

Predicting turn-ends in discourse context

Ruth E. Corps, Martin J. Pickering & Chiara Gambi

To cite this article: Ruth E. Corps, Martin J. Pickering & Chiara Gambi (2019) Predicting turn-ends in discourse context, *Language, Cognition and Neuroscience*, 34:5, 615-627, DOI: [10.1080/23273798.2018.1552008](https://doi.org/10.1080/23273798.2018.1552008)

To link to this article: <https://doi.org/10.1080/23273798.2018.1552008>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 03 Dec 2018.



[Submit your article to this journal](#)



Article views: 170



[View Crossmark data](#)

Predicting turn-ends in discourse context

Ruth E. Corps^a, Martin J. Pickering^a and Chiara Gambi^{a,b}

^aDepartment of Psychology, University of Edinburgh, Edinburgh, UK; ^bSchool of Psychology, Cardiff University, Cardiff, UK

ABSTRACT

Research suggests that during conversation, interlocutors coordinate their utterances by predicting the speaker's forthcoming utterance and its end. In two experiments, we used a button-pressing task, in which participants pressed a button when they thought a speaker reached the end of their utterance, to investigate what role the wider discourse plays in turn-end prediction. Participants heard two-utterance sequences, in which the content of the second utterance was or was not constrained by the content of the first. In both experiments, participants responded earlier, but not more precisely, when the first utterance was constraining rather than unconstraining. Response times and precision were unaffected by whether they listened to dialogues or monologues (Experiment 1) and by whether they read the first utterance out loud or silently (Experiment 2), providing no indication that activation of production mechanisms facilitates prediction. We suggest that content predictions aid comprehension but not turn-end prediction.

ARTICLE HISTORY

Received 11 May 2018
Accepted 15 November 2018

KEYWORDS

Prediction; turn-taking;
dialogue; turn-end
prediction; prediction-by-
production

Introduction


During conversation, interlocutors' contributions are so finely coordinated that there is often little gap between their utterances (around 200 ms on average; Stivers et al., 2009). Most theories agree that listeners achieve such timing by predicting the content of the speaker's utterance (i.e. what the speaker is going to say; e.g. Garrod & Pickering, 2015; Levinson & Torreira, 2015). However, so far turn-taking research has largely focused on predictions based on the content of a single utterance, and so we do not know whether or how the context of wider discourse influences the timing of turn-taking. To address this question, we conducted two experiments, in which participants listened to two-utterance sequences: Participants listened to the first utterance, which provided the wider discourse context, and then indicated when they thought the speaker of the second utterance would reach the end. Crucially, first utterances were either constraining or unconstraining with regard to the content of the second utterance, so we could test whether a constraining discourse context leads to better-timed responses.

In addition, comprehenders may use language production mechanisms to generate content predictions (Pickering & Gambi, 2018; Pickering & Garrod, 2013). To

test whether predictions during turn-taking are supported by activation of the production system, participants in Experiment 1 listened to two-utterance sequences that either mimicked monologue (i.e. both utterances were produced by the same speaker) or dialogue (i.e. the utterances were produced by different speakers), while those in Experiment 2 either produced or heard someone else produce the first utterance before indicating the end of the second. Thus, participants were never required to engage their production system to generate a response, but we manipulated the activation of the production system either indirectly (i.e. by exposing participants to a dialogic context) or directly (i.e. by having them read the first utterance out loud) to test whether this would lead to a larger effect of context constraint on the timing of participants' responses.

There is substantial evidence that listeners make content predictions during comprehension (see Huettig, 2015, Pickering & Gambi, 2018, for reviews) and research has begun to consider its role in turn-taking (see Corps, Gambi, & Pickering, 2018, for a review). Some studies investigating the role of content predictability during turn-taking have used a button-press paradigm, in which participants listen to single utterances and press a button when they expect the

CONTACT Ruth E. Corps  rcorps@exseed.ed.ac.uk

 Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2018.1552008>.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

speaker to reach the end of their utterance. For example, De Ruiter, Mitterer, and Enfield (2006) found that although flattening the pitch of utterances had no effect on button-press times, responses were earlier (i.e. they occurred before the actual turn-end) when the words of the utterance were unintelligible. In a further study, Magyari, Bastiaansen, De Ruiter, and Levinson (2014) found that participants responded before the end of utterances whose final words were predictable in content (assessed using a gating paradigm; e.g. *I live in the same house with four women and another man*), but after the end of utterances in which the final words were unpredictable (e.g. *She was again alone in the north*). Concurrent electroencephalography (EEG) recordings showed a power decrease in the beta band, which has been associated with movement preparation (e.g. Alegre et al., 2006) and semantic and syntactic processing (e.g. Weiss & Mueller, 2012). This power decrease started around 1250 ms before the end of predictable but not unpredictable utterances. Together, these results support the idea that, in conversation, listeners use predictions of utterance content (along with other factors such as turn-final completion cues; e.g. Bögels & Levinson, 2017; Bögels & Torreira, 2015) to determine the speaker's turn-end and time response articulation.

But not all evidence suggests that listeners predict turn-endings by predicting utterance content. For example, Corps, Crossley, Gambi, and Pickering (2018) used a similar method to Magyari et al. (2014) and manipulated the content predictability of the final word(s) of the speaker's question, so that it was either predictable (e.g. *Are dogs your favourite animal?*) or unpredictable (e.g. *Would you like to go to the supermarket?*) given the preceding context. Unlike previous studies, we found no effects of content predictability on the timing of button-press responses, suggesting that participants may not have used this information to determine the speaker's turn-end. In contrast, content predictability affected the timing of participants verbal responses to the questions (i.e. yes or no), suggesting that content predictions may instead be helpful because they allow listeners to prepare earlier.

It is possible that the discrepancy in the results of studies investigating predictability effects on button press responses can be explained by other variables that may affect turn-end prediction. In particular,

previous research demonstrates that button-press times are strongly influenced by utterance duration: Shorter utterances tend to elicit later button-presses than longer utterances (e.g. Corps, Crossley, et al., 2018; De Ruiter et al., 2006). This effect may occur because the probability of the reaction stimulus (the turn-end) continuously increases (Sanders, 1966), and so the listener is more likely to respond earlier when the utterance is longer rather than shorter.

This durational effect may explain why button-press times in Corps, Crossley, et al.'s (2018) study were not affected by content predictability, even though Magyari et al. (2014) had previously demonstrated that participants responded earlier when utterances were predictable rather than unpredictable in content. Specifically, Corps, Crossley, et al. accounted for effects of duration in a post-hoc manner, by including it as a covariate in their analyses. Magyari et al. matched the average duration of their two predictability conditions so they did not significantly differ. But since their predictable utterances were still 410 ms longer than their unpredictable utterances on average, it is possible that participants responded earlier to the predictable compared to the unpredictable utterances simply because they were longer. In other words, it is not clear whether their predictability effect occurred independently of a duration effect.

Nevertheless, neither of these studies fully controlled for duration (e.g. by using the same utterances in all conditions, as in De Ruiter et al., 2006); such control is of course very difficult when the manipulation of interest is predictability, because it requires using different stimuli across conditions. In our studies, we instead used two-utterance sequences, in which participants listened to the first utterance and then indicated (by button-press) the end of the second. Second utterances were identical in the constraining (examples 1a and 1b in Table 1) and unconstraining (examples 2a and 2b) conditions, and we manipulated predictability while keeping utterance duration constant by varying the first utterance.

Importantly, research suggests content predictions can span multiple utterances, and it is therefore very likely that participants in our experiments will generate predictions on the basis of wider discourse. For example, Van Berkum, Brown, Zwisterlood, Koojiman,

Table 1. Example stimuli for each of the four stimuli conditions in Experiment 1.

Number	Constraint	Sequence type	Initial statement/question
1a	Constraining	Monologue	I listen to a lot of music
1b		Dialogue	What music do you listen to?
2a	Unconstraining	Monologue	There's one thing you need to know about me
2b		Dialogue	Is there anything specific I should know about you?

Note that the second utterance is the same across the four examples (*I really like Taylor Swift*).

and Hagoort (2005) created two sentence stories, in which the predictability of the speaker's final noun (e.g. *painting*) depended on the speaker's previous sentence (e.g. *The burglar had no trouble locating the family safe. Of course it was situated behind a ...*). Listeners displayed a positive deflection in event-related potential (ERP) waveforms when they encountered an adjective that did not agree in syntactic gender with the predictable noun. This effect disappeared when participants heard the same utterance in the absence of the speaker's previous sentence, suggesting they used the content of the first utterance to generate predictions about the content of the second.

Moreover, across-sentence predictability is likely to be particularly high in conversational dialogue. Dialogues are typically organised around predictable utterance sequences (such as question-answer pairs; e.g. Sacks, Schegloff, & Jefferson, 1974). In addition, interlocutors tend to align their representations and therefore repeat sentence structures and words previously used by their partner (e.g. Branigan, Pickering, & Cleland, 2000; Garrod & Anderson, 1987; see Pickering & Garrod, 2004). As a result, listeners may often be able to use the content of one utterance to predict the content of following utterances. Importantly, if across-utterance content predictability facilitates the timing of button-press responses, as in previous studies (e.g. Magyari et al., 2014), we expect participants to respond earlier when the content of the second utterance is constrained rather than unconstrained by the content of the first.

In the button-pressing task, the instructions typically encourage participants to time their response precisely, and so a few studies have analysed not only the timing of responses (i.e. how quickly participants responded) but also how precisely they respond (i.e. how close to the turn-end participants responded; Bögels & Torreira, 2015; Corps, Crossley, et al., 2018; De Ruiter et al., 2006). For example, Bögels and Torreira considered the proportion of button-presses occurring within a one second interval of the turn-end. De Ruiter et al. measured the entropy of response times, with low entropy indicating that button-presses clustered together and high entropy indicating they were distributed across a wider range. Finally, Corps et al. analysed the absolute difference between participants' response times and the actual turn-end and found that although response precision was not affected by content predictability, it was affected by utterance duration. In particular, responses to longer utterances tended to be less precise than responses to shorter utterances. We adopted the same measure of precision here, thus providing an additional opportunity to test whether predictability affects precision when duration is identical in the two predictability

conditions. If precision is affected by predictability, then we expect participants to respond closer to the turn-end when the second utterance is constrained by the first.

Using two-utterance sequences also allowed us to investigate whether participants use their production mechanisms to make turn-end predictions. Some theoretical accounts (e.g. Dell & Chang, 2014; Pickering & Garrod, 2013) suggest that the mechanisms that the listener uses to prepare a verbal response are the same as those used to predict what the speaker is going to say. Consistent with these accounts, Martin, Branzi, and Bar (2018) found that the N400 effect for articles in unexpected noun phrases (e.g. *a hat*) embedded in highly predictive contexts (e.g. *The king wore on his head ...*) was reduced when participants simultaneously produced the syllable /ta/ (thus preventing the use of inner speech) compared to when they tapped their tongue or listened to their own voice producing /ta/, thus suggesting that participants are worse at predicting when they are simultaneously using their language production system. In another study, Drake and Corley (2015) presented participants with high-Cloze sentence fragments that predicted a particular completion (e.g. *tap* after *When we want water, we just turn on the ...*). They found that when participants named a picture phonologically related to the predicted word (e.g. *cap*) after such sentence contexts, articulation diverged more from a control condition in which participants named the pictures without any sentence context, compared to when they named the predicted picture (e.g. *tap*) instead. Thus, predictions made during comprehension influenced later speech production, suggesting that prediction and production share a common mechanism.

Together, these studies demonstrate that production interferes with concurrent prediction, suggesting comprehenders can use their production system to generate predictions during comprehension. We investigated this issue in our experiments, but instead of asking whether using the production system interferes with concurrent prediction, we asked whether boosting activation of the production system beforehand facilitates subsequent turn-end prediction. To our knowledge, no study has looked at whether listeners use production-based mechanisms to generate turn-end predictions during dialogue.

In Experiment 1, we therefore presented participants with two utterances that were either produced by the same speaker (monologue condition) or by two different speakers (dialogue condition), under the assumption that overhearing dialogue activates the production system to a greater extent than overhearing monologue because comprehenders can also be speakers in dialogue but not in monologue (Gambi &

Pickering, 2016; Pickering & Garrod, 2013). In Experiment 2, we strengthened this manipulation by asking participants to either overtly produce or listen to a first utterance before predicting the end of a second utterance. If listeners use their production mechanisms to make turn-end predictions, then we expect an interaction between constraint and sequence type (Experiment 1) or participant role (Experiment 2). In particular, we expect stronger effects of constraint when participants listen to dialogue sequences or produce first utterances compared to when they listen to monologue sequences or listen to someone else produce first utterances, as their production system should be activated to a greater extent.

Experiment 1

Experiment 1 used a button-press paradigm with four conditions to investigate turn-end prediction. Participants listened to two utterances and were instructed to press a button when they expected the speaker to reach the end of their second utterance. We manipulated the degree to which the first utterance constrained the second, with the content of the second utterance being either constrained (and thus predictable) or unconstrained (and thus unpredictable) given the content of the first, to determine whether listeners can make content predictions across multiple utterances. Additionally, the two utterances were either produced by the same speaker (monologue conditions) or by two different speakers (dialogue conditions).

If across-utterance predictability influences the timing of button-press responses, then we expect participants to respond earlier when the content of the second utterance is predictable rather than unpredictable given the content of the first. If listeners use their production mechanisms to make turn-end predictions, and hence the dialogue conditions activate such mechanisms to a greater extent than the monologue conditions, we also expect an interaction between predictability and sequence type, such that the predictability effect will be stronger in the dialogue than the monologue conditions.

Method

Participants

Forty-eight native English speakers were recruited from the undergraduate student population at the University of Edinburgh. Participants had no known speaking, reading, or hearing impairments.

Design

Both constraint (constraining vs. unconstraining) and sequence type (monologue vs. dialogue) were manipulated within participants and items, and so there were four versions of each stimulus. We created four experimental lists (each containing 80 items) using a Latin Square procedure, so that all participants saw one version of each item and 20 items from each condition.

Materials

We constructed 80 two-utterance sequences, which consisted of either a question-answer sequence (dialogue conditions) or a statement-statement sequence (monologue conditions); see Appendix B in the Supplementary Material. First utterances were either constraining of the content of the second utterance, or unconstraining (see Table 1). There was a mistake with list creation, however, such that two items appeared twice in two of the lists (rather than once in each of the four). These items were used in the experiment, but excluded from all analyses. Thus, we report stimulus characteristics for the 78 items.

Stimuli were recorded by two female native speakers of English. One speaker recorded both utterances in the monologue condition and answers for the dialogue conditions, while the other recorded the questions. First utterances were between 829 and 2951 ms, and second utterances were between 928 and 3316 ms. Although we used the same second utterances in the constraining and unconstraining conditions, thus controlling for durational differences, second utterances were different in the monologue and dialogue conditions because 37 (47%) of the second utterances in the monologue condition had to be changed slightly so they were appropriate for the dialogue condition. For example, although *I can't watch anything that's*

Table 2. Means (*M*) and standard deviations (*SD*) for second utterance duration (ms; answers in the dialogue conditions; statements in the monologue conditions), Cloze and entropy values for stimuli in Experiment 1.

Constraint	Sequence type		Second utterance duration	Cloze ^a	Entropy ^b
Constraining	Monologue	<i>M</i>	1540	60%	1.29
		<i>SD</i>	385	35%	0.98
	Dialogue	<i>M</i>	1744	62%	1.37
		<i>SD</i>	435	34%	1.56
Unconstraining	Monologue	<i>M</i>	1540	27%	2.19
		<i>SD</i>	385	32%	1.02
	Dialogue	<i>M</i>	1744	28%	2.03
		<i>SD</i>	435	34%	0.97

^aCloze percentages of the final word of second utterances. If Cloze percentage is higher, then participants converged on a completion.

^bEntropy of second utterance fragments presented to participants in the Cloze task. If entropy is lower, then participants converged on a completion.

scary was an appropriate second utterance after *I didn't sleep well* (monologue condition), it was changed to *Not at all, I can't watch anything that's scary* so that it worked as answer to the question *Did you sleep well?* (dialogue condition). Using different second utterances in the monologue and dialogue conditions meant that they were longer in the monologue than dialogue condition ($p < .001$; see Table 2), and so duration may influence the effect of sequence type. We return to this issue in the Data Analysis section.

We assessed the constraint of stimuli using a Cloze post-test, in which 64 further native English speakers (16 per list) were visually presented with each adjacency pair with the final word missing from the second utterance (i.e. the final word *Swift* was missing from *I really like Taylor Swift*). Participants were instructed to "Complete the utterance using the word that you think is most likely to follow given the context of the first utterance".

The constraint of second utterances was assessed using Shannon entropy (i.e. $-\sum p_i \log_2(p_i)$, where p_i is the proportion of times each completion occurred for a given fragment; Shannon, 1948), which is low (a minimum of 0) when completions are the same across participants (i.e. content is predictable), and high (a maximum of 4 when each of the 16 participants for each list provided a different continuation) when completions are different. In addition, we used Cloze probability (Taylor, 1953) to calculate the percentage of participants who provided a particular continuation. Stimuli in the constraining conditions had significantly higher Cloze and lower entropy values (both $ps < .001$) than stimuli in the unconstraining conditions. In contrast, these values did not differ for the monologue and dialogue conditions and there were no interactions ($ps > .05$; see Table 2). There were also no differences in the Cloze or entropy values between the adjusted and original second utterances (all $ps > .18$), suggesting that such adjustments did not affect predictability.

Procedure

Stimulus presentation and data recording were controlled using E-Prime (version 2.0). To make stimulus onset salient, a fixation cross (+) appeared 500 ms before the onset of audio playback of the first utterance. The screen turned red 500 ms after the offset of the first utterance, and audio playback of the second utterance began simultaneously. Participants were told to

Press the button on the response box to indicate when the second statement/answer will end. Do not wait until the speaker has finished and there is silence. Instead, you should press the button as soon as you expect them to be finished talking.

These instructions were a translation of those used by De Ruiter et al. (2006). Thus, participants were encouraged to predict the turn-end, rather than simply wait for the speaker to reach the end of their utterance. Participants responded by pressing the middle button of a SR-box and audio playback stopped as soon as a response was recorded. The next trial began automatically after 1000 ms.

Participants completed two initial practice trials to familiarise themselves with the experimental procedure. The 80 experimental stimuli were individually randomised for each participant.

Data analysis

Response times were defined with respect to second utterance offset and were negative when participants responded before the end of the speaker's second utterance and positive when they responded after the end. Of the 3744 trials, we discarded one (0.03%) greater than 10,000 ms because it was a clear outlier. We replaced 110 (2.94%) responses falling at least 2.5 standard deviations above the by-participant mean and 173 (4.62%) responses below the by-participant mean with the respective cut-off value. We evaluated the effects of constraint and sequence type on response times with linear mixed effects models (LMM; Baayen, Davidson, & Bates, 2008) using the *lmer* function of the *lme4* package (version 1.1–12; Bates, Maechler, Bolker, & Walker, 2015) in RStudio (version 0.99.896) with a Gaussian link function.

Response precision was defined as the absolute value of response time (see Corps, Crossley, et al., 2018). Before taking the absolute value, we first standardised response time to have a mean of zero, so that we could assume a half-normal distribution or, equivalently (Leone, Nelson, & Nottingham, 1961), a normal distribution truncated at zero. As a result, the distributional assumptions of *lmer* were not met. Therefore, we used Bayesian mixed effects models (BMM) as implemented in the *brms* package (version 1.6.1; Bürkner, 2017). We initially fitted models using a normal distribution truncated at zero. However, such models did not converge and so we modelled our data using three other distribution families (log-normal, gamma, and Weibull). The Weibull was the best fitting model (assessed using LOO comparisons), and so we report parameters and credible intervals from this model. We ran 4 chains per model, each for 1600 iterations, with a burn-in period of 800, and initial parameter values set to zero. All of the reported models converged with no divergent transitions (all \hat{R} values ≤ 1.1); the number of effective samples for each

estimate is reported in Appendix A in the Supplementary Material.

Although the parameterisation of the Weibull distribution implemented in *brms* is based on a scale and shape parameters, we report and interpret only scale parameters (but full models are reported in Appendix A in the Supplementary Material). The shape parameter is most often used to model failure rates, which is not relevant to our analysis. On the other hand, the scale parameter quantifies the spread of the distribution and is thus informative of the degree of precision in participants' responses. Note that scale parameters were fitted on the log scale (reported in Appendix A in the Supplementary Material), but we report exponentiated estimates in the Results section as they are easier to interpret: The larger the exponentiated value of the scale parameter, the more spread out the probability mass of the distribution. All distributions were fitted using default *brms* priors.

In all analyses, we fitted models using the maximal random effects structure justified by our design (Barr, Levy, Scheepers, & Tily, 2013), but correlations among random effects were fixed to zero to aid model convergence. We fitted the full model where response times or precision were predicted by Constraint Condition (reference level: unconstraining vs. constraining), Sequence Type (reference level: dialogue vs. monologue), and their interaction. These predictors were contrast coded (−0.5, 0.5) and centred. Utterances in the constraining and unconstraining conditions had exactly the same duration, and thus duration cannot explain any effects of Constraint Condition. However, second utterances were longer in the monologue condition, and so we included Second Utterance Duration in our analyses. This predictor was centred, and was included only as a main effect.

For the LMM analyses, we report coefficient estimates (b), standard errors (SE), and t values for each predictor. We assume that an absolute t value of 1.96 or greater indicates significance at the 0.05 alpha level (Baayen et al., 2008). In addition, we computed the Bayes Factors (reported as BF in the Results section) for null effects in the LMM analysis by fitting BMM models using a normal distribution with 10,000 iterations. To get accurate estimates, we defined means and standard deviations for priors based on coefficients and standard errors reported in the LMM analysis (see Appendix A for full models), with the exception that means for null effects were set to zero. In all instances, we compared the full model to a model excluding the relevant (null) predictor(s). Following Dienes (2014), we interpret a Bayes factor (i) greater than 3 as strong evidence for the alternative hypothesis over the null, (ii) less than

0.33 as strong evidence for the null hypothesis over the alternative, and (iii) between 0.33 and 3 as weak evidence.

For the BMM analyses, we report coefficient estimates of effect size (b), estimate errors (SE), and the 95% credible interval (Crl; i.e. under the model assumptions, there is a 95% probability that the parameter estimate is contained in this interval) for each predictor. If zero lies outside the credible interval, then we conclude there is sufficient evidence to suggest the estimate is different from zero.

Results and discussion

Response times

On average, participants responded 35 ms before the end of the speaker's second utterance (see Figure 1 for a breakdown by condition) and 95% of responses occurred within 1000 ms of the speaker's turn-end (see Figure 2).

We found a significant effect of Constraint Condition: Participants responded earlier when second utterances were constrained by the context of the previous utterance than when they were not ($b = -43.66$, $SE = 11.83$, $t = -3.69$), suggesting that listeners can use across-utterance predictability to predict turn-endings. However, there was no effect of Sequence Type ($b = -13.12$, $SE = 16.34$, $t = -0.80$; $BF = 0.88$) and no interaction between Sequence Type and Constraint Condition ($b = -13.01$, $SE = 33.63$, $t = -0.39$; $BF = 0.64$). Note that although Figure 1 does show a difference between average response times in the dialogue and monologue conditions, these means are not adjusted for Second Utterance Duration, which was a negative predictor of response times ($b = -115.03$, $SE = 9.30$, $t = -12.36$).

These results may suggest that listeners did not use their production mechanisms to predict turn-endings, since we expected such mechanisms to be activated to a greater extent in the dialogue conditions (which may require a response) than the monologue conditions (which requires no response). However, it is possible we found no evidence for prediction-by-production because participants in this experiment did not actually activate their production system. Thus, Experiment 2 investigates this issue further by asking participants to produce questions (and they should thus be more likely to activate their production system) before predicting turn-endings.

Response precision

Unlike our analysis of response times, we found no evidence that Constraint Condition ($b = 1.01$, $SE = 1.04$, Crl [−0.07, 0.10]), Sequence Type ($b = 1.04$, $SE = 1.06$, Crl

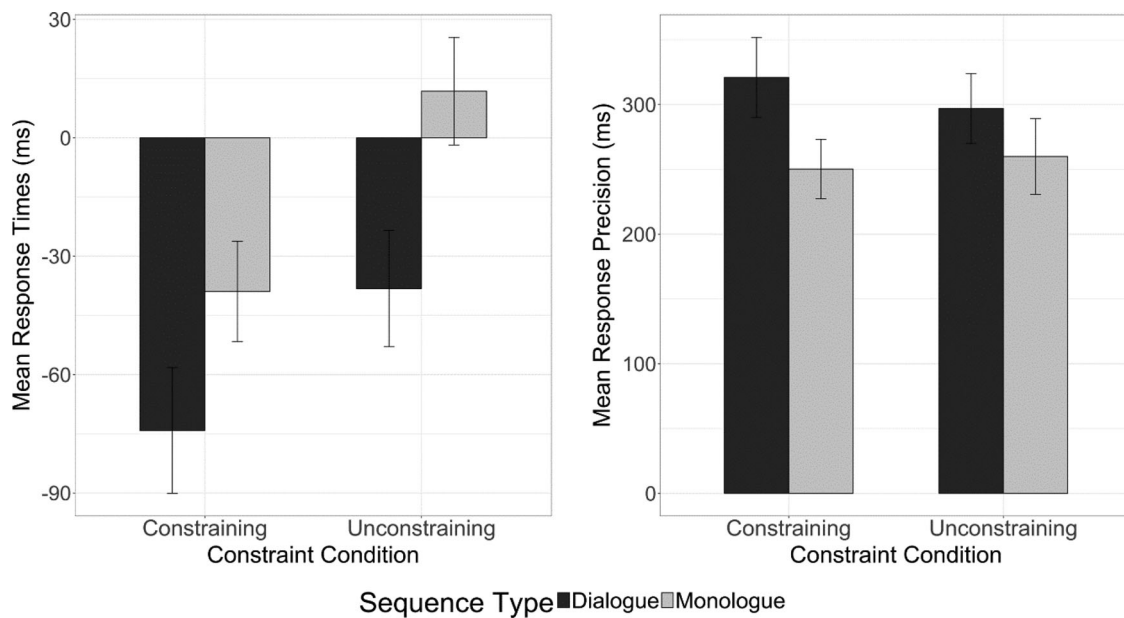


Figure 1. Observed means of response times (left) and precision (right) for the four conditions in Experiment 1. Error bars represent ± 1 standard error from the mean.

[$-0.16, 0.07$]), or their interaction ($b = 1.05, SE = 1.08, CrI [-0.21, 0.10]$) influenced response precision. However, much like previous research (Corps, Crossley, et al., 2018), we found that Second Utterance Duration had a positive effect on scale ($b = 1.22, SE = 1.03, CrI [0.15, 0.25]$), such that the spread of the distribution was greater when second utterance were longer, perhaps because longer utterances contain earlier potential completion points (cf. Bögels & Torreira, 2015). These results are consistent with previous research, which has found no effects of the content predictability of single utterances on response precision (Corps, Crossley, et al., 2018).

Experiment 2

Experiment 1 showed that listeners responded earlier in a button-pressing task when the content of the first utterance constrained predictions about the content of the second, but this constraint did not influence response precision (i.e. how closely participants responded to the turn-end). Since previous research has focused on the predictability of single utterances, we conducted Experiment 2 to investigate whether the effect of constraint on response timing replicated. To do so, we used the dialogue conditions from Experiment 1.

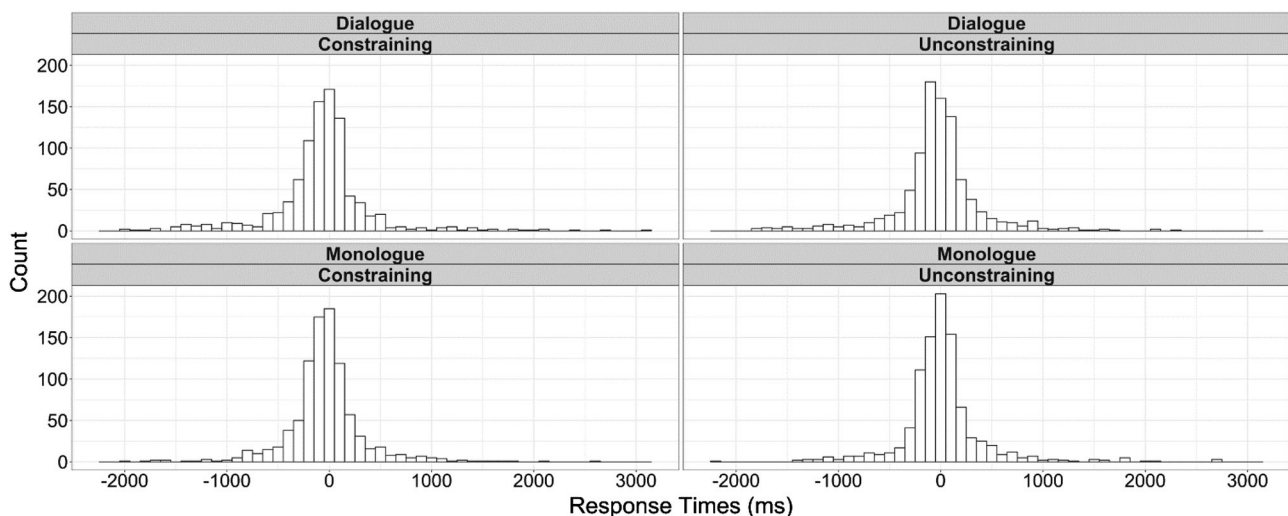


Figure 2. The distribution of response times in the four conditions in Experiment 1. Trials are placed into 100 ms time bins.

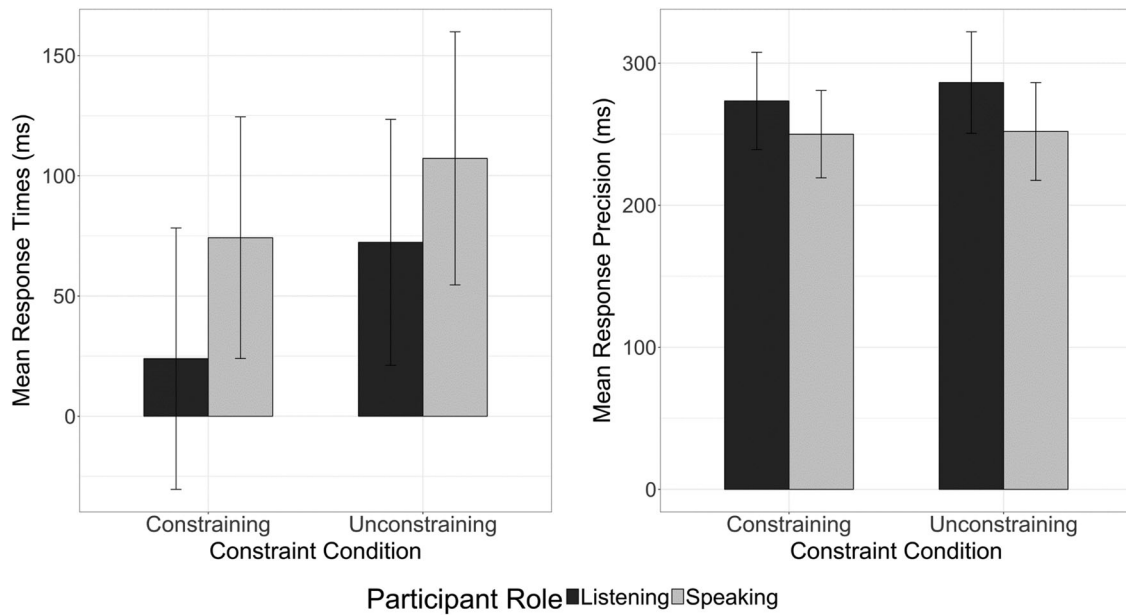


Figure 3. Observed means of response times (left) and precision (right) for the four conditions in Experiment 2. Error bars represent ± 1 standard error from the mean.

Experiment 1 also showed that listeners were no better at predicting turn-endings when utterances mimicked dialogue compared to monologue, suggesting they did not use their production system to predict turn-ends. However, it is possible that we found no evidence for prediction-by-production because participants did not actually activate their production system. Thus, we asked participants in Experiment 2 to either produce the question and then predict the end of a pre-recorded answer (speaking conditions) or listen to another speaker produce the question (listening conditions). If listeners use their production system to make turn-end predictions, we expect constraint effects to be larger when the production system has been recently activated (i.e. in the

speaking conditions) than when it has not (i.e. in the listening conditions).

Method

Participants

Forty-eight further native English speakers participated on the same terms as Experiment 1.

Materials and Design

We used the adjacency pairs from the dialogue conditions in Experiment 1 (see Table 1). The predictability of the initial question (constraining vs. unconstraining) was manipulated both within participants and items, so there were two versions of each stimulus. We used the same

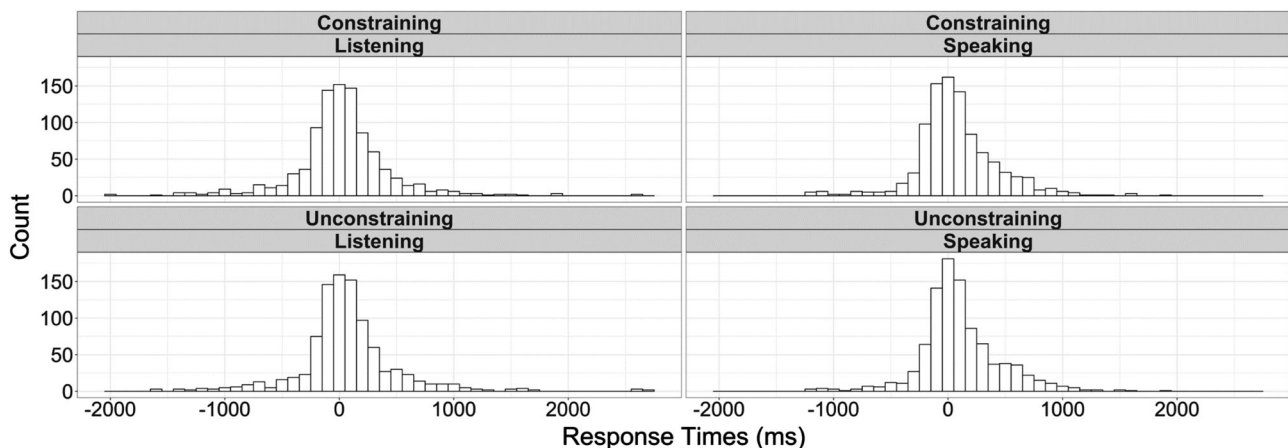


Figure 4. The distribution of response times in the four conditions in Experiment 2. Trials are placed into 100 ms time bins.

stimulus lists as Experiment 1. Participant role (speaking vs. listening) was manipulated within items but between participants, so that participants either produced the initial question or listened to a pre-recorded speaker.

Procedure

The procedure was identical to that used in Experiment 1, with the exception that the question visually appeared on-screen after the presentation of the fixation cross. Participants in the speaking condition were instructed to read the question aloud, and participants in the listening condition were instructed to listen to the speaker. Since previous research suggests that there is a lag of between 500–600 ms between reading and speaking (e.g. Inhoff, Solomon, Radach, & Seymour, 2011; Laubrock & Kliegel, 2015), we assumed that there would be a delay between the text appearing on-screen and the moment when participants in the speaking condition began producing the question. To ensure this delay was comparable across the two role conditions, the question in the listening condition began 600 ms after the text appeared on-screen. In both conditions, participants pressed the middle button on the SR-box either when they had finished producing the question (speaking conditions) or when the pre-recorded speaker had finished producing the question (listening conditions). The screen turned red 500 ms after this button-press, and answer playback began simultaneously. After this moment, the rest of the procedure was identical to Experiment 1, and participants pressed a button on the response box when they expected the speaker to reach the end of their second utterance.

Data analysis

Response times and precision were calculated using the same procedure as Experiment 1. We discarded four trials (0.11%) greater than 10,000 ms and replaced 65 (1.7%) responses at the upper limit and 70 (1.9%) at the lower limit. We fitted models using the same procedure as Experiment 1, but Sequence Type was replaced by Participant Role (reference level: listening vs. speaking). Since the latter predictor was between-participants, we included random slopes by items only.

Results and discussion

Response times

On average, participants responded 70 ms after the end of the speaker's utterance (see Figure 3 for a breakdown by condition). Note that this average is slower than Experiment 1, in which participants responded 33 ms before the end of the speaker's second utterance. It is

possible this discrepancy occurred because participants in Experiment 2 were instructed to press the button after the first utterance to begin playback of the second. Thus, participants in Experiment 2 may have adopted a different strategy to those in Experiment 1. Much like Experiment 1, the great majority of responses (98%) occurred within 1000 ms of answer end (see Figure 4).

Consistent with Experiment 1, participants responded earlier when questions were constraining rather than unconstraining ($b = -40.86$, $SE = 9.43$, $t = -4.33$). Figure 3 shows that participants in the listening condition were slightly faster than those in the speaking condition. This difference in response times could be attributed to differences in cognitive load in the two conditions: Participants in the speaking condition had to switch between two overt tasks (speaking and button-pressing) and speaking is generally more cognitively demanding than listening (e.g. Cook & Meyer, 2008). Nevertheless, there was no effect of Participant Role ($b = 42.61$, $SE = 71.81$, $t = 0.59$, $BF = 0.96$), and no interaction between Participant Role and Constraint Condition ($b = 15.06$, $SE = 18.86$, $t = 0.80$, $BF = 0.80$). Thus, producing the question prior to predicting the turn-end did not influence prediction. As in Experiment 1, Second Utterance Duration was again a significant negative predictor of response times, such that longer second utterances elicited earlier responses than shorter second utterances ($b = -72.00$, $SE = 11.37$, $t = -6.33$).

Response precision

As in Experiment 1, we found no significant effects of Constraint Condition ($b = 1.01$, $SE = 1.04$, $CrI [-0.06, 0.09]$), Participant Role ($b = 1.05$, $SE = 1.21$, $CrI [-0.42, 0.31]$), and no interaction between the two ($b = 1.01$, $SE = 1.08$, $CrI [-0.15, 0.16]$). However, the spread of the distribution was greater when second utterances were longer ($b = 1.11$, $SE = 1.03$, $CrI [0.04, 0.16]$).

General Discussion

In two experiments, we investigated whether listeners can use across-utterance predictability to predict turn-endings during dialogue. In Experiment 1, we manipulated the predictability of two utterance sequences, so that the content of the second utterance (e.g. *I really like Taylor Swift*) was either constrained (e.g., *What music do you listen to?*) or unconstrained (e.g. *Is there anything specific I should know about you?*) by the content of the first. We found that listeners responded earlier, but not more precisely (i.e. closer to the speaker's turn-end), when second utterances were constrained rather than unconstrained. This effect occurred regardless of

whether sequences mimicked monologue (i.e. both utterances were produced by the same speaker) or dialogue (i.e. the utterances were produced by different speakers), suggesting that listeners were not more likely to activate their production system, and thus predict more, when listening to dialogue than monologue. Experiment 2 replicated the predictability effect, but showed that response times and precision were not affected by whether listeners had recently engaged their production mechanisms (by producing first utterances) or not (by listening to first utterances), suggesting that engaging the production system directly also failed to elicit more prediction.

The effect of constraint on response timing is consistent with previous research, which has found that listeners press a button to indicate a turn-end earlier when the final word of a single utterance is predictable rather than unpredictable on the basis of the sentence context (e.g. Magyari et al., 2014; Magyari & De Ruiter, 2012). But this result is inconsistent with our previous research (Corps, Crossley, et al., 2018), which found no effects of the predictability of single utterances on button-pressing times. It is possible that effects of content predictability were not detected in that study because they were masked by large differences in utterance duration. In contrast, differences in duration cannot explain the effects in the present study because second utterances were identical in the constraining and unconstraining conditions. Our findings thus suggest that listeners can generate turn-end predictions on the basis of discourse (e.g. Van Berkum et al., 2005), and independently of utterance duration. Such across-utterance predictability is of course particularly important during conversational turn-taking, since the content of one speaker's utterance is likely related to the content of the previous speaker's utterance (e.g. adjacency pairs; Sacks et al., 1974).

However, it is unclear whether the current findings actually demonstrate that participants were predicting the turn-end. Although previous research has typically interpreted effects of utterance predictability on response timing as demonstrating that people use this information to predict turn-ends (e.g. Magyari, De Ruiter, & Levinson, 2017), elsewhere we have argued that precision is a better measure of turn-end prediction compared to response timing (see Corps, Crossley, et al., 2018). After all, participants in the button-pressing task are encouraged to respond when they think the speaker will reach the end of their utterance (i.e. precisely at that moment). An earlier (negative) response (i.e. before the end of the utterance) in the button-pressing task would mean that listeners expected an earlier turn-end than actually occurred. Conversely, a later response would mean that listeners expected a later

turn-end. It is thus unclear in these instances whether earlier responses are actually preferable.

Importantly, we replicated our earlier findings (Corps, Crossley, et al., 2018) and found no effects of constraint on response precision, suggesting that content predictability does not help listeners predict turn-endings more precisely. But if this is the case, then why did participants respond faster to more predictable second utterances? One possibility is that constraint affects response timing, but not precision, because the first utterance speeds up listeners' understanding of the second utterance. When the content of the first utterance constrains the content of the second, listeners can use their prediction to determine whether the second utterance satisfies the semantic expectations set up by the first utterance before the end of the utterance (e.g. is the second utterance about food when the speaker has previously spoken about food?). In other words, the processing system runs in a top-down "verification mode" and the utterance does not need to be processed extensively (e.g. Rommers & Federmeier, 2018; Van Berkum, 2010). When the second utterance is unconstrained, however, listeners do not have any specific predictions about the content of the utterance, and so must allocate more resources to processing the utterance. In a similar vein, the timing of button-press responses might be sensitive to the detection of completion points: listeners may respond earlier when the utterance is constrained rather than unconstrained because in the former case they are able to determine whether the utterance is pragmatically and/or semantically complete earlier in time (e.g. Bögels & Torreira, 2015).

This account could also explain effects of predictability on response times in previous experiments using single utterances (Magyari et al., 2014; Magyari & De Ruiter, 2012): Listeners may respond earlier when the final words of single utterances are predictable rather than unpredictable because they can more easily identify that the final word(s) satisfies the predictions based on the (same-utterance) context. However, neither Magyari and De Ruiter (2012) nor Magyari et al. (2014) included duration as a control variable in their analyses, and in our previous study (Corps, Crossley, et al., 2018), in which we fully controlled for duration, we did not find evidence to suggest that predictability influences response timing. Thus, it remains unclear whether predictability influences the timing of button-press responses to single utterances presented in isolation. Since duration is a strong predictor in the button-press task, it is likely that future studies wishing to separate effects of predictability from duration will need to follow the same procedure as our studies and ensure that the duration of utterances in the two conditions are identical.

Although some studies have adopted this method by using identical utterances across conditions, they have not investigated the role of content predictability. For example, De Ruiter et al. (2006) found that button-presses occurred further from the turn-end (too early) when the words of an utterance were unintelligible, suggesting that the lexical content of a speaker's utterance is important for prediction, but they did not assess the predictability of these stimuli. Similarly, Riest, Jorschick, and De Ruiter (2015) found that participants responded earlier when the order of words in an utterance was scrambled compared to when it was not. They suggested that scrambling word order prevented participants from using the preceding words of the speaker's utterance to predict subsequent words. However, it is equally possible that scrambling made integration of words into the preceding discourse very hard. Thus, although these studies clearly show that being able to interpret the content of an utterance is important for turn-end prediction, they do not specifically show that the predictability of that content is important.

In sum, our results suggest that listeners may not use content predictions to determine the speaker's turn-end. If confirmed, this proposal would be inconsistent with theories of conversational turn-taking that suggest turn-end prediction plays a central role in timing response articulation (e.g. the late-planning hypothesis; see Bögels & Levinson, 2017). But if listeners do not use content prediction to predict turn-ends, do they instead use an alternative strategy to ensure responses are articulated at the appropriate moment (i.e. without overlap nor long gaps)? One possibility is that listeners launch articulation of their response reactively, after encountering one or more turn-final cues (e.g. a falling boundary tone). This strategy may lead to short inter-turn intervals if listeners prepare their response in advance of the turn-end, because launching articulation does not take as long as preparing a response from scratch (the articulatory component of single-word production takes around 145 ms; Indefrey & Levelt, 2004). Note that listeners are likely sensitive to multiple cues (e.g. Bögels & Torreira, 2015) and could use them together to determine points of possible utterance completion.

Before concluding, we briefly discuss the lack of evidence that listeners used production mechanisms to make predictions in our task: Response times and precision were unaffected by whether participants listened to sequences produced during monologue or dialogue (Experiment 1), or whether or not they produced first utterances before predicting the end of the second (Experiment 2). These results appear inconsistent with previous research showing that participants use their production mechanisms to make predictions during comprehension (e.g. Martin et al., 2018). However, there are a number of

possible explanations for this lack of effect. First, Martin et al. instructed participants to carry out a production task (syllable production) while they simultaneously listened to sentences, and so participants were simultaneously using their production system while predicting. In contrast, participants in our study used their production system before comprehending the second sentence and predicting its end, and the activation of production mechanisms may have decayed before the end of the second sentence. Second, participants in the speaking condition in Experiment 2 read sentences from the screen, meaning that they did not generate the message nor formulated the utterance themselves, and so they may not have activated the early stages of language production. Finally, it is possible that the production system was activated to the same extent in all conditions (even when participants were comprehending) because prediction is relevant to the button-pressing task in general.

In conclusion, we have shown that listeners in a button-pressing task can use the wider discourse of previous utterances to make predictions during conversational dialogue. In particular, participants responded earlier, but not more precisely, when utterances were constrained by the preceding sentence. These findings suggest that content predictions may not help listeners predict the turn-end more precisely, but instead aid comprehension. In addition, we did not find any evidence to suggest that prediction was enhanced by activation of the production system. These findings have important implications for understanding how interlocutors coordinate their contributions during dialogue.

Acknowledgements

Ruth Corps was supported by the Economic and Social Research Council [grant number ES/J500136/1]. Martin Pickering is supported by Leverhulme Research Project Grants [RPG-2017-239; RPG-2018-259]. Chiara Gambi was supported by a Leverhulme Research Project Grant [RPG-2014-523] to Martin Pickering and Hugh Rabagliati. We thank Holly Abercrombie, Fraser Caldwell, Alix Dobbie, Nicole Emslie, Stevie Goodrich, Charlotte Jones, Maurane Limeux, Laura McCabe, Luke Raben, David Sonski, Arved-Sebastian Stucke, and Laura Wilson for assisting with data collection.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Economic and Social Research Council [grant number ES/J500136/1]; Leverhulme Trust [grant number RPG-2014-523, RPG-2017-239, RPG-2018-259].

References

- Alegre, M., Imirizaldu, L., Valencia, M., Iriarte, J., Archocha, J., & Artieda, J. (2006). Alpha and beta changes in cortical oscillatory activity in a go/no go randomly-delayed-response choice reaction time paradigm. *Clinical Neurophysiology*, *117*, 16–25.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed-effects models using "Eigen" and S4* (R package version 1.1-12). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*, 71–89.
- Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–B25.
- Bürkner, P.-C. (2017). *Brms: Bayesian regression models using Stan* (R package version 1.6.1). Retrieved from <https://CRAN.R-project.org/package=brms>
- Cook, A. E., & Meyer, S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886–899.
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition*, *175*, 77–95.
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*, 230–240.
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B*, *369*. doi:10.1098/rstb.2012.0394
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*, 515–535.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00781
- Drake, E., & Corley, M. (2015). Effects in production of word pre-activation during listening: Are listener-generate predictions specified at a speech-sound level? *Memory & Cognition*, *43*, 111–120.
- Gambi, C., & Pickering, M. J. (2016). Predicting and imagining language. *Language, Cognition and Neuroscience*, *31*, 60–72.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181–218.
- Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00751
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
- Inhoff, A. W., Solomon, M., Radach, R., & Seymour, B. (2011). Temporal dynamics of the eye voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, *23*, 543–558.
- Laubrock, J., & Kliegel, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01432
- Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, *3*, 543–550.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00731
- Magyari, L., Bastiaansen, M. C. M., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*, 2530–2539.
- Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00376
- Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.00211
- Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*. doi:10.1038/s41598-018-19499-4
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*, 1002–1044.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–190.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 329–347.
- Riest, C., Jorschick, A. B., & De Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00089
- Rommers, J., & Federmeier, D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16–30.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, *50*, 696–735.
- Sanders, A. F. (1966). Expectancy: Application and measurement. *Acta Psychologica*, *25*, 293–313.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
- Stivers, T., Enfield, N. J., Brown, P., Engelert, C., Hayashi, M., Heinemann, T., ... Levinson, S. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*, 10587–10592.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.
- Van Berkum, J. J. (2010). The brain is a prediction machine that cares about good and bad – Any implications for neuropragmatics? *Italian Journal of Linguistics*, *22*, 181–208.

Van Berkum, J. J., Brown, C. M., Zwisterlood, P., Koojiman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs during reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.

Weiss, S., & Mueller, H. M. (2012). “Too many beta do not spoil the broth”: The role of beta brain oscillations in language processing. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00201