Journal of Experimental Child Psychology 261 (2026) 106358

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

From age two, children use pronouns to predict who will speak next in conversation

Imme Lammertink [a,b,*], Maartje de Vries [c], Caroline Rowland [b,d], Marisa Casillas [d,e]

[a] Centre for Language Studies Radboud University, the Netherlands
[b] Donders Institute for Brain, Cognition, and Behaviour, the Netherlands
[c] Groningen University Medical Centre, the Netherlands
[d] MPI for Psycholinguistics, the Netherlands
[e] University of Chicago, United States

ARTICLE INFO

Keywords:
Interactional development
Lexical cues
Prosodic cues
Child language development
Conversational structure
Anticipatory gaze

ABSTRACT

Children's exposure to language is shaped by interactional needs in conversation. Prior research has largely focused on how such language influences word learning and long-term lexical knowledge. However, the effects of interactional language are likely to go far beyond word learning. Earlier studies showed that in conversation, from around age 2;6 children who are watching a conversation are more likely to spontaneously switch their gaze to an upcoming responder when they hear a question compared to when they hear a non-question. However, what information drives these predictions remains unclear. Tracking the eye gaze behavior of Dutch children (1–4-year-olds) and adults, the purpose of this research was to examine whether participant's predictions are driven by individually informative linguistic cues, comparing two cues associated with interrogatives: one lexical cue (subject pronoun) and one canonically associated prosodic cue (utterance-final intonation). We find that from age children 2;0 make more and earlier anticipations of an upcoming addressee response when hearing the early lexical cue (you vs. I subject pronouns), but we have no evidence that their predictions are changed by the later prosodic cue. Further, we investigated how cue use depends on linguistic context by comparing semantically meaningful (Study 1) and non-meaningful (Study 2) context. Only in meaningful contexts did participants show a pronoun advantage in predicting conversational structure. This suggests that using these cues relies on broader linguistic context. The findings take us a step closer to understanding how linguistic and interactional skills become intertwined in development.

Introduction

Young children gain access to language during spontaneous interactions involving the people around them. Children's exposure to language is, therefore, shaped by the needs of interactants in conversation: establishing joint attention, making and fulfilling (or denying) requests, sharing objects and events of interest, and more (see Clark, 2022 for a recent overview). Such interactional language, whether it involves the child or is simply a part of their ambient language environment, gives critical information regarding the pragmatic, grammatical, lexical, phonological, and distributional features of their home language(s) (e.g., Clark, 2007; Cristia, 2020;

* Corresponding author.
E-mail address: imme.lammertink@ru.nl (I. Lammertink).

https://doi.org/10.1016/j.jecp.2025.106358
Received 18 April 2025; Received in revised form 25 August 2025;
Available online 15 September 2025

Donnelly & Kidd, 2021; Estes et al., 2007; Goldstein & Schwade, 2008; Knightly et al., 2003; Rowe & Snow, 2020).

Interactional language—especially when it directly engages the child—has been argued to benefit language development, particularly vocabulary development (e.g., Anderson et al., 2021; Endevelt-Shapira et al., 2024; Rowe, 2012; Tamis-LeMonda et al., 2014). After all, multimodal caregiver-child coordination helpfully constrains speaker meaning, both in a single here-and-now moment and across the course of repeated, ordinary home routines (e.g., Clark & Estigarribia, 2011; Rowe & Goldin-Meadow, 2009; Roy et al., 2015; Suarez-Rivera et al., 2022; Tamis-LeMonda et al., 2019; Yu & Smith, 2013). Much past work on (directed) interactional language focuses on how it influences children's fleeting comprehension of word meanings and, over time, their long-term lexical knowledge. However, the effects of interactional language likely go far beyond word learning.

One advantage of interactional language is that its contingent, predictable, and repetitive format affords children the opportunity to track, anticipate, and contribute to interaction without requiring full comprehension (Bruner, 1985; Ferreira et al., 2002; Ratner & Bruner, 1978; Shatz, 1978). Consider the following sequence between a US English-acquiring child (1;6.7) and their mother:

```
MOT: What babe?
CHI: [unintelligible].
MOT: Can you- can you explain what you mean?
CHI: Me- please?
MOT: Please what?
CHI: Please (.) down (gestures to the next seat over).
MOT: You want me to sit down?
CHI: Down.
MOT: Okay, I'm coming.
```

The child recognizes and responds appropriately to the series of clarification questions, even though it is unlikely that they parsed and understood every word their mother said (and vice versa, Meylan et al., 2023). Doing so requires the child to repeatedly recognize that a question is being posed, that it requires a specific type of response, and perhaps even that some response types are preferred over others, e.g., confirmation over disconfirmation (Pomerantz & Heritage, 2013; Robinson, 2020; Sacks, 1973/1987). Adults display these same behaviors in everyday conversation, but we would typically expect adults to fully (or almost fully) parse and understand the content of each utterance. We cannot hold the same expectation for young children. However, young children still manage to engage. This ability to meaningfully participate in interaction despite asymmetries in comprehension may ultimately help children to scaffold their own linguistic and social development. The present work examines how this process of conversational tracking develops alongside language.

Keeping track of conversation

The integral role of contingent responding—responding immediately and appropriately—in everyday conversation places a high priority on quickly determining whether a response is needed and, if so, what type of response to give (Bögels et al., 2015; Levinson & Torreira, 2015; Shatz, 1978). In human conversations, there exist a range of methods for eliciting contingent responses: questions bring about answers, complaints bring about accounts, imperatives bring about action responses, and so on (Schegloff, 2007; Stivers, 2013). In both adult-adult and (Western) adult-infant conversation, these methods are employed frequently (Snow, 1977; Stivers & Rossano, 2010) and occur at predictable moments (Bruner, 1975; Nelson, 1985; Ratner & Bruner, 1978). As we argue below, among all the types of stereotyped prompts and responses in interaction, request-response pairs (e.g., question–answer) may play an especially pronounced role in early development. Across languages, questions are often overtly linguistically marked (e.g., question-markers, such as "where"; Ultan, 1978; Siemund, 2001, but see Estigarribia (2010), and are associated with stereotyped response types (e.g., a locative response after a "where" question; Clark, 2006; Clark & Lindsey, 2015; Schegloff, 2007; Stivers, 2013). From a very young age, children already appear to be aware of the social imperative that questions (more than non-questions) require an answer.

Support for this idea comes from the finding that, from around age 2;6, children who are watching a conversation are far more likely to spontaneously switch their gaze to a potential upcoming responder when they hear a question compared to when they hear a non-question (see, e.g., Casillas & Frank, 2017; Lammertink et al. (2015); see also Keitel et al., 2013, Keitel et al., 2015, and Kalashnikova & Kember, 2020). This effect arises despite the fact that we can, on average, expect questions to be syntactically more complex than non-questions (via use of question words, inversion, etc.), at least in the languages tested so far (English, German, Dutch). Instead, the predictive advantage appears to derive from the interactional implications of question-formatted utterances: the viewer can expect that one person will now cease talking and, shortly thereafter, another person will begin talking.

It remains an open question how the above described question-advantage in children's gaze behavior while observing videos of conversation translate to natural first-party conversations, where children may play a more active role. On the one hand, one may expect the question-advantage to be smaller in natural first-party conversation because in these contexts, children may have less cognitive resources available to anticipate a turn end given that they also need to formulate and articulate a response. Indeed, Casillas et al. (2016) observed a positive relation between the linguistic complexity of a child's answer and the time it takes a child to answer: the more complex the answer, the longer it takes a child to respond. Relatedly, the conversations that children watched in the experiments mentioned above display clear, well scripted conversations with relatively long pauses in between utterances. All these factors likely facilitate the anticipation of turn ends whereas natural conversations are messier (Holler & Kendrick, 2015). On the other hand, one could also argue the question-advantage to be larger in natural conversation because of the children's increased degree of engagement (Holler & Kendrick, 2015). The fact that children make predictions about possible turn transitions, even as mere observers of interaction, however, attests to a high priority for response-seeking behaviors in interactional contexts (see also Fitneva, 2012;

Stivers et al., 2018).

It is yet unclear which cues children rely on to make real-time predictions about upcoming responses. With respect to non-linguistic cues, observational evidence from adult spontaneous conversation demonstrates that speaker gaze and interactional gestures are associated with greater pressure to respond, and may therefore also serve as important predictive cues for upcoming response (e.g., Kendrick et al., 2023; Latif et al., 2017; Stivers & Rossano, 2010). However, the use of these same cues by children is less well attested (Keitel et al., 2015; Stivers et al., 2018) and they cannot explain the benefit for questions in prior work because those stimuli featured no character gaze switches or gestures (Casillas & Frank, 2017; Lammertink et al., 2015).

With respect to linguistic cues, the evidence to-date suggests that, for adults, lexical access paves the way to accurate turn-end prediction, with prosodic cues serving as a critical final (dis)confirmatory cue (Bögels & Torreira, 2015, 2021; de Ruiter et al., 2006). For children, Casillas and Frank (2017) showed children videos in languages they did not understand (ages 3;0–5;11) as well as videos with acoustic manipulations that removed access to prosodic, lexical, or (all) verbal information (1;0–6;11). While a clear benefit for questions emerged, there was no advantage for prosodic over lexical information or vice versa in making anticipatory gaze switches to the addressee (see also Keitel et al., 2013). Crucially, most prior work has investigated the use of individual linguistic cues by limiting participant's access to entire channels of linguistic information (e.g., flattening all intonational contours) rather than manipulating individual cues within a channel (e.g., adding a high-rise terminal intonational cue). Complete removal of an entire channel, however, is not very representative of what children (or adults) actually perceive and leads to unnatural sounding speech (Casillas & Frank, 2017). A significant limitation of this channel-based approach is thus that it prevents insight into the specific linguistic cues that children (and adults) might use to identify questions during real-time interaction. It may well be that a few individual cues to an upcoming response (e.g., subject-verb inversion) are reliable enough to allow observers to predict a switch in talkers even *without* full linguistic comprehension. Taking a step in this direction, Lammertink and colleagues (2015) pitted incomplete syntactic and prosodic phrases against each other by using a splicing technique similar to that of Bögels and Torreira (2015), with English- and Dutch-acquiring children aged 2;6. They found strong evidence that complete syntactic phrasing (in this case, with interrogative syntax) supports children's spontaneous next-turn predictions. Prosodically complete (vs. incomplete) utterances showed a minor additional advantage in prediction. However, because the phrasally complete stimuli were always interrogative and the phrasally incomplete stimuli were never interrogative, the findings cannot speak to what specifically aids children's predictions—the general advantage for questions (as in Casillas & Frank, 2017) or the individual syntactic and prosodic cues used.

The current study

In the present study, we therefore explored whether an individual, potentially interactionally informative linguistic cue—the second-person subject pronoun (e.g., "you") —influences children's and adult's predictions about the likelihood of an upcoming response. We tested this prediction in Dutch, a Germanic Indo-European language. Our overarching hypothesis was that children who are not yet linguistically sophisticated enough to fully parse incoming speech may rely on individual cues to successfully predict upcoming turn structure. For example, in the absence of complete sentence comprehension, children may reasonably guess what is about to happen once they hear a question word (e.g., "Where…"), a rising terminal intonation contour, a verb-first sentence ("Is…"), or any epistemic cue to questionhood. This last cue type (epistemic) is the focus of the present study.

Epistemics in conversation relates to the distribution of knowledge among interactants; someone who knows less about a topic than their fellow interactant(s) is more likely to ask questions than make assertions about that topic. Thus, who knows what—and how that is understood between interactants—is key to establishing likely roles as "questioner" and "responder" as people navigate different topics in conversation. Following this rationale, we argue that utterances with "you" as the subject are typically epistemically tilted toward the recipient because you tend to know more about "you" than "I" do (see Discussion for alternative explanations). We thus expect to observe such general epistemic asymmetry between the second-person pronoun "you" subject and first-person pronoun "I" subject across all utterance types. However, an association between "you" subjects and an immediate response by the addressee is likely to be most prominent in questions. While questions often function to reduce epistemic asymmetry, non-question sequences serve a much wider array of interactional functions, many of which do not seek response at all. Following this, we predicted a positive statistical association between "you" and interrogatives in conversational speech and verified this basic pattern in a small corpus study of Dutch subject-verb inversions in child-directed speech [see Supplementary Materials S4 for details].

Consequently, we argue that "you"-subject utterances are typically more likely than "I"-subject utterances to receive an immediate, contingent response (e.g., "Are you/Am I sleepy?"). Importantly, and as explained above, this natural alignment between person, epistemic asymmetry, and questionhood does not obligate all utterances with "you"-subjects to be interpreted as questions, nor utterances with "I"-subjects to be interpreted as statements. As such, we do not claim that the second-person pronoun is an epistemic marker in and of itself, nor the only marker of epistemic asymmetry that would drive contingent response. On the contrary, a definite interpretation of speaker meaning and relevant response requires access to the full semantic and sentential context of an utterance. This semantic and sentential context, however, may not be accessible to young children, because of their limited language knowledge. Therefore, we hypothesize that in cases where the semantic and sentential context is not accessible, individual epistemically tilted cues like subject pronoun ("you" vs. "I") may help to predict upcoming responses. Of course, use of this cue would be based on a heuristic – distributional association between the use of a subject pronoun ("you") and questionhood – rather than a rule. Whether such heuristic is strong and reliably enough to pick up on our hypothesized epistemic effect will be further investigated in the present study.

In doing so, we investigate whether children make more anticipatory looks to an addressee (i.e., a potential responder) following turns with a second-person subject pronoun (Dutch: *jij* /jɛi/) compared to turns with a first-person subject pronoun (*ik* /ɪk/). Second-person subject pronouns in Dutch interrogatives present a challenging test case for the idea that children track and react to individual cues to upcoming responses because they are prosodically and phonetically subtle (i.e., brief and non-utterance-initial, as in *ben jij slaperig?* [are you sleepy?]), and because their association with interrogatives is not categorical (i.e., not all utterances with the second-person pronoun are interrogative). We therefore expected this cue to provide a stringent test of the idea that children track simple cues to reliably predict turn structure. To the best of our knowledge, the use of epistemic cues—and specifically the second-person subject pronoun—have not yet been studied as real-time cues for predicting upcoming response. Therefore, in the present experiment we tested both children and adults, with the aim of examining whether either group makes use of this cue in their prediction of upcoming responses. Given the heuristic nature of the cue, we expect that adults – differently from the children – would not rely as much on the second-person pronoun as an individual, interactionally informative cue to questionhood. That is, in adults contextual information may serve a critical role in confirming the information-seeking function of the utterance, and in confirming that the addressee is indeed more of an epistemic authority on the topic at hand when the speaker uses a second-person subject pronoun. In comparison, because of their limited linguistic knowledge, young children may not have access to this contextual information and may thus be more inclined to use the heuristic, epistemically tilted cue of subject pronoun in their anticipation of upcoming responses.

In what follows, we describe the results of two eye-tracking studies that we conducted to test our hypotheses. The first study investigated whether turns with second-person pronoun subjects (*jij*) elicit more anticipatory looks to the addressee than turns with first-person pronoun subjects (*ik*) in Dutch conversation. Here we predict that adults and children 3;0 and older would use the pronoun-based epistemic cue ("you" vs. "I" subjecthood) to predict upcoming responses. The second study investigated whether children also use the pronoun-based epistemic cue ("You" vs "I" subjecthood) to predict upcoming responses when they do not have full access to lexical information. To do so, we used the same procedure and analysis—but this time with a nonsense Dutch-like language (Jabberwocky Dutch)—to check for the same expected second-person pronoun advantage. We use Jabberwocky Dutch to simulate the use of the second-person cue very early in learning—when the child does not have sufficient lexical knowledge to make predictions on the basis of utterance content. Here we predict that children, but not necessarily adults would use the pronoun-based epistemic cue ("you" vs. "I" subjecthood) to predict upcoming responses, and would thus make anticipatory gaze switches to the addressee earlier and more often when hearing an utterance with a second-person subject pronoun (versus a first-person subject pronoun). In both studies, we focus on the effect of second-person pronoun subject (versus first-person pronoun subject) in questions rather than non-questions because (1) we expect the link between subject pronoun and contingent response to be most prominent in questions (see above) and (2) because research has shown that anticipatory gaze switches are much less common in non-questions than questions. As such, designing our study around declaratives would have increased the likelihood of null results, not because there is no "you" effect but because declaratives offer a less predictable environment in which to observe response anticipation.

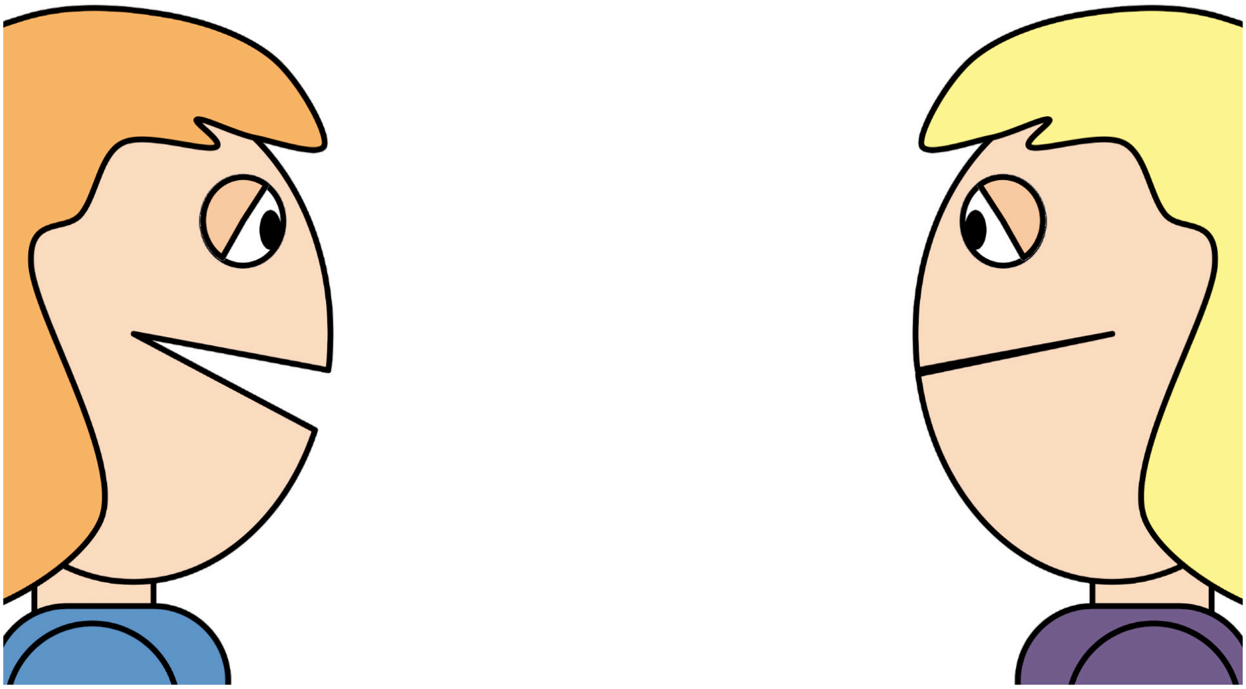

Fig. 1. Video frame from a dialogue stimulus. The same set of puppet images was used to create video stimuli for Study 1 and Study 2.

Study 1

Methods

All methods for participant recruitment, compensation, data collection, and subsequent data storage were pre-approved by the Radboud University Social Sciences Ethics Committee (ECSW/MB17U.000160). Note that we did not pre-register this study because the study began in 2014 when practices of pre-registration were not quite as common as they are now. The study plan (including hypotheses and analyses) though was documented in a department-internal project proposal approved in 2018. In this original plan the analyses focused on logistic mixed-effects models analyzing participants' proportion of anticipatory looks. Additional analyses – including cluster-based permutation tests to examine the timing of predictive gaze shifts (reported in the main manuscript and Supplementary Analyses S2 and S3), as well as exploratory analyses of filler items (Supplementary Analysis S1) – were conducted post hoc to further explore and contextualize the main findings.

Materials

We created videos of eight short dialogues between two animated female characters (Fig. 1). All animations can be found in the Materials folder at: https://osf.io/7urkt/. The same characters were used across all dialogues. Note that we purposely decided to work with animations of puppets rather than videos of real human interaction, because the primary goal of our study is to isolate children's sensitivity to specific linguistic cues to turn anticipation. Having videos of real human interaction would increase the ecological validity of our study, because in real life humans make use of visual cues (e.g., eye gaze, body movements, breaths etc.) to turn anticipation as well, but it would have made it more difficult to determine the extent to which children's anticipatory behavior is driven by linguistic cues alone.

We wrote the dialogue scripts to test the effect of the subject pronoun on response prediction, in this case specifically comparing first-person *ik* vs. second-person *jij* subject pronouns in constructions with subject-verb inversion (e.g., *Heb ik/jij ze al laten zien?* [Have I/you already shown them?]). The full dialogue scripts can be found in the Materials folder at: https://osf.io/7urkt/. We limited our target utterances to this single inversion construction for two reasons: (1) to limit the duration of the experiment and (2) because the interrogative syntactic context is more likely to support anticipatory looks in this kind of task (e.g., Casillas & Frank, 2017).

Each dialogue was 11 sentences long, and focused on a child-friendly topic (e.g., feeding and petting animals, visiting the pool, building a snowman, etc.). The 11 sentences included 4 target sentences containing the contrasting cue of interest, as well as 7 filler sentences. The target sentences (32 in total; 4 x 8 dialogues) were composed as follows: a verb (*moeten* [should], *kunnen* [can], *zullen* [shall], *mogen* [may], *hebben* [have], *nemen* [take], *komen* [come], or *zijn* [be]) followed by either the first-person singular pronoun *ik* ([I], 16 sentences) or the second-person singular *jij* ([you], 16 sentences), then followed by the rest of the sentence.

As mentioned above, we had previously conducted a small corpus study to confirm that the second-person pronoun in subject-verb inverted utterances is a statistically reliable cue to questionhood (see Supplementary Materials S4). In that study, we observed that some verbs demonstrate a preference for one or the other target pronoun (e.g., *Mag ik* [May I] is much more common than *Mag jij* [May you]). To avoid a confound between verbs and pronouns in our stimuli, we used each pronoun (*ik/jij*) at least once with each verb over the eight dialogues. Each dialogue character also used each verb at least once. The 56 filler sentences were all declarative statements with the following subjects: *ik* ([I], 19), *jij/je* ([you], 9), other person pronouns (*hij/we/het* [he/we/it], 12), possessive pronouns (*mijn X* [my X/mine], 4), and other noun phrase types (e.g., *het huis/Mama* [the house/Mommy], 12).

The dialogues were structured so that participants could not know whether a character would continue speaking or not at the end of each sentence. Across the whole set of eight dialogues, target sentences were equally likely to be followed by a speaker switch (i.e., a response by the other speaker) or by a speaker continuation (e.g., the character who had previously spoken resumes talking). Individually, each character contributed approximately half of the sentences (pronouns: 51 % and 49 % per character) and was approximately equally likely to transition or continue speaking at the end of her sentences (pronouns: 51 % and 49 % transitions per character). An example dialogue is shown below in Dutch (blue text) with a loose English translation in square brackets (black text). Target sentences are bolded. We recorded two versions of each target sentence; one with rising final intonation and one with falling final intonation.

A: Ik heb twee konijnen gekregen. [I got two rabbits.].
A: **Zul <u>jij</u> ze leuk vinden?** [**Would you find them nice?**].

B: Ik vind konijnen lief. [I think rabbits are sweet.]

B: Ik heb zelf ook een konijn [I myself also have a rabbit.]

A: Ik zal mijn konijnen ophalen. [I'll go get my rabbits.].
A: **Neem <u>ik</u> dan mee naar binnen?** [**Should I bring them inside?**].

B: **Mag <u>jij</u> ze binnen zetten?** [**Are you allowed to put them inside?**]

B: Mijn konijn mag nooit naar binnen. [My rabbit may never come inside.]
B: Het huis wordt dan vies. [Then the house would get dirty.]

`A:` **`Zal ik ze dan buiten laten? [Should I leave them outside then?]`**.

`A: Het is toch warm buiten. [It's warm out after all.].`

In sum, all (and only) the target sentences were questions, half of which used *ik* as the subject pronoun and half of which used *jij* as the subject pronoun, with all questions formatted as subject-verb inverted polar interrogatives. The scripts were child friendly and as balanced as possible without losing a feeling of dialogic coherence, as judged by two native Dutch speakers.

A separate pair of adult female native Dutch speakers voiced the dialogues. Stimuli were recorded in two sessions. In the first session, both speakers were present to record the dialogues together. Recordings were made with a Marantz PMD 661 solid-state field recorder, with each talker wearing a corded Shure WL185 cardioid lavalier microphone. This first recording session yielded examples of each conversation as it sounds in an actual dialogic context. For example, capturing the prosodic alignment across turns that can occur in real-time dialogue (see, e.g., Figueroa et al., 2023) and Kruyt et al., 2023).

In a second session, each speaker was brought back to record a "clean" version of each sentence, free of overlapping speech from the other talker, disfluencies produced by the speaker, and other issues that arose in the first session. For each sentence, the speaker listened back to their own recording from the dyadic context and imitated the original prosody while eliminating disfluencies. Using Praat (Boersma & Weenink, 2024) we then clipped the disfluency-free utterances to their phonetic onsets and offsets, then volume-normed and concatenated each speaker's utterances back into a stereo, dialogue format. In the final dialogue audio, each speaker's voice was associated with a single audio channel (left or right), and all inter-sentence silences were set to 500 msec. The inter-turn silence clips were sampled from the recording sessions to match the profile of background noise audible during the speech clips.

The audio file for each dialogue was converted using custom scripts into a simple animation of two female characters talking (https://github.com/marisacasillas/AnimatingConversation). The scripts take as input the audio files, a Praat TextGrid file (Boersma & Weenink, 2024) demarcating the syllable boundaries for each utterance, and a set of simple drawings for the characters. The materials used to create the videos are available in the "Materials/Materials to create videos" folder at: https://osf.io/7urkt/. The pair of animated speakers included two pale-skinned female cartoon characters with distinct hair and shirt colors; a red-headed character wearing a purple shirt and a blonde character wearing a blue shirt. We created four versions of each dialogue, swapping the position of the characters (blond left or redhead left) and the terminal intonation contour of the target sentences (rising or falling). To limit the number of possible video versions, the audible voices associated with the visual characters stayed the same across all video versions. For example, in the four versions of Dialogue 1, the voice of Speaker B is always on the right, and that same Speaker B always speaks the first line, but she is shown as the blonde in two of the videos and the redhead in two of the videos. Within a single test session, a participant saw only one of the four version types for the eight dialogues (e.g., all dialogues featuring the blond character with speaker A voice on the left, and with rising intonation on all target sentences). The final videos were each approximately 25 s long.

Recruitment

Adult participants were recruited and compensated via the Max Planck Institute for Psycholinguistics participant pool, which includes students from the nearby Radboud University but also other adults in and around the Nijmegen region. We invited local Dutch participants between ages 18 and 30 who had no known hearing or language impairments. Those with corrected vision were invited to participate, as were those fluent in a second or further language (most of them), so long as Dutch was their primary language. Adult participants were compensated €6 for their participation. Child participants were recruited via the Baby and Child Research Center database, which maintains secure metadata for a wide range of families in and around the Nijmegen area who had previously expressed interest in participating in infant and child research. We contacted legal guardians of monolingually Dutch-learning children with no known hearing or language impairments. If the guardian expressed interest in participating after hearing about the procedure and basic purpose of the study, the experimenter scheduled an in-person visit. We recruited and scheduled children so that they would be within two weeks of our target ages when they visited the lab (1;0, 2;0, 3;0, and 4;0 each $\pm$ 2 weeks). Guardians and children were compensated with €10 or a children's book. We recruited 50 participants for each of the four versions of the videos (200 in total). The participants were divided into groups of 10 for each of the five age groups for each version (1, 2, 3, and 4 years old, plus adults).

Procedure

Upon arrival, adult participants were shown to the testing room, where they were given written information about the study and a consent form to sign. After participants gave their consent, the experimenter helped them place a target-shaped eye-tracking sticker on their forehead, just above the eyes. The experimenter then adjusted the tracker height and angle as needed to prepare for calibration.

Child participants, upon arrival, were taken first to a warm-up room with toys, books, and drawing materials. During the warm-up period, the experimenter explained the procedure and went through the consent and information forms with the child's guardian. In some cases, the guardian had already completed this process in advance, via forms sent through the ail. The guardian, experimenter, and child then played together until the child appeared comfortable with the experimenter. At that point, all three entered the testing room. Children were placed in a high-chair in front of the eye tracker. One-year-olds were seated in a high chair with a built-in tray (Ikea Antilop). Older children were seated in a less restrictive high chair (Stokke Tripp Trapp). In all cases, the guardian was seated immediately behind their child in a second chair. The experimenter distracted the child while quickly applying the target-shaped eye-tracking sticker to their forehead, and then swiftly began a short animated film to keep children's attention while the experimenter adjusted the tracker height and angle as needed to prepare for calibration. If the child had hair hanging over their forehead, the experimenter used a hair clip before applying the sticker. Guardians were given headphones to wear, over which the experimenter played language-masking music. With the guardian's consent, participating children were given a bread stick (or similar) to eat while watching the videos. Pilot testing had revealed that giving a bread stick improved children's general attention to the videos.

All participants were tested in the same office, using the same eye-tracking equipment (Eyelink 1000 set up (Research, 2005) with a 890 nm illuminator and 25 cm lens) attached by an ergonomic arm to a table on which the tracker computer sat. The videos were displayed on a large monitor (55 cm, 1920 x 1080 pixels) and the audio was played back over stereo speakers (Genelec 8010A Studio Monitor) attached just behind the screen, to its left and right sides. The participants' view of the experimenter was blocked by a tall sound-dampening screen (BuzziSpace). The tracker was run from one computer that uses the tracker's custom operating system (Eyelink, 5.13). The experiment itself was run from a second computer using Presentation (Neurobehavioral Systems, 21.1 09.05.19). A video of the session was captured by a webcam placed on the stimulus monitor and recorded using OBS Studio (21.1.0) on a third computer.

Once a participant satisfactorily completed a 5-point calibration, the experiment immediately started. The order of the eight dialogue videos for that participant was randomized, with short, attention-getting filler videos after every two target videos (e.g., a ladybug zipping back and forth across the screen). Eye-tracking data were collected from the left eye continuously throughout the experiment, with the x- and y- measures logged for each sample and written to an output file for later analysis.

Participants who did not achieve successful calibration after multiple tries were shown the experiment stimuli without actual use of the tracker—the experimenter marked the metadata for these sessions as lacking eye tracking data. In total, 12 (6 %) participants were excluded from our analysis on this basis (5 one-year-olds; 2 two-year-olds; 3 three-year-olds; 2 four-year-olds). We note that future manual coding of the webcam data for those "no data" sessions is possible. When a participant successfully completed calibration but some time later became too fussy to continue attending to the stimuli, the experimenter ended the experiment early. These sessions were highly variable in overall participant attention and the total number of stimuli seen by the child. We therefore included these early-ended sessions in the primary analysis below, knowing that instances of very low-attention during video playback would be systematically excluded during data pre-processing (see below).

After the experiment ended, adult participants and guardians of child participants were debriefed on the purpose of the study and were given the opportunity to ask questions or make comments before receiving their compensation and departing.

Results

We analyzed participants' gaze behavior in two ways. First, we examined the likelihood that participants made an anticipatory gaze switch from the current speaker to their addressee across conditions. Second, we examined whether gaze switches begin earlier or later across conditions. We had predicted that adults and children 3;0 and older would make more and earlier (anticipatory) gaze switches to the addressee when hearing a second-person pronoun subject (*jij*) sentence than a first-person pronoun (*ik*) subject sentence. All data (for Study 1 and Study 2) as well as the the Rmarkdown scripts that were used to analyse the data are available at: https://osf.io/7urkt/
.

Analysis 1: Likelihood of an anticipatory gaze switch.

Following Casillas and Frank (2017), we first evaluated, for each gap, whether participants predicted an upcoming response. We did so by checking whether participants initiated an eye movement from the current speaker to the addressee some time around the end of the current speaker's turn. Critically, this had to happen before the participant could register any response by the addressee. We defined an anticipatory gaze as follows: if a participant fixated on the current speaker in the last 300 ms of her turn for at least 100 ms, and subsequently began fixating for at least 100 ms on the potential responding speaker before 200 ms of the post-gap utterance had elapsed, the participant was marked as having made an anticipatory gaze for that utterance (score = 1). If the participant fixated on the current speaker in the last 300 ms of her turn but then did not transition as described, they were marked as not making an anticipatory gaze for that utterance (score = 0). If the participant did not fixate on the current speaker toward the end of her turn, then they did not meet the pre-conditions for making an anticipatory switch and were thus given an "NA" for that utterance (score = NA). NAs were not included in the statistical analysis. Thus, we analyze the likelihood of making a gaze switch given attention to the current speaker in the critical moments before a response might be given.

Following prior work (Kamide et al., 2003), we assume that it takes adults at least 200 ms to plan an eye movement, and so we attribute fixations starting up to 200 ms after the onset of post-gap speech as still anticipating a possible response—not yet as reactive to the onset of speech. While this saccadic planning window is likely longer for young children, we lack a clear estimate for the increased time needed to plan an eye movement, and therefore we use the adult threshold of 200 ms for children as well. This is a conservative decision, in the sense that this threshold makes it likely that we will miss some slow, but anticipatory gaze switches from young children (i.e., underestimate children's slow anticipatory behavior). In brief, using an 1100 ms analysis window (300 ms on either side of the 500 ms target gaps), we were able to measure how frequently each participant made an anticipatory gaze switch to the current addressee.

We pre-processed the raw data to remove tracker samples where gaze location was invalid and to classify remaining gaze measurements as within the left or right character's region of interest (the left- and right-most 250 pixels of the screen). We then measured the number of valid gaze samples for each stimulus video and removed any cases where a participant had valid eye tracking data for less than 25 % of the total video duration (25 s). That is, we imposed a minimum of $\sim$ 6.25 s of valid looking time for each stimulus video; our intention to measure anticipatory gaze while the participant was at least partly following the interaction requires a lower threshold of looking time. A minimum of 25 % is an inclusive cut-off. However, in this free-viewing task, which lacked direct participant prompts (i.e., no "look at the X!" prompts), participants often gazed off screen—at the equipment, at their shoes, at their parent, etc.. Ultimately, participant gaze data from 287 videos was excluded on this criterion, leaving 82.1 % of the original dataset for analysis. The excluded videos had very little on-screen looking, with a mean of 8.28 % valid gaze data for each video; about 2.07 s of viewing time out of the 25 s of video duration. Exclusions affected 80 of the 200 participants, with a mean of 3.59 (median = 2.5) eliminated videos per participant (range = 1–8).

Anticipatory gaze switch rate is illustrated in Fig. 2. We modeled the child data using a generalized logistic mixed-effects regression of anticipatory gaze switches (0/1; binary outcome) that included the following fixed effects: utterance subject pronoun (first/second; contrast-coded factorial variable), utterance-terminal intonation (falling/rising; contrast-coded factorial variable), participant age (1–4; scaled and centered), and their two-way interactions. We also included participants and utterances as random effects. We modeled the adult data separately from the child data by using the same model structure, but omitting the participant age variable.[1]

The model of children's anticipatory switches ($N_{obs} = 2484$; $N_{ptcp} = 142$; $N_{utt} = 32$; LogLik = -1419.40; AIC = 2856.70; Table 1) revealed a significant increase in anticipatory switching with age ($B = 0.82$, $SE = 0.089$, $z = 9.18$, $p < .001$), consistent with prior work on these age groups (Casillas & Frank, 2017; Lammertink et al., 2015). We also observed a main effect of subject pronoun ($B = 0.77$, $SE = 0.12$, $z = 6.53$, $p < .001$) in the expected direction: children were more likely to make an anticipatory gaze switch after hearing a sentence with a second-person pronoun (*jij*), compared to a first-person pronoun (*ik*). The model suggested no significant effects of intonation nor any two-way interactions (all p's > 0.1).

We ran post-hoc regressions on each individual age group to find the age at which the pronoun effect first emerges. Note that partitioning the data by age group in this way limits repeated observation of some target utterances, such that we had to drop the random effect of utterance in these models (though when we do include this random effect, the (non-convergent) models give qualitatively identical outcomes). We saw that the second-person subject pronoun predicted significantly greater likelihood of anticipatory gaze switches at ages 2;0 ($B = 0.49$, $SE = 0.14$, $z = 3.45$, $p < .001$), 3;0 ($B = 0.65$, $SE = 0.13$, $z = 5.14$, $p < .001$), and 4;0 ($B = 0.56$, $SE = 0.12$, $z = 4.61$, $p < .001$). Meanwhile, children at 1;0 did not show an effect of subject pronoun ($B = 0.035$, $SE = 0.21$, $z = 0.17$, $p = .87$). From this we infer that the pronoun effect is detectable by children's second birthday.

The model of adults' anticipatory switches ($N_{obs} = 1189$; $N_{ptcp} = 40$; $N_{utt} = 32$; LogLik = -680.80; AIC = 1373.60; Table 2) revealed no significant effects of any predictor on anticipatory gaze switching. That said, we did observe an effect of subject pronoun at $p = .08$ in the expected direction ($B = 0.5$, $SE = 0.29$, $z = 1.74$, $p = .081$), in alignment with the child data.

Analysis 2: Timing of gaze switches

To investigate the timing of children's looking behavior to the addressee, we carried out cluster-based permutation analyses using the eyetrackingR package (Dink & Ferguson, 2015). Note that in these second set of analyses we are interested to see how participant's gaze changes over time. Therefore these analyses included all looks to the addressee, whether gaze met the criteria of being anticipatory (Analysis 1) or not. For simplicity we collapse data across the two utterance-terminal intonation conditions (falling/rising); time course analyses including both pronoun and intonation splits of the data can be found in the Supplementary Materials S1.

We ran the cluster-based permutation analyses on a 1500 ms analysis window (500 ms on either side of the 500 ms target gaps). Before running the permutation analyses we removed individual target utterances during which children looked less than 25 % of the time. Note that these individual targets are removed in addition to the target utterances that were already removed by the full dialogue (video) exclusion criteria, as described in the pre-processing for Analysis 1. This led to an exclusion of an additional 270 target utterances for one-year-olds, 189 for two-year-olds, an 99 for three-year-olds, 127 for four-year-olds, and 14 for adults. Additionally, individual datapoints (gaze samples) were removed if they occurred at a time interval smaller than 0.0015 ms (which is.0005 ms smaller than the tracker's set sampling rate). This criteria led to an exclusion of 24,611 (6.1 %) individual datapoints for the one-year-olds, 35,515 (5.7 %) for the two-year-olds, 37,104 (5.6 %) for the three-year-olds, 40,683 (6 %) for the four-year-olds, and 49,028 (5.4 %) for the adults.

After excluding these data, the 1500 ms time window was split into 50 ms time bins. For each age group and time bin we calculated a *t*-statistic to determine whether participants looked significantly more towards the addressee for second- vs. first-person subject pronouns in the target utterances. Adjacent time bins for which there were significant differences between the two pronouns were clustered together and the *t*-values of the bins were summed. Then, the data were shuffled 1000 times, to see how likely the observed clustered and summed statistics would be if the data were randomly generated. The timecourse data is visualized in Fig. 3.

As detailed below, we found evidence supporting a second-person pronoun advantage in the looking data for all age groups except 1-year-olds. For 1-year-olds, no clusters were detected in which children looked more at the addressee after hearing second- versus first-person pronouns within the targeted 1500 ms time window. In 2-year-olds, the analysis detected three clusters with a second-person pronoun advantage, of which one showed a statistically significant difference in the proportion of looks to the addressee (sum $t = -36.01$; $p = .008$). This significant cluster started 450 ms post gap onset and ran until 500 ms post gap offset (duration = 550 ms). The other two clusters—one from 400 ms to 300 ms pre-gap onset (duration = 100 ms; sum $t = -4.24$; $p = .195$) and one from 150 ms pre-gap onset to 150 ms post-gap onset (duration = 300 ms; sum $t = 14.92$; $p = .061$)—did not show statistical evidence of a second-person pronoun advantage. In 3-year-olds, the analysis detected one cluster with a second-person pronoun advantage from 350 ms pre gap onset till 500 ms post gap onset, which was also found to statistically differ from zero (duration = 1350 ms; sum $t = -130.41$; $p < .001$). In 4-year-olds, the analysis detected one cluster with a second-person pronoun advantage from 500 ms pre gap onset to 400 ms post gap offset, which was found to statistically differ from zero (duration = 1400 ms; sum $t = -131.74$; $p < .001$).

[1] glmer(Switch_binary ~ Age (in years) + Subject_pronoun (first/second) + Intonation (rising/falling) + Age:Subject_pronoun + Age:Intonation + Subject_pronoun:Intonation + (1|Participant) + (1|Utterance), data = Child_data, family = binomial). The adult dataset was modeled identically, minus age effects.

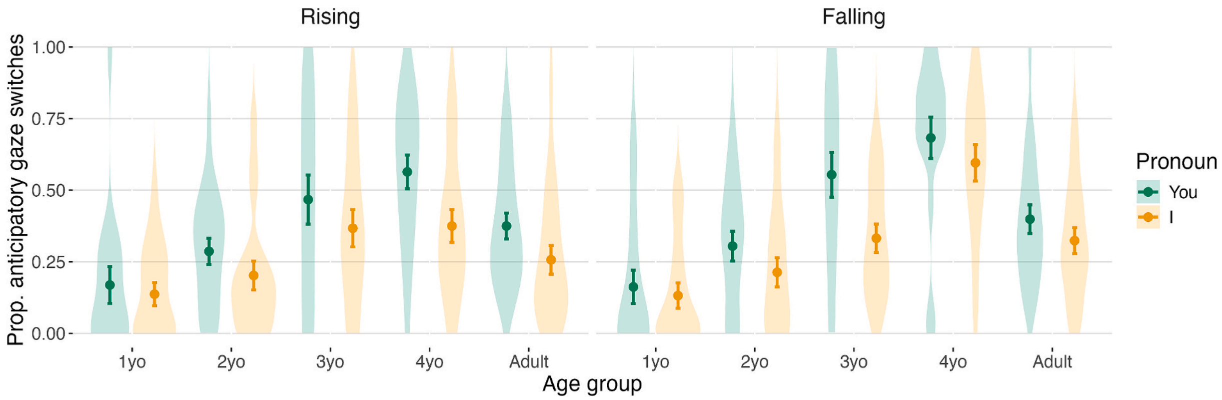

Fig. 2. Proportion anticipatory gaze switches to the addressee (y-axis) by participant age (x-axis), utterance-terminal intonation (rising = left panel; falling = right panel), and subject pronoun (you = green; I = orange) in Dutch dialogue stimuli. Points represent the mean and vertical bars represent standard error around the mean per participant age, pronoun, and intonation condition. Violins represent the distribution of proportion anticipatory gaze switches per participant age, pronoun, and intonation condition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Regression estimates from the model of children's anticipatory gaze behavior in the Dutch dialogues.

Predictor	B	Std. Error	t	p
(Intercept)	−0.62	0.10	−6.34	<.001
Age group	0.82	0.09	9.18	<.001
Pronoun	0.77	0.12	6.53	<.001
Intonation	−0.18	0.18	−1.02	0.31
Age:Pronoun	0.15	0.10	1.49	0.135
Age:Intonation	−0.29	0.18	−1.61	0.107
Pronoun:Intonation	−0.24	0.19	−1.24	0.216

Table 2
Regression estimates from the model of adults' anticipatory gaze behavior in the Dutch dialogues.

Predictor	B	Std. Error	t	p
(Intercept)	−0.87	0.20	−4.23	<.001
Pronoun	0.50	0.29	1.74	0.081
Intonation	−0.25	0.32	−0.78	0.438
Pronoun:Intonation	0.42	0.28	1.52	0.128

Finally, for adults, the analysis detected three clusters with a second-person pronoun advantage, two of which showed a statistically significant difference from zero: from 300 ms pre gap onset to 200 ms post gap onset (duration = 500 ms; sum t = -28.80; p = .005), and from 250 ms post gap onset to 400 ms post gap offset (duration = 650 ms; sum t = -40.27; p < .001). The third and final adult cluster, from 450 ms post gap offset to 500 ms post gap offset, did not show a statistically significantly effect of pronoun (duration = 50 ms, sum t = -2.23; p = 0.32).

Visual inspection of Fig. 3 suggests that the second-person pronoun advantage for 3-year-olds and 4-year-olds already starts before the onset of our 1500 ms analysis window (i.e., more than 500 ms before the gap "pre-gap onset"). This could reasonably occur given that the pronouns occur early in the utterances—an average of 1150 ms pre-gap onset (range: 1713–665 ms pre-gap onset). To explore this further, we conducted a second set of cluster-based permutation analyses on a 650 ms time window following the pronoun onset, divided over 13 smaller bins from 50 ms each. Using this approach, we find evidence that 4-year-olds start to show a second-person pronoun advantage 450 ms after pronoun onset until the end of our time window (duration = 150 ms, sum t = -15.93, p = .008). In all other age groups, we find no evidence for a second pronoun advantage in 650 ms time window following pronoun onset. For more details see Supplementary Materials S2.

Discussion

This first study was designed to investigate Dutch children's (1;0, 2;0, 3;0 and 4;0) and adults' real-time use of subject pronouns and utterance-terminal intonation to predict upcoming responses during conversation. We hypothesized that both child and adult observers would make anticipatory looks more often—and switch their gaze earlier—when hearing an utterance with a second-person

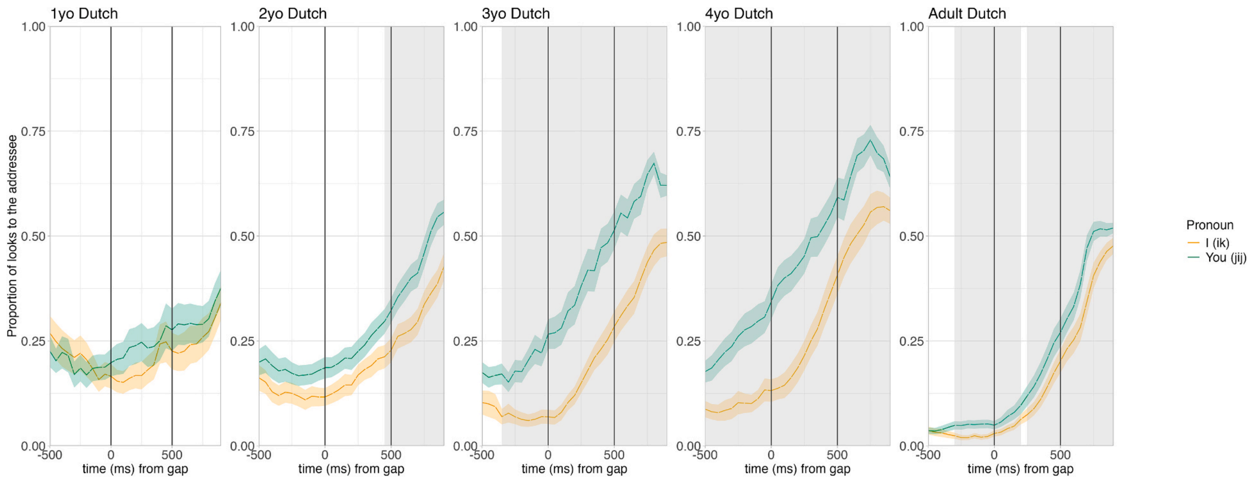

Fig. 3. Mean proportion of looks to the addressee during the 1500 ms analysis windows in the Dutch dialogues, split by subject pronoun (you = green; I = orange) and age group (separate panels). The 500 ms gap begins at 0 on the x-axis, and gray areas indicate clusters with statistically greater gaze to the addressee. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subject pronoun *jij* (vs. a first-person pronoun *ik*) and when hearing an interrogative utterance ending in rising intonation (vs. falling intonation).

We found no impact of intonation on the likelihood or timing of children's anticipatory looks to the addressee, but we did find evidence that children use subject pronouns to anticipate an upcoming response. Child age was also an important factor: older children made more anticipatory gazes to the addressee. The adult data overall mirrored these findings: no evidence for an effect of intonation, and a benefit for the second-person pronoun (albeit a weaker benefit than seen in the child data).

No evidence for an effect of utterance-terminal intonation contour

An utterance-terminal rising intonational contour is prototypically associated with polar questions in Dutch, as it is in English, and other languages (Siemund, 2001; Ultan, 1978), so why didn't it influence anticipatory looks? In past work, the lack of a clear effect for prosody in turn-taking predictions has been attributed to the low informativeness and timing of the information it contains about turn ends and speech acts Casillas and Frank (2017): intonational contours carry a great deal of information (e.g., about syntax, pragmatics, word segmentation, stance, valence, etc.), and so they are mapped many-to-many with different social moves in conversation. In running speech, intonation contours carry useful information for word segmentation, but may not be sufficient on their own to segment single lexical units from larger phonological phrases (Christophe et al., 1997; Davis, 2000). Finally, by its nature, the high-rise terminal intonational cue comes at the very end of the utterance—it may be too late to aid in anticipatory looking. For these reasons, listeners may prioritize prediction on the basis of lexico-syntactic cues (de Ruiter et al., 2006)—and, in our design, all the target utterances are equally lexico-syntactically marked as questions. So, even though intonation patterns in connected speech carry a great deal of information, they are seldom sufficient as cue to word segmentation, and—as we see here and elsewhere—they may have limited impact in anticipating upcoming speaker switches Casillas and Frank (2017).

However, multiple factors may be at play in participants' use of prosodic cues to make real-time predictions about upcoming turn taking. Adult listeners have shown clear sensitivity to utterance-final prosody when it cancels the otherwise-predicted end of an ongoing turn (Bögels & Torreira, 2015, 2021), and listeners may fall back on prosody if the lexico-syntactic cues are less reliable. In Study 2 (below), we were able to observe whether participants rely more on utterance-terminal intonation when lexical meaning was no longer available.

Sensitivity to an epistemically derived cue

Participants age 2 years and older showed sensitivity to the second-person singular pronoun as a cue to questionhood. Their use of this cue is remarkable considering that this cue is: (a) not deterministically coupled with interrogatives in Dutch, (b) syntactically non-initial in the stimuli, and (c) not phonetically or prosodically salient in running speech. Because past work shows that the question effect reliably emerges between 2;0 and 3;0, we had expected to only find sensitivity to this cue from age 3;0 and up. But our analyses suggest that sensitivity to this cue can be demonstrated earlier than we had hypothesized—by age 2;0. That said, the gaze timing analysis (Analysis 2) suggested that children's predictive looks to the addressee for second-person pronoun subjects were made earlier and more robustly from age 3;0 onward, lending support to the idea that the question effect becomes more prominent between the second and third birthdays (Fig. 3). We found no evidence for differences in the second-person pronoun subject effect between ages 3;0, 4;0, and the adult sample, which also suggests that the impact of these cues is relatively stable after age 3;0.

A primary motivation in conducting this work was to determine whether children who are relatively limited in their linguistic skills might latch onto individual, interactionally informative linguistic cues during unfolding talk. Interactionally informative cues may keep young children afloat when following conversations that they do not fully understand, allowing them to engage beyond their current linguistic capacities to get more out of interaction (Shatz, 1978; Warlaumont et al., 2014; Zhang et al., 2024). The data from Study 1 suggest that subject pronouns may play such a role—at least, in the subject-verb inverted syntactic environment and when participants have access to the full linguistic signal. Children were reliably able to use the subject pronoun to predict upcoming responses by 2;0, an age when many of them are just beginning to combine words (Schlichting, 1996; Verrips & Wijnen, 1998) and when their vocabulary is noticeably growing and diversifying (Frank et al., 2021; McMurray et al., 2012; Nelson, 1973). Notably, the adults' data suggests a similar, but weaker advantage for second-person pronoun utterances, which may indicate that these cues are still marginally useful even with a fully adult-like linguistic processing capacity.

A possible amendment to the above explanation is that individual cues are not sufficient—listeners may require a *suite* of accessible and sufficiently informative cues to make predictions during real-time conversation. In Study 1, children encountered controlled manipulations of subject pronoun and intonation contour within a context of ordinary spoken Dutch. Their ability to pick up on these challenging, non-salient cues is highly likely to be affected by the semantically meaningful context, enabling them to track the topic at hand, the immediately preceding lexical context, and so on. This content-based information provides indirect semantic and pragmatic cues to questionhood that may alter participants' predictions. An even more stringent test of our target cues (subject pronouns and utterance-terminal intonation) would then be to examine their effect on anticipatory gaze switches in the absence of semantic cues, which is the aim of Study 2.

Study 2

The second study was designed to test the influence of subject pronoun (*ik/jij*) and terminal intonation contour (rising/falling) in Dutch conversation, this time with minimal supporting semantic information (removal of open-class nouns, verbs, adjectives etc.). The stimuli for Study 2 were the same as in Study 1, but translated into a nonsense (Jabberwocky) Dutch-like language. The design and analyses for Study 2 are identical to Study 1 to maximize comparability of the datasets.

Methods

The methods for Study 2 were identical to those of Study 1, with the exception that the dialogues were spoken in Jabberwocky Dutch—a nonsense, Dutch-like language which we call "Nepperlands". Our methods for creating this novel language followed other work using nonsense English- and Dutch-like languages (Fedorenko et al., 2016; Johnson & Goldberg, 2013; Kaufeld et al., 2020). This type of modified language is often referred to as "Jabberwocky" in recognition of Lewis Carroll's (1871) poem of the same name.

Materials

We first created a Dutch-to-Jabberwocky translation dictionary for all the open-class words in the Study 1 scripts (e.g., verbs, nouns, adjectives, adverbs, proper names). We created a comprehensive list of the unique open-class lemmas (e.g., "give" for "gives", "given", "giving") and generated multiple phonotactically matched non-words for each using Wuggy (v0.2.0b2, Keuleers and Brysbaert (2010): https://wuggycode.github.io/wuggy/; e.g., *geven* [give] yielded *kesen*, *besen*, and *pesen*).

We then pre-tested each candidate Jabberwocky word with adult Dutch speakers to ensure that they could not reliably guess the original Dutch meaning, neither (a) in isolation nor (b) in its Jabberwocky sentence context. To do this, we recruited 32 adult native Dutch speakers with no known language or hearing impairments using the Max Planck Institute for Psycholinguistics participant database. Participants were seated in front of a computer and were asked to (a) judge whether each text string reminded them of a word in Dutch and (b) if so, to type that word into a text box. In a first block, participants saw 211 non-words in randomized order. In a second block, participants saw 192 sentences using candidate Jabberwocky words from the scripts, also in randomized order. Participants were compensated €12 for their time.

The first round of testing showed that 35 Dutch words did not have sufficiently opaque Jabberwocky candidates: the initial candidates for these words were correctly guessed more than a third of the time when in isolation (6) or more than half of the time in a sentence context (33). Thus, in a second round of testing, we hand-created additional non-word candidates for these 35 words, trying to maintain similar phonotactic similarity as generated via Wuggy. We then asked an additional 20 adult native Dutch speakers to guess the meanings of each new word candidate in isolation and in a sentence context, this time finding a suitable candidate for each word.

Using this technique, we were able to create semantically opaque stimuli, defined here as a set of words that were never correctly guessed more than a third of the time in isolation or more than half of the time in a sentence. Consider, too, that our adult pre-test participants were able to reflect for several seconds about each word before giving their answer. In comparison, our eye-tracking study participants below only heard these words fleetingly, as they unfolded during fluent Jabberwocky speech.

With this mini Dutch-to-Jabberwocky dictionary in hand, we then fully translated the scripts to Jabberwocky Dutch, inflecting verbs and nouns with regular Dutch morphology (e.g., the Jabberwocky Dutch word *loktel*, derived from the Dutch word *wortel* [carrot], received the regular plural form *loktels* [carrots]). The remainder of the words in the script, including the pronouns, remained in ordinary Dutch. An example dialogue is shown below in Jabberwocky Dutch (red text), original Dutch (blue text), and a loose English translation (black text). Target sentences are indicated in bold. As before, two versions of each target sentence were recorded—one with rising and one with falling final intonation.

A: Ik vib twee kopenen gebroegen.

Ik heb twee konijnen gekregen. [I got two rabbits.].
A: **Fut jij ze lerf ganden?**
Zul jij ze leuk vinden? [Would you find them nice?]

B: Ik gand kopenen liem.
Ik vind konijnen lief. [I think rabbits are sweet.]
B: Ik vib zelf ook een kopeen.
Ik heb zelf ook een konijn. [I myself also have a rabbit.]

A: Ik fut mijn kopenen opdaren.
Ik zal mijn konijnen ophalen. [I'll go get my rabbits.].
A: **Zees ik dan mee naar ginben?**
Neem ik dan mee naar binnen? [Should I bring them inside?]

B: **Hoom jij ze ginben nitten?**

Mag jij ze binnen zetten? [Are you allowed to put them inside?]

B: Mijn kopeen hoomt nooit naar ginben.
Mijn konijn mag nooit naar binnen. [My rabbit may never come inside.]

B: Het hies mort dan zies.
Het huis wordt dan vies. [Then the house would get dirty.]

A: **Fut ik ze dan vieten zoten?**
Zal ik ze dan buiten laten? [Should I leave them outside then?]
A: Het kijnt toch warf vieten.
Het is toch warm buiten. [It's warm out after all.].

All methods for recording and processing the stimuli were identical to Study 1 with two exceptions. First, the voice actors (the same as in Study 1) needed more time to rehearse the Jabberwocky stimuli in order to achieve comparable fluency and naturalness to the real Dutch stimuli. Second, in rehearsing the stimuli, the speakers used the original Dutch sentences as a reference during Jabberwocky sentence production. The Dutch reference sentences helped to ensure that prosody (e.g., speech rate, intonation contour) was similarly produced, sentence by sentence, across the language conditions. Because Studies 1 and 2 were planned together, the voice recording sessions—and all other steps of stimulus preparation and data collection—took place for both studies simultaneously.

Recruitment and procedure

All procedures for recruiting and testing the 200 participants for Study 2 were identical to those used and described for Study 1. Again we excluded children with whom we were unable to complete calibration. For Study 2 this resulted in the exclusion of 13 (6 %) participants (3 one-year-olds; 4 two-year-olds; 4 three-year-olds; 2 four-year-olds).

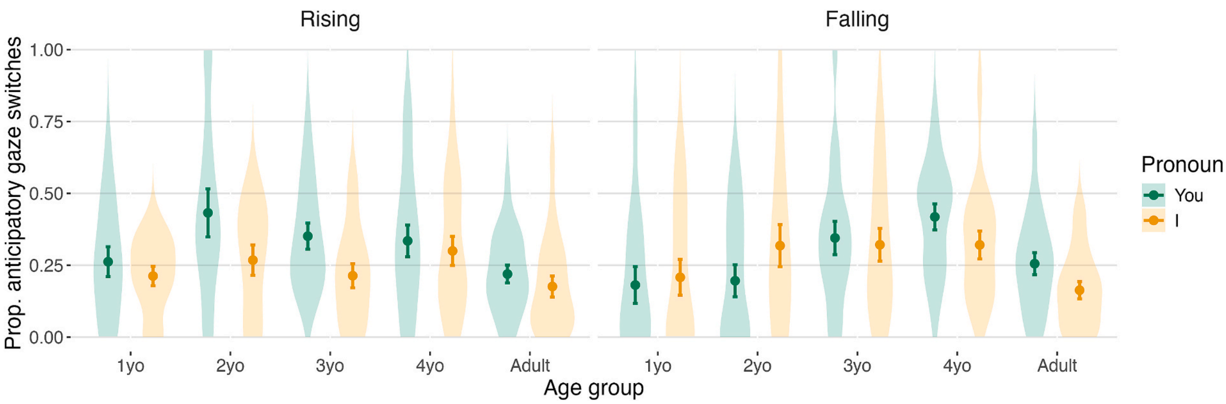

Fig. 4. Proportion anticipatory gaze switches to the addressee (y-axis) by participant age (x-axis), utterance-terminal intonation (rising = left panel; falling = right panel), and subject pronoun (you = green; I = orange) in Jabberwocky Dutch dialogue stimuli. Points represent the mean and vertical bars represent standard error around the mean per participant age, pronoun and intonation condition. Violins represent the distribution of proportion anticipatory gaze switches per participant age, pronoun and intonation condition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Results

Analysis 1: Likelihood of an Anticipatory Gaze Switch

The pre-processing pipeline of the raw eye-tracking data was identical to that described for Study 1. Our minimum cut-off for gaze to a video of 25 % (6.25 s) resulted in the exclusion of participant gaze data from 410 videos, leaving 74.4 % of the original dataset for analysis. We note that this exclusion rate is higher than that of Study 1, in which 82.1 % of the original data was retained for analysis. The mean valid gaze data for excluded videos was 6.18 % (1.55 s of looking time to the 25-second videos). Exclusions affected 103 of the 200 participants, with a mean of 3.98 (median = 3) eliminated videos per participant (range = 1–8)—both measures of exclusion are higher than in Study 1 (see Discussion later on). We used the same regression formulae as in Study 1 with both the child and adult Jabberwocky datasets.

Anticipatory gaze switch rate is illustrated in Fig. 4. The model of children's anticipatory switches ($N_{obs} = 1953$; $N_{ptcp} = 133$; $N_{utt} = 32$; LogLik = -1154.00; AIC = 2326.00; Table 3) revealed a significant increase in anticipatory switching with age ($B = 0.21$, $SE = 0.080$, $z = 2.64$, $p < .01$), consistent with prior work on this age group (Study 1, also Casillas and Frank (2017) and Lammertink et al. (2015)). We observed no influence of pronoun ($B = 0.18$, $SE = 0.13$, $z = 1.34$, $p = .18$), intonation ($B = -0.015$, $SE = 0.16$, $z = -0.094$, $p = .92$), or any of the two-way interactions (all p's > 0.12) on children's likelihood of anticipatory gaze switch.

Similarly, the model of adults' anticipatory switches ($N_{obs} = 1201$; $N_{ptcp} = 40$; $N_{utt} = 32$; LogLik = -580.7; AIC = 1133.30; Table 4) revealed no significant effects of any predictor on anticipatory gaze switching (all p's $> .12$).

Analysis 2: Timing of gaze switches

As in Study 1, we conducted an analysis of gaze timing using cluster-based permutation analyses. In line with Study 1, before segmenting participants' looking data within our 1500 ms time window into 50 ms time bins, we excluded trials at the individual participant level if the percentage of included looks during a trial was less than 25 %. We also removed individual datapoints (gaze samples) if they occurred at time interval smaller than 0.0015 ms. This led to an exclusion of an additional 240 target utterances, and 20,899 individual datapoints for one-year-olds; 138 target utterances and 24,593 datapoints for two-year-olds, 185 target utterances and 30,803 datapoints for three-year-olds, 163 target utterances and 41,668 individual datapoints for four-year-olds and 12 target utterances and 49,289 individual datapoints for adults. These exclusion rates are slightly higher for three-year-olds and four-year-olds than what we saw for Study 1.

After clustering adjacent time bins that showed significant differences in the proportions of looks to the addressee by pronoun condition (*jij* vs. *ik*), the data were shuffled 1000 times. As before, for simplicity we collapse data across the two utterance-terminal intonation conditions (falling/rising); analyses including both pronoun and intonation splits of the data can be found in the Supplementary Materials S1.

As detailed below, we found only sparse evidence supporting a second-person pronoun advantage in the looking data: one significant cluster for 2-year-olds and one significant cluster for adults. For the 1-year-olds, no clusters were detected in which the children looked more at the addressee for second-person pronouns than for first-person pronouns. In 2-year-olds, one cluster was found—from 100 ms pre gap onset until the gap offset—which showed a pronoun effect significantly different from zero (duration = 600 ms; sum $t = -37.17 p = .008$). In 3-year-olds, one cluster was found—from 300 ms post gap offset to 350 ms post gap offset—but it was not statistically significantly different from zero (duration = 50 ms; sum $t = -2.94$; $p = 0.25$). In 4-year-olds, three clusters were found, but none showed a pronoun advantage statistically significantly different from zero (cluster 1: 450 ms pre gap onset to 300 ms pre gap onset [duration = 150 ms, sum $t = -6.49$; $p = .15$]; cluster 2: gap onset till 50 ms post gap onset [duration = 50 ms; sum $t = -2.44$; $p = .24$]; cluster 3: from 100 ms post gap offset till 350 ms post gap offset [duration = 250 ms; sum $t = -11.91$; $p = .070$]). Finally, in adults the analyses detected two clusters, but only one—from 100 ms post gap offset to 450 ms post gap offset—showed a second-person pronoun advantage statistically significantly different from zero (duration = 350 ms; sum $t = –23.75$; $p = .011$). The other adult cluster detected—from 450 ms post gap onset to gap offset—did not show a significant pronoun effect (duration = 50 ms; sum $t = -2.23$; $p = .32$). See Fig. 5 for a visual representation of these outcomes.

Similarly to Study 1 we also explored whether there was evidence for a second-person pronoun advantage earlier in the time window, namely the 650 ms following pronoun onset. In none of our groups did we find evidence for a second-person pronoun advantage in this early time window though. For more details see Supplementary Materials S2.

Table 3
Regression estimates from the model of children's anticipatory gaze behavior in the Jabberwocky dialogues.

Predictor	B	Std. Error	t	p
(Intercept)	−0.93	0.09	−10.16	<.001
Age group	0.21	0.08	2.64	<.01
Pronoun	0.18	0.13	1.34	0.18
Intonation	−0.02	0.16	−0.09	0.925
Age:Pronoun	0.14	0.11	1.28	0.201
Age:Intonation	−0.25	0.16	−1.54	0.124
Pronoun:Intonation	0.16	0.21	0.77	0.439

Table 4
Regression estimates from the model of adults' anticipatory gaze behavior in the Jabberwocky dialogues.

Predictor	B	Std. Error	t	p
(Intercept)	−1.71	0.22	−7.84	<.001
Pronoun	0.52	0.33	1.56	0.12
Intonation	−0.02	0.31	−0.07	0.947
Pronoun:Intonation	−0.36	0.31	−1.17	0.24

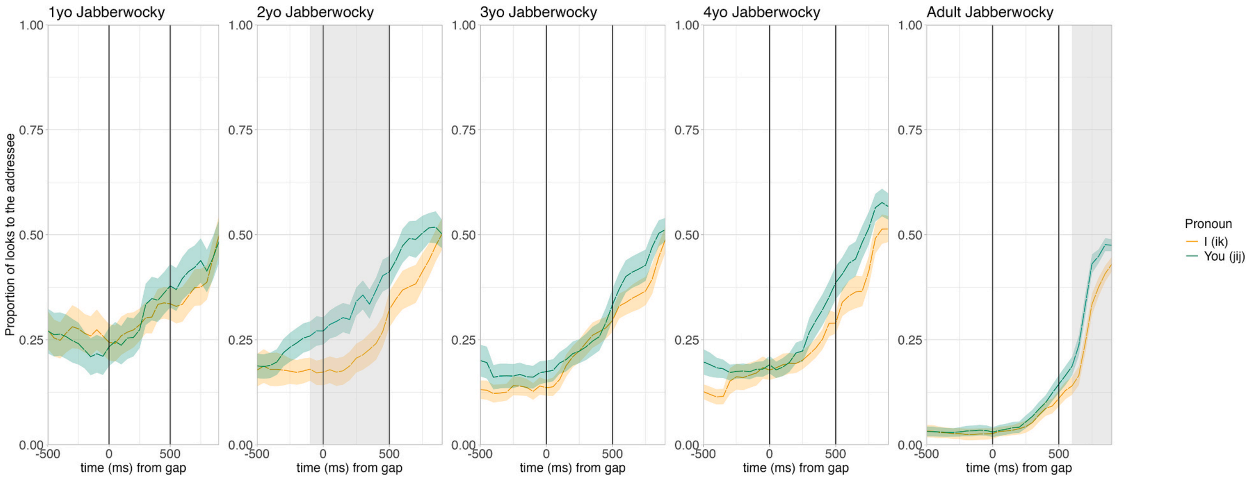

Fig. 5. Mean proportion of looks to the addressee during the 1500 ms analysis windows in the Jabberwocky Dutch dialogues, split by subject pronoun (you = green; I = orange) and age group (separate panels). The 500 ms gap begins at 0 on the x-axis, and gray areas indicate clusters with statistically greater gaze to the addressee. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Discussion

Study 2 was designed to test whether participants used the two cues of interest—subject pronoun and utterance-terminal intonation contour—to predict upcoming speaker switches in the absence of semantically meaningful information. The design and analysis of Study 2 paralleled that of Study 1, the only difference being that the spoken language in the video stimuli was a Dutch-like nonsense language rather than actual Dutch.

In short, we found that participants relied little on either cue in the absence of semantic cues to speaker meaning. In the anticipation likelihood analyses (Analysis 1) we found no evidence for use of either cue (subject pronoun or utterance-terminal intonation contour). In the cluster-based timing analyses (Analysis 2), we found spotty evidence of use for the pronoun, only by two-year-olds and adults. As a reminder, an effect of the pronoun was evident in Study 1 across all age groups 2;0 and older. As before, we do see clear increases in anticipation likelihood across age (Casillas & Frank, 2017; Lammertink et al., 2015), there was no consistent evidence for an interaction between pronoun use and age.

We had considered the possibility that participants would be more likely to use intonation here in Study 2, when the lexical cues were uninformative (i.e., as a "plan B" of sorts), but this did not appear to be the case. Rather, the findings lend support to the idea proposed in the discussion of Study 1: that the late timing of the intonational cue we used may undermine its use for anticipation of upcoming responses.

In Study 1 children from age 2;0 showed sensitivity to a non-salient pronoun cue in predicting upcoming responses—the results of Study 2 suggest that use of this cue requires a semantically meaningful context. We can think of a few reasons why the lexical context of open class words might matter in these cases. First, there is the matter of recognizing the target utterances as question-formatted. Morphosyntactic, and to some extent lexico-syntactic, information is maintained in Jabberwocky languages. So, in theory, participants could identify the subject-verb inversion construction used in our target sentences. However, participants' ability to syntactically project this construction in real time depends on their recognition of the first word as a verb. Recognition of the first word as a verb is likely impaired by the use of non-words, leading to a later (or unsuccessful) identification of the intended syntactic construction which would then downgrade the perceived likelihood that the sentence is a question. This downgrading would reduce participants' gaze to the addressee and thereby minimize observable differences across the pronoun conditions. Second, accurate initial recognition of the syntactic construction may be key to catching onto these fleeting subject pronouns in time to make a prediction about upcoming response. That is, if participants do not process the sentence structure fast enough, they might miss the (non-salient) subject pronouns, and then cannot use them for making a prediction. Third, participants' broader sense of the topic and direction of conversation may have indirectly supported their sentence prediction in Study 1 to help in directing their attention to the critical subject pronoun cue.

Adult and child participants in past work have consistently demonstrated that they make turn taking predictions best when the whole linguistic signal is available (e.g., Casillas & Frank, 2017; de Ruiter et al., 2006).

There may be another explanation for the lack of a robust pronoun effect in Study 2: the Jabberwocky Dutch stimuli were more difficult and/or less interesting to watch, leading to disengagement. We mention this possibility given the somewhat higher rates of exclusion for Study 2 compared to Study 1. By design, Study 2 simulates the learning scenario in which we are most interested: what very young children do in conversation when they are faced with a linguistic signal that they don't understand. We predicted that they would use the accessible informative cues (intonation and pronoun), if only because it's the best information they have. However, it may be that children were simply not interested in closely attending to this linguistically challenging signal (Foushee et al.; Kidd et al., 2012). For example, they may have tracked the overall structure of the interaction reactively without attending closely to the anticipatory cues available in the unfolding utterances. Riest et al. (2015) propose that such reactive turn tracking could serve as a backup strategy in case timely turn-end anticipation fails. However, even though this is a possibility, we consider it less likely an explanation because even though we had somewhat higher rates of exclusion for Study 2 compared to Study 1, data loss in Study 2 was still relatively low.

The adult data can help us reason about these different explanations for the lack of a pronoun effect. Adults in both studies were closely attending to the stimuli: of the 160 adults across the two studies, only 3 had any videos excluded from a lack of valid gaze data (two in Study 1; one in Study 2; eight videos total between them). These adult native Dutch speakers, paying full attention to the stimuli, should have the best chance of anyone in the study when it comes to picking up on our subtle cues of interest. And yet we see no evidence that adults capitalize on the pronoun (or intonation) cue in the absence of a semantically meaningful context—even when that cue on its own can still weakly inform participants' predictions about what will happen next.

General Discussion

From around age 2;6, children begin to demonstrate a question advantage in their anticipation of upcoming responses when observing dyadic conversations (Casillas & Frank, 2017; Kalashnikova & Kember, 2020; Keitel et al., 2015, 2013; Lammertink et al., 2015). The present study aimed to better understand what drives this advantage in Dutch children and adults. In doing so we assessed children's and adult's use of specific, interactionally informative cues (rising versus falling intonation; first- versus second- person subject pronoun). We tested children at ages 1;0, 2;0, 3;0 and 4;0 to investigate developmental changes in the use of these cues, both in semantically rich (Study 1) and semantically limited (Study 2) contexts. To our knowledge, this was the first study in adults or children to look at the role of epistemic cues (first- versus second-person pronoun) in real-time response prediction during conversation.

Our first study shows that, from their second birthday, children start to use subject pronouns to anticipate upcoming conversational structure in semantically meaningful contexts. By age 3;0, their use of the pronoun cue in these contexts is similarly efficient to that of four year olds and adults. Thus, when presented with real Dutch sentences, only the youngest children (1;0) did not seem to benefit from the presence of subject pronouns in their anticipation of speaker switches. This pattern nicely coincides with the ages at which Dutch children are reported to start using subject pronouns in their spontaneous speech (2;0 for the first-person pronoun and 2;6 for the second-person pronoun; Bol and Kuiken (1986) and Verrips and Wijnen (1998)). In Study 2, when participants had reduced access to semantic content, we observed no stable evidence for a pronoun advantage in any of our participant groups, suggesting that semantic context plays an important role in participant's prediction of conversational structure. We also found no evidence that participants use terminal intonational contours in their anticipation of conversation structure, neither when semantic content was available (Study 1) nor when it was absent (Study 2).

In what follows we discuss these findings by first considering the challenges presented by different informative cues to upcoming turn structure. Then we assess what further evidence could be gathered to examine children's use of these particular cues over development.

Different cues, different challenges

The study findings support the idea that adults and children attend to individually informative linguistic cues when making real-time predictions about upcoming turn structure. However, the findings also suggest that not all cues are equally useful for such predictions. The two cues tested here—subject pronouns and utterance-terminal intonation contours—present different challenges to real-time prediction during conversation.

Subject pronouns. There are two points in favor of using subject pronouns to predict upcoming responses in conversation. First, as emphasized in the introduction, there is an epistemic asymmetry between the first-person and second-person subjects; use of a second-person subject pronoun can imply that the following utterance information is within the addressee's epistemic domain and is thus more likely to be used in an interrogative context than a first-person subject pronoun would be (for more see Heritage, 2012). The theoretical epistemic asymmetry then results in an empirical cue to questionhood: second-person subject pronouns are more closely associated with interrogative utterances than first-person subject pronouns are (Supplementary Materials S4). Subject pronouns also typically occur early in the utterance. Their early utterance position, combined with this epistemic asymmetry make them a potentially useful cue for timely utterance prediction. Yet subject pronouns also present some challenges in making such predictions. First, subject pronouns are often (and in our stimuli, always) prosodically unmarked. Second, they are only loosely associated with the need for a response in Dutch child-directed speech (Supplementary Material S4). Finally, pronouns are deictic words which may have impacted our outcomes in different ways.

First, as a reviewer suggested, the second-person pronoun "you" may draw children's attention not because it serves as an epistemic cue per se, but simply because it directly refers to the addressee. In this view, hearing "you" might lead children to look at the addressee automatically without necessarily anticipating turn structure. To explore this alternative explanation, we conducted an exploratory analysis on a subset of our non-question filler utterances (declaratives) that included either first- or second-person pronouns (see Supplementary Materials S3 for details). In these declaratives, children showed a second-person pronoun advantage in semantically and syntactically rich contexts (i.e., real Dutch), but not in the Jabberwocky condition (though a similar trend was observed). In our view, these findings are compatible with both accounts. The presence of a second-person advantage in declaratives with full semantic context aligns with our epistemic account, which predicts that "you" becomes informative about turn structure particularly when children can access meaning (see our explanation in the Introduction). At the same time, the trend toward a second-person advantage in Jabberwocky Dutch for children suggests that referential salience alone, i.e., hearing "you" and recognizing it as a cue to the addressee, may also contribute to their behavior. Notably, the absence of such an effect in adults under Jabberwocky conditions may indicate that adults rely more on semantic and syntactic information in order to treat "you" as a reliable cue to turn transition. In this light, our supplementary analyses (S2) on the 650 ms time window following pronoun onset are also informative. If children were simply responding to the referential salience of you, one would expect them to shift their gaze toward the addressee immediately after the pronoun was spoken – akin to a word recognition effect observed in classic looking-while-listening paradigms. However, we find no clear evidence for such an effect in our data. This suggests that children's predictive looking patterns are more likely driven by interactional parsing processes than by an immediate response to deictic cues.

Second, the deictic nature of the personal pronouns may make the cue challenging for children because it depends on a contextual frame as well. That is, the meaning of the pronoun depends on a contextual frame, in this case, the speaker's personal perspective (e.g., Fillmore, 1997). As it happens, perspective-taking and full theory of mind skills are mastered relatively late in development, long after children begin producing subject pronouns (Kuiper et al.,2021); for more on theory of mind see this recent review: Rakoczy, 2022). Cumulatively, these factors help explain why pronouns make for challenging predictive cues, and may clarify why we did not observe a pronoun advantage outside of semantically meaningful contexts—even by adults. The adults, who have fully developed psycholinguistic capacities and whose attention didn't waver in this more challenging linguistic context, still did not show a pronoun advantage without recognizable lexical content. Semantic cues may also moderate epistemic ones for adult listeners; for example, contextual information may serve a critical role in confirming that the addressee is indeed more of an epistemic authority on the topic at hand when the speaker uses a second-person subject pronoun.

Utterance-final intonation. In contrast, we have the case of intonation. The major point in favor of intonation is that rising terminal intonation is canonically associated with polar interrogatives in Dutch (as it is in English). However, not all polar questions feature rising terminal intonation (Caspers, 1998; Englert, 2010; van Heuven & Haan, 2000). Perhaps more importantly, this cue comes at the very end of the utterance, leaving little time for the participant to integrate it and launch a predictive gaze switch before the next utterance begins. Indeed, intonational cues have been shown in prior work to effectively aid Dutch-speaking participants when revising inaccurate turn-end anticipations (Bögels & Torreira, 2015), but not in predicting the precise timing of turn-ends (de Ruiter et al., 2006, but also see Bögels & Torreira, 2021) or upcoming turn structure (Casillas & Frank, 2017; Lammertink et al., 2015). Here we similarly find no impact of terminal intonation contour on children or adults' anticipatory gaze switches, on its own or in combination with pronominal cues, with or without a semantically meaningful context.

The difference we've observed in participants' use and non-use of these cues across age and semantic contexts gives the impression that participants both (1) attend to individually informative cues, and (2) rely on a collection of additional information sources to support their use of those cues. We had originally wondered whether participants indeed use individual cues and, if so, whether those cues could serve as an important foot in the door to children's linguistically challenging conversations. For now, it appears that if such a cue appears early, and if it's sufficiently reliable, children (and adults) will use it. What is less clear is whether such cues really serve as a 'foot in the door' for children—if participants depend on access to the semantic and syntactic context in order to use the cue, that also implies that the same participants can already process a good portion of that context (when it's available). If they can process most of the lexical and syntactic content already, the cue may only provide minimal additional help in making predictions. We should then perhaps test our 'foot-in-the-door' prediction with cues that are less conceptually complex (e.g., concrete, non-deictic referents) or that take advantage of natural cue combinations in conversation (e.g., verb-pronoun combinations that reliably cue interrogativity, like *mag ik* [may I]).

Future work

The current work takes a first step toward understanding whether children's capacity to make predictions about conversation is driven primarily by a few, small (but key) cues in the unfolding speech. The answer to this question will help us reason about the developmental link between linguistic knowledge and interactional skill. Our present findings suggest that children require a suite of accessible and sufficiently informative cues to make predictions about turns. However, our findings also highlight important issues to pursue in future work, including a better understanding of why this is the case. For example, does the absence of semantic and/or syntactic context lead to more disengagement and therefore less attentive and precise tracking of conversations? Or is semantic context needed to give meaning to epistemic cues to questionhood? Future work might target the psycholinguistic interplay between individually informative linguistic cues and their semantic, syntactic, and prosodic context in predicting conversational structure. For example, the absence of an effect for children (and adults') use of utterance-final intonation in the present study does not mean that participants ignore this cue; participants likely use utterance-final intonation as a critical (dis)confirmatory cue to turn-end, especially if the utterance is perceived as syntactically incomplete (Bögels & Torreira, 2015, 2021; de Ruiter et al., 2006; Keitel et al., 2015, 2013; Lammertink et al., 2015). Furthermore, future work would ideally find ways to examine children's use of these cues while the children themselves are participants in the interaction and/or take part or observe multimodal conversations. The present studies rely on third-party observation to understand children's comprehension process in the absence of pressure to formulate their own responses and in the absence of non-linguistic cues like body movement, breaths and eye gaze. However, in order for these results to generalize to children's everyday language use, we should find ways to examine these cues in first-person multimodal interaction. Finally, we would like to emphasize that an utterance's subject pronoun, as studied here, is only one example of a potential, individual linguistic cue that children can use to make predictions at the interactional level of conversation. In other languages—particularly pro-drop languages, where subject pronouns are often omitted—such a cue would rarely be useful for real-time predictions about upcoming turn structure. We expect that, across languages, the kinds, quantities, and reliabilities of linguistic cues to questionhood will vary. To the extent that reliable cues are available in a given language, we expect to find predictive use of these cues for upcoming turn structure in diverse languages.

Conclusion

Beginning around age 2;6, children begin to consistently anticipate upcoming responses when observing conversation. We examined whether such predictions might be driven by individually informative linguistic cues, comparing two cues associated with interrogatives: one originating in epistemic asymmetries and occurring early in the utterance (subject pronoun) and one canonically associated with questions and occurring late in the utterance (utterance-final intonation). We find that children from age 2;0 make more and earlier anticipations of an upcoming addressee response when hearing the early epistemic cue (you vs. I subject pronouns), but we find no evidence that their predictions are changed by the later prosodic cue. Participants' use of these predictive cues depends on speakers' use of semantically meaningful content, suggesting that the use of these individually informative cues still depends on a basis of broader linguistic context to support participant predictions. The findings take us a step closer to understanding how linguistic and interactional skills become intertwined in development.

CRediT authorship contribution statement

Imme Lammertink: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Maartje de Vries:** Writing – review & editing. **Caroline Rowland:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Marisa Casillas:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are deeply grateful to the hundreds of families who participated in this work, the stellar research assistants who collected data (Roos Greven, Elske Gros, Cielke Hendriks, Daphne Jansen, Inge Stok, Liz Tollenaar, Sam Teunissen), and the lab managers who organized this team (Patricia Manko and Nienke Rulkens-Dijkstra). We thank Marjolein Ankone and Maartje Weenink for lending us their voices for the stimuli. Parts of this work were presented at IASCL 2017 and 2024, and we are grateful for the audience comments and questions, which helped improve this work

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2025.106358.

Data availability

I have shared the data/code via an OSF-link in the manuscript

References

Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis. *Child Development, 92*(2), 484–501.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer [computer program]*. http://www.praat.org/.

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.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics, 52*, 46–57.

Bögels, S., & Torreira, F. (2021). Turn-end estimation in conversational turn-taking: The roles of context and prosody. *Discourse Processes, 58*(10), 903–924.

Bol, G. W., & Kuiken, F. (1986). Het gebruik van pronomina bij kinderen van een tot vier. *Toegepaste Taalwetenschap in Artikelen, 24*, 47–58.

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language, 2*(01), 1–19.

Bruner, J. S. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy, 1*(1), 111–114.

Carroll, L. (1871). *Through the looking-glass and what Alice found there*. Macmillan.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337.

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.

Caspers, J. (1998). Who's next? the melodic marking of question vs. *Continuation in Dutch. Language and Speech, 41*(3–4), 375–398.

Christophe, A., Guasti, T., Nespor, M., Dupoux, E., & Ooyen, B. (1997). Reflections on prosodic bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes, 12*, 585–612.

Clark, E. V. (2006). Color, reference, and expertise in language acquisition. *Journal of Experimental Child Psychology, 94*(4), 339–343.

Clark, E. V. (2007). Young children's uptake of new words in conversation. *Language in Society, 36*(2), 157–182.

Clark, E. V. (2022). Language is acquired in interaction. In S. Lappin, & J.-P. Bernardy (Eds.), *Algebraic structures in natural language* (pp. 77–94). CRC Press.

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture, 11*(1), 1–23.

Clark, E., & Lindsey, K. (2015). Turn-taking: a case study of early gesture and word use in answering Where and Which questions. *Frontiers in Psychology, 6*.

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review, 57*, Article 100914.

Davis, M. H. (2000). *Lexical segmentation in spoken word recognition*. University of London. PhD thesis.

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.

Dink, J W, & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*. Retrieved from http://www.eyetrackingr.com.

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development, 92*(2), 609–625.

Endevelt-Shapira, Y., Bosseler, A. N., Mizrahi, J. C., Meltzoff, A. N., & Kuhl, P. K. (2024). Mother–infant social and language interactions at 3 months are associated with infants' productive language development in the third year of life. *Infant Behavior and Development, 75*, Article 101929.

Englert, C. (2010). Questions and responses in Dutch conversations. *Journal of Pragmatics, 42*(10), 2666–2684.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260.

Estigarribia, B. (2010). Facilitation by variation: Right-to-left learning of English yes/no questions. *Cognitive Science, 34*(1), 68–93.

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Word learning emerges from the interaction of online referent selection and slow associative learning. *Proceedings of the National Academy of Sciences, 113*(41), E6256–E6262.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.

Figueroa, C., Beňuš, Š., & Skantze, G. (2023). Prosodic alignment in different conversational feedback functions. *20th International Congress of Phonetic Sciences (ICPhS)*.

Fillmore, C. J. (1997). *Lectures on deixis. CSLI Publications*.

Fitneva, S. (2012). Beyond answers: Questions and children's learning. In J.-P. de Ruiter (Ed.), *Questions: Formal, functional, and interactional perspectives* (pp. 165–178). Cambridge, UK: Cambridge University Press.

Foushee, R., Srinivasan, M., & Xu, F. (in revision). Selective attention to spoken language in preschoolers based on speech complexity and learning rate. *PsyArxiv Preprints*, 1–8.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. Cambridge, MA: MIT Press.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*(5), 515–523.

Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction, 45*(1), 1–29.

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: Optimizing recipiency. *Frontiers in Psychology, 6*(98), 1–14.

Johnson, M. A., & Goldberg, A. E. (2013). Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes, 28*(10), 1439–1452.

Kalashnikova, M., & Kember, H. (2020). Prosodic cues in infant-directed speech facilitate young children's conversational turn predictions. *Journal of Experimental Child Psychology, 199*, Article 104916.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156.

Kaufeld, G., Rutger Bosker, H., ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience, 40*(49), 9467–9475.

Keitel, A., Daum, M. M., et al. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology, 6*(108), 265–273.

Keitel, A., Prinz, W., Friederici, A. D., Von Hofsten, C., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology, 116*(2), 264–277.

Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B, 378*(1875), Article 20210473.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One, 7*(5), Article e36399.

Knightly, L. M., Jun, S.-A., Oh, J. S., Kit-fong, Au., & T.. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America, 114*(1), 465–474.

Kruyt, J., de Jong, D., D'Ausilio, A., & Beňuš, Š. (2023). Measuring prosodic entrainment in conversation: A review and comparison of different methods. *Journal of Speech, Language, and Hearing Research, 66*(11), 4280–4314.

Kuiper, S. J. M., Hartman, C. A., & Hendriks, P. (2021). Children's pronoun interpretation problems are ralated to theory of mind and inhibition but not working memory. *Frontiers in Psychology, 12*(610401), 1–16.

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), 274–291.

Latif, N., Alsius, A., & Munhall, K. (2017). Seeing the way: The role of vision in conversation turn exchange perception. *Multisensory Research, 30*(7–8), 653–679.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 831–877.

Meylan, S. C., Foushee, R., Wong, N. H., Bergelson, E., & Levy, R. P. (2023). How adults understand what young children say. *Nature Human Behaviour, 7*(12), 2111–2125.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1/2), 1–135.

Nelson, K. (1985). *Making sense: The acquisition of shared meaning*. Academic Press.

Pomerantz, A., & Heritage, J. (2013). Preference. In J. Sidwell, & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 210–228). Blackwell.

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology, 1*(4), 223–235.

Ratner, N., & Bruner, J. S. (1978). Games, social exchange and the acquisition of language. *Journal of Child Language, 5*(03), 391–401.
Riest, C., B, A., & de Ruiter, J.-P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology, 6*(495), 62–75.
Robinson, J. D. (2020). Revisiting preference organization in context: A qualitative and quantitative examination of responses to information seeking. *Research on Language and Social Interaction, 53*(2), 197–222.
Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.
Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*(5916), 951–953.
Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language, 47*(1), 5–21.
Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences, 112*(41), 12663–12668.
Sacks, H. (1973/1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button, & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 54–69). Multilingual Matters.
Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.
Schlichting, J. E. P. T. (1996). *Discovering syntax: An empirical study in dutch language acquisition*. Nijmegen University Press.
Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive Psychology, 10*(3), 271–301.
Siemund, P. (2001). Interrogative constructions. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 1010–1027). Walter de Gruyter.
Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*(1), 1–22.
SR Research Ltd. (2005). EyeLink 1000. [Apparatus and software]. https://www.sr-research.com/.
Stivers, T. (2013). Sequence organization. In J. Sidwell, & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 191–209). Blackwell.
Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction, 43*(1), 3–31.
Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics, 124*, 14–30.
Suarez-Rivera, C., Schatz, J. L., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy, 27*(2), 232–254.
Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development, 90*(6), 2135–2152.
Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23*(2), 121–126.
Ultan, R. (1978). Some general characteristics of interrogative systems. In J. H. Greenberg (Ed.), *Universals of human language, volume 4: syntax* (pp. 211–248). Stanford University Press.
van Heuven, V. J., & Haan, J. (2000). When and how do we hear whether a dutch speech utterance is a statement or a declarative question? *Linguistics in the Netherlands, 17*(1), 93–104.
Verrips, M., & Wijnen, F. N. K. (1998). The acquisition of dutch. In S. Gillis, & A. de Houwer (Eds.), *The acquisition of dutch syntax* (pp. 233–299). John Benjamins.
Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science, 25*(7), 1314–1324.
Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One, 8*(11), Article e79659.
Zhang, V. H., Elmlinger, S. L., & Goldstein, M. H. (2024). Developmental cascades of vocal turn-taking connect prelinguistic vocalizing with early language. *Infant Behavior and Development, 75*, Article 101945.