SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 202–203.
Los Angeles, California, 21–23 September 2011.

Timing in turn-taking: Children's responses to their parents' questions

Marisa Tice
Stanford University
Margaret Jacks Hall
Stanford, CA 94305-2150
middyp@stanford.edu

Susan C. Bobb
University of Göttingen
Goßlerstraße 14
14D-37073 Göttingen, Germany
sbobb@gwdg.de

Eve V. Clark
Stanford University
Margaret Jacks Hall
Stanford, CA 94305-2150
eclark@stanford.edu

Abstract

In this study we track the development of timing in children's answers to their parents' questions. We find that over the ages of 1;8 to 3;4, children's response timing decreases, converging to adult norms. Overall, their responses are faster to simpler questions (e.g. yes/no questions vs. wh-questions) and when the answer includes information that was stated in the preceding two utterances. Parents, on the other hand, remain relatively stable over this period, showing similar response times to all types of questions their children ask them.

1 Introduction

When adults converse, they observe a convention of 'one speaker at a time' (Sacks et al, 1974; Stivers et al. 2009), and when one speaker's turn ends and another's begins, the transition time is minimized with little resulting overlap in speech (e.g., Levinson 1983).
By contrast, young children are chronically late in turn-taking. This is particularly apparent in triadic conversations where two-year-olds often come in up to two turns too late (e.g., Dunn & Shatz 1989). We hypothesized that it requires considerable practice to retrieve the words and structures needed in planning an appropriate turn, and that children should therefore become faster with age until they match adult timing. We also expected that in responding to questions, children would be able to respond more quickly to simple questions (yes/no) than to more complex ones (wh-).

2 Methods

We analyzed patterns of turn-taking in the recordings of five mother-child pairs from the Providence corpus of the CHILDES database (MacWhinney 2000; Demuth et al., 2006). The five children and their parents were filmed and audio recorded approximately twice per month while performing their daily activities at home from the ages of one to three years of age. Sampling at six evenly-spaced time periods from 1;8 to 3;4, we extracted from the recordings the first 15 questions asked by the child and answered by the mother, and the first 15 questions asked by the mother and answered by the child. Our data constitute a total of 180 question-answer (Q-A) pairs per mother-child group, with 900 Q-A pairs overall.
The duration of silence (or overlap) between the end of the question and the onset of the response was measured from the audio recording using Praat acoustic analysis software (Boersma & Weenink, 2011). Measurements were made by a phonetically trained undergraduate naïve to the purpose of the study.
Each Q-A pair was coded for a range of properties hypothesized to affect response timing, including question length (in clauses and morphemes), question familiarity (has it already been stated, part of a sequence, or part of a routine?), and question complexity (question type, e.g. yes-no, X or Y, wh-, etc.)

3 Results

Our results show that adults' response timing to child questions remains consistent regardless of the child's age[1], while children gradually reduce in the time it takes them to respond to questions (See Figures 1a and 1b). By 3;4, children approach adult Q-A response timing, but their timing is not uniform: they took longer to answer more complex questions, so were slower replying to wh- questions than to yes/no questions (*p*<.05). Within wh-questions, they were slower to answer *who* questions than *what*/*where* (Figures 2a-b). This is consistent with children's order of acquisition for wh-question words (Ervin-Tripp, 1979). These results were confirmed using a linear mixed model for gap duration, with child's age, question type (wh-, yes/no, X or Y), and informational overlap of the answer with the preceding two utterances as fixed effects, and the child as random effect. Both question type and informational overlap in the preceding two utterances were found to be significant predictors of gap duration (t=5.062 and -2.15, respectively), with age coming out marginally significant as well (t=-1.891). This indicates that with age, children's gap durations in responding to their parents' questions shrank, but that the duration of their response was significantly affected overall by the complexity of the question type—wh-

[1] If anything, the adults' response times are slightly increasing with time, averaging above norms for adult-adult conversation (Stivers et al., 2009).

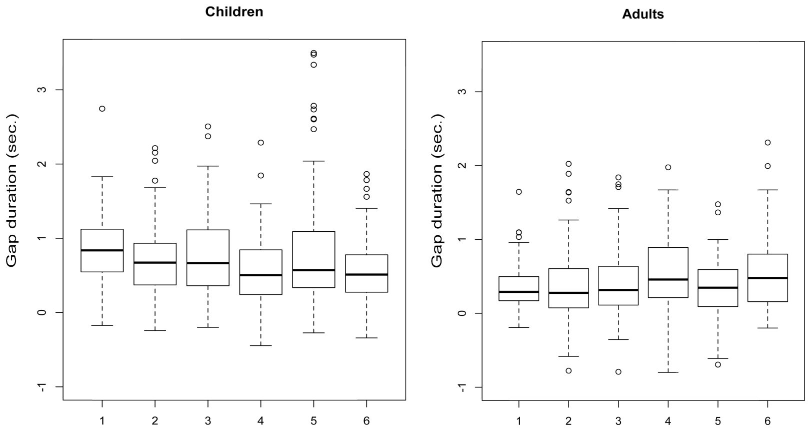

Figure 1: Average gap durations of (a) children and (b) adults at each of the evenly-spaced six time slices from 1;8 to 3;4. Children's responses over time begin to decrease, while adults stay stable, possibly with a slight increase over time.

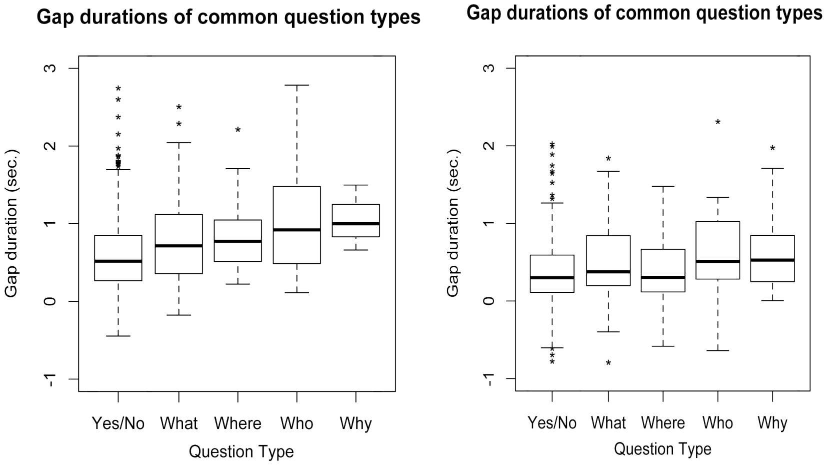

Figure 2: Average gap durations of (a) children and (b) adults for the common question types in our data (tokens >20). Children's responses show differentiation between yes/no and several different wh-question types, while adults do not. When yes/no and wh- types are compared as groups, the difference in timing is significant for children but not for adults.

questions take longer to respond to than yes/no questions—and the informational overlap of the answer—it takes longer to respond when the child has to come up with all new material.

We hypothesize that these findings support the view that children’s ability to take turns on time is largely determined by their ability to retrieve the right words for the information that they wish to convey as they plan an utterance for the next turn.

In yes/no questions, and answers in which the relevant information has been recently stated, response access needs are minimized, but in wh-questions, children must find the relevant information outside the actual question itself. Moreover, different wh-forms call for different kinds of information, e.g., what—category label, where—place label, who—person label, etc. Some kinds of answers appear easier to access than others, resulting in variable response timing by question complexity.

References

Boersma, Paul & Weenink, David (2011). Praat: doing phonetics by computer. Retrieved 3 August 2011 from http://www.praat.org/

Demuth, K., Culbertson, J. & Alter, J. 2006. Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech*, 49, 137-174.

Dunn, J. & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, 60.

Ervin-Tripp, S.M. (1979). “Children's verbal turn-taking”. In *Developmental Pragmatics*. NY:Academic.

Levinson, S. (1983). *Pragmatics*. CUP.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. *Language*, 50.

Stivers, T., Enfield, N., Brown, P., Englert, C., Hayashi, M., Heinemann, T. (2009). Universals and cultural variation in turn taking in conversation. *PNAS*, 106.

SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 204–205.
Los Angeles, California, 21–23 September 2011.

The eye gaze of 3rd party observers reflects turn-end boundary projection

Marisa Tice
Dept. of Linguistics, Stanford University
Margaret Jacks Hall
Stanford, CA 94305-2150
middyp@stanford.edu

Tania Henetz
Dept. of Psychology, Stanford University
Jordan Hall
Stanford, CA 94305-2150
thenetz@stanford.edu

Abstract

We show that when observers watch a dialogue, their eye gaze is a viable measure of online turn processing. Third-party listeners not only track the current speaker with their gaze, but they look anticipatorily to the next speaker during question-answer pairs. Eye gaze is a measure of turn-boundary projection that has all the benefits of previous measures, but does not require the participant to make explicit judgments, and so provides a natural alternative for exploring turn-end boundary cues.

1 Introduction

Speakers in conversation take turns with remarkably little delay or overlap (Sacks et al., 1974; Stivers et al, 2009). To accomplish this, potential next speakers must comprehend the present utterance while simultaneously planning a contribution and projecting when the current turn will end. There are a number of candidate cues to turn-completion including pragmatic, prosodic, or lexicosyntactic cues (e.g., Ford and Thompson, 1996; de Ruiter et al., 2006), but little is known about the role of these cues in online turn projection. We attempt to investigate this practice by employing a continuous measure of online processing: gaze tracking.

In a recent study, de Ruiter et al. (2006) addressed turn-end boundary projection experimentally using a non-continuous response measure. They asked Dutch speakers to listen to spontaneous speech fragments and press a button at the moment they anticipated the speaker would finish her utterance. The speech fragments were phonetically manipulated to investigate projection cues such as intonation, lexicosyntax, and rhythm. Their results suggest that speakers rely primarily on lexicosyntax to identify upcoming turn-end boundaries.

But the speech signal is continuously unfolding so listeners' use of particular types of cues may change over the course of an utterance. Eye gaze provides a continuous measure of projection that could detect these potential changes. Since the stimuli for gaze measures can be manipulated in the same ways as the stimuli used by de Ruiter et al. (2006), tracking observer gaze may provide a natural, passive, and continuous method for exploring how interlocutors manage the timing of turns.

To establish observer gaze as a measure of turn-end projection, we show that observers
(1) track current speakers with their gaze, and
(2) look anticipatorily to next speakers.

2 Methods

Thirty-two volunteers (*females* = 17) watched two short "split-screen" dialogues from a recent motion picture (*Mean Girls*, Paramount Pictures, 2004) while we recorded their eye movements[1].

Participants watched the clips *with* or *without* sound (N=16 each). Participants in the *without* sound condition were warned that they would not hear sound while the clips were playing.

We report data from the first film clip. The dialogue's five question-answer (Q-A) pairs were selected for analysis because Q-A pairs are reliable as adjacency pairs and provide a linguistically diverse sample of turns. Each participant's gaze was coded for gaze direction (right, left, center, blink) every 50ms by two coders: one of the authors and one trained coder naïve to our hypotheses (96% agreement).

3 Results

Observers in the sound condition consistently tracked the current speaker with their gaze: over 70% of looks were directed at the current speaker (Speaker 1=72.6%, Speaker 2=77.5%). Without sound, under 50% of looks were to the current speaker (Speaker 1 = 42.2%, Speaker 2=

[1] Two-thirds of participants in each condition reported having seen the film. These participants were less likely to look at the main character overall, but reliably tracked the current speaker.

41.9%). This was confirmed using generalized linear mixed effects models of gaze direction (looking at Speaker 1/not looking at Speaker 1) with current speaker, condition, and their interaction as fixed effects, and subject and turn as crossed random effects. For Speaker 1, there was a significant interaction such that the likelihood of looking at her during her turn differed across conditions (β=2.75, Z=2.5 p=.01). Looks to Speaker 1 increased during her turns in the condition with sound (β= 2.55, Z=5.5, p<.001), but not without sound (p=.9). Results for Speaker 2 were similar: there was a marginal speaker by condition interaction (β=2.49, Z=1.84, p=.06), and a significant effect of current speaker with sound (β = 2.5, Z=3.47, p <.001), but not without sound (p =.9).

Observers tended to shift their gaze from current to next speaker during the inter-turn gap. Figure 1 shows the average gaze trajectories from current to next speaker for each condition.

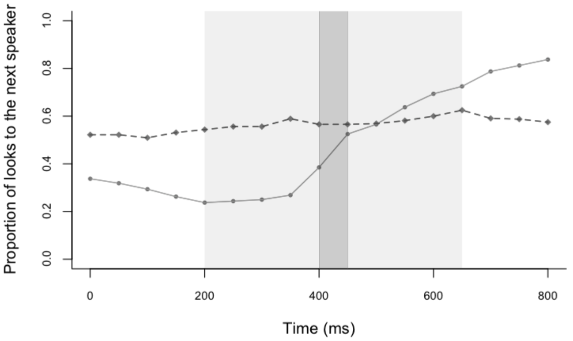

Figure 1: Average gaze trajectory across Q-A pairs with (solid) and without (dashed) sound. The dark shaded region represents the average inter-turn gap and the light shaded regions represent the 200ms before and after the gap.

To assess whether observers anticipate turn-transitions, we compared the proportion of looks to the next speaker in the 200ms surrounding the inter-turn gap. Since eye movements must be planned at least 200ms in advance, an increase in looks to the next speaker during this time would indicate that observers are looking to the next speaker *before* she speaks.

We used linear mixed models to predict gaze direction (current/next speaker), with position (pre-gap/post gap) and condition as fixed effects, and subject and Q-A pair as crossed random effects. There was a significant interaction between position and condition such that the increase in looks to the next speaker across the inter-turn gap was greater for the sound than the without sound condition (β=1.83, Z=3.49, p<.001). This increase in looks was significant only for the sound condition, showing anticipation (β = 2.7, Z = 2.76, p = .006).

4 Discussion

Previous methods for measuring anticipatory turn behavior were unable to track continuous changes in boundary projection and required explicit judgments that are not a part of typical turn-taking. Observer gaze has all the benefits of these methods, but is a passive task that collects continuous, online data.

Here we show that observers not only gaze at the current speaker, but they often look anticipatorily to the next speaker, especially when sound is available. This suggests that gaze in our task is primarily driven by linguistic information.

We are now extending this method to dialogues where the audio is phonetically manipulated to control the linguistic cues that are available (similar to de Ruiter et al., 2006) using spontaneous dialogues from the Meet a Friend corpus (Tice & Henetz, 2011). We are also replicating the current study with still images accompanying the dialogue instead of film. In the future, this method will lend itself well to examining turn processing in an understudied population: children. We expect that observer gaze will provide opportunities for many studies of turn processing that would otherwise not be possible without this natural, continuous measure.

References

Ford, C. & Thompson, S. (1996). “Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn-completion.” In *Interaction and Grammar*. CUP.

de Ruiter, J., Mitterer, H., & Enfield, N. (2006). Projecting the end of a speaker’s turn: A cognitive cornerstone of conversation. Language, 82.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. Language, 50.

Stivers, T., Enfield, N., Brown, P., Englert, C., Hayashi, M., Heinemann, T. (2009). Universals and cultural variation in turn taking in conversation. PNAS, 106.

Tice, M. & Henetz, T. (2011). The Meet a Friend Spontaneous Speech Corpus. Accessed at www.stanford.edu/~middyp/Meet-a-Friend