Routledge
Taylor & Francis Group

Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody

Sara Bögels & Francisco Torreira

To cite this article: Sara Bögels & Francisco Torreira (2021) Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody, Discourse Processes, 58:10, 903-924, DOI: 10.1080/0163853X.2021.1986664

To link to this article: https://doi.org/10.1080/0163853X.2021.1986664

DISCOURSE PROCESSES
2021, VOL. 58, NO. 10, 903–924
https://doi.org/10.1080/0163853X.2021.1986664

Routledge
Taylor & Francis Group

Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody

Sara Bögels[a,b,c] and Francisco Torreira[b,d]

[a]Department of Communication and Cognition, Tilburg University; [b]Language and Cognition Department, Max Planck Institute for Psycholinguistics; [c]Donders Institute for Brain, Cognition, and Behaviour, Radboud University; [d]Department of Linguistics, McGill University

ABSTRACT

This study investigated the role of contextual and prosodic information in turn-end estimation by means of a button-press task. We presented participants with turns extracted from a corpus of telephone calls visually (i.e., in transcribed form, word-by-word) and auditorily, and asked them to anticipate turn ends by pressing a button. The availability of the previous conversational context was generally helpful for turn-end estimation in short turns only, and more clearly so in the visual task than in the auditory task. To investigate the role of prosody, we examined whether participants in the auditory task pressed the button close to turn-medial points likely to constitute turn ends based on lexico-syntactic information alone. We observed that the vast majority of such button presses occurred in the presence of an intonational boundary rather than in its absence. These results are consistent with the view that prosodic cues in the proximity of turn ends play a relevant role in turn-end estimation.

Introduction

Turn-end estimation

In conversation, turn transitions between speakers are often smooth (Sacks et al., 1974): the most frequently observed turn transitions are about 100–300 ms long, regardless of the language (e.g., Heldner & Edlund, 2010; Stivers et al., 2009). Such short gaps are important not just for maintaining the flow of conversation, but also for precluding any unwanted inferences regarding potentially unexpected or unwanted responses (e.g., Bögels, Kendrick et al., 2015; Kendrick & Torreira, 2015). Given that language planning generally involves much longer time spans (e.g., 600 ms for single-word picture naming; Indefrey & Levelt, 2004), it seems puzzling that interlocutors often produce such short gaps when exchanging turns (Levinson & Torreira, 2015; Sacks et al., 1974). From a cognitive point of view, it can be posited that responding to a turn in conversation involves at least two complex processes, namely (1) planning the content of one's next turn (e.g., Barthel et al., 2016; Bögels, Magyari et al., 2015; Sjerps & Meyer, 2015) and (2) estimating the end of the ongoing turn so as to be able to produce one's turn in a timely manner. The present study focuses on the second of these processes.

Different proposals have been put forward regarding how turn-end estimation is achieved. Several observational studies have shown that turn ends correlate with the end of syntactic constituents and with certain prosodic phenomena, such as specific phrase-final intonation contours and the lengthening of phrase-final segments, syllables, and words (e.g., Ford & Thompson, 1996; Gravano & Hirschberg, 2011; Local & Walker, 2012; Rühlemann & Gries, 2020; Wells & Macfarlane, 1998).

CONTACT Sara Bögels s.bogels@tilburguniversity.edu Department of Communication and Cognition, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

These studies suggest that listeners could use syntactic and prosodic information as cues when determining possible turn ends. A series of experimental psycholinguistic studies, for the most part using a button-press paradigm (e.g., Bögels & Torreira, 2015; De Ruiter et al., 2006) have also investigated how listeners estimate the end of the current speaker's turn. These studies generally assume or conclude that lexico-syntax plays an important role in turn-end estimation (but see Corps et al., 2019), although they have not always yielded converging conclusions regarding the role of prosody. A second point of uncertainty arising from most of these studies stems from the fact that their experimental trials consisted of turns presented in isolation outside of conversational contexts. Because of this, we do not know whether and how previous results would hold under more naturalistic conditions where the conversational context is available to listeners. The present study investigates the roles of both prosody and conversational context in turn-end estimation by means of two button-press experiments.

Previous button-press studies on turn-end estimation

The role of lexico-syntactic information

The first online experimental study on turn-taking was performed by De Ruiter et al. (2006). They asked participants to listen to turns containing five words or more extracted from a corpus of conversational Dutch. Participants were asked to press a button anticipating when they thought the speaker would finish their turn. The authors argued that the obtained distribution of button-press times relative to turn ends was similar to that of turn transitions in the corpus (i.e., centered around zero), and suggested that the cognitive processes involved in the button-press task are similar to those of turn-end estimation by participants in real conversation (cf. Levinson, 1983; Sacks et al., 1974; Stivers et al., 2009). The experimental items were manipulated in several ways in an attempt to disentangle the effects of lexico-syntactic and intonational information. To "remove" lexico-syntactic information, turns were low-pass filtered such that words were unintelligible, but acoustic information below 500 Hz (i.e., including pitch information) was preserved. To suppress tonal information, pitch was flattened in a different condition. Average button-press times in the flattened-pitch condition were found to be similar to those in the non-manipulated condition, but button-press times in the low-pass filtered condition ("no words") were significantly less accurate than both of these. The authors concluded that lexico-syntactic information is necessary and possibly sufficient for turn-end projection, whereas intonational information is not. However, as pointed out in Local and Walker (2012) and in Bögels and Torreira (2015), since the procedure of pitch flattening does not suppress duration and intensity cues, other prosodic cues must have been available to participants in the no-pitch condition. Furthermore, since phrase-final prosodic cues are not independent of each other (e.g., Holzgrefe-Lang et al., 2016 among many others), and participants could easily notice that pitch had been artificially flattened, they may have strategically focused on non-pitch cues only, in the context of the experiment.

Several more recent button-press studies have investigated lexico-syntactic anticipation. Three of these studies claimed to have found neural evidence of early turn-end projection based on EEG data (i.e., beta power reduction: Magyari et al., 2014; Readiness potential: Jansen et al., 2014; Wesselmeier et al., 2014). These neural signatures appeared more than a second before turn ends, suggesting that turn-end projection can start at an early stage during an ongoing turn. Wesselmeier et al. (2014) found that the readiness potential was disrupted in turns featuring semantic or syntactic violations, and concluded that both types of information are relevant for turn-end anticipation. A behavioral button-press study (Riest et al., 2015) also investigated the role of semantic and syntactic information in turn-end anticipation by low-pass filtering either all function words (as a proxy for syntactic information) or all content words (as a proxy for semantic information) in turns extracted from a German conversational corpus. This study found less accurate button-presses in both of these conditions compared to non-manipulated turns, especially so in the "no-semantics" condition, suggesting that semantic information is particularly important for accurate turn-end anticipation. This study also

included a condition in which participants read the full transcript of the target turn before they heard the fragment and pressed a button upon its end. No difference was observed between the data in this condition and that in the baseline condition without transcripts. Based on this finding, the authors argued that listeners can anticipate turn ends, and that anticipation is probably a predominant strategy in turn-end estimation. However, an alternative explanation is that acoustic cues close to turn ends were sufficient for accurate button-pressing in both conditions.

Taken together, the studies discussed so far suggest that anticipation of lexico-syntactic information is crucial for smooth turn-taking. However, other psycholinguistic studies have come to a different conclusion (Corps et al., 2018, 2019). These studies used tasks where participants responded to predictable and unpredictable turns with button presses or short answers (i.e., 'yes' or 'no'). Neither of these studies found an effect of predictability on the accuracy of button presses, although Corps et al. (2018) did find faster spoken answers for predictable questions. The latter study concluded that, although lexico-syntactic anticipation may speed up response planning, it does not necessarily allow for more accurate estimation of turn ends.

The role of context

As discussed above, no earlier studies have directly investigated whether the conversational context preceding a turn can facilitate turn-end estimation. Why could conversational context help with this process? It is known that listeners can predict upcoming words based on contextual information from the preceding discourse context (e.g., Van Berkum et al., 2005). However, as discussed above, it is unclear whether a greater predictability of words in a turn helps with turn-end estimation. While some studies (see above) suggest that predictability facilitates this process (Magyari et al., 2014; Riest et al., 2015), others found that it does not (Corps et al., 2018, 2019). Most relevantly, Corps et al. (2019) also found that predictability of the last word in a turn based on the context of the turn was unrelated to the accuracy of button-presses.

Provided that turn-end estimation relies on local cues to turn completion, the previous context of a turn could also facilitate turn-end estimation by making it easier to decide whether a turn is pragmatically complete at a certain point. It is known that context can play an important role in disambiguating pragmatically ambiguous utterances (Levinson, 1983) in conversation. For example, the utterance "I have a credit card." may either be an answer to a question about how the speaker will pay, or an offer to pay for the interlocutor's ticket (Gisladottir et al., 2015). This is of course especially relevant in natural dialogue, where the pragmatic import of utterances generally depends on context, especially when they are part of adjacency pairs (Sacks et al., 1974). If the previous conversational context can indeed help disambiguate the pragmatic import of utterances, this may help conversational participants determine whether a specific word in the current turn is the last word of that turn. The present study investigates whether the availability of the context preceding a turn indeed facilitates turn-end estimation.

The role of intonational phrasing

As discussed above, most psycholinguistic studies devoted to turn-taking have focused on the role of lexico-syntax in turn-end estimation, ignoring the potential role of prosodic information. Bögels and Torreira (2015) investigated the role of intonational phrasing in turn-end estimation using a button-press task. The stimuli were questions asked in a semi-spontaneous interview produced in two different versions: short (e.g., "So you are a student?") and long (e.g., "So you are a student at the Radboud University?"). The long question always contained an early syntactic and pragmatic completion point (i.e., "student"), lacking phrase-final prosodic cues (i.e., in this case final rising intonation and lengthening), as the end of the intonational phrase occurs later in the question (i.e., "University"). Participants in the button-press task heard original and manipulated stimuli. Original short questions led to shorter button-press times (around 100 ms on average) than manipulated versions of the long questions cut short at the turn-medial syntactic completion point, which lacked phrase-final prosodic cues (around 400 ms). This result indicated that the latter type of turn ends were unexpected to

participants in comparison with turn ends coinciding with the end of an intonational phrase. Also, when the short question was cross-spliced with the latter part of the long question (i.e., "at the Radboud University"), about 30% of participants pressed the button within 250 ms of the syntactic completion point (featuring phrase-final prosodic cues), whereas no button presses were found around the same position (without phrase-final prosodic cues) in the original long questions. Thus, faced with the exact same lexico-syntactic information, participants behaved differently on the basis of the presence or absence of phrase-final prosodic cues. Such phrase-final cues thus appear to be necessary for timely turn-end identification. Another study using a list-completion paradigm (Barthel et al., 2017) found that the presence of a late intonational cue to turn finality led to faster response times, similarly concluding that speech initiation is prompted by late cues to turn completion.

The present study

The studies reviewed above appear to provide some evidence for the relevance of lexico-syntactic anticipation and phrase-final prosodic cues in turn-end estimation. Only one study to our knowledge has explicitly manipulated the previous conversational context (Corps et al., 2019). This study found that higher predictability due to contextual information did not help listeners estimate turn ends more accurately. Still, the preceding context may facilitate turn-end estimation by disambiguating the pragmatic meaning of the turn (see section 'The role of context' above). An important limitation of many earlier studies is that they used carefully constructed stimuli (i.e., recorded by a speaker following precise instructions), or natural stimuli which were manipulated in conspicuous ways.

In the present study we aim to address these issues with two button-press experiments using turns sampled from a corpus of spontaneous telephone conversations. In Experiment 1, we present participants with written transcripts of the turns, whereas in Experiment 2 we present the natural auditory versions. We investigate (a) the potential facilitatory role of context on turn-end estimation, and (b) the generalizability of previous results regarding the role of intonation (cf. Bögels & Torreira, 2015).

The role of context

To investigate the role of prior context, we manipulate the availability of the (auditory) prior context to participants and examine whether this improves the accuracy of turn-end estimation, both when participants have access to lexico-syntactic information alone (Experiment 1: visual), and when they have access to acoustic information as well (Experiment 2: auditory). We hypothesize that having access to the prior context may improve the accuracy of button presses for the reasons argued above.

The role of intonational phrasing

To address issues of generalizability in earlier studies, and especially in Bögels and Torreira (2015), we built our stimuli using turns randomly selected from a corpus of spontaneous spoken Dutch telephone conversations. Bögels and Torreira (2015) concluded that phrase-final prosodic cues play an important role in turn-end estimation, but it is unclear how well these results can be extended to more varied conversational materials. First, the speaker in their study read the stimuli in a careful speech register, which differs from that of most conversations in that the latter often involves substantial reduction and blurring of phonetic contrasts (Johnson, 2004). Second, their stimuli consisted exclusively of questions of a specific syntactic format, and it is possible that phrase-final prosodic cues may be more relevant in the case of questions than for other types of turns, as the former often make use of salient pitch rises in final position.

In the present study, we revisit the role of intonational phrasing without introducing (prosodic) manipulations that may render stimuli conspicuously unnatural. To do so, we first identify points of possible lexico-syntactic completion in our stimuli based on orthographic transcriptions only. This is done based on the results of a visual version of the button-press task (Experiment 1), in which participants

are presented with word-by-word transcripts of the turns and are asked to press a button when they think they have encountered the last word of the turn. We then annotate the presence of intonational phrase boundaries at such points of lexico-syntactic turn completion. In an auditory version of the button-press experiment (Experiment 2), we address the hypothesis that an intonational phrase boundary at such points of lexico-syntactic completion will increase the likelihood of button presses in their vicinity. As we expect turn ends to generally co-occur with intonational phrase boundaries, in this analysis we will focus on turn-medial positions, as these are expected to provide more variability in terms of intonational phrasing. This will provide a close replication of Bögels and Torreira (2015), where the condition involving turn ends spliced into utterance-medial positions led to a button-press rate of around 30%.

Experiment 1

In Experiment 1, we investigate to what extent participants can estimate the end of conversational turns based on lexico-syntactic information alone, that is, in the absence of acoustic phonetic information, while we manipulate the availability of the previous conversational context. As explained above, we hypothesize that having heard the previous context will make it easier for participants to estimate turn ends more accurately.

Given that we use stimuli randomly drawn from a corpus of spontaneous speech, we control for turn length in our statistical models. Several studies have shown a relation between turn length and response timings, but results were inconsistent. Some studies found a negative correlation (i.e., later responses for shorter questions; Bögels, 2020; Magyari et al., 2017), while one study found later responses for longer questions (in individuals at clinical high risk for psychotic disorders; Sichlinger et al., 2019). A large-scale corpus study by Roberts et al. (2015) found a non-linear relationship between turn-length and response times. Longer turns were associated with longer response times, but very short turns (< 700 ms) also showed longer response times. Based on these findings, we consider linear and quadratic trends for turn length in our regression models.

Methods

Participants

Twenty-four native speakers of Dutch without reading problems (mean age 21.8, range 19–26; 6 males, 18 females) were recruited from the participant database of the Max Planck Institute for Psycholinguistics. They all gave written informed consent and were paid €6 for their participation.

Materials

We chose the Corpus Gesproken Nederlands (CGN, Corpus of spoken Dutch; Oostdijk, 2000), as our source of speech materials. We used part c of the corpus, which consists of 358 ten-minute telephone calls between friends and acquaintances. In the Context condition we wanted to present an amount of context before the critical turn that would not be too long (as not to make the experiment needlessly long), but still provide enough information so that the position of the critical turn within the conversation would be understood. For this reason we selected critical turns and contexts using the following procedure: First, we identified the first floor change (i.e., speaker switch) between 10 and 20 seconds from the start of the call, which would later constitute the end of the critical turn. If no speaker switch occurred there, the call was discarded. Next, the last complete turn before this floor change was selected as a critical turn. The start of the critical turn constituted either a speaker change or a pause at a point of syntactic and prosodic completion within the speech of the same speaker. By prosodic completion we mean that the utterance before the boundary exhibited at least a pitch accent and a boundary tone resulting in a characteristic phrase-final nuclear contour (i.e., falling, low rising, high rising, fall-rising). Note that a floor change was thus always present after the critical turn, but not necessarily before. The context of each critical turn consisted of the start of the telephone conversation until the start of the critical turn (i.e., lasting about 10–20 seconds).

We did not consider the following cases as full floor changes when selecting the floor changes that would constitute the ends of our critical turns: backchannels, repair initiations (e.g., the Dutch equivalents of *huh?* and *what?*), simultaneous starts (when both interlocutors start speaking in overlap closely to each other), and floor changes with gaps longer than one second after the critical turn. We discarded critical turns with laughter around the floor change, since this complicated the segmentation of the turn. We also discarded critical turns that included substantial overlapping speech from the other speaker, since this might affect the turn design of the current speaker. Overlapping backchannels and terminal overlaps within the final syllable of the critical turn were not considered cases of substantial overlap and were thus included in the data. However, we never included cases in which any amount of overlap was audible in the channel of the critical turn (i.e., we discarded all cases of cross-talk between channels).

This selection procedure resulted in 96 critical turns (including only one of the two channels from the stereo signals available in the corpus) with their contexts from the beginning of the phone call (including both channels). The transition between the context and the critical turn constituted a speaker change in 71 cases; the same speaker kept the floor in the other 25 cases (in these cases an intonational phrase boundary and a pause always occurred before the critical turn). The selected critical turns ranged from 306 to 6515 ms in duration (M = 1833 ms, SD = 1431 ms), and contained between 1 and 21 words (M = 8.05, SD = 4.31). The duration of the previous context ranged between 9.3 and 19.7 s (M = 13.4, SD = 2.7).

Procedure

To make sure that participants had access to the words incrementally, as in the auditory modality (Experiment 2), we presented the critical turns word-by-word rather than in full form. Each word in a turn appeared in white on a black background in the middle of the screen for one second, without any punctuation marks. Each critical turn was preceded by a white fixation cross on a black background for one second at the position where the first word would occur next. Participants were instructed to press a button when they thought that the word currently on the screen was the last word of the turn. When they pressed the button, the trial stopped and the next trial was presented. If they did not press within 1000 ms after the onset of the last word of the turn, they saw 'XXX' on the screen, which indicated that the trial was over. Trials were considered successful if participants pressed the button during the presentation of the last word of the turn (i.e., within 1000 ms from its onset), otherwise they were considered failures. Trials in the context condition were preceded by an auditory presentation of the preceding context of the turn in the corpus. Participants were informed that the context constituted the start of the actual phone call, and that the critical turn immediately followed the context in the original phone call. After the context had been played over headphones, participants were instructed to press a button to start a word-by-word visual presentation of the critical turn. It was also indicated on the screen whether the last speaker in the auditory context would continue speaking or whether the other speaker would take the floor, since both options were possible at this stage, and the identity of the next speaker could not be reliably recovered from the written words.

Participants were tested in a sound-proof booth. They encountered 96 trials divided into two blocks of 48 items each. Items were presented with and without context in separate blocks. Twelve participants heard items 1–48 with context and the other twelve participants heard items 49–96 with context. Within each of these groups of twelve participants, six received the context block first and the other six received the context block second. Within a block, the items were presented in a random order generated separately for each participant. Each block was preceded by five practice items after which participants could ask questions. The experiment lasted about 45 minutes in total.

Results

The percentage of successful button-presses (targeting the actual last word of the turn) was generally higher in the context condition (42%) than in the no-context condition (34%). A similar proportion of button presses occurred before the last word of the turn (36% for the context condition and 36% for

the no-context condition). Time-out responses (i.e., where participants did not press the button within 1000 ms from the onset of the last word of the turn) were somewhat less frequent (22% for the context condition, and 30% for the condition without context).

To assess the role of context in turn-end estimation, we used mixed-effects regression modeling as implemented in the LME4 package in R. We fitted a logistic regression model with a binary dependent variable (success vs failure) as the response, Context and Turn Length in number of words (both linear and quadratic trends) as fixed predictors, as well as interactions between each term of Turn Length and Context. Regarding the random structure of the model, we specified intercepts for Participant and Item, as well as an interaction between Item and Context, our main variable of interest. We did not include an interaction between Participant and Context in the model, as this produced a singular fit. This regression model (Model 1) yielded a statistically significant interaction between the quadratic term of Turn Length and Context. Table 1 presents a summary of the model, whereas Figure 1a illustrates its predictions as a function of Context and Turn Length. It can be seen in this figure that Context is predicted to provide an advantage only in short turns of up to five words. To further explore the non-linear interaction between Context and Turn Length, we transformed the continuous variable Turn Length into a discrete variable with four levels ((1) 1–4 words: 432 observations, 18 items; (2) 5–8 words: 1008 observations, 42 items; (3) 9–12 words: 480 observations, 20 items; (4) 13–21 words: 384 observations, 16 items), and inspected the proportions of successful responses for each of these levels. These data are shown in Figure 1b. It can be seen in this figure that the percentage of successful responses was highest in turns consisting

Table 1. Experiment 1: Summary table of Model 1. Formula: Success ~ Context * poly(Length, 2) + (1 | Participant) + (1+ Context | Item). Statistically-significant p-values in bold.

Number of obs: 2304; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−1.26	0.23	−5.47	**< .0001**
Context	0.61	0.17	3.62	**< .0005**
Length	−39.9	12.4	−3.22	**< .005**
Length ^2	−54.37	13.09	−4.15	**< .0001**
Context:Length	3	10.05	−0.3	.76
Context:Length^2	30.87	10.96	−2.82	**< .005**
Random effects:				
ICC: 0.53				
Grouping	Effect	SD		
Item	Intercept	1.87		
	Context	0.84		
Participant	Intercept	0.19		

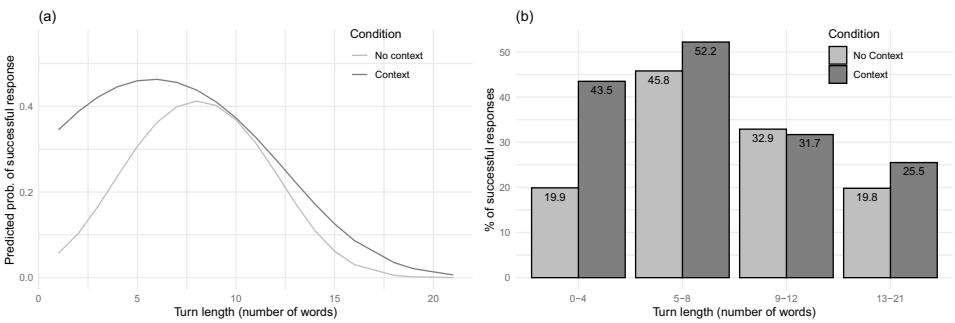

Figure 1. Experiment 1. (a) Model 1: Predicted probability of success as a function of Context and Turn Length. (b) Percentage of successful responses as a function of Context and Turn Length (divided into four bins). See text for details.

of five to eight words, and that the difference in percentage of successful responses across Context conditions is the greatest for short turns of up to four words, smaller for turns of five to eight words, and even smaller for longer turns. To further examine these data, we fitted an additional regression model with a similar structure to that of the model discussed above except for the inclusion of the discrete instead of the continuous version of Turn Length (Model 2; see Table A1 in Appendix for a detailed summary). We then ran a series of pairwise comparisons between the Context conditions within each discrete level of Turn Length using the Tukey procedure as implemented in the emmeans R package (Lenth, 2020). We observed a facilitatory effect of Context for the shortest turns (19.9% vs 43.5% correct; estimate: −1.49; $z = -4.46$; $p < .0001$), and a positive trend for Context in turns of 5 to 8 words (45.8% vs 52.3% correct; estimate: −0.38; $z = -1.88$; $p = .06$). No statistical effect was observed in either of the other two cases ($p > .2$). Taken together, these results show an advantage for the Context condition in turns of short length only. Moreover, Turn Length appears to have a non-linear effect on response accuracy, with turns of medium length leading to a greater likelihood of success in participants' responses than turns of either short or long lengths.

Discussion

In Experiment 1, participants were presented with turns transcribed in written form in an incremental manner (i.e., word-by-word) with or without their preceding contexts, and were asked to press a button when they thought that the word being presented on the screen was the last word in its turn. We observed a quadratic effect of turn length on the accuracy of button presses. Turns of medium length led to more accurate button-presses than short or long turns. This effect is similar in nature to the quadratic effect described by Roberts et al. (2015) for response times in a corpus study. We come back to this in the general discussion. Regarding the role of context, we found a facilitatory effect on button-press accuracy for short turns of one to four words only. This may be explained by the fact that short turns do not provide much information in themselves, and that in such cases the context provides useful additional cues.

Another remarkable finding is that successful button presses targeting the actual last word of the turn constituted only over one third of all button presses. This suggests that turn-end estimation is not very accurate when it is based on lexico-syntactic information alone.

Experiment 2

Experiment 2 consists of an auditory button-press task where participants have access to the auditory recordings of the turns transcribed in Experiment 1. We first investigate the role of the previous conversational context in the presence of acoustic information by presenting it auditorily before the target turn in one of the two conditions, using the same procedure as in Experiment 1.

In Experiment 2 we also investigate the role of intonational phrasing in turn-end estimation. We first use the results of Experiment 1 (collapsed over context and no-context conditions) to identify the position in each turn that constitutes the most plausible end in terms of lexico-syntactic information alone. We then examine whether the presence of an intonational phrase boundary at such positions is related to the occurrence of button presses. Based on previous experimental results (Bögels & Torreira, 2015), we hypothesize that participants will be more likely to press the button close to such positions when an intonational phrase boundary is present.

Methods

Participants

Twenty-four native speakers of Dutch without uncorrected hearing problems (mean age 21.6, range 18–27; 5 males, 19 females) were recruited from the participant database of the Max Planck Institute for Psycholinguistics. None of them had participated in Experiment 1. They all gave written informed consent and were paid €6 for their participation.

Materials

Experimental items were similar to those used in Experiment 1 except for the fact that critical turns were presented in their original auditory form.

Item annotation

For each item, we identified the word whose visual stimulus had received the highest number of button presses within 1000 ms after its onset in Experiment 1 (collapsing over context and no-context conditions), which we refer to as the *target word* from now. Target words constituted the final word in their turns in 53 of the 96 items (55.2%), and thus occurred earlier in the turn in all remaining cases (43 items). For each target word, we annotated whether it coincided with the end of an intonational phrase in the auditory version of the turn. By intonational phrase we mean a prosodic unit ending in a characteristic tonal movement beginning in its last accented syllable (e.g., fall, rise-fall-rise, low-rise, high-rise, or L* L%, H* LH%, L* H%, H* H%, respectively, in autosegmental-metrical notation; Ladd, 2008). Note that a pause was never present after the intonational phrase boundary, as this would have constituted a turn boundary according to our sampling criteria. Initially, both authors annotated the sound files separately. Eight cases (8.3%) of disagreement in this initial annotation were resolved after a brief discussion between the authors.[1] In 74 items (75%), the target word coincided with the end of an intonational phrase. In target words occurring in turn-final position (n = 53), all but three items coincided with an intonational boundary. These three cases ended in a disfluency or filler word (e.g., *uh*). In medial position, however, we observed more variability, in that roughly half of the target words coincided with an intonational boundary (n = 23 out of 43; 53.5%).

We wanted to ensure that any effect of Intonational Phrasing could not be explained by lexico-syntax alone. We controlled for this factor by means of an estimate of plausibility of turn completion based on the results of Experiment 1, which we refer to as *lexico-syntactic completion index*. We normalized the number of button presses on the target word by the number of participants that saw this word (i.e., excluding participants that pressed the button earlier in the turn, since trials ended as soon as a participant pressed the button). If, for instance, the fifth word of a turn was the target word and 12 button presses occurred on this fifth word, five on the second word, and seven on the last word of the turn, the five presses on the second word were excluded (leaving 19 participants that saw the target word), such that the lexico-syntactic completion index on the fifth word was 12/19 (.63). Alternatively, as in the example in Figure 4, if no button presses occurred before the target word ('proefwerkweek,' *test week* in Figure 4), the number of button presses on the target word (20 in Figure 4) was divided by the total number of button presses (24; leading to a lexico-syntactic completion index of 20/24 or .83). Finally, we annotated whether the critical turn ended in a question or not, as we were interested in seeing if previous experimental effects observed with items exclusively involving questions (Bögels & Torreira, 2015) could be observed in other turn types as well. Twenty-two questions were identified on the basis of their syntax (i.e., subject-verb inversion, *wh*-constructions). Twelve statements directly referring to the addressee's mental states or knowledge (or B-events, see Labov & Fanshel, 1977) were also considered questions, as they clearly appeared to function as requests for confirmation rather than as plain assertions. Thus, 34 items in total (35%) were coded as questions.

Procedure

The procedure was the same as described in Experiment 1, except for any differences reported in this section. The instruction to participants was the same as in previous button-press studies (Bögels & Torreira, 2015; De Ruiter et al., 2006). Specifically, participants were instructed to anticipate the moment that the speaker would finish speaking and press a button at that precise moment. Trials without context started with a white fixation cross on a black screen for one second, after which the turn started to play over the headphones. Whenever the participant pressed the button, the audio file stopped playing immediately and the fixation cross disappeared for one second, after which the next trial began. Trials with context started with a fixation cross for one second, after which the context of

the target turn was played. When the context section was finished, participants saw a written instruction on the screen asking them to press a button in order to start the critical turn. After pressing the button again, the audio file stopped playing and the fixation cross disappeared for one second, after which the next trial (i.e., the next context) began. The experiment lasted about half an hour in total.

In order to investigate participants' button-press behavior, we fitted pairs of logistic regression models with a binary response variable encoding whether button presses were produced within a window of fixed length centered around positions of interest in each item. In the first model of each pair the window had a length of 250 ms (i.e., with boundaries 125 ms before and 125 ms after the end of the target word), whereas in the second model the window had a length of 500 ms (i.e., with boundaries 250 ms before and after the end of the target word). The wide response window (500 ms) was the same as that used by Bögels and Torreira (2015). In addition, we used a narrow response window (250 ms) to test our hypotheses more extensively. The structures of these models are discussed below in the corresponding sections.

Results

Distribution of button-press times relative to turn ends

Figure 2 shows the distribution of button-press timings relative to the end of critical turns. Looking at this figure, it can be seen that button presses most frequently occurred between 200 and 300 ms after the turn end. This distribution of button-press timings is similar to that of turn-transition times in conversational corpora (cf. Heldner & Edlund, 2010; Levinson & Torreira, 2015; Roberts et al., 2015; Torreira et al., 2015). It is notable too that the mean and median button-press timings (217 and 247 ms, respectively) are of similar magnitude to the turn transition times that the experimental stimuli had in the corpus (mean = 251 ms; median = 210 ms). The button-press timings in the present study do appear later than those in the first button-press study on turn-taking by De Ruiter et al. (2006), which also used turns from a corpus of natural conversations (i.e., −186 ms on average). However, the average actual turn-transition times for these turns in the corpus was also negative (−78 ms). A possible explanation for these relatively small (negative) timings may be that they only included stimuli that consisted of minimally five words, whereas the present study also included very short turns, which may be more likely to lead to long transition/button-press times (see effects of turn length below).

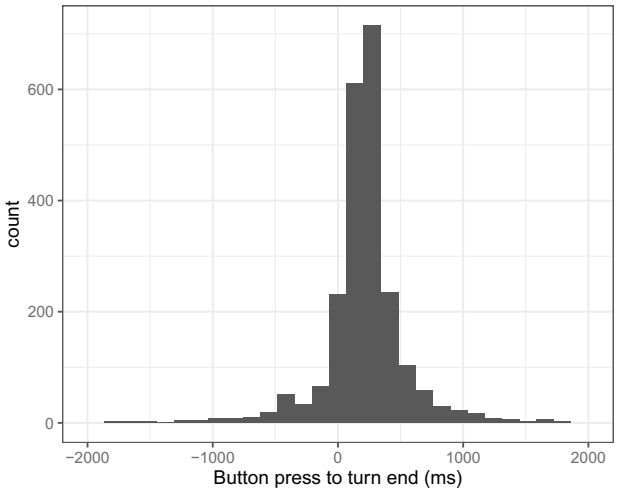

Figure 2. Histogram of button-press times relative to turn end (Experiment 2). Extreme values (i.e., those lower than −2000 ms and greater than 2000 ms; 1.7% of the data) have been excluded from the figure for the sake of conciseness.

Table 2. Experiment 2: Summary table for Model 3A (250-ms success window): Model formula: Success ~ Context * poly(Length, 2) + (1 | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs.: 2249; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−2.76	0.3	−9.13	**< .0001**
Context	0.06	0.22	0.28	.78
Duration	−15.97	12.35	−1.29	.2
Duration^2	−46.61	14.54	−3.2	**< .005**
Context: Duration	−4.2	10.78	−0.39	.7
Context: Duration^2	28.1	12.88	2.18	**< .05**
Random effects:				
ICC: 0.50				
Grouping	Effect	SD		
Participant	Intercept	1.1		
	Context	0.63		
Item	Intercept	1.3		

Table 3. Experiment 2: Summary table for Model 3B (500-ms success window). Model formula: Success ~ Context * poly(Length, 2) + (1 + Context | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs: 2249; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−0.48	0.28	−1.67	<.1
Context	0.09	0.2	0.45	.65
Duration	0.54	6.72	0.08	.94
Duration^2	−25.93	6.98	−3.71	**< .0005**
Context: Duration	−7.35	6.26	−1.17	.24
Context: Duration^2	16.63	6.52	2.55	**< .05**
Random effects:				
ICC: 0.47				
Grouping	Effect	SD		
Participant	Intercept	1.49		
	Context	0.52		
Item	Intercept	1.11		
	Context	0.66		

Table 4. Experiment 2: Summary table for Model 5A (250-ms window around turn-medial positions): Model formula: Button-press ~ Intonational Boundary + (1 | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs: 568; Items: 24; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−6.39	1.24	−5.14	**< .0001**
Intonational boundary	2.29	1.11	2.06	**< .05**
Random effects:				
ICC: 0.36				
Grouping	Effect	SD		
Participant	Intercept	1.09		
Item	Intercept	0.83		

Table 5. Experiment 2: Summary table for Model 5B (500-ms window around turn-medial positions): Model formula: Button-press ~ Intonational Boundary + (1 | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs: 568; Items: 24; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−5.78	0.97	−5.94	**< .0001**
Intonational boundary	2.5	0.86	2.89	**< .005**
Random effects:				
ICC: 0.48				
Grouping	Effect	SD		
Participant	Intercept	1.28		
Item	Intercept	1.15		

Role of context: analysis of button-press times relative to turn ends

As explained above, we fitted a pair of mixed-effects logistic regression models with a binary response variable (success vs failure) encoding whether button presses were produced within a window of fixed length centered around the end of each item (i.e. 250 ms in Model 3A; 500 ms in Model 3B). The fixed predictors were Context, our main variable of interest, Turn Duration (linear and quadratic terms), and their interaction. The initial random structures of both models included intercepts for Participant and Item, as well as interactions between each of these factors and Context. The interaction between Context and Item was dropped from Model 3A, as its inclusion in the model led to a singular fit. Model summaries are presented in Tables 2 and 3. Both regression models yielded a non-linear interaction between Context and Turn Duration. These interactions are illustrated in Figure 3a and 3b, which show model predictions of successful responses as a function of these factors. To further investigate this interaction, we transformed the continuous Duration variable into a discrete variable with four levels ((1) 0–1 s: 432 observations, 21 items; (2) 1–2 s: 1071 obs., 46 items; (3) 2–3 s: 444 obs., 19 items; (4) 3 s or longer: 239 obs., 10 items), and inspected the proportions of successful responses at each of these levels. These data are shown in Figure 3c and 3d. The effect of Context appears to follow the same trend as in Experiment 1, in that Context appears to provide some advantage in turns of short (0–1 s) and medium short length (1–2 s), although the difference across Context conditions is less clearly discernible than in Experiment 1. Another difference with the results of Experiment 1 is that in turns of mid to long duration Context is associated with a decrease rather than an increase in successful responses. In order to assess the statistical significance of these observations, we fitted additional regression models with similar structures to the models discussed above except for the inclusion of the discrete version of Turn Duration instead of its original continuous version (Models 4A and 4B; see Tables A2 and A3 in Appendix). We then ran a series of Tukey pairwise comparisons between the Context conditions at each of the four levels of Turn Duration. Regarding the first model, where success involved a 250-ms response window, we observed a positive trend for Context in the shortest turns only (8.8 vs 15.3% correct; estimate = −0.64; $z = -1.78$, $p = .07$), and a negative effect of Context in turns of 2 to 3 s (20 vs 14.1% correct; estimate = 0.7; $z = 2.06$, $p < .05$). Regarding the second model, where success involved a 500-ms window, we did not observe any statistical effect ($p > .15$ in all cases). Furthermore, it appears from the figures (especially Figure 3d) that, as in Experiment 1, the percentage of successful responses was highest in turns of medium duration.

Role of intonational phrasing: analysis of button-press times relative to the end of target words

In order to investigate the role of intonational phrasing, we examined our participants' button-pressing behavior in the vicinity of the end of target words (i.e., those that received the highest number of button presses in the visual task in Experiment 1) as a function of the presence of an intonational boundary. Since, as we expected, the vast majority of target words in turn-final position coincided with an intonational boundary, and the few that did not ended in a disfluency, we focused

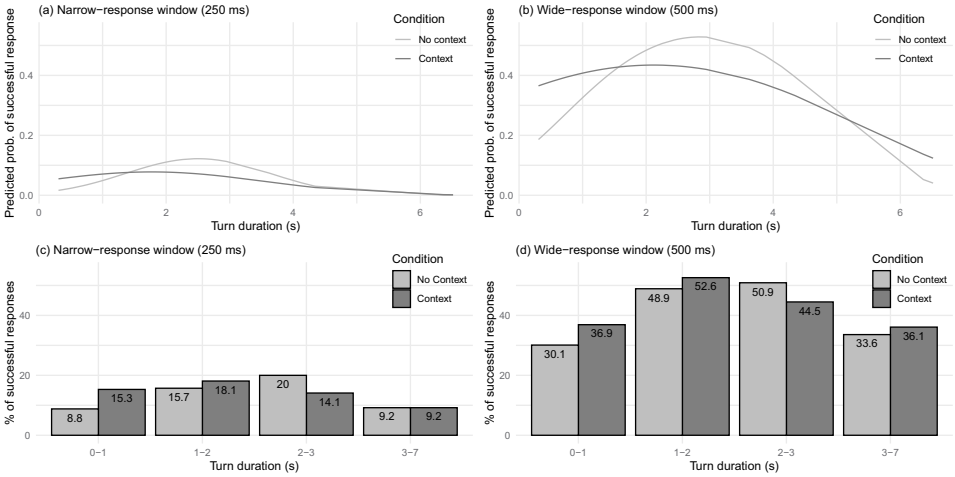

Figure 3. Experiment 2. Predicted probabilities for models 3A (a) and 3B (b) as a function of Context and Turn Duration, where success involves 250-ms and 500-ms response windows, respectively. Percentages of successful responses as a function of Context and Turn Duration based on a 250-ms response window (c) or on a 500-ms response window (d). See text for details.

only on those items with target words in turn-medial position, where we did observe variation in intonational phrasing. Moreover, we excluded items where the target words ended within 500 ms of the turn end, as button presses in this time region could have plausibly targeted the turn end rather than the end of the target word. The resulting dataset (24 items of which 9 were questions; 568 observations; mean offset to turn end: −1446 ms; SD: 911 ms) featured a low but non-negligible number of button presses close to target words (250 ms window: n = 12; 500 ms window: n = 31). As predicted by our hypothesis, the vast majority of these button presses were produced when the target word coincided with an intonational boundary (250-ms window: 11/12; 500-ms window: 28/31). Figure 4 illustrates participants' responses for one of the items. In this particular case, we observed two button presses within 125 ms and four within 250 ms of the end of the target word *proefwerkweek* ("exam week"). We fitted two regression models with a binary response (i.e., the occurrence of a button press within a 250- and 500-ms window centered around the end of the target word; Models 5A and 5B, respectively), where the main fixed predictor of interest was Intonational Boundary. The following control variables were also included in the models: Lexico-syntactic Completion Index, Context, Question (whether the turn ended in a question), and linear and quadratic terms for Turn Duration up to the end of the target word. Moreover, we considered interactions between each of these fixed predictors and Intonational Boundary. Based on our previous analyses in Experiments 1 and 2, we also considered an interaction between Context and Turn Duration up to the end of the target word. The initial random structures of both models included intercepts for Participant and Item, as well as an interaction between Participant and Intonational Boundary, our main variable of interest. This interaction could not be retained in any of the final models due to issues of singular fit.

None of the predictors other than Intonational Boundary were statistically significant ($p > .1$ in all cases), and were therefore removed from the models. Intonational Boundary yielded statistically significant effects in both models (Model 5A: $p < .05$; Model 5B: $p < .005$). As predicted by our hypothesis, button presses in turn-medial positions were significantly more common in the presence of an intonational boundary (250-ms window: n = 11, 4.7%; 500-ms window: n = 28, 11%) than in its absence (250-ms window: n = 1, 0.003%; 500-ms window: n = 3, 0.01%). Tables 4 and 5 show summaries of these models.

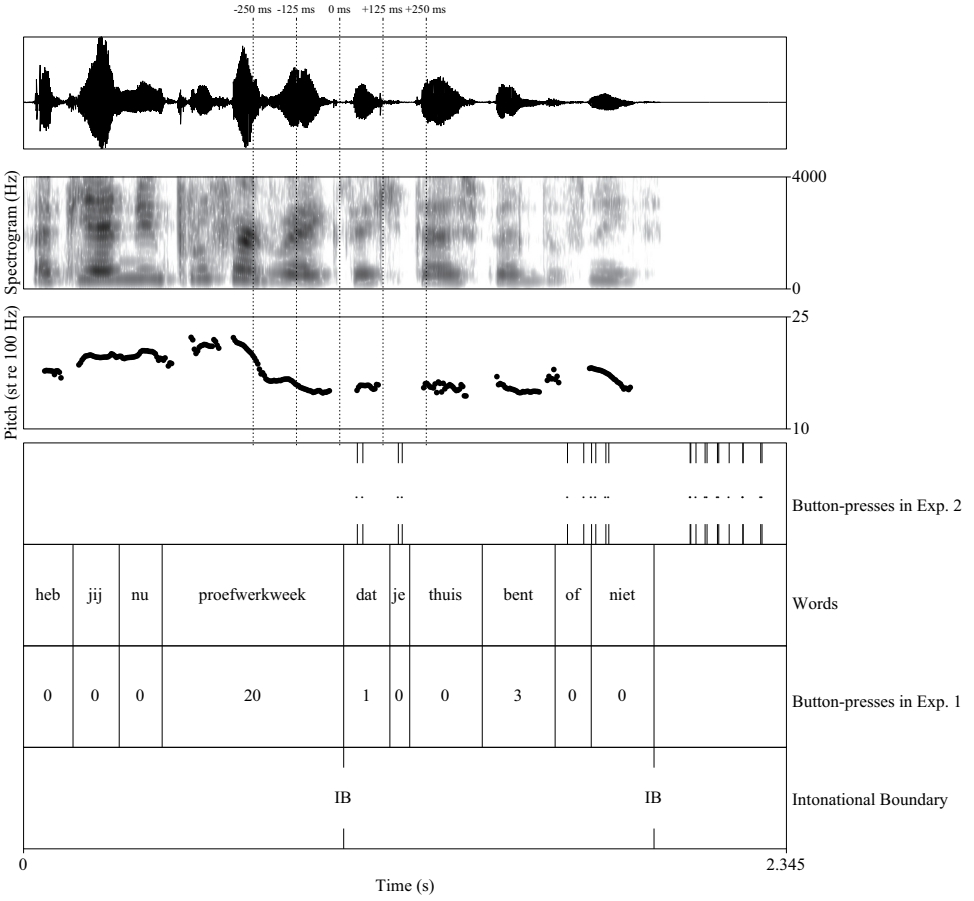

Figure 4. Waveform, spectrogram, pitch track, annotation (Words and Intonational Boundary), and participants' responses (in both experiments) for example, item fn008160 (number corresponding to the phone call in the corpus of spoken Dutch, CGN). The Dutch utterance *Heb jij nu proefwerkweek dat je thuis bent of niet?* in the item can be translated as "Do you have exam week now that you are at home or not?".

Discussion

In Experiment 2 we first investigated the role of context in turn-end estimation by means of an auditory button-press task. As in Experiment 1, which involved a visual version of the task, in Experiment 2 we found some evidence of a facilitatory effect of context in turns of short duration (0–1 s) in the form of a statistical trend. Interestingly, we also observed a clearer *inhibitory* effect in turns of 2–3 s, where the presentation of context led to reduced accuracy in turn end estimation.

In Experiment 2 we also investigated the role of intonational boundaries in turn-end estimation. As turn ends in our data almost always coincided with intonational boundaries, we focused on turn-medial positions that could have been turn ends based on lexico-syntactic information alone (i.e., based on the visual task in Experiment 1). Although only a limited proportion of button presses were observed at such locations (5% within 125 ms of the end of the target word; 11% within 250 ms), the vast majority of these button presses occurred in items where such positions coincided with an intonational boundary (i.e., 92% of those within 125 ms of the end of the target word; 90% of those within 250 ms).

General discussion

The experiments presented above were aimed at testing two hypotheses regarding how conversational participants estimate turn ends during floor exchanges. First, we hypothesized that having access to the context preceding a conversational turn before a floor change would facilitate turn-end estimation. We observed a facilitatory effect of context when participants did not have access to acoustic information (i.e., when they viewed visual transcripts of the turns; Experiment 1) for short turns only. When acoustic information was available (i.e., when they listened to the turns; Experiment 2), we observed a trend in the same direction for the shortest turns. Second, we hypothesized that listeners use phrase-final prosodic cues (together with lexico-syntactic information) in order to estimate turn ends (cf. Bögels & Torreira, 2015), as opposed to lexico-syntactic information alone (De Ruiter et al., 2006). Our findings, based on data comprising a variety of turns sampled from a corpus of natural conversation, suggest that listeners take intonational phrasing into account when projecting the end of conversational turns. We discuss these two findings and their implications in the following sections.

Effect of context

When participants had access to lexico-syntactic information only (Experiment 1) in short turns (< 1 s), they were more than twice as accurate in targeting the last word of the turn when the prior context had been presented than when it had not (44 vs 20%, respectively). When participants had access to both lexico-syntactic *and* acoustic information (Experiment 2), the facilitatory effect of context for short turns was less clear and only showed a trend in the analysis with the narrow response window (and no effect in the wide response window). In turns of medium length (2–3 s) having access to context even led to fewer button presses in the narrow response window.

The fact that we find an effect of context for short turns in Experiment 1 (and a trend in Experiment 2) shows that there is information in the previous context that *can* help determine which word in the current turn constitutes a turn end. Whether this information is also useful to listeners may depend on other factors. One relevant factor appears to be the length of the turn. If the turn is short, the turn itself does not provide much information, and the preceding context may provide useful information. In contrast, when the turn is long, the turn itself may provide enough information for correct turn-end estimation and any information provided by the context may be redundant.

In Experiment 2, we also found a surprising detrimental effect of context in turns of medium length. One speculative explanation of this effect is that participants were slightly less focused on the task in context blocks, since they intermittently listened to longer contexts and shorter critical turns that they needed to respond to, while the no-context blocks only featured the latter. While this explanation may apply to turns of all length, a potential distraction of the intervening contexts may have been most detrimental in turns of medium length, where participants' button-press accuracy was highest.

In the Introduction we brought forward that context may facilitate turn-end estimation because it may assist listeners with long-range lexico-syntactic anticipation. In Experiment 1, where we found the clearest effect of context (in short turns), participants had ample time for reading each word (1000 ms; cf. EEG reading studies presenting one word every 300–700 ms, e.g., Nigam et al., 1992; Sereno et al., 2020), rendering strict anticipation an unnecessary strategy in this set-up. We also conjectured that the previous context may help listeners decide whether the turn is pragmatically complete at any given point (local anticipation). Given that especially short turns may have been pragmatically ambiguous in isolation, the previous context may have been helpful in disambiguating these turns, making it easier for participants to determine the actual turn end. Future studies could look into this by controlling for this type of ambiguity and examining whether more ambiguous turns benefit more from the preceding context. Overall, the effect of context only appeared in a subset of turns, and was not very strong in the auditory experiment, which is closest to natural turn-taking. One may argue that we did not include a very long context in our experimental trials (the duration of our contexts ranged from 10 to

20 seconds at most). It is possible that the effect of context would have been stronger and/or present in longer turns as well if we had included a longer stretch of context. It should be noted, however, that numerous turns in everyday conversation occur before a context longer than 10 to 20 seconds is available (and many conversations may not even reach this length).

Effect of prosody

In Experiment 2 we observed that button presses occurred more often close to turn-medial plausible points of syntactic completion if these were accompanied by an intonational phrase boundary. More specifically, button presses close to such points were significantly rarer in the absence of an intonational boundary (0.01% or less in both analyses) than in its presence (5 and 11% depending on the analysis). These findings complement the button-press experiment by Bögels and Torreira (2015) in that the results of both studies are in line with the idea that phrase-final prosodic acoustic cues are relevant for turn-end projection. Interestingly, the effect of intonational phrasing in the present study was not confined to questions, but was present across a random sample of conversational turns. We should note here that, contrary to Bögels and Torreira (2015), the present study provides correlational evidence only, as we did not manipulate the presence or absence of an intonational phrase boundary. The reason for this was that it is not straightforward to achieve the same level of experimental control as in Bögels and Torreira (2015) when using spontaneous conversational data without conspicuously distorting the experimental materials. Yet, since the evidence for a role of prosodic information is correlational, we cannot exclude that another variable, correlated with the presence of intonational phrase boundaries, could explain the observed effect on turn-end estimation. However, it is unlikely that the effect of intonational boundary can be fully attributed to lexico-syntactic plausibility, (i.e., positions with a intonational boundary also being more plausible turn-ends based on lexico-syntactic grounds) since the effect of lexico-syntactic plausibility was not statistically significant in both models we ran. Other than this, we should note that it is now well known that intonational phrase boundaries, besides exhibiting pitch cues associated with final boundary tones (e.g., pitch rises before high boundary tones, pitch falls before low boundary tones), are often marked with non-pitch cues as well, such as final lengthening (e.g., Turk & Shattuck-Hufnagel, 2014) and changes in the phonetic characteristics of segments (Cho, 2016; Cole, 2015). From the present study we cannot determine whether and how such cues are used in turn-end estimation, as this would require manipulating these cues experimentally. Future studies could in principle address this issue, although as noted above, experimental control of this kind may compromise ecological validity. For the time being, therefore, we believe that converging evidence from both experimental and observational studies can offer valuable insights into the cognitive processes underlying turn-taking (cf. Meyer et al., 2018).

Additional support for the idea that acoustic information is relevant for turn-end estimation comes from the observation that lexico-syntactic information alone does not appear to be a sufficient source of cues. In Experiment 1, where acoustic information was not available to participants, button presses targeting the actual last word of the turn constituted only 38% of the responses, whereas 36% targeted an earlier word. This shows that, based on lexico-syntactic information alone, numerous earlier points in the turns constituted potential turn ends.

Another relevant observation made in this study is that turn-medial positions that were syntactically plausible turn ends *and* coincided with an intonational phrase boundary received only around 11% of button presses in the analysis using a 500-ms target window and 5% in the analysis using a 250-ms target window. These are substantially lower proportions than those reported by Bögels and Torreira (2015), who, using also a target window of 500 ms, observed 30% of button presses in turn-medial positions when participants encountered cross-spliced original short questions featuring phrase-final prosodic cues (e.g., Are you a student?) immediately followed by the second half of longer questions ([. . .] at the Radboud University?). Although it is not possible to draw a direct comparison between experiments, we can say that they notably differ in that the turn-medial positions in Bögels and Torreira (2015) constituted actual turn ends (i.e., they had been cross-spliced across

items), whereas in the present study they were turn-medial by all standards. The former, since they were originally followed by silence when they were spoken, may have contained utterance-final phonetic characteristics detectable by listeners and may thus have constituted extra cues to turn finality. Still, even in such conditions the turn ends cross-spliced into turn-medial positions in Bögels and Torreira (2015) received only 30% of button-presses. In our view, these findings suggest that listeners are often able to inhibit their response when they hear further incoming information instead of silence. We are not suggesting by this that silence alone, regardless of the linguistic information that precedes it, is the main cue to a turn's end, but rather that its absence or presence may complement earlier cues in signaling turn completion or turn continuation (since turn ends usually coincide with structural points of linguistic completion *and* silence; cf. Sacks et al., 1974).

Effect of turn length

Given earlier reports of effects of turn length on button-press and/or response times, we took account of turn length in our models of button-press accuracy. Model predictions in both experiments showed a quadratic effect of turn length such that turns of medium length generally received more accurate button-presses than either short or long turns.

The observation that short turns lead to less accurate turn-end estimation appears compatible with earlier findings of a negative correlation between turn length and response times for verbal responses (Bögels, 2020; Magyari et al., 2017; Roberts et al., 2015). However, in these studies this effect may be related to the fact that participants have a longer time to plan their response during longer turns. Since this is unlikely to play a role in button-press responses, we are tempted to speculate that the effect in the present study may simply be due to participants being less ready to respond after only a short amount of time. It is also possible that short turns contain fewer cues assisting turn projection, or that such cues occur closer to turn ends, making turn-end projection more challenging.

The other side of the effect, that is, the smaller probability of accurate button presses for the longest turns, appears consistent with two earlier studies on response times in verbal responses. Sichlinger et al. (2019) found later answers to longer questions (in individuals at high clinical risk of psychotic disorders) and Roberts et al. (2015) found longer turn-transition times after long turns in a telephone corpus (as part of a more general quadratic effect). Both of these effects may be explained by longer turns being more complex and more difficult to understand, leading to more time being needed to respond to them. This may also partly affect button-press timings since comprehension is part of the button-press task, but the effect may be less strong since no answer has to be formulated or planned. The effects observed in our experiment may also be partly due to the fact that longer turns are likely to provide more points of potential completion before their end than short turns, leading to a greater number of early button-presses in the long turns. Similarly, De Ruiter et al. (2006) report a negative correlation between turn length and button-press timings relative to turn end (referred to as BIAS by them), showing a more negative BIAS for longer turns, which is also most likely due to more early responses in long turns.

Broader implications for models of turn-taking

As indicated at the start of the Introduction, responding to a turn in conversation involves at least two processes in order to produce smooth turn transitions: planning the content of one's turn, and estimating the end of the interlocutor's turn. In the present study, we have focused on the second of these processes: turn-end estimation. Our results suggest that intonational phrasing offers relevant cues contributing to turn-end estimation, presumably along many other cues, such as lexico-syntactic completion, pragmatic completion, and perhaps even silence. Note that all these "completion" cues are relatively local cues, providing information about the extent to which the turn is "complete" at a certain moment in time. These cues could be contrasted with other cues allowing anticipatory

behavior of longer range, such as those offered by syntactic constituent structure at the sentence level and/or the previous conversational context. The present study does not show a very clear, overall effect of context, especially when acoustic information is also present, suggesting that long-range anticipation is less prevalent in turn-end estimation than more local anticipatory or reactive processes taking place toward the ends of turns. In contrast, the first process mentioned above, response planning, may depend much more on the available contextual information (within or before the turn), and may require some degree of long-range anticipation in many cases. The results from Corps et al. (2018) discussed in the Introduction support this idea, since they showed that better predictability led to faster, but not more accurate answers to questions. This indeed suggests that predictability speeds up response planning but does not necessarily facilitate turn-end estimation. It is possible that contextual information has a similar effect, such that it may help participants with response planning rather than with turn-end estimation. More research is needed to further elucidate these and other questions regarding the relevance of local and global contextual information for both response planning and turn-end estimation.

Conclusion

The present study showed a facilitatory effect of context on turn-end estimation for short turns when only lexico-syntactic information was available to participants (in a visual task), as well as a statistical trend of the same nature when turns were presented auditorily. We propose that the previous conversational context may provide useful information for turn-end estimation in some cases, but may often be redundant for this purpose when sufficient information is present in the turn itself, such as when turns are sufficiently long. We also observed that turn-medial positions are more likely to be considered a suitable turn end when an intonational phrase boundary is present than when it is not. This finding is consistent with previous claims that listeners use lexico-syntactic information *together with* prosodic information when estimating turn ends (Barthel et al., 2017; Bögels & Torreira, 2015; Casillas & Frank, 2017; Ford & Thompson, 1996; Lammertink et al., 2015).

Note

1. All annotations as well as other experimental materials and data can be obtained from the authors on request.

Acknowledgments

We would like to thank Ruben van den Bosch and Mart Lubbers for their assistance during the experiment. We are grateful to Stephen C. Levinson, Elliott Hoey, and the members of the INTERACT project at the Max Planck Institute for Psycholinguistics for extensive discussion of this work. We also thank Dan Dediu for useful comments about our statistical analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the ERC under ERC Advanced Grant 269484 INTERACT to Stephen C. Levinson.

ORCID

Sara Bögels http://orcid.org/0000-0002-4945-5765

References

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals." *Frontiers in Psychology*, *8*, 393. https://doi.org/10.3389/fpsyg.2017.00393

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*, 1858. https://doi.org/10.3389/fpsyg.2016.01858

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, *203*, 104347. https://doi.org/10.1016/j.cognition.2020.104347

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no . . . How the brain interprets the pregnant pause in conversation. *PloS One*, *10*(12), e0145474. https://doi.org/10.1371/journal.pone.0145474

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(1), 12881. https://doi.org/10.1038/srep12881

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, *92*, 234–253. https://doi.org/10.1016/j.jml.2016.06.013

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass*, *10*(3), 120–141. https://doi.org/10.1111/lnc3.12178

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

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. https://doi.org/10.1016/j.cognition.2018.01.015

Corps, R. E., Pickering, M. J., & Gambi, C. (2019). Predicting turn-ends in discourse context. *Language, Cognition and Neuroscience*, *34*(5), 615–627. https://doi.org/10.1080/23273798.2018.1552008

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*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In Ochs, E., Schegloff, E. A., Thompson, S.A. (Eds.), *Studies in Interactional Sociolinguistics 13: Interaction and Grammar* (pp. 134–184). Cambridge University Press.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PloS One*, *10*(3), e0120068. https://doi.org/10.1371/journal.pone.0120068

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, *25*(3), 601–634. https://doi.org/10.1016/j.csl.2010.10.003

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

Holzgrefe-Lang, J., Wellman, C., Petrone, C., Raling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience*, *31*(7), 904–920. https://doi.org/10.1080/23273798.2016.1157195

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jansen, S., Wesselmeier, H., de Ruiter, J. P., & Mueller, H. M. (2014). Using the readiness potential of button-press and verbal response within spoken language processing. *Journal of Neuroscience Methods*, *232*, 24–29. https://doi.org/10.1016/j.jneumeth.2014.04.030

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Proceedings of the Workshop on Spontaneous Speech: Data and Analysis* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289. https://doi.org/10.1080/0163853X.2014.955997

Labov, W., & Fanshel, D. (1977). *Psychotherapy as conversation*. Academic Press.

Ladd, R. D. (2008). *Intonational phonology*. Cambridge University Press.

Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology*, *6*, 495. https://doi.org/10.3389/fpsyg.2015.00495

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means*. R package (Version 1.4.8) [Computer software]. https://CRAN.R-project.org/package=emmeans

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731

Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, *42*(3), 255–280. https://doi.org/10.1017/S0025100312000187

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539. https://doi.org/10.1162/jocn_a_00673

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*, 211. https://doi.org/10.3389/fpsyg.2017.00211

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, *9*, 525. https://doi.org/10.3389/fpsyg.2018.00525

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*(1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Oostdijk, N. (2000). Het corpus gesproken Nederlands. *Nederlandse Taalkunde*, *5*(3), 280–284.

Riest, C., Jorschick, A. B., & De Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, *6*, 89. https://doi.org/10.3389/fpsyg.2015.00089

Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology*, *6*, 509. https://doi.org/10.3389/fpsyg.2015.00509

Rühlemann, C., & Gries, S. T. (2020). Speakers advance-project turn completion by slowing down: A multifactorial corpus analysis. *Journal of Phonetics*, *80*, 100976. https://doi.org/10.1016/j.wocn.2020.100976

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.1353/lan.1974.0010

Sereno, S. C., Hand, C. J., Shahid, A., Mackenzie, I. G., & Leuthold, H. (2020). Early EEG correlates of word frequency and contextual predictability in reading. *Language, Cognition and Neuroscience*, *35*(5), 625–640. https://doi.org/10.1080/23273798.2019.1580753

Sichlinger, L., Cibelli, E., Goldrick, M., & Mittal, V. A. (2019). Clinical correlates of aberrant conversational turn-taking in youth at clinical high-risk for psychosis. *Schizophrenia Research*, *204*, 419. https://doi.org/10.1016/j.schres.2018.08.009

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304–324. https://doi.org/10.1016/j.cognition.2014.10.008

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Frontiers in Psychology*, *6*, 284. https://doi.org/10.3389/fpsyg.2015.00284

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1658), 20130395. https://doi.org/10.1098/rstb.2013.0395

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech*, *41*(3–4), 265–294. https://doi.org/10.1177/002383099804100403

Wesselmeier, H., Jansen, S., & Müller, H. M. (2014). Influences of semantic and syntactic incongruence on readiness potential in turn-end anticipation. *Frontiers in Human Neuroscience*, *8*, 296. https://doi.org/10.3389/fnhum.2014.00296

Appendix

Table A1. Experiment 1: Summary table for Model 2 (discrete Turn Length): Model formula: Success ~ Context * poly(Length, 2) + (1 | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs: 2304; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−2	0.5	−4.03	**< .0001**
Context	1.49	0.33	4.46	**< .0001**
Length (5–8)	1.8	0.59	3.06	**< .005**
Length (9–12)	0.72	0.68	1.06	0.29
Length (13–21)	−0.42	0.74	−0.57	0.57
Context:Length (5–8)	−1.12	0.39	2.86	**< .005**
Context:Length (9–12)	−1.56	0.46	3.4	**< .001**
Context:Length (13–21)	−1.01	0.51	−1.97	**< .05**
Random effects:				
ICC: 0.53				
Grouping	Effect	SD		
Participant	Context	0.19		
Item	Intercept	1.87		
	Context	0.78		

Table A2. Experiment 2: Summary table for Model 4A (discrete Turn Duration; 250-ms success window): Model formula: Success ~ Context * Turn Duration + (1 + Context | Participant) + (1 | Item). Statistically-significant p-values in bold.

Number of obs: 2249; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−3.42	0.47	−7.32	**< .0001**
Context	0.64	0.36	1.78	0.07
Duration (1–2 s)	1.03	0.46	2.22	**< .05**
Duration (2–3 s)	1.26	0.54	2.34	**< .05**
Duration (3–7 s)	0.06	0.69	0.08	.93
Context:Duration (1–2 s)	−0.62	0.37	−1.67	0.09
Context:Duration (2–3 s)	−1.34	0.43	−3.1	**< .005**
Context:Duration (3–7 s)	−0.9	0.58	−1.55	0.12
Random effects:				
ICC: 0.51				
Grouping	Effect	SD		
Participant	Intercept	1.09		
	Context	0.63		
Item	Context	1.31		

Table A3. Experiment 2: Summary table for Model 4B (discrete Turn Duration; 500-ms success window): Model formula: Success ~ Context * Turn Duration + (1 + Context | Participant) + (1 + Context | Item). Statistically-significant p-values in bold.

Number of obs: 2249; Items: 96; Participants: 24				
Fixed effects:				
Effect	Estimate	Std Error	*z* value	Pr(≥|z|)
Intercept	−1.32	0.39	−3.39	**< .0001**
Context	0.46	0.32	1.44	0.15
Duration (1–2 s)	1.19	0.35	3.37	**< .001**
Duration (2–3 s)	1.26	0.43	2.96	**< .005**
Duration (3–7 s)	0.32	0.51	−0.6	0.53
Context:Duration (1–2 s)	−0.35	0.33	−1.05	0.29
Context:Duration (2–3 s)	−0.89	0.4	−2.22	**< .05**
Context:Duration (3–7 s)	−0.41	0.48	−0.85	0.4
Random effects:				
ICC: 0.47				
Grouping	Effect	SD		
Participant	Intercept	1.22		
	Context	0.72		
Item	Intercept	1.11		
	Context	0.69		