Journal of Memory and Language 92 (2017) 234–253

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The development of children's ability to track and predict turn structure in conversation

Marisa Casillas [a,*], Michael C. Frank [b]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Department of Psychology, Stanford University, United States

ARTICLE INFO

Article history:
Received 3 October 2014
revision received 27 June 2016

Keywords:
Turn taking
Conversation
Development
Questions
Eye-tracking
Anticipation

ABSTRACT

Children begin developing turn-taking skills in infancy but take several years to fluidly integrate their growing knowledge of language into their turn-taking behavior. In two eye-tracking experiments, we measured children's anticipatory gaze to upcoming responders while controlling linguistic cues to turn structure. In Experiment 1, we showed English and non-English conversations to English-speaking adults and children. In Experiment 2, we phonetically controlled lexicosyntactic and prosodic cues in English-only speech. Children spontaneously made anticipatory gaze switches by age two and continued improving through age six. In both experiments, children and adults made more anticipatory switches after hearing questions. Consistent with prior findings on adult turn prediction, prosodic information alone did not increase children's anticipatory gaze shifts. But, unlike prior work with adults, lexical information alone was not sufficient either—children's performance was best overall with lexicosyntax and prosody together. Our findings support an account in which turn tracking and turn prediction emerge in infancy and then gradually become integrated with children's online linguistic processing.

Introduction

Spontaneous conversation is a universal context for using and learning language. Like other types of human interaction, it is organized at its core by the roles and goals of its participants. But what sets conversation apart is its structure: sequences of interconnected, communicative actions that take place across alternating turns at talk. Sequential, turn-based structures in conversation are strikingly uniform across language communities and linguistic modalities. Turn-taking behaviors are also cross-culturally consistent in their basic features and the details of their implementation (De Vos, Torreira, & Levinson, 2015; Dingemanse, Torreira, & Enfield, 2013; Stivers et al., 2009).

Children participate in sequential coordination (proto-turn taking) with their caregivers starting at three months of age—before they can rely on any linguistic cues (see, among others, Bateson, 1975; Hilbrink, Gattis, & Levinson, 2015; Jaffe et al., 2001; Snow, 1977). However, infant turn taking is different from adult turn taking in several ways: it is heavily scaffolded by caregivers, has different inter-turn timing, and lacks semantic content (Hilbrink et al., 2015; Jaffe et al., 2001). But children's early, turn-structured social interactions are presumably a critical precursor to their later conversational turn taking, establishing the protocol by which children come to use language with others. How then do children integrate linguistic knowledge with these preverbal turn-taking abilities?

In this study, we investigate when children begin to make predictions about upcoming turn structure in

* Corresponding author at: Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
E-mail addresses: marisa.casillas@mpi.nl (M. Casillas), mcfrank@stanford.edu (M.C. Frank).

http://dx.doi.org/10.1016/j.jml.2016.06.013
0749-596X/

conversation and how online linguistic processing becomes integrated into their predictions as they grow older. We first give a basic review of turn-taking research and the state of current knowledge about adult turn prediction. We then discuss recent work on the development of turn-taking skills before presenting the details of the present study.

Adult turn taking

Turn taking itself is not unique to conversation. Many other human activities are organized around sequential turns at action. Traffic intersections and computer network communication both use turn-taking systems. Children's early games (e.g., give-and-take, peek-a-boo) have built-in, predictable turn structure (Ratner & Bruner, 1978; Ross & Lollis, 1987). Even monkeys take turns: Non-human primates such as marmosets and Campbell's monkeys vocalize contingently with each other in both natural and lab-controlled environments (Lemasson et al., 2011; Takahashi, Narayanan, & Ghazanfar, 2013). In all these cases, turn taking serves as a protocol for interaction, allowing the participants to coordinate with each other through sequences of contingent action.

Conversational turn taking distinguishes itself from other turn-taking behaviors by the complexity of the sequencing involved. Conversational turns come grouped into semantically-contingent sequences of action. The groups can span turn-by-turn exchanges (e.g., simple question–response, "How are you?"–"Fine.") or sequence-by-sequence exchanges (e.g., reciprocals, "How are you?"–"Fine, and you?"–"Great!"). Compared to other turn-taking behaviors, the possible sequence and action types in everyday talk are diverse and unpredictable.

Despite this complexity, conversational turn taking is precise in its timing. Across a diverse sample of conversations in 10 languages, one study found a consistent average inter-turn silence of 0–200 ms at points of speaker switch (Stivers et al., 2009). Experimental results and current models of speech production suggest that it takes approximately 600 ms to produce a content word, and even longer to produce a simple utterance (Griffin & Bock, 2000; Levelt, 1989). In order to achieve 200 ms turn transitions, speakers must begin formulating their response before the prior turn has ended (Levinson, 2013; Levinson, 2016). Moreover, to formulate their response early on, speakers must track and anticipate what types of response might become relevant next. They also need to predict the content and form of upcoming speech so that they can launch their articulation at exactly the right moment. Prediction thus plays a key role in timely turn taking.

Adults have a lot of information at their disposal to help make accurate predictions. Lexical, syntactic, and prosodic information (e.g., *wh*-words, subject-auxiliary inversion, and list intonation) can all inform addressees about upcoming linguistic structure (De Ruiter, Mitterer, & Enfield, 2006; Duncan, 1972; Ford & Thompson, 1996; Bögels & Torreira, 2015). Non-verbal cues (e.g., gaze, posture, and pointing) often appear at turn-boundaries and can sometimes act as late indicators of an upcoming speaker switch (Rossano, Brown, & Levinson, 2009; Stivers & Rossano, 2010). Additionally, the sequential context of a turn can make the next action obvious: answers after questions, thanks or denial after compliments, etc. (Schegloff, 2007).

Prior work suggests that adult listeners primarily use lexicosyntactic information to accurately predict upcoming turn structure. De Ruiter et al. (2006) asked participants to listen to snippets of spontaneous conversation and to press a button whenever they anticipated that the current speaker was about to finish his or her turn. The speech snippets were controlled for the amount of linguistic information present; some were normal, but others had flattened pitch, low-pass filtered speech, or further manipulations. With pitch-flattened speech, the timing of participants' button responses was comparable to their timing with the full linguistic signal. But when no lexical information was available, participants responded significantly earlier within the turn. The authors concluded that lexicosyntactic information[1] was necessary and possibly sufficient for turn-end projection, while intonation was neither necessary nor sufficient. Congruent evidence comes from studies varying the predictability of lexicosyntactic and pragmatic content: adults anticipate turn ends better when they can more accurately predict the exact words that will come next (Magyari & De Ruiter, 2012; see also Magyari, Bastiaansen, De Ruiter, & Levinson, 2014). They can also identify speech acts within the first word of an utterance (Gísladóttir, Chwilla, & Levinson, 2015), allowing them to start planning their response at the first moment possible (Bögels, Magyari, & Levinson, 2015).

Despite this body of evidence, the role of prosody for adult turn prediction is still a matter of debate. De Ruiter et al.'s (2006) experiment focused on the role of intonation, which is only a partial index of prosody. Prosody is tied closely to the syntax of an utterance, so the two linguistic signals are difficult to control independently (Ford & Thompson, 1996). Bögels and Torreira (2015) used a combination of button-press and verbal responses to investigate the relationship between lexicosyntactic and prosodic cues in turn-end prediction. Critically, their stimuli were cross-spliced so that each item had full prosodic cues to accompany the lexicosyntax. Because of the splicing, they were able to create items that had syntactically-complete units with no intonational phrase boundary at the end. Participants never verbally responded or pressed the "turn-end" button when hearing a syntactically-complete phrase without an intonational phrase boundary. And when intonational phrase boundaries were embedded within multi-utterance turns, participants were tricked into pressing the "turn-end" button 29% of the time. These findings suggest that listeners actually do rely on prosodic cues to execute a response, and that their use of prosodic cues interacts with their predictions about the unfolding syntactic structure (see also De Ruiter et al., 2006, 525). These experimental findings corroborate other corpus

[1] The "lexicosyntactic" condition only included flattened pitch and so was not exclusively lexicosyntactic—the speech would still have residual prosodic structure, including syllable duration and intensity.