguistic formulation choices (reference form, connector use) and measures of utterance planning (fixations to the visual stimuli and latency to begin speaking). We examine planning measures to test whether fixations in our paradigm reflect overall utterance planning in a discourse context, and whether fixations and latency reveal differences in response planning across the two groups of participants.

2 Methods
2.1 Participants
49 native English speakers participated for class credit. 6 participants were excluded for technical reasons (file corruption, audio failure), and 6 due to track loss, leaving 37 participants in the analysis.

2.2 Materials and Design
We used a narrative production paradigm designed by Rosa & Arnold (under review; for stimuli see jaapstimuli.unc.edu), in which subjects were given the role of a tabloid photographer. In a background video, participants were told that they witnessed a murder in a mansion and they happened to capture pictures of the events from that day. Participants learned about six characters in the story: three males (the duke, the butler, and the driver), and three females (the duchess, the maid, and the chef). Participants were asked to describe these pictures to a “detective” (a voice in the computer) to help solve the crime. This engaging task encouraged participants to develop a rich discourse representation, and critically allowed us to test utterance planning within a discourse context.

Participants viewed 53 pairs of “evidence photos,” and heard the detective describe the first picture. In critical trials, the context sentence described a transfer event with two characters (one goal, one source; see Figure 1). The participants’ task was to describe the second picture, which pictured the target character. The target was either the goal or the source of the previous sentence, manipulated between items. We manipulated the detective’s sentence between-subjects to control for the grammatical placement of the Target character, either in the subject or the non-subject position, e.g. “The Duchess handed a painting to the Duke” vs. “The Duke received a painting from the Duchess.”

Filler trials (29) had between one to three characters. All trials had to be presented in the same order for each participant to create a coherent story in which all utterances are related. As a
control, half of the critical trials in each condition had two characters of the same gender, and the other half were of different genders. There were four possible conditions of character continuation: Goal/Subject, Goal/Non-Subject, Source/Subject, and Source/Non-Subject.

2.3 Procedure

Participants sat at a computer, wearing an Eyelink II head-mounted Eyetracker (SR Research) and a headset headphones/microphone. Then they were shown a preview slideshow in PowerPoint, which played the background video (described above), introduced them to the six characters, and showed all 53 pairs of pictures in order (5 s for each pair). The purpose of this preview was to familiarize the participant with the series of the events and the outcome of the story. This imitates the characteristics of natural language production, in which speakers typically relate information they already know. Then participants heard a sample trial from the narrator, and were instructed to describe the second picture after hearing the detective describe the first. They saw all six characters again and named them aloud to the experimenter to check that they knew the characters’ names well.

For each trial in the main task, the pair of pictures appeared on the screen next to each other for 500 ms (Preview period), and stayed on screen while the context sentence played (average 2330 ms duration) and the participant responded. Participants then clicked on a green circle for the next trial (see Figure 3).

3 Results

3.1 General analytic procedure

3.1.2 Response coding

Trials were included only if the participant referred to the target character in the subject position of their utterance. 51 trials did not meet this criterion, and 2 more were excluded due to corrections at the target. This left 835 trials in the analysis (94% inclusion). Research assistants coded responses for type of reference (pronoun/zero, name) use of connective (and, then, and then, and now), and disfluencies (e.g., uh, um).

3.1.3 Audio data coding

Three research assistants analyzed the audio files of included critical trials in Praat (Boersma & Weenink, 2015) to measure latency to begin speaking for each trial, which was calculated as the time between the end of the context sentence and the beginning of fluent speech. If there were any disfluent fillers prior to the response, they contributed to the latency measure.
Inter-rater correlations were calculated between each of the three research assistants for two of the participants, with a range of 0.91-0.99.

### 3.1.4 Eyetracking data coding

We identified three areas of interest: 1) Panel 1 Target; 2) Panel 1 Non-Target, 3) Panel 2 Target; see Figure 3. Ports were rectangles approximately around the entire character’s body, and sometimes also contained parts of objects in the character’s hands. Any area besides these three were considered as “other.” Empirical logits were calculated for each area with the following transformation formula (example area is Panel 2 Target character): \( \log(\text{Sum of looks to Panel 2 Target} + 0.5) / \log(\text{Total sum of looks} - \text{Sum of looks to Panel 2 Target} + 0.5) \) based on Barr’s (2008) methods. These were calculated separately over each window of time, described below in Section 3.3.

Figure 3 here

### 3.1.5 Model-building procedure

Generalized linear-mixed effects models were used to account for the dependencies in the repeated measures in SAS 9.4. PROC GLIMMIX was used for analyses of dichotomous outcomes with a logit link (reference form, connectors), and PROC MIXED was used for analyses of continuous outcomes (empirical logit of looks to each character and latency to begin speaking). In our approach, models of each dependent variable were constructed with random intercepts of participant and item to account for these differences. Effects coding was used for binary predictors, and is reported in each model as comparison group vs. reference group. Random slopes by participant and item were included when appropriate, but if a slope was estimated to be zero it wasn’t included (Searle, Cassella, & McCullouch, 1992). The final models’ fixed and random effects are reported in tables 1-9 in the Appendix.

For each model, first a main effects model was built with just our primary predictors of interest (Goal continuation, Subject continuation, Reference Form Type, and Context Group). Log Latency was included in models that examined the Latency time period. Then, two-way interactions between Goal continuation and Subject continuation, Context Group and Goal continuation, and Context Group and Subject continuation were added to this model. Interactions were only retained in the final model if nearly significant at \( t > 1.5 \). Non-significant interactions are noted under each table. To avoid overfitting each model, no three-way interactions were tested.
3.2 Linguistic measures

The linguistic measures (referential form, connector use, and latency) show that individuals differed in their use of the discourse context when planning an utterance. 24 of the 37 participants only produced descriptions on critical trials; we call them the Context-Ignorers. Even on the subset of Filler trials where only one character acts in both pictures, these participants primarily used descriptions (9 pronouns total out of 252 trials), which supports the conclusion that these participants adopted a response strategy of ignoring the context.

The other 13 participants used some number of reduced forms (pronouns or zeros); we call them the Context-Users. These are participants who appeared to be considering the discourse context as they planned their utterance. Consistent with prior findings (Rosa & Arnold, under review), they used significantly more pronouns/zeros for subject than non-subject continuations, and marginally more for goal than source continuations (see Figure 4, Table 1).

Figure 4 here

Context-Users also used far more connector words (and, and then, so) than Context-Ignorers (33.7% vs. 1.5%; see Table 2, Figure 5). There were also significantly more connectors for utterances containing a pronoun or a zero than for utterances containing a description name. Connector words help signal the relation between two utterances. The fact that connectors were virtually absent from the Context-Ignorers’ speech highlights our conclusion that they were making their utterance in isolation, ignoring its relation to the context.

Figure 5 here

3.3 Planning measures

3.3.1 Latency

Our first question was whether consideration of the context would lead to faster or slower latency (utterance initiation), measured from the offset of the context sentence. In our task, where both pictures were available for the entire trial, the latency measure reflects two processes: a) variation in the time needed to prepare a response, and b) variation in the degree of response planning that occurred during the detective’s speech.

In support of the group differences noted above, we found that Context-Users had a shorter latency to begin speaking fluently than Context-Ignorers (1461 ms vs. 1778 ms; see
This finding is not due to differences in the formulation of the responses; although Context-Users produced more connector words, this difference holds if we restrict the analysis to utterances without connector words, in which case the utterance always started with a reference to the target character. In addition, this finding cannot be explained by disfluency: the Context-Users tended to be more disfluent during the latency period (7.6% of trials) than the Context-Ignorers (2.2%). Disfluency might be expected to increase latency, but instead we observed that Context-Users tended to have shorter latencies.

The tendency for Context-Users to respond more quickly is surprising, because one could expect that processing the context involves creating an abstract discourse model, which might slow the planning process down. Instead, we think it may be that Context-Users were faster to plan their responses/utterances overall because utterance formulation is facilitated in the context of strong discourse representations. Alternatively, it may be that Context-Users engaged in a greater degree of pre-planning. If the speaker were to think about the response while also processing the context sentence, it would require having both events co-activated. Co-activation may support the activation of relations between events, increasing the use of linguistic devices to mark coherence.

We have further evidence that latency is related to utterance planning from the analysis of the experimental conditions predictors. Both groups of participants were faster to begin their utterance for goal than source continuations, resulting in a main effect of goal continuation. This replicates earlier reports (Rosa & Arnold, under review; Zerkle, Rosa, & Arnold, 2016), and suggests that goal continuations, which are more predictable, may be easier to conceptualize. When a referent is more likely to be mentioned (as are goals), it may be that the referent’s representation is activated more quickly at the message level. Alternatively, predictability may correlate with more unified representations between events at the conceptual level (cf. Gillespie, 2011). There was also a significant interaction between the goal/source and subject/non-subject manipulations, such that the goal benefit was stronger for non-subjects than for subjects.

3.3.2 Eyetracking measures

We analyzed fixations during the task to answer two questions. 1) Does our discourse production task reveal systematic fixations that reveal utterance planning processes? and 2) Do
our eyetracking measures reveal planning differences between our two participant groups? To do this, we looked at four different time windows for each trial (see Figure 7):  
• The **Preview** period, the first 700 ms (500 ms + 200 ms for saccade delay) when participants can see both pictures but the context sentence hasn’t begun,  
• The **Detective Sentence** period (+200 ms for saccade delay),  
• The **Latency** period, the time between offset of detective sentence and onset of the participant’s speech,  
• The **Response** period, when the participant utters their continuation sentence.

**Figure 7 here**

Because we were interested in planning effects, we focused our attention on the regions prior to the response. There were no significant effects during the Detective Sentence window, perhaps because it was long. Here we report analyses of the Preview Period, which reflects participants’ initial looks to the stimuli at trial onset, and the Latency period, which reflects the participants’ attention right before utterance articulation.

### 3.3.3 Evidence for Planning

The first question was whether our eyetracking measure would detect planning strategies overall. We found that it did, in both the Preview and the Latency windows.  
**Preview Window:** In the first 700 ms, all speakers looked to the character that they were planning to refer to, that is, fixating the Panel 1 Target more than the Panel 1 Non-Target (see Figure 8, Tables 4 and 5). Given that the target was the goal half the time, and the source half the time, we assessed this pattern in two models, which asked 1) Is the Panel 1 goal character fixated more when it is the target than not? And 2) Is the source character fixated more when it is the target than not? The models in Tables 4 and 5 confirm that yes, both goal and source characters were fixated more in the Preview period when they were the target than when not. Although we also found group differences (see below for discussion), the Context Group predictor did not interact with the target predictor. This implies that all speakers were planning which character they will refer to in their upcoming utterance in this very early period. Speakers were able to easily identify the Target character, either from seeing the entire story beforehand or their peripheral vision to Panel 2. In a separate model, we compared goal Targets with source Targets, and found that there was no difference in the amount of time spent looking at each (p =0.6617).
Latency Window: Both groups tended to look at the Panel 2 Target character right before their response (see Figure 9). To assess this pattern, we compared looks to the Target during the detective sentence with looks during the Latency period for each trial. In both groups, Panel 2 Target looks were greater in the latency period than the detective sentence period (Table 6). This supports evidence that speakers tend to look at objects that they are about to refer to, especially right before speaking when visual attention and utterance planning are tightly coordinated (Griffin & Bock, 2000). There were also more Panel 2 Target looks when there was a pronoun or zero in the utterance, a main effect of log latency (which indexes the size of the latency window), and a main effect of Context Group, but it did not interact with Window of analysis.

3.3.4 Group Differences

The second question was whether our eyetracking measures would reveal differences in utterance planning across our Context Use groups. We tested attention to each panel (Context vs. Target) in the Preview and Latency windows, and found differences in both cases. Preview Window: The Context-Users looked more at the characters in the Context panel than the Context-Ignorers, who spent relatively more time looking at the Target panel in this period. This shows up in two analyses. First, the Panel 1 Target character was fixated more by the Context-Users than the Context-Ignorers (see Figure 10A, Table 7). There was also a main effect of subject continuation, and an interaction between subject continuation and Context group. Figure 10A suggests that Context-Users directed more attention to non-subject Targets than subject Targets, but the Context-Ignorers were not influenced by grammatical role. Conversely, the Panel 2 Target character was fixated more by the Context-Ignorers than the Context-Users (see Figure 10B, Table 8). Together, these results suggest that compared to Context-Ignorers, the Context-Users were paying more attention to the characters in the first Context panel, especially the one they would soon talk about (see also Figure 7).
Latency Window: The Context-Users looked significantly more at the Panel 2 Target character during the Latency period than the Context-Ignorers did. There was also a main effect of the log latency predictor, which represents the size of the latency window: participants looked at the Target for a greater percentage of the window when the window was longer (see Figure 11, Table 9). This finding is not due to differences in disfluency or connector use across groups: when we excluded trials with disfluencies or connectors, the results remained the same.

Given that Context-Ignorers tended to have longer latencies than the Context-Users, the fact that they spent less time looking at the target immediately prior to their response is striking. This suggests that Context-Users more quickly shift attention towards the panel 2 Target as they planned the response sentence, after having ignored it during the preview period. This suggests that Context-Users employed a more systematically incremental planning strategy, while the Context-Ignorers may comprise a heterogeneous set of planning behaviors.

Response Window: The pattern observed during the Latency window is also visually (but not statistically) present in the Response window: Context-Users looked at the Panel 2 Target more than the Context-Ignorers throughout their Response periods.

4 Discussion

This study is the first to show that we can understand individual differences in sensitivity to the discourse context in terms of utterance planning and reference form choice. It focused on a type of variation that has received little attention, namely overall variation in the use of linguistic cohesion markers. Our participants fell into two groups: those who produced utterances that were connected with the discourse context, and those who didn’t. The Context-Users used more reduced forms and more explicit connector words than the group of participants of Context-Ignorers throughout this task. Reduced references and connector use correlated, as in other studies (Arnold & Griffin, 2007; Arnold & Nozari, under review).

Eyetracking measures revealed that we can use this paradigm to examine planning in a discourse context, in that we observed systematic examination of the context and target pictures overall. In addition, eyetracking measures suggested that our groups used different planning strategies. The Context-Users looked more at the Context picture during the Preview period than the Context-Ignorers, suggesting that they were doing more initial processing of the discourse context. They also initiated utterances more quickly, and looked more at the Target picture during their Latency periods than the Context-Ignorers, revealing more systematic planning of their
reference to this character. Additional work is needed to assess whether context-use strategy guides the planning strategies recruited for this task, or if participants adopted a planning strategy for this task that resulted in the observed linguistic choices.

Our findings suggest that our groups differed in their strategy of planning responses. The Context-Users paid more attention to the context, but also were likely engaging in greater pre-planning of the response. This simultaneous processing of context and response promotes a mental representation of the discourse in which the events are connected, increasing the speaker’s propensity to mark this connection linguistically. A similar finding is reported by Arnold & Nozari (under review), who found that both pronoun/zero use and connector production were more likely when the utterance was planned simultaneously with the articulation of the previous utterance.

Additional work is also needed to link variation in planning strategy with other measures of individual differences, such as language background or cognitive differences. We speculate that individuals responded differently to the competing goals of the task. One goal was to describe the Target picture, while a second was to put the description within a discourse story context. The importance of the story may have been weakened by the fact that their “interlocutor” was a voice in a computer. In a social setting, there may be social motivation to signal discourse coherence more explicitly. Nevertheless, we know that some language deficits, like autism, result in a deficit in the appropriate use of language in context (Tager-Flusberg, 1999; Kjelgaard & Tager-Flusberg, 2001). This study provides a first step toward characterizing individual differences in terms of utterance planning.

Acknowledgements

This work was funded by the National Science Foundation under Grant 1348549 to Jennifer E. Arnold. All procedures were performed in compliance with relevant laws and institutional guidelines, and the University of North Carolina at Chapel Hill Institutional Review Board has approved them. We thank many colleagues for helpful discussions reflected in this paper, and two anonymous reviewers for their feedback on our manuscript. We would also like to thank Ana Medina Fetterman, Liz Reeder, Samuel Adam Smith, Jacob Pascual, Jenna Roller, Brianna Torres, and Kristen Bubak for their help collecting and coding the data.
References


Appendix

Table 1. Critical predictors of pronoun/zero use within the Context-Users group*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.7027</td>
<td>0.3913</td>
<td>12.9</td>
<td>-4.35</td>
<td>0.0008</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>0.6702</td>
<td>0.3753</td>
<td>23.03</td>
<td>1.79</td>
<td>0.0874</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>2.5225</td>
<td>0.4151</td>
<td>285</td>
<td>6.08</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

* SE = standard error. Random intercepts of participant and of item. There was no significant interaction between Goal continuation and Subject continuation in a model with this included.

Table 2. Critical predictors of connector use*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.9875</td>
<td>0.9017</td>
<td>25.7</td>
<td>-2.2</td>
<td>0.0367</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.2106</td>
<td>0.4486</td>
<td>21.62</td>
<td>-0.47</td>
<td>0.6434</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>0.3112</td>
<td>0.4458</td>
<td>823</td>
<td>0.7</td>
<td>0.4853</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>2.3676</td>
<td>0.6003</td>
<td>823</td>
<td>3.94</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>-3.9097</td>
<td>1.3629</td>
<td>44.15</td>
<td>-2.87</td>
<td>0.0063</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slope of Context Group by item.

Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).
Table 3. Critical predictors of log latency*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.1339</td>
<td>0.03011</td>
<td>0</td>
<td>104.07</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.07194</td>
<td>0.02584</td>
<td>822</td>
<td>-2.78</td>
<td>0.0055</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>0.01684</td>
<td>0.01096</td>
<td>822</td>
<td>1.54</td>
<td>0.1249</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>-0.0179</td>
<td>0.02189</td>
<td>822</td>
<td>-0.82</td>
<td>0.4138</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>0.07891</td>
<td>0.03454</td>
<td>822</td>
<td>2.28</td>
<td>0.0226</td>
</tr>
<tr>
<td>Goal continuation *Subject continuation</td>
<td>0.04388</td>
<td>0.0214</td>
<td>822</td>
<td>2.05</td>
<td>0.0407</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slope of Goal continuation by participant, and random slopes of Context Group and Subject continuation by item. Interactions were not significant when added to the model (Context Group x Goal continuation; Context Group x Subject continuation).

Table 4. Critical predictors of looks to Panel 1 Goal character in Preview window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5327</td>
<td>0.05886</td>
<td>0</td>
<td>9.05</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>0.2307</td>
<td>0.1005</td>
<td>822</td>
<td>2.3</td>
<td>0.0219</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>-0.02231</td>
<td>0.02884</td>
<td>822</td>
<td>-0.77</td>
<td>0.4393</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>-0.01556</td>
<td>0.06069</td>
<td>822</td>
<td>-0.26</td>
<td>0.7977</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>-0.1383</td>
<td>0.05111</td>
<td>822</td>
<td>-2.71</td>
<td>0.007</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slopes of Goal continuation and Context Group by participant. Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).
Table 5. Critical predictors of looks to Panel 1 Source character in Preview window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5206</td>
<td>0.06081</td>
<td>0</td>
<td>8.56</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.2045</td>
<td>0.0921</td>
<td>823</td>
<td>-2.22</td>
<td>0.0266</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>-0.01489</td>
<td>0.05128</td>
<td>823</td>
<td>-0.29</td>
<td>0.7717</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>0.1017</td>
<td>0.1002</td>
<td>823</td>
<td>1.01</td>
<td>0.3106</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>-0.1061</td>
<td>0.06373</td>
<td>823</td>
<td>-1.66</td>
<td>0.0964</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slopes of Subject continuation and Context Group by participant, and random slopes of Goal continuation and Context Group by item. Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).

Table 6. Critical predictors of looks to Panel 2 Target character in comparing the Latency window with the Detective Sentence window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.1054</td>
<td>0.83</td>
<td>0</td>
<td>-2.54</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>0.05744</td>
<td>0.1067</td>
<td>1649</td>
<td>0.54</td>
<td>0.5904</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>-0.1621</td>
<td>0.08896</td>
<td>1649</td>
<td>-1.82</td>
<td>0.0686</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>0.5134</td>
<td>0.2286</td>
<td>1649</td>
<td>2.25</td>
<td>0.0248</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>0.2368</td>
<td>0.1175</td>
<td>1649</td>
<td>2.02</td>
<td>0.0439</td>
</tr>
<tr>
<td>Log Latency (ms)</td>
<td>0.5919</td>
<td>0.2633</td>
<td>1649</td>
<td>2.25</td>
<td>0.0247</td>
</tr>
<tr>
<td>Window: Latency Period vs. Detective Sentence Period</td>
<td>0.4701</td>
<td>0.1365</td>
<td>1649</td>
<td>3.45</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slope of Window by participant, and random slopes of Window and Reference Form by item. Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).
Table 7. Critical predictors of looks to Panel 1 Target character in the Preview window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6616</td>
<td>0.05913</td>
<td>0</td>
<td>11.19</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.01401</td>
<td>0.1028</td>
<td>821</td>
<td>-0.14</td>
<td>0.8916</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>-0.1698</td>
<td>0.07792</td>
<td>821</td>
<td>-2.18</td>
<td>0.0296</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>0.04987</td>
<td>0.08115</td>
<td>821</td>
<td>0.61</td>
<td>0.539</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>-0.1694</td>
<td>0.057</td>
<td>821</td>
<td>-2.97</td>
<td>0.0031</td>
</tr>
<tr>
<td>Context Group*Subject continuation</td>
<td>0.1977</td>
<td>0.0884</td>
<td>821</td>
<td>2.24</td>
<td>0.0256</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slopes of Goal continuation and Context Group by both participant and item. Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation).

Table 8. Critical predictors of looks to Panel 2 Target character in the Preview window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.1407</td>
<td>0.02651</td>
<td>0</td>
<td>-5.31</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.03605</td>
<td>0.03497</td>
<td>823</td>
<td>-1.03</td>
<td>0.3029</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>-0.00654</td>
<td>0.03446</td>
<td>823</td>
<td>-0.19</td>
<td>0.8496</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>0.01258</td>
<td>0.05976</td>
<td>823</td>
<td>0.21</td>
<td>0.8333</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>0.195</td>
<td>0.05483</td>
<td>823</td>
<td>3.56</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slopes of Subject continuation, Context Group, and Reference Form by participant, and random slop of Context Group by item. Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).
Table 9. Critical predictors of looks to Panel 2 Target character in the Latency window*.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3922</td>
<td>0.1276</td>
<td>0</td>
<td>3.07</td>
<td>.</td>
</tr>
<tr>
<td>Goal vs. Source continuation</td>
<td>-0.04515</td>
<td>0.03939</td>
<td>822</td>
<td>-1.15</td>
<td>0.252</td>
</tr>
<tr>
<td>Subject vs. Non-Subject continuation</td>
<td>0.01274</td>
<td>0.01325</td>
<td>822</td>
<td>0.96</td>
<td>0.3366</td>
</tr>
<tr>
<td>Pronoun/Zero vs. Description</td>
<td>-0.02498</td>
<td>0.02702</td>
<td>822</td>
<td>-0.92</td>
<td>0.3554</td>
</tr>
<tr>
<td>Context Ignore vs. Use Group</td>
<td>-0.0376</td>
<td>0.01616</td>
<td>822</td>
<td>-2.33</td>
<td>0.0202</td>
</tr>
<tr>
<td>Log Latency (ms)</td>
<td>0.1498</td>
<td>0.04012</td>
<td>822</td>
<td>3.73</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

* Random intercepts of participant and of item, random slope of Context Group by item.

Interactions were not significant when added to the model (Goal continuation x Subject continuation; Context Group x Goal continuation; Context Group x Subject continuation).
Figure 1

Figure 1. Context and Target event illustration panels.
Figure 2. Variability in the rate of pronoun/zero use across participants in 3 experiments.
Figure 3

Figure 3. Example areas of interest around three characters for one trial: Panel 1 Non-Target, Panel 1 Target, and Panel 2 Target.
Figure 4. Within the Context-Users, percentage of items by each condition type where participants used a Pronoun/Zero to refer to the Target character.
Figure 5. Percentage of items by each group of participants where connectors are used.
Figure 6

Figure 6. Average latency to begin speaking fluently by each group of participants.
Figure 7. Four time windows of a trial, averaged over all items, split by two Context Use groups. On the x-axis is time, starting at the beginning of the trial for the Preview/Detective Sentence graphs, and centered around the onset of participants’ response in the Latency/Response graphs. Grayed out boxes in Latency/Response graphs show average length of the detective sentence that is still playing, and Latency window is average by group.
Figure 8. All speakers initially look at the Panel 1 Target in the Preview window, the character who they will have to mention in their upcoming utterance. This is evidence for planning done by all speakers.
Figure 9

Figure 9. All speakers look to the Panel 2 Target more than the Panel 1 characters in the seconds before and after speaking.
Figure 10. A) Context-Users look more to the Panel 1 characters in the Preview period. B) Context-Ignorers look more to the Panel 2 Target character in the Preview period.
Figure 11

Figure 11. Context-Users look at the Panel 2 Target in the Latency period (0 ms on x-axis is when response begins) more than the Context-Ignorers do (note: y-axis scale starts at 0.3).