Predicting conversational turns: Signers' and non-signers' sensitivity to language-specific and globally accessible cues

Connie de Vos, Tilburg Center for Cognition and Communication, Tilburg University, clgdevos@uvt.nl

Marisa Casillas, Department of Comparative Human Development, the University of Chicago, mcasillas@uchicago.edu

Tom Uittenbogert, Center for Language Studies, Radboud University, tom.uittenbogert@let.ru.nl

Onno Crasborn, Center for Language Studies, Radboud University, onno.crasborn@ru.nl

Stephen C. Levinson, Language and Cognition Department, Max Planck Institute for Psycholinguistics, Stephen.Levinson@mpi.nl

Predicting conversational turns: Signers’ and non-signers’ sensitivity to language-specific and globally accessible cues

Predicting conversational turns: Signers' and non-signers' sensitivity to language-specific and globally accessible cues*

Precision turn-taking may constitute a crucial part of the human endowment for communication. If so, it should be implemented similarly across language modalities, as in signed vs. spoken language. Here in the first experimental study of turn-end prediction in sign language, we find support for the idea that signed language, like spoken language, involves turn-type prediction and turn-end anticipation. In both cases, turns eliciting specific responses like questions accelerate anticipation. We also show remarkable cross-modality predictive capacity: non-signers anticipate sign turn-ends surprisingly well. Finally, we show that despite non-signers' ability to intuitively predict signed turn-ends, early native signers do it much better by using their access to linguistic signals (here, question markers). As shown in prior work, question formation facilitates prediction, and age of sign language acquisition affects accuracy. The study thus sheds light on the kind of features that may facilitate turn-taking universally, and those that are language-specific.

Keywords: turn-taking, turn-end anticipation, interactional linguistics, conversation analysis, discourse processing, Sign Language of the Netherlands, gesture

* With a project this size, we inevitably have many people to thank for their contribution to this paper. First and foremost, we would like to thank all our participants, deaf and hearing, for their time and attention in participating in these perception experiments. Our experiments were carried out in a mobile lab at various locations and we are highly indebted to the hospitality and generosity we received in each of these locations (Stichting Welzijn Doven Rotterdam, Stichting Welzijn Doven Amsterdam, Utrecht University of Applied Sciences, De Gebarenkorf, Stichting Welzijn Doven Drenthe, Koninklijke Kentalis Zoetermeer), in addition to the deaf and hearing families who allowed us to park on their sidewalk (Fam. Hartzema, Fam. Uittenbogert, Iris Wijnen & Nico Borst, and Mirjam-Iris Crox & Thomas op de Coul). We also thank all four deaf signers, who preferred not to be mentioned here by name, for allowing us to record their

spontaneous conversations as a basis for the stimulus materials, and Nick Wood for editing them. We also thank Merel van Zuilen for modeling the images in in Figure 1, Ellen Nauta and Marjolein Ankone for their assistance in data collection, and Frouke van Winsum for her help in data transcription.

**1.** INTRODUCTION Everyday conversation is the primary mode of language use across human societies. During conversation, we take turns at talking by responding to each other contingently. A turn at talk in conversation is produced by a single interactant, is composed of components such as words or clauses, is variable in length, typically ends with an opportunity for turn-transition (or else a lapse in conversation or the end of the interaction), and generally isn't pre-specified in advance with respect to what will be produced or who will take a turn next (Sacks et al. 1974). Despite this variability in turn form, length, and allocation between interactants, turns in conversation are typically taken swiftly. Transitions between turns in adult conversation average around 200 msec cross-linguistically (Stivers et al. 2009; de Vos et al. 2015). This average transition time is fast—at one third the time it takes to plan a single word and one eighth the time it takes to plan a simple transitive sentence, everyday conversational turn-taking is only made possible by the orchestration of multiple linguistic processes engaged in real time as the current turn unfolds (see Levinson & Torreira 2015 for a review). Such a brief temporal window arises naturally as a consequence of coordinating who talks when (Sacks et al. 1974), and aids in the early identification and management of delayed responses (e.g., hesitations) and communicative breakdowns (Bögels et al. 2015; Kendrick & Torreira 2015; Pomerantz & Heritage 2012).

While the typical timing for a contingent response varies greatly across the animal kingdom (Pika et al. 2018), the human turn-taking system is special in maintaining a rapid response norm while *also* supporting immense flexibility in what is talked about. Human conversations are not pre-formulated, but rather collaboratively organized in both content and timing. Each turn progresses the interaction forward, bit by bit, making different kinds of responses relevant along the way: some turns demand specific next actions (e.g., "How many days will they be away?"), others fulfill these demands (e.g., "Four."), refer to points of potential communicative breakdown (e.g., "Who, our neighbors?"), or make many other kinds of interactive plays (see e.g. Schegloff 2007 for an overview of conversational sequencing). Given the shifting and flexible nature of relevant responses in human conversation, the observed pattern of rapid transitions between turns at talk presents a truly impressive psycholinguistic puzzle.

**1.1.** THE ROLE OF PREDICTION IN RAPID TURN TRANSITIONS This ability to quickly and smoothly transition between speakers requires the responder to both (a) track incoming speech for cues as to when a response might be needed and, if so, what sort, as well as to actually (b) plan that response such that it is ready to launch immediately at the end of the current speaker's turn (Levinson & Torreira 2015). If the ongoing turn contains overt linguistic or gestural cues that indicate the need for a response (e.g., "Where is …", "..., isn't it?", or a point at the addressee), those cues alone may allow the addressee to effectively and immediately judge that a response is required, and, in many cases, allow the inference of what type of response it should be (Stivers & Rossano 2010). Turn-structure prediction, which encompasses both predicting the ends of unfolding turns as well as their hoped-for responses, is thus facilitated by the use of response-eliciting cues by the current speaker. The repertoire of cues contributing to response elicitation includes both linguistic and non-linguistic signals (Stivers & Rossano 2010), but work exploring the psycholinguistic processes underlying conversation has typically privileged the former.

In a landmark study, de Ruiter and colleagues (2006) developed an experimental method for investigating the linguistic cues by which listeners could predict the upcoming ends of ongoing turns. They recorded dyadic conversations between Dutch speakers, extracted individual turns at talk, and manipulated them to control for the availability of lexicosyntactic and prosodic cues, among others. Other Dutch-speaking participants then listened to the manipulated turns and pressed a button at the moment they felt each turn was about to end, being encouraged to anticipate that point. They found that participants, above all, required lexicosyntactic information to accurately predict upcoming turn ends, whereas intonational information on its own did not lead to accurate prediction. A number of studies since then have used similar paradigms or eye-tracking to identify other linguistic cues that may aid online turn-structure prediction. For example, Bögels and Torreira (2015) used a similar paradigm to establish that prosodic cues are crucial in the final moments of the turn to disambiguate whether the speaker will continue on with another increment of talk or will finish their speaking turn. This role for prosody is important because many turns in natural conversation contain multiple points of potential completion (i.e., multiple turn-constructional-units (TCUs); Sacks et al. 1974), where each 'unit' is a prosodically, syntactically, and pragmatically complete clause of any length).[1] So while lexicosyntax demonstrably plays a crucial role in identifying the precise moment of turn end (i.e., the end of the last TCU), prosody also plays a pivotal role in multi-TCU turns for anticipating whether the current turn-constructional unit is intended as the last one and, therefore, whether speaker transition is likely to take place soon.

Linguistic cues are only part of the picture when it comes to turn-related predictions. Similar experimental paradigms to that used by de Ruiter and colleagues (2006) have also been used to investigate the use of non-linguistic visual information in predicting upcoming turn structure, primarily finding that visual information aids accurate prediction for spoken conversation (Latif et al. 2018; Keitel & Daum 2015). These experimental findings are complemented by those from systematic analyses of naturalistic observational data, including data from head-worn eye-trackers in spontaneous conversation (Holler et al. 2018; Kendrick & Holler 2017; Holler & Kendrick 2015; Rossano et al. 2009; Stivers et al. 2009). In fact, several studies have noted that, even when the observer does not understand the conversational speech (e.g., because it is in an unfamiliar language), they can still spontaneously predict upcoming responses (Casillas & Frank 2017) and reliably identify turn-end boundaries (Carlson et al. 2005; Fenlon et al. 2007)—abilities which may be attributed to the integration of multiple non-linguistic cues and/or adaptation of related linguistic knowledge. The bigger picture emerging from this literature is that conversational participants are experts at tracking and capitalizing on cues to upcoming response, be they linguistic or not.

The ability to track and capitalize on cues to upcoming response is also observed in passive experimental studies of turn-taking, in which participants spontaneously predict upcoming responses when observing a conversational exchange. Despite not being participants in the exchange themselves, participants appear to employ similar anticipatory processes to what they use in first-person interaction, perhaps as the outcome of simply attempting to follow the conversation (Casillas & Frank 2017). Across multiple such studies, questions have shown a privileged status in participants' predictions. For example, when asked to watch a video of dyadic conversation, adults and children ages two and above are significantly more likely to look at the upcoming addressee when they hear a question than when they hear a non-question (Casillas & Frank 2017; Lammertink et al. 2015; see also Keitel & Daum 2015; Keitel et al. 2013). Indeed, returning to observations of natural interaction, many of the documented rapid responses for adult conversation across typologically diverse languages come from the transitions between questions and their answers (Stivers et al. 2009; Holler et al. 2018). Taking all this evidence together, language users appear to be both motivated and well-equipped to track the incoming signal for signs of response elicitation, giving questions and other similarly response-eliciting turn-types priority in the multi-threaded psycholinguistic process of comprehending and producing during conversation.

While research investigating the psycholinguistic processes underlying turn-taking has built on a typologically and methodologically rich collection of findings, few studies have experimentally investigated these processing effects in signed languages, and there are none as far as we know that have experimentally investigated the ability of non-signers to parse the turn-structure of sign language. Meanwhile, the INTERACTION ENGINE HYPOTHESIS (IEH; Levinson 2006, 2019) proposes that a human-specific propensity for coordinating with others on joint activities should lead to interactional skills (like turn-taking) emerging early in development and in similar ways across the world's language communities, including signing communities. By using both native signers and non-signers in the current study, we are able to tease apart the relative contribution of linguistically coded and non-linguistically coded response-eliciting cues within the same visual modality, potentially giving substantial support for the special role that a universal interactive competence may play in communicative interaction.

Our central research question in the current study is whether similar predictive processes underlie turn-end anticipation in signed and spoken languages. Specifically we investigate whether: (a) response-seeking utterances initiate advantaged predictive processing in signed turns, as they do in spoken turns, (b) cues with a shared basis in sign and co-speech gesture are utilized by signers and non-signers alike in turn-end prediction, and (c) cues specific to the sign language are only leveraged for turn-end prediction by its signers, with an advantage for signers who began learning it earlier in life.

To test these predictions, we develop a modified version of de Ruiter and colleagues' classic (2006) button-press experiment, specifically adapting it to evaluate turn-end prediction in Sign Language of the Netherlands (Nederlandse Gebarentaal, henceforth NGT). As described above, this method places experimental participants as 'stand-in' addressees who respond to turns extracted from a pre-recorded spontaneous conversation. They are asked to press a button at the moment they feel each turn is about to end, being encouraged to anticipate that point; the outcome is thereby purely a measure of participants' response to the unfolding linguistic structure of the current turn. We chose this controlled experimental method over, e.g., analysis of spontaneous signed conversation (de Vos et al. 2015), because it allows us to analyze an overt indicator of real-time turn-end prediction that is similarly executed across turns and participants (i.e., the button press), to re-use the same stimuli across multiple participants and thus identify how specific types of linguistic cues consistently influence predictions, and to test the same stimuli with both signing and non-signing participants. We recruited a diverse sample of NGT signers and a matched sample of hearing, non-signing Dutch speakers to participate in the experiment. We also used highly naturalistic stimuli (described below) to analyze the role of linguistically coded (henceforth 'language-specific') and 'globally-accessible' (i.e., not language specific; further defined below) response-eliciting cues for turn-end prediction by signers and non-signers.

**1.2.** VISUAL SIGNALS TO TURN BOUNDARIES AND TRANSITIONS Across time and space, sign languages have arisen spontaneously from the social interaction among deaf individuals and between deaf and hearing individuals. Despite the fact that they are produced gesturally and perceived visually and/or tactically, signed languages, both urban and rural, display evidence of linguistic organization parallel to the phonological (i.e., sub-morphemic) level (Stokoe 1960; Brentari 1998; van der Kooij 2002), the morphosyntactic level (Wilbur 1987; Sandler & Lillo-Martin 2006), and the prosodic (or paralinguistic) level (Sandler 1999; Brentari & Crossley 2002; Russel et al. 2011) of spoken languages (see also Senghas & Coppola 2001, de Vos & Nyst 2018, de Vos & Pfau 2015; Zeshan & Palfreyman 2017). This also likely extends to aspects of language in use that are associated with social interaction: turn-taking and communicative act development (Casillas & Hilbrink 2020).

Earlier work on sign language interaction has suggested that the signed modality may allow for more overlap between consecutive turns than in the auditory modality, because, unlike the interference effects of speaking and listening, signed language is perceived visually but produced motorically (Baker 1977; Emmorey et al. 2009). For this reason, overlap between consecutive turns at talk may not be as problematic as it would be in spoken conversations. In line with this assumption, Coates and Sutton-Spence (2001) claimed that a group of four British signers, who were close friends, allow for more overlapping talk in signed conversation than has typically been found for spoken conversation. In spoken conversation, too, visible cues used in the coordination of turns at talk appear to influence interactional timing, typically associated with speedier responses, even when prosodic patterns are taken into account (Stivers et al. 2005; Holler et al. 2018; Holler & Levinson 2019). If our ultimate aim is to reveal the cognitive mechanisms that allow us to smoothly coordinate during everyday conversation, we must take these observations into account; the affordances of visual signals might have substantially shifted the functioning of the turn-taking system.

Notwithstanding these initial observations, recent corpus analyses demonstrate sign language users are in fact as sensitive to turn-boundaries as spoken language users. In making a comparison between signed and spoken modalities, however, allowance must be made for a key difference: whereas preparatory motor articulations are mostly not visible inside the mouth, they *are* visible on the hands: signed languages use heavy articulators, with a lot of inertia, which move a large distance between rest position and signing space. Therefore signers display early preparatory and late retracting movements, engendering smooth transitions in and out of a conversational turn, leaving the impression of overlapping talk, while new content is not (yet) provided (McCleary & Leite, 2013 for Brazilian Sign Language; Groeber & Pochon-Berger 2014 and Girard-Groeber 2015 for Swiss German Sign Language; Manrique & Enfield 2015 on Argentinean Sign Language, and Byun et al. 2018 on cross-signing).

In further support of the IEH, corpus analyses of NGT have revealed that when the physical preparation and retraction phases needed for sign articulation are taken into account—that is, when we only look at so-called STROKE-TO-STROKE TURN BOUNDARIES (i.e., the linguistic signal and not its preparation or fade-out)—turn-timing in NGT conversations looks remarkably similar to turn-timing in a diverse set of spoken languages (cf. Stivers et al. 2009, and de Vos et al. 2015). Casillas and colleagues (2015) then experimentally tested this observed pattern by asking NGT signers to watch signed turns and press a button at the moment they thought the turn was about to end; even when cutting off the turns at the end of the final stroke, participants were able to accurately and reliably anticipate the turn end. Beukeleers and colleagues (2020) provide further converging evidence for the stroke-to-stroke boundary hypothesis by showing that signers of Flemish Sign Language spontaneously produce anticipatory gaze shifts towards turn-end boundaries in a mobile eye-tracking study. It is yet unknown what kinds of linguistic cues drive signers' predictions about upcoming turn ends.

As predicted by the IEH, cross-linguistic work on spoken languages indicates that when hearing individuals are asked to listen to an unfamiliar spoken language, they can still identify some aspects of turn structure, and even predict upcoming responses by a third party to some degree (Carlson et al. 2005; Casillas & Frank 2017). This indicates that at least some cues to turn structure do not require language-specific knowledge. In line with this hypothesis, although focusing on monologue rather than dialogue, Fenlon and colleagues (2007) showed two story retellings in British Sign Language (BSL) and in Swedish Sign Language (SSL) to four signers of BSL and to four speakers of British English. After two practice sessions, the participants were asked to view the narrative twice more and press a button whenever they spotted a sentence boundary. The valid window for which responses were included as correct was defined as 500 msec before and 1000 msec after the final frame in which the handshape of the sentence-final sign was still held (i.e., the end of the linguistic signal). There were no significant group differences, with participants identifying "pauses, drop hands, and holds" as the most reliable cues to the sentence boundaries (Fenlon et al. 2007:192ff). All these manual prosodic cues involve the cessation (or preparation for cessation) of hand movements and are effectively equivalent to inter-utterance silence in spoken language data.

Brentari and colleagues (2011) asked eight ASL signers and eight non-signers (Purdue undergraduates) to watch constructed stimuli which were recorded as infant-directed signing (IDS) in the presence of a 16-month-old addressee (for more on IDS in ASL, see Holzrichter & Meier 2000). Participants first watched a long segment of IDS and were then subsequently asked whether or not a sentence break was present between the two target signs presented as stills on their paper form. All signs were controlled for syllable count and identical sets of target signs were contrasted within the stimulus set, but with different intonational phrasing. Participants then indicated on a Likert scale how confident they were of a sentence break between the two signs. Again, there were no significant differences between the signers and non-signers in terms of their ability to identify prosodic breaks and non-breaks between the signs. A further analysis of prosodic cues indicated that whereas signers rely mostly on pauses to identify the breaks, the non-signers were additionally sensitive to the dropping of hands and holds. Due to the nature of these cues, Brentari and colleagues suggested that it is possible that adult non-signers could do this segmentation reliably given their lifelong experience with gesture. In order to determine whether non-signers segment on the basis of such gestural competence, they also tested 24 hearing non-signing nine-month-olds in a visual fixation procedure and found similar results on the same stimuli, suggesting that prosodic boundaries can be identified on a non-linguistic basis and even by those with little experience with co-speech gesture.

While prior work has therefore shown that speakers and signers alike can detect utterance segments on the basis of cues that are, for them, either linguistic or non-linguistic, few studies have measured whether this ability is exercised in real time during interaction, and none have investigated what role these cues play in predicting upcoming turn structure. In the current study, we probe the extent to which linguistic knowledge is required to reliably anticipate turn-ends in signed conversation. We gather data from signing participants and a hearing, non-signing control group that was matched to the signers as well as possible for participant age, gender, and education. We analyze the data with respect to differences in anticipation patterns across these groups.

While it is clear that sign language grammars are not in any way dependent on or derived from spoken language grammar, there are transparent and active links between the expressive repertoire of sign languages and the co-speech gesture systems in which they have emerged (e.g., Janzen & Schaffer 2002, Le Guen 2012, de Vos 2015, Tano & Nyst 2018). A general finding across sign languages is that, while these gestural forms are still used by hearing individuals alongside speech, signers use them with a higher degree of systematicity to the extent that they have in some cases become obligatory grammatical markers (e.g., Janzen & Schaffer 2002, Pfau & Steinbach 2006). An example of this is the use of eyebrow movements in concert with either speech or manual signing. Ekman (1979) notes that brow frowns are associated with the expression of puzzlement, while brow raises are associated with surprise for speakers of American English. In NGT, polar questions are grammatically marked by brow raises, while content questions feature a frown (Coerts 1992; de Vos, van der Kooij & Crasborn 2009). Hence, while both speakers and signers use these signals communicatively, they have become an intricate aspect of linguistic question-marking in NGT, as well as many other signed languages (Zeshan 2004). Because of the gestural or paralinguistic origins of these signs, in the present study we assume some degree of accessibility to non-signers as well, due to their experience with such cues in spoken conversations (see also Mondada 2007; Brentari et al. 2011).

In addition to these prosodic markers, which function on par with question intonation in spoken languages, NGT has multiple lexical signs associated with questions. On the one hand, there are a set of signs with gestural origins, e.g., palm-up and pointing, which may be accessible to non-signers as shown in Figure 1. Prior work on NGT, had already shown that both index finger points and palm-up gestures frequently occur in phrase-final position (Crasborn et al. 2012), and thus are potential markers of TCUs. On the other hand, there are lexical question signs (HOW, HOW-MANY, WHO etc.; see Figure 1 for examples), for which no conventional gestures or emblems are attested in the surrounding Dutch co-speech gesture system (Crasborn & Akkermans 2020). Hence, this group of question cues clearly relies purely on linguistic convention, and hence is unlikely to be semantically accessible to non-signers. In NGT these question signs may either occur in sentence-final position, sentence-initial position, or in both (Coerts 1992; de Vos et al. 2009). Moreover, when the question sign is used at the beginning of the question, it is often combined with a palm-up in sentence-final position. For both reasons, we hypothesized both palm-ups and index finger points would function as turn-final cues that invite an addressee response. In sum, the specific combination of cues with a clear gestural origin and cues that arise from linguistic convention makes questions in NGT of particular interest in this study, which includes NGT signers (who have access to both types of cues) and non-signers (who only have access to the first type). We hereafter refer to these two types of cues as 'globally-accessible' vs. 'language-specific' with respect to our study population (NGT signers and non-signing Dutch speakers).

<INSERT FIGURE 1 ABOUT HERE>

FIGURE 1. Lexical cues to questionhood present in the stimulus set. These images were produced with the permission of Merel van Zuilen.

By investigating turn prediction in NGT conversation with signers and non-signers, we experimentally extend prior work suggesting cross-linguistic competence in identifying boundaries in upcoming turn structure (Brentari et al. 2011; Carlson et al. 2005; Casillas & Frank 2017; Fenlon et al. 2007) while also systematically testing how linguistic knowledge (cue type) influences participants' ability to anticipate what will come next. Specifically, we hypothesized that NGT signers and non-signers would be able to predict turn ends in NGT conversation, that both would show an advantage for questions over non-questions, and that response accuracy would benefit both groups when turns contained globally-accessible response-eliciting cues, but that only signers would show a benefit of questions when the critical cue was specific to NGT. By capitalizing on visual cues shared between NGT signers and non-signing Dutch speakers in this way, the current study critically examines the extent to which linguistic knowledge contributes to turn-end prediction.

**2.** MATERIALS AND METHODS Following de Ruiter and colleagues' (2006) study, we asked participants to view short video clips from a dyadic conversation between two native NGT signers and to press a button at the moment they thought the ongoing turn was about to end. We introduced several modifications to the basic experiment design to adapt it for signed conversation as illustrated by Figure 2. First, target turns were presented after a short period of conversational context displaying both signers, after which only the target signer was shown, frozen for a moment, before their target turn began. Second, to ensure that the entire signing space, including non-manuals, was optimally comprehensible to participants, we recorded the stimulus conversations in a specialized set-up that allows participants to make eye contact during the conversation, despite them being seated in separate recording rooms. This specialized setup thereby gives experimental participants the frontal view they would have if they were the addressee in a normal dyadic conversation. Third, each target turn ended at the end of the final stroke (the last conventional linguistic signal), and not at the end of the retraction of the final sign, in line with corpus evidence for perceptual turn-end boundaries for signed conversation (de Vos et al. 2015), as further explained below.

We made special efforts to recruit a diverse sample of NGT signers to ensure that our results would generalize to as much of the community as possible. In particular, we note that the NGT community is heterogenous in age of acquisition, for both deaf and hearing signers. That is to say, it is estimated that only between five and ten percent of deaf children in Western societies acquire sign language from adults who are themselves native signers (Schein & Delk 1974; Kyle & Woll 1985), but the number could be even lower in the case of smaller deaf communities (Costello et al. 2008). Owing to the historical educational policy in the Netherlands, deaf children have oftentimes not received any sign language input until they were old enough to attend a deaf boarding school (Tijsseling 2014). On the other hand, availability of interpreter training and sign language teacher programs in recent times has boosted the acquisition of NGT among (mostly) hearing, young adults. Thus, there are large differences in the age at which NGT signers began to experience a sign-immersive environment. In what follows we give details on the NGT community as well as our methods of participant recruitment and stimulus selection, and a description of the structure of each trial.

<INSERT FIGURE 2 ABOUT HERE>

FIGURE 2. Stimulus creation and trial structure: (A) the video-chat recording set-up for each participant used a one-way mirror to enable mutual gaze during the conversation while recording high-quality front-angle footage of each participant, resulting in two aligned video streams that were; (B) mapped into the left and right regions of a joint video file, with one of the participants masked in a second (left-only) and third (right-only) version of that joint file, from which clips were extracted for each button press trial. Panel C illustrates the structure of an individual button press trial including these clips.

**2.1** PARTICIPANTS We recruited 64 NGT signers and 53 non-signing Dutch-speaking control participants for the current study. Because the NGT community, like many signing communities, is characterized by inter-individual variability in the onset and quality of sign language input, we sampled signers across a broad age range and with diverse language backgrounds and educational profiles. The NGT signers were classified as “early learners” (N = 32; started learning NGT before age 5) or "late learners" (N = 32; started learning NGT at age 5 or later) for present analyses (see Appendix Figure 1 for a demographic overview of the tested participants).

The early learners (Females = 18; 56%) all self-reported as deaf (deafness onset for all before age three), were 47.25 years old on average (range: 10–77; median: 47), and ranged in their linguistic input experience from hearing and/or deaf signers in their household (N = 7), other children and teachers at primary school (N = 14), or a combination of these (N = 11). Early signers also ranged in completed education from primary school to a professional bachelor's equivalent, and predominantly represented one of the three Western dialects of NGT, though many reported fluency in multiple dialects.

The late learners (Females = 21; 66%) included participants who self-reported as deaf (N = 20; 15 with deafness onset before age three, two after adolescence, and the rest in-between) and hearing (N = 12), were 41.56 years old on average (range: 19–76; median: 29), and ranged in their linguistic input experience from hearing and/or deaf signers in both their household and school (N = 4), other children and teachers at primary school (N = 11), professional education as an interpreter (N = 13; all the hearing late learners and one deaf late learner), or via social interactions in the NGT community (N = 4). Late signers also ranged in completed education from primary school to a professional and/or research bachelor's equivalent, and predominantly represented the Western dialects of NGT, with slightly more Northern signers than the early learner group.

The non-signing Dutch-speaking participants (Females = 37; 71%) were 44.77 years old on average (range: 13–80; median: 44) and ranged in completed education from primary school to a professional and/or research bachelor's equivalent.

In order to recruit participants from such a diverse sample, we created an ad hoc experiment room inside the back of a utility van and then drove this mobile lab to multiple NGT sub-communities around the Netherlands. We reached out to participants in advance through personal contacts and via their responses to study advertisements. When several participants in a single region were interested, we scheduled a community visit during a time when participants could be tested back-to-back in a single day. Experimental sessions were run in the back of a large utility van, from which the seats had been removed and the windows covered to keep the visual testing environment similar between locations. NGT signers were recruited and tested over the course of three months (12 testing days) in seven locations around the Netherlands until 32 early learners and 32 late learners were tested. Non-signing participants were then recruited and tested over the course of six months (12 testing days) in three locations around the Netherlands using the same basic procedure for recruitment. The only difference in recruiting the hearing, non-signing participants is that they were selected beforehand to, as best as possible, match the age, educational profile, and gender balance of the signing participants who had already been tested. We aimed for 64 Dutch-speaking controls, and were able to recruit 53; something of a feat given the strict sampling restrictions we put on age and education combined with a constrained data-collection period due to the primary recruiter's availability (limited to her master's thesis work period). Non-signers were all native Dutch speakers, and all spoke English as a second language with some degree of fluency.

**2.2** MATERIALS We created the stimuli by recording two completely spontaneous, unscripted conversations in NGT and then splicing out 80 fragments from each one. We invited two dyads of NGT signers, each previously acquainted (i.e., friend pairs) but not recently in contact with each other, to come and catch up on camera. The first dyad was a close friend pair catching up after the summer holidays who discussed, among other things, a recent trip and a night out with friends. The second dyad were acquaintances who talked, among other things, about favorite travel destinations and what they would do if they were to win the lottery. All four signers (one male, three females) were in their twenties and thirties, and reported using one of the three Western dialects predominantly. The conversations were recorded with the two signers sitting in separate rooms and chatting over a specialized video chat-like set-up with one-way mirrors so that they could engage in mutual gaze as depicted in panel A of Figure 2. Again, this special set-up means that the resulting stimuli allow the experimental participants to have full frontal-view access to each signer's body movements, facial expressions, and manual signs, as is typically afforded to addressees. Each pair was recorded for 90 minutes, at which point the recording was stopped. As confirmed by a native signer of NGT, the first conversation had a natural and even back-and-forth from the beginning, while it took the second dyad approximately ten minutes to arrive at a similarly paced conversation. These recording sessions resulted in two separate but synchronized 90-minute video streams for each dyad. We created a joint video file for each conversation by embedding the two video streams into the left and right regions of the frame, visible in panel B of Figure 2. We then also created two alternate versions of each joint video file in which either (1) the left signer or (2) the right signer was masked such that only the other signer was visible.

The first author then used the joint video files—which display both signers simultaneously—to identify 80 turns from each conversation for use in the button press experiment, using the next-turn proof procedure (Hutchby & Wooffitt 1998:15). We first narrowed our focus to turns that were immediately followed by a response from the addressee with non-overlapping stroke-to-stroke timing—this criterion helps to ensure that the target turn contains sufficient cues to speaker transition to have successfully elicited a well-timed response under natural circumstances. It also allowed us to use the same turns in a separate experiment, not reported here (cf. Casillas et al. 2015). Among those turns with the required response timing, we focused on turns that were comprehensible given only a few seconds of prior context (e.g., excluding "inside" references to places/events/people known only to the interlocutors), at least one second long, and free of significant mid-turn self-repair or other distractions (e.g., the addressee drinking from a water bottle). These turns, while fully comprehensible without extensive context, still contained markers of natural utterance production (e.g., minor pauses and self repairs) and were highly variable in content, duration, and the number of TCUs they contained. From the remaining turns under consideration, we selected approximately equal numbers from each signer, and then split them into a set of turns used for training (i.e., for practice in getting used to the task) and turns used for testing (i.e., those used in our analyses). Ultimately we identified 28–32 test turns and 9–11 training turns for each of the four signers, summing to 80 turns per dyad (see Appendix Figure 3 for details on each target turn). We also then extracted the immediately preceding context for each turn, limiting this context to what was minimally necessary to fully comprehend the target turn. We extracted the context video clips from the joint-view video, which showed both signers. We then extracted the target turn video clips from the masked version of the video that only showed the signer producing the target turn. We cropped the end of the target turn to the end of the final stroke, in line with the stroke-to-stroke turn boundary hypothesis (cf. de Vos et al. 2015). Between the two dyads, this process resulted in 160 conversation fragments (each clipped into context clip and target turn clip), thus resulting in 320 video clips for use during training and testing trials in the experiment.

ANNOTATION With the indispensable contributions of three native NGT signers and two

NGT experts, we then annotated each of the target turns for its communicative act type (question vs. non-question), number of potential end points (single- vs. multi-TCU), as well as any lexical or prosodic cue it contained that could be used for indicating upcoming signer transition/upcoming turn end (see Appendix Figure 3 for the full set ultimately considered). Given the spontaneous nature of these conversational stimuli, we did not know in advance which cues would be used, nor how often. We therefore started by creating an exhaustive list of potential cues and then annotated each cue when it occurred in a target turn. The cues annotated include: brow movements, including frowns, raises, and mixed categories (de Vos et al. 2009 on NGT; Ekman 1979), body leans (van der Kooij et al. 2006), head movements, and eye blinks (Ormel & Crasborn 2012). Notably, these non-manual prosodic cues abounded in our naturalistic dataset, but because of their multifunctional nature, these cues occurred across the question and non-question stimuli with similar frequencies. The complex and covarying use of these cues in naturalistic interaction prevented us from using them to systematically investigate predictions in the current experiment. In contrast, the different lexical manual cues to questionhood in our dataset were salient, highly associated with questions, and included both language-specific and globally-accessible forms (see below).

We thus identified two groupings of commonly used lexical signs that indicate questionhood: cues accessible to both the signers and non-signers as potentially question marking because of their use in Dutch co-speech gesture ("globally-accessible" cues) and cues only available to signers as potentially question marking ("language-specific" cues). The category of globally-accessible cues included points to the addressee (NGT gloss: YOU, YOU-DUAL) and the palm-up motion, with either or both hands. The category of language-specific cues included lexical signs for information questions (NGT gloss: HOW, HOW-LONG, WHY, HOW-MANY, QUESTION-MARK). Details on these signs can be found in Figure 1.

**2.3** PROCEDURE Upon arrival, participants were taken through the informed consent process with a native NGT signer (signing participants) or a native Dutch speaker (non-signing participants), after which they were briefly interviewed for further information about their language background. Each participant was randomly assigned to see video clips from either the first dyad or the second dyad (they saw the complementary dyad's turns in a second experiment not reported here). Participants then saw an instruction video in NGT explaining what to expect and how to perform the task (subtitled in Dutch for non-signing participants). The NGT-signing experimenter (signers) or the Dutch speaker (non-signers) then conversationally checked whether the participant understood the instructions. Participants then tried the task out on the 20 practice trials, after which they consulted with the experimenter once more. If they needed further clarification, they completed the (same) 20 practice trials again. A total of 45% of signing participants and 17% of non-signing participants opted for a second round of practice before beginning the test trials. Participants then conducted the button-press task with the 60 test trials, and were given the opportunity to take a short break after 30 trials. This experiment took 20 minutes, and was the second of three in the same test session (Casillas et al. 2015), which altogether typically took between 1 and 1.5 hours per participant. Participants were compensated with a €20 voucher for their time. Recruitment, informed consent, data collection, and data archiving were all done in accordance with ethical oversight by the Radboud University Ethical committee under the research program The structure and development of signed conversations (De Vos and Levinson; project code ECG2012-1304-098).

**2.3.1** TRIAL STRUCTURE We programmed the experiment so that each button press trial had the same structure as shown in panel C of Figure 2. Participants first saw the context video with both signers. This was followed by 500 msec in which they saw only the first frame of the target turn (i.e., a "frozen" view of the target speaker; with the addressee having disappeared)—this cue was used to indicate to participants that the context was over and that they should now focus on predicting the end of the current turn. Then participants saw the target turn. If participants had not pressed the button by the time the target turn ended, the screen froze on the final frame of the target turn for up to two seconds. We added these final two seconds to ensure that our design matched that of de Ruiter and colleagues' (2006), which used two seconds of silence after the turn offset in spoken Dutch. These seconds of silence indicated that the turn had ended, and gave participants a change to respond reactively (not anticipatorily) to the turn end; a response more likely when turns are less predictable (Magyari & de Ruiter 2012). In our experiment, participants knew the turn had ended when they saw the final "freeze", and had up to two seconds to react to this end before the trial terminated automatically. The 20 practice trials and 60 test trials in each experiment were presented in a randomized order for each participant.

**2.4** EXCLUSIONS We excluded a total of 14 participants before conducting any analysis for the following reasons: task misunderstanding (N = 2; 1 early learner, 1 late learner of NGT), non-completion of the task (N = 2; 1 early learner, 1 late learner of NGT), significant motor problems affecting the button-press response (N = 1; early learner NGT), lack of NGT fluency (N = 1 late learner), non-native Dutch speaker (N = 1 non-signer), and experimenter error (N = 7 non-signers).

Despite our adaptations to the instructions and multiple practice sessions, we noted that a handful of participants found the task instructions rather complicated, which resulted in their giving very early, very late, and/or multiple-press responses that indicated their lack of understanding or the difficulty they faced in executing the task as instructed. We intended to analyze the data under the assumption that all participants understood and were able to reliably execute the task as instructed, so we therefore made further systematic exclusions to remove these cases of participants using unusual button press response. Given that our method and its use with this participant community is novel, we have no prior guidelines for which patterns of button-press indicate task understanding. Instead, we considered two diagnostic indicators of non-compliance with the task instructions: too-early and too-late button presses. For each indicator, we established an exclusion threshold by examining the distribution of participant responses for a cut-off point between the typical (i.e., distributional peak) and atypical (i.e., long tail) cases (see Appendix Figure 2). First, we excluded participants who pressed the button too early—that is, in response to the context videos and therefore even before the target turn had begun. The overwhelming majority of participants made early button presses on 2 or fewer trials, so we excluded participants who did so on 3 or more trials (5%+ of the time). This exclusion criterion resulted in the removal of data from 12 participants (5 early learners, 2 late learners, and 5 non-signers) who made early responses on an average of 18.3% of test trials (median: 11.7%; range = 6.7%–55%). Second, we excluded participants who pressed the button too late—that is, more than 500 msec after the end of the turn. The vast majority of participants made late button presses on 5 or fewer trials, so we excluded participants who did so on 6 or more trials (10%+ of the time). We note that late button presses are more likely than early ones, even for participants who can understand and execute the instructions well, because the final freeze is sometimes ambiguous as to whether it is a ‘hold’ by the signer or simply the end of the turn; for this reason, the typical response pattern resulted in a higher threshold for late responses. The late response exclusion criterion resulted in the removal of data from an additional 6 participants (1 early learner and 5 non-signers) who made their late responses on an average of 16.7% of test trials (median: 15.8%; range = 11.7%–23.0%).[2] The remaining data therefore included 85 participants (23 early learners, 27 late learners, and 35 non-signers). Although this loss of data of 32 participants from the 117 originally tested is substantial, it reflects our balance of recruiting a diverse sample of participants while also systematically imposing limits to better ensure that our assumptions about the button press behavior are adequately met for analyzing the experimental outcomes. In fact, 5 of the 18 participants excluded on the basis of the too-late and too-early criteria would have been excluded on both counts, suggesting that our thresholds were effective in identifying divergent response patterns. We also note that the exclusion criteria cumulatively affected a similar number of participants in all three participant groups (early learners, late learners, and non-signers).

Among the remaining participants' data, we did a final pass of exclusions for individual trials with uninterpretable responses, including: trials with no response and trials where a response came within the first 720 msec of the stimulus (the minimal time needed for turn-end informative information across our items). We also excluded one item with an unusually long target turn; at 10.28 seconds, it was more than 4 seconds longer than all other target turns and thus systematically elicited false early responses (other target turn durations: mean = 2.58, median = 2.28, range = 0.88–6.16 seconds). After these trial- and item-level exclusions, we maintained 94.3% of the verified participant data for analysis.

**3.** RESULTS All analyses were conducted in R (version 3.6.1; R Core Team 2017), with analyses and plots generated using the lme4 and ggplot2 packages (Bates et al. 2015; Wickham 2016). Because this is the first study using an experimental measure of turn-end prediction in sign language, our initial analyses tested whether participants were indeed able, on average, to reliably anticipate the end of the turns in the stimuli. Our second set of analyses aimed to test whether questions (i.e., a specific type of response-eliciting turn) maintain a privileged status in prediction of upcoming turn structure, and whether such an effect varies depending on the participant's linguistic background. Our third and final set of analyses aimed to test the role of NGT-specific cues to questionhood in participants' (timely) anticipations. The high variability in individual cue use across items prevents individual cue analysis in the present study (see

Appendix Figure 3). We analyze the likelihood of anticipation in each of these analyses; reaction time alone shows almost no differences between participant groups, likely due to the wide variability in timing within each group (see the linked analysis scripts for more information). In what follows, we report significant effects from each model; non-significant effects are overtly marked as such when mentioned. Full model output tables can be found in the Appendices, and anonymized data and analysis scripts can be found at https://github.com/marisacasillas/NGT-Turn_end_prediction.

**3.1** ANALYSIS 1: OVERALL ANTICIPATION Overall, participants responded before the end of the turn (i.e., anticipated) turn-ends 71.7% of the time, with somewhat fewer anticipatory responses for participants with less signing experience (early learners = 75.8%; late learners = 73%; matched non-signers = 68%)—differences between groups in average anticipation rate were minimal. To test whether participants reliably anticipated turn-ends, we built a mixed-effects logistic regression for each group with anticipation as a binary dependent variable (1 = pressed the button before the end of the turn; 0 = pressed the button after the end of the turn), only including random effects of participant and item.[3] A positive and significant model intercept in this case indicates that anticipation values are significantly different from zero for that group (i.e., that participants in that group reliably anticipated turn ends). All three participant groups—early learners ($\beta = 1.774$; SE = 0.314; $z = 5.653$; $p < 0.001$), late learners ($\beta = 1.462$; SE = 0.245; $z = 5.980$; $p < 0.001$), and matched non-signers ($\beta = 1.145$; SE = 0.250; $z = 4.571$; $p < 0.001$)—significantly differed from zero and therefore reliably anticipated turn ends. These regression findings are also found if we instead test each group's anticipation rate as different from 0 in a series of one-tailed t-tests (all $p$ values < 0.001).

**3.2** ANALYSIS 2: QUESTION STATUS AND LANGUAGE BACKGROUND IN ANTICIPATION We next tested whether anticipation was more likely when seeing a question vs. a non-question, and whether this effect varied across participant groups. We first further limited the data to utterances with only a single turn-constructional unit; that is, we only analyzed responses to utterances with one possible turn end (76.6% of the items in analysis 1; Appendix Figure 3). We introduced this extra limitation for interpretational clarity, given that the button presses in this subset of the data should theoretically come in anticipation of a single syntactic unit perceivable as a question or non-question. Consider, for example, a question turn in which the first turn-constructional unit is not interrogative (“It was the day before yesterday, did you go?”). Participants pressing the button in response to the first TCU are responding to a non-question unit while those responding to the turn-end make their response to a question. We do not know a priori which TCU end participants were aiming for, so we cannot analyze the item as a question or non-question; it is ambiguous with respect to the participant's response. Note, however, that an identical model to what we present below, only using the entire analysis 1 dataset (i.e., both multi- and single-TCU trials), shows weaker but qualitatively similar results (see linked analysis scripts).

We tested effects of question status and participant group with a mixed effects logistic regression, using anticipation as a binary dependent variable (same as before) and participant group (factorial; early learner/late learner/non-signer), question status (factorial; question/non-question), and their interaction as predictors of interest. We additionally included three fixed effects that may predictably affect response patterns but are not of theoretical interest: duration of the turn (numeric; in seconds), trial number (numeric, to control for any order effects), and signer dyad featured in the stimulus (factorial; A/B). The model also included random effects of participant and item (N = 3802, log likelihood = -1867.5). [4]

There was no evidence for significant pairwise differences between participant groups (early learner vs. late learner: $\beta = -0.232$, SE = 0.372, $z = -0.623$, $p = 0.533$; early learner vs. non-signer: $\beta = -0.603$, SE = 0.352, $z = -1.712$, $p = 0.087$), but strong evidence for an overall question benefit on anticipation ($\beta = 0.373$, SE = 0.140, $z = 2.663$, $p = 0.008$), and significant pairwise interactions between participant group and question status for early learners vs. non-signers (early learner vs. late learner: $\beta = -0.009$, SE = 0.118, $z = -0.080$, $p = 0.936$; early learner vs. non-signer: $\beta = -0.258$, SE = 0.111, $z = -2.328$, $p = 0.020$). The primary outcomes here are illustrated by Figure 3: (a) turn ends were much more likely to be anticipated for questions than for non-questions overall, (b) early learners are overall only marginally more likely to anticipate than non-signers, and are statistically indistinguishable from late learners, and (c) early learners show a significantly larger benefit of question status compared to non-signers, but are statistically indistinguishable from late learners in question status effects.

<FIGURE 3 ABOUT HERE>

FIGURE 3. Average proportion of turns with anticipatory responses across participants, divided by participant group (early learner/late learner/non-signer) and turn type (question/non-question).

To pairwise test the difference between late learners and non-signers, we ran a second, identical model to the first one, only now with late learners as the reference level for participant group and found that late learners and non-signers were overall statistically indistinguishable in their anticipation rate (β = -0.371, SE = 0.333, $z = -1.113$, $p = 0.266$), but late learners showed a significantly larger benefit of question status compared to non-signers (β = -0.249, SE = 0.103, $z = -2.402$, $p = 0.016$). In a nutshell, these analyses find no evidence for difference in anticipation rate between signing groups, but do suggest a difference with non-signers, particularly with respect to the benefit in anticipation from question-formatted turns.

In addition to these effects of interest, there was a strong positive effect of turn duration: even though all turns in this analysis had only one possible end point (i.e., single-TCU turns, see above), longer turns were still associated with a higher likelihood of anticipation (β = 0.476, SE = 0.092, $z = 5.185$, $p < 0.001$). There were no significant effects of trial number or signer dyad.

**3.3** ANALYSIS 3: LINGUISTIC ACCESS IN ANTICIPATION In our final set of analyses we investigated the extent to which some specific cues produced by signers might have supported early button presses for response-eliciting turns. We differentiate between turns that contain any of the lexicalized NGT cues to questionhood presented in Figure 1 and turns that only contain lexicalized cues that might be apparent to non-signers as being response-eliciting (i.e., that can be used as response-eliciting Dutch co-speech gestures). We predicted a linguistic advantage in anticipation for the signers only in the first case, the NGT-lexicalized response elicitation cues. If a turn contained a palm-up gesture or an index-finger second person pronoun, we considered it to have a salient response eliciting cue that was available to both signers and non-signers (hereafter "global" response-elicitation cue; these have lexical status in NGT but are apparent to both groups as response eliciting). For turns with just these "global" clues to questionhood, we predicted no difference between signers and non-signers.

To test whether signers maintained an advantage over non-signers due to linguistic access to cues to questionhood, we first restrict the data to only include trials where at least one lexicalized cue to questionhood cue was present (see Figure 1; 41 items; this subset represents 44.5% of the items in analysis 2). We take this subset approach as a way of running a semi-controlled experiment within our highly variable stimuli; this subset "experiment" tests the proposition that, for turns with at least one lexicalized NGT question cue, signers have a predictive advantage over non-signers. If so, we expect to see a significant difference between signing and non-signing participant groups for this part of the dataset.[5]

We analyzed this subset of the data with a mixed effects logistic regression with anticipation as a binary dependent variable (same as before) and participant group (factorial; early learner/late learner/non-signer) as the predictor of interest. We additionally included the same three control predictors as before: duration of the turn (numeric; in seconds), trial number (numeric), and signer dyad featured in the stimulus (factorial; A/B). The model also included random effects of participant and item (N = 1782, log likelihood = -822.8).[6]

There was a significant pairwise effect of participant group between early learners and non-signers (early learner vs. late learner: $\beta = -0.292$, SE = 0.410, $z = -0.710$, $p = 0.477$; early learner vs. non-signer: $\beta = -0.768$, SE = 0.389, $z = -1.972$, $p = 0.049$) in addition to a significant overall effect of turn duration ($\beta = 0.347$, SE = 0.149, $z = 2.332$, $p = 0.020$) and no effects of trial number or dyad. We constructed an identical second model, only now again with late learners as the reference group, and found no evidence for a difference between late learners and non-signers in this subset of the data ($\beta = -0.476$, SE = 0.364, $z = -1.310$, $p = 0.190$), though note that their anticipation rate was indistinguishable from early learners in the first model, suggesting that their response patterns in this context fall somewhere between the other two groups.

To test whether non-signers showed an equal benefit for global response-eliciting cues, we again use a subsetting approach, this time focusing exclusively on trials where at least one global cue was used, but no sign-specific response-eliciting cue was used (i.e., use of YOU/YOU-DUAL/PALM-UP, but none the other signs in Figure 1; 34 items; this subset represents 36.9% of the items in analysis 2). In other words, we test the proposition that, for turns with a global cue to questionhood but no sign-specific cue, signers still have an advantage over non-signers; if so, we expect to see a significant difference between signing and non-signing participant groups. Model structure was identical to the previous analysis.

There was no significant main effect of participant group (early learner vs. late learner: β = -0.219, SE = 0.367, $z$ = -0.598, $p$ = 0.550; early learner vs. non-signer: β = -0.534, SE = 0.347, $z$ = -1.536, $p$ = 0.125); only a significant effect of turn duration (β = 0.426, SE = 0.091, $z$ = 4.678, $p$ < 0.001) and no effects of trial number or dyad. We again constructed an identical second model with late learners as the reference group, and found no evidence for a difference between late learners and non-signers in this subset of the data (β = -0.314, SE = 0.329, $z$ = -0.955, $p$ = 0.339).

**4.** DISCUSSION Our findings reveal (i) that response-eliciting features aid in the recognition of turn-type and thus turn-ending, regardless of whether the conversation is taking place in a language participants can understand (here, NGT), (ii) that accurate prediction is faster when the turn contains response-eliciting cues (like a question), and (iii) that linguistic access to the unfolding turn nevertheless yields an advantage for cues that are otherwise not salient as cues to imminent upcoming response. The first finding—that both signers and non-signers can reliably and accurately predict the upcoming turn ends of unfolding turns—provides further evidence that participants can make accurate real-time judgments about upcoming turn ends, even when they do not understand the language being used. This finding is in line with those using participants' judgments of cross-linguistic turn- and phrase-end identification in signed and spoken conversations (Carlson et al. 2005; Fenlon et al. 2007) and bolsters previous findings from gaze-based measures of response prediction in unfamiliar languages (Casillas & Frank 2017). The second finding—that participants are more likely to make an anticipatory button press during a question turn than a non-question turn—is in line with prior observational and experimental work on turn-taking suggesting that questions lead to a higher likelihood of anticipation and potentially faster responses. Notably, the benefit for questions was significantly larger for early learners compared to non-signers, suggesting that while questions were sufficiently marked with globally-accessible cues to cause a prediction benefit in both signers and non-signers, linguistic access to NGT-specific question cues renders an additional advantage. The third finding—that NGT-specific cues result in an anticipatory advantage limited to signers but that the advantage disappears for globally-accessible cues—underscores the fact that participants, be they signers or non-signers, are highly competent in noting and acting on the response-eliciting cues they have access to during real-time turn change prediction. Our results accord with Brentari and colleagues' (2018) observation that German and American non-signers alike, are able to identify commands and other speech acts on the basis of globally-accessible, or universal, cues such as head nods, head tilts and eye aperture.

This pattern of findings aligns with the interaction engine hypothesis in that we see relatively few differences across groups and, in our third analysis, evidence that linguistic differences in prediction derive mainly from cases where globally-accessible cues are not present to help the observer. In these contexts, we found only limited evidence for a difference between our early and late signer groups, despite the fact that these two groups represented rather different linguistic profiles with respect to the age of acquisition and input types for NGT; the late learners incorporate linguistic coding of questions into turn-end prediction, but do not statistically pair with either the early learner or the non-signer group. Late signers may therefore sometimes have rapid access to linguistic question-marking cues, but in other cases rely more heavily on the non-linguistic, globally-accessible cues to turn-ending.

Our finding of a benefit for questions in a button-press task also suggests that, across languages, participants may prioritize response-eliciting cues; an effect that can be attributed to their importance for coordinating who speaks next at the (otherwise vulnerable) points for possible floor transition (Stivers & Rossano 2010; Sacks et al. 1974). Together with the other evidence gathered on typologically distinct languages, our results support the idea that these basic interactional skills are integral to human communication at large, and thereby are likely to play a major role in shaping the patterns of everyday language use from infancy to adulthood.

While our findings support the idea of a species-wide capacity for interaction that influences the way unfolding turns are processed during conversation, we also see strong evidence for language specificity in those predictive processes. Specifically, we found that NGT signs that have no historical link with co-speech gesture in Dutch were effective in aiding anticipation for signers, but not for non-signers; this language-specific effect presumably explains the overall greater benefit of question turns for early learners compared to non-signers. A general hypothesis therefore is that, when cues are language specific, non-iconic, and not otherwise conventionally used, they will lead to specific benefits for fluent users of the language. Otherwise fluent and non-fluent participants will be comparably good at exploiting more globally-accessible cues in conversational prediction, even in the context of processing language in real time. A caveat here, though, is that while some of globally-accessible cues may be universal, others may be culturally-specific, e.g. where gestures accompanying local spoken languages have been incorporated into proximal sign languages (e.g. Janzen & Schaffer 2002, Pfau & Steinbach 2006, Le Guen 2012, de Vos 2015, Tano & Nyst 2018).

Extending our findings to the prosodic domain is a crucial next step in understanding potential parallels between spoken and signed turn prediction; while lexicosyntactic cues appear crucial for precisely identifying turn ends in spoken language (de Ruiter et al. 2006), prosodic cues may provide critical disambiguating information at potential turn ends (Bögels & Torreira 2015) and may very well contribute to the response-elicitation privilege documented in the present study. As in spoken languages, the boundary between linguistically coded aspects of prosody and the expressive or gestural ones is theoretically disputed in sign linguistics. Nevertheless, we note that some prosodic cues in sign (e.g., blinks used to signal turn boundaries in NGT) seem to be more ‘digital’, discrete events than many prosodic cues in spoken language (e.g., intonation contour): blinks for example may occur multiple times in a single turn. Moreover, there is evidence, at least in some sign languages, that such signals are consistently used to mark utterance boundaries (Nespor & Sandler 1999, Herrmann 2010). Thus prosody may play a somewhat different role in the online prediction of upcoming turn-ends and upcoming responses in sign conversation not just because of the discrete nature of these signals, but also the consistency with which they appear to be used. That is to say, there may be differences not so much in the kinds of visual signals that play a role in face-to-face interaction in either language modality, but rather in the degree to which such signals have developed grammatical consistency. This typological difference between spoken and signed languages could have major implications for psycholinguistic models of turn prediction and real-time language processing in sign languages, but this requires further investigation. That said, we have not ruled out the possibility that other unexamined properties of the turns with linguistic coding of questions drove early responses for the signers. This potential confound, driven in part by the naturalistic nature of our stimuli, could be systematically investigated using more controlled stimuli in follow-up work.

**4.1.** LIMITATIONS AND NEXT STEPS Following the 2006 study by de Ruiter and colleagues, we have tried to combine experimental rigor with high levels of ecological validity by using spontaneous NGT dialogue between friends. We also recruited participants from across the NGT community, reaching out to signers and non-signers who would otherwise be unlikely to participate in a psycholinguistic study because of their limited education, age, linguistic background, and/or location (cf. Henrich et al. 2010). Our study demonstrates that it is possible to gather reliable data on turn-end prediction using a button-press paradigm with both signers and non-signers, and mirrors de Ruiter and colleagues' (2006) primary finding that lexical cues are important. At the same time, the variable stimuli and diverse participant pool required us to use strict exclusion criteria that resulted in substantial data loss. Future work can build on these strengths and weaknesses by combining our general experimental approach with more controlled linguistic stimuli. For example, further studies on turn prediction in NGT could use the button press method as altered by Bögels & Torreira (2015) to test whether NGT signers (and non-signers) are similarly sensitive to prosodic cues at points of possible turn completion (e.g., by manipulating prosodic cues such as blinks and brow movements while keeping lexicosyntactic information identical). Manipulating specific cues in sign would, however, require advanced methodology such as video manipulation, or the use of sign language avatars (cf. Wolfe et al. 2011). Further, the controlled nature of our experiment and the scope of our current research questions limits our insight into other factors that may facilitate prediction in real signed interaction, including linguistic and processing advantages that could make differences between signing and non-signing participants more apparent (e.g., in making predictions during multi-party conversation, making content-specific predictions, and in integrating subtle contextual cues to make accurate predictions earlier on). Future work can follow-up on these questions with a combination of corpus study and further experimentation along the lines we present here. The use of scripted stimuli or carefully selected cue-specific stimuli from naturalistic conversation in future work would also help to overcome naturalistic variability that, in the present study, resulted in our focus on groups of cues rather than individual cues relating to prediction. In any case, we hope that this study will contribute to a line of research exploring the similarities and differences in both the cue types and timelines of spoken and signed languages in their natural ecologies (cf. Hosemann et al. 2013; Sehyr et al. 2020).

**4.2.** CONCLUSIONS We asked NGT signers and non-signing Dutch speakers to predict the ends of unfolding turns extracted from spontaneous conversation in NGT. We found that signers and non-signers alike were able to reliably anticipate upcoming turn ends. We also found that both groups were more likely to anticipate turn ends when the unfolding turn was a question, but that this advantage was greater for early learners of NGT than non-signers. When we looked closer at the use of language-specific and globally-accessible cues to questionhood, we found that signers were significantly more likely to benefit from the language-specific cues in making their predictions, but that both groups benefited equally from globally-accessible cues. Our findings support the idea that participants, whether they have access to the language or not, predict upcoming turn ends and track both linguistic and non-linguistic cues that may aid in that prediction, meanwhile linguistic cues still provide an advantage over and above globally-accessible ones.

The current findings demonstrate that the ability to accurately predict upcoming turn structure extends across language modalities and can even be implemented, to some extent, without linguistic cues, underscoring the idea that our capacity for language is first and foremost grounded in our ability to predict and produce relevant social actions (Levinson 2006). While linguistic cues offer us an answer to the query of how the turn-taking system manages to be both consistently fast and consistently precise (Levinson & Torreira 2015), multi-modal accounts of turn-taking may offer critical insights into communicative resources that can be used to coordinate interaction across interactants who don't share a language, including infant-caregiver interactions (Casillas & Hilbrink 2020), cross-signing between deaf individuals who do not know a common signed language (Byun et al. 2018), translanguaging between speakers and signers of distinct languages (Kusters et al. 2017), and even more recently, the study of homesign interactions between deaf people and their hearing relatives in the absence of conventional language input (Haviland 2020). The basic communicative resources and abilities that feature in these interactions may help to understand the foundations of human conversational interaction and therefore to shed light on the evolutionary processes by which language came to be.

DECLARATION OF INTEREST STATEMENT No potential conflict of interest was reported by the authors.

DATA AVAILABILITY STATEMENT Anonymized experiment data and scripts to replicate the reported analyses are available at https://github.com/marisacasillas/NGT-Turn_end_prediction. All analyses were completed in R, an open-source software (R Core Team 2017).

DATA DEPOSITION In addition to the anonymized experiment data and scripts available at https://github.com/marisacasillas/NGT-Turn_end_prediction, the stimuli and experiment scripts are available on request.

## APPENDICES

We here provide: (a) full output for all statistical models reported in the main text (Tables 1–9) and (b) a collection of supplementary figures illustrating spread in participant demographics, response patterns, and target turn properties.

TABLE 1. Model output for mixed-effects logistic regression of overall anticipation within the early learner group (N = 1378; AIC = 1274.4; BIC = 1290.1; log likelihood = -634.2).

<TABLE 1 ABOUT HERE>

TABLE 2. Model output for mixed-effects logistic regression of overall anticipation within the late learner group (N = 1617; AIC = 1571.9; BIC = 1588.1; log likelihood = -783.0).

<TABLE 2 ABOUT HERE>

TABLE 3. Model output for mixed-effects logistic regression of overall anticipation within the matched non-signer group (N = 2099; AIC = 2130.0; BIC = 2146.9; log likelihood = -1062.0).

<TABLE 3 ABOUT HERE>

TABLE 4. Model output for mixed-effects logistic regression of anticipation given participant group, question status, and control predictors, with early learners as the reference level for group (N = 3802; AIC = 3757.0; BIC = 3825.6; log likelihood = 3735.0).

<TABLE 4 ABOUT HERE>

TABLE 5. Model output for mixed-effects logistic regression of anticipation given participant group, question status, and control predictors, with late learners as the reference level for group (N = 3802; AIC = 3757.0; BIC = 3825.6; log likelihood = 3735.0).

<TABLE 5 ABOUT HERE>

TABLE 6. Model output for mixed-effects logistic regression of anticipation of turns with at least one sign-specific cue to transition, given participant group and control predictors, with early learners as the reference level for group (N = 1782; AIC = 1661.6; BIC = 1705.5; log likelihood = -822.8).

<TABLE 6 ABOUT HERE>

TABLE 7. Model output for mixed-effects logistic regression of anticipation of turns with at least one sign-specific cue to transition, given participant group and control predictors, with late learners as the reference level for group (N = 1782; AIC = 1661.6; BIC = 1705.5; log likelihood = -822.8).

<TABLE 7 ABOUT HERE>

TABLE 8. Model output for mixed-effects logistic regression of anticipation of turns with at least one global cue to transition but no sign-specific cues, given participant group and control

predictors, with early learners as the reference level for group (N = 3696; AIC = 3669.1; BIC = 3718.8; log likelihood = -1826.6).

<TABLE 8 ABOUT HERE>

TABLE 9. Model output for mixed-effects logistic regression of anticipation of turns with at least one global cue to transition but no sign-specific cues, given participant group and control predictors, with late learners as the reference level for group (N = 3696; AIC = 3669.1; BIC = 3718.8; log likelihood = -1826.6).

<TABLE 9 ABOUT HERE>

<APPENDIX FIGURE 1 ABOUT HERE>

APPENDIX FIGURE 1. Demographic overview of the present participant sample. Panel A: Number of participants by age (x-axis) and education level (y-axis); darker tile color indicates more participants for that age and education level combination. Participants are split into men (left) and women (right) and by signer (top row) and non-signer (bottom row) categories. Levels of education, using standard Dutch abbreviations, are defined as follows: 1 = LBO/LHNO or some MAVO/(V)MBO/VWO or Other; 2 = HAVO/MAVO/(V)MBO/VWO; 3 = HBO/bachelor. Panel B: More detailed demographic information for the signing participant group, including; TOP: Number of participants by age (x-axis), education level (y-axis), gender (men = first and third panels from the left; women = second and fourth panels from the left), and learner group (early exposure = left two grids; later exposure = right two grids).; BOTTOM-LEFT: Type of linguistic input (color) by learner group (early exposure = left graph; later exposure = right graph) and auditory status (deaf = upper row; hearing = lower row); BOTTOM-RIGHT: Distributions of age of onset for exposure to NGT by learner group (early exposure = light; later exposure = dark), showing group means (solid line) and medians (dashed line).

<APPENDIX FIGURE 2 ABOUT HERE>

APPENDIX FIGURE 2. Distribution of button behaviors used to make participant exclusions. Left panel: Number of participants displaying different rates of pressing the button during the context videos; we excluded participants who pressed the button during three or more test trials (i.e., all 12 participants represented on the right of the 5% vertical line), the clear point of separation between the main group of participants and the long tail of outliers. Cut-off points at 10% and 15% of test trials are also shown for reference (dashed and dotted vertical lines). Right panel: Number of participants displaying different rates of pressing the button very late during the target videos; we excluded participants who pressed the button late on six or more test trials (i.e., all 6 participants represented on the right of the 10% vertical line), the point of separation between the main group of participants and the long tail of outliers. Cut-off points at 15% and 20% of test trials are also shown for reference (dashed and dotted vertical lines).

<APPENDIX FIGURE 3 ABOUT HERE>

APPENDIX FIGURE 3. Detailed overview of annotated linguistic characteristics of each target turn. Single-unit turns are shown on the left panel and multi-unit turns on the right panel. Each turn is labeled by the signer dyad featured (MM or YR) and item number. For each turn, the following linguistic features are shown (in left-to-right order): duration (in seconds; darker = longer; numeric value shown in each cell); utterance type (shaded = polar/wh/alternative question; white = declarative); “NGT-only” question marker (shaded = yes; white = no regarding use of a question marker that is not conventionally associated with questions in Dutch); “NGT & Dutch” manual question marker (shaded = yes; white = no regarding use of a manual question marker that is conventionally associated with questions in Dutch); brow raise used (shaded = yes; white = no); brow frowning/furrowing used (shaded = yes; white = no); head tilt used (shaded = yes; white = no); blink used (shaded = yes; white = no); backchannel used (shaded = yes; white = no); nonmanual prosodic cue used (shaded = yes; white = no); manual prosodic cue used (shaded = yes; white = no); and tag marker used (shaded = yes; white = no).

REFERENCES

BAKER, CHARLOTTE. 1977. Regulators and turn-taking in American Sign Language discourse. ***On the other hand: New perspectives on American Sign Language***, ed. by Lynn A. Friedman, 215–36. New York, NY: Academic Press.

BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. ***Journal of Statistical Software*** 67.1–48.

BEUKELEERS, INEZ; GEERT BRÔNE; and MYRIAM VERMEERBERGEN. 2020. Unaddressed participants' gaze behavior in Flemish Sign Language interactions: Planning gaze shifts after recognizing an upcoming (possible) turn completion. ***Journal of Pragmatics*** 162.62–83.

BÖGELS, SARA; KOBIN H. KENDRICK; and STEPHEN C. LEVINSON. 2015. Never say no … How the brain interprets the pregnant pause in conversation. ***PLoS ONE*** 10. Online: https://doi.org/10.1371/journal.pone.0145474.

BÖGELS, SARA; and FRANCISCO TORREIRA. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. ***Journal of Phonetics*** 52.46–57.

BRENTARI, DIANE. 1998. ***A prosodic model of sign language phonology***. Cambridge, MA: MIT Press.

BRENTARI, DIANE; and LAURINDA CROSSLEY. 2002. Prosody on the hands and face: Evidence from American Sign Language. ***Sign Language & Linguistics*** 5.105–30.

BRENTARI, DIANE; JOSHUA FALK; ANASTASIA GIANNAKIDOU; ANNIKA HERRMANN; ELISABETH VOLK; and MARKUS STEINBACH. 2018. Production and comprehension of prosodic markers in sign language imperatives. ***Frontiers in Psychology*** 9. Online: https://doi.org/10.3389/fpsyg.2018.00770.

BRENTARI, DIANE; CAROLINA GONZÁLEZ; AMANDA SEIDL; and RONNIE WILBUR. 2011. Sensitivity to visual prosodic cues in signers and nonsigners. ***Language and Speech*** 54. 49–72.

BYUN, KANG-SUK; CONNIE DE VOS; ANASTASIA BRADFORD; ULRIKE ZESHAN; and STEPHEN C. LEVINSON. 2018. First encounters: Repair sequences in cross-signing. ***Topics in Cognitive Science*** 10.314–34.

Carlson, Rolf; Julia Hirschberg; and Marc Swerts. 2005. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. ***Speech Communication*** 46.326–33.

Casillas, Marisa; Connie de Vos; Onno Crasborn; and Stephen C. Levinson. 2015. The perception of stroke-to-stroke turn boundaries in signed conversation. ***Proceedings of the 37th annual meeting of the Cognitive Science Society***, ed. by David C. Noelle, Rick Dale, Anne S. Warlaumont, Jeff Yoshimi, Teenie Matlock, Carolyn D. Jennings, and Paul R. Maglio, 315–20. Austin, TX: Cognitive Science Society.

Casillas, Marisa, and Michael C. Frank. 2017. The development of children's ability to track and predict turn structure in conversation. ***Journal of Memory and Language*** 92.234–53.

Casillas, Marisa, and Elma Hilbrink. 2020. Communicative act development. ***Developmental and clinical pragmatics***, ed. by Klaus P. Schneider and Elly Ifantidou, 61–88. Berlin: De Gruyter Mouton.

Coates, Jennifer, and Rachel Sutton-Spence. 2001. Turn-taking patterns in deaf conversation. ***Journal of Sociolinguistics*** 5.507–29.

Coerts, Jane. 1992. ***Nonmanual grammatical markers: An analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands: Academisch proefschrift …*** Amsterdam: Universiteit van Amsterdam dissertation.

Costello, Brendan; Javier Fernández; and Alazne Landa. 2008. The non- (existent) native signer: Sign language research in a small deaf population. ***Sign languages: Spinning and unraveling the past, present, and future. TISLR9, forty five papers and three posters from the 9th Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006***, ed. by Ronice M. de Quadros, 77–94. Petrópolis/RJ: Editora Arara Azul.

Crasborn, Onno, and Noortje Akkermans. 2020. ***EmblemsNL dataset in Global Signbank***. Nijmegen: Radboud University. Online: http://signbank.cls.ru.nl.

Crasborn, Onno; Els van der Kooij; and Johan Ros. 2012. On the weight of phrase-final prosodic words in a sign language. ***Sign Language & Linguistics*** 15.11–38.

de Ruiter, Jan P.; Holger Mitterer; N. J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. ***Language*** 82.515–535.

**Formatted:** English (United States)
**Field Code Changed**
**Formatted:** English (United States)
**Formatted:** English (United States)

DE VOS, CONNIE. 2015. The Kata Kolok pointing system: Morphemization and syntactic integration. ***Topics in Cognitive Science*** 7.150–68.

DE VOS, CONNIE, and VICTORIA NYST. 2018. The time depth and typology of rural sign languages. ***Sign Language Studies*** 18.477–87.

DE VOS, CONNIE, and ROLAND PFAU. 2015. Sign language typology: The contribution of rural sign languages. ***Annual Review of Linguistics*** 1.265–88.

DE VOS, CONNIE; FRANCISCO TORREIRA; and STEPHEN C. LEVINSON. 2015. Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00268.

DE VOS, CONNIE; ELS VAN DER KOOIJ; and ONNO CRASBORN. 2009. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. ***Language & Speech*** 52.315–39.

EKMAN, PAUL. 1979. About brows: emotional and conversational signals. ***Human ethology: Claims and limits of a new discipline***, ed. by Mario von Cranach, Klaus Foppa, Wolf Lepenies, and Detlev Ploog, 169–202. Cambridge: Cambridge University Press.

EMMORY, KAREN; RAIN BOSWORTH; and TANYA KRALJIC. 2009. Visual feedback and self-monitoring of sign language. ***Journal of Memory and Language*** 61.398–411.

FENLON, JORDAN; TANYA DENMARK; RUTH CAMPBELL; and BENCIE WOLL. 2007. Seeing sentence boundaries. ***Sign Language & Linguistics*** 10.177–200.

GIRARD-GROEBER, SIMONE. 2015. The management of turn transition in signed interaction through the lens of overlaps. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00741.

GROEBER, SIMONE, and EVELYNE POCHON-BERGER. 2014. Turns and turn-taking in sign language interaction: The role of frozen signs. ***Journal of Pragmatics*** 65.121–36.

HAVILAND, JOHN. 2020. Signs, interaction, coordination, and gaze: Interactive foundations of "Z"—an emerging (sign) language from Chiapas, Mexico. ***Emerging Sign Languages of the Americas***, ed. by Oliver Le Guen, Josefina Safar, and Marie Coppola, 35–96. Berlin: De Gruyter Mouton.

HENRICH, JOSEPH; STEVEN J. HEINE; and ARA NORENZAYAN. 2010. Beyond WEIRD: Towards a broad-based behavioral science. ***Behavioral and Brain Sciences***, 33.111–135.

HERRMANN, ANNIKA. 2010. The interaction of eye blinks and other prosodic cues in German Sign Language. ***Sign Language & Linguistics*** 13.3–39.

HOLLER, JUDITH, and KOBIN H. KENDRICK. 2015. Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00098.

HOLLER, JUDITH; KOBIN H. KENDRICK; and STEPHEN C. LEVINSON. 2018. Processing language in face-to-face conversation: Questions with gestures get faster responses. ***Psychonomic Bulletin & Review***, 25.1900–8.

HOLLER, JUDITH, and STEPHEN C. LEVINSON. 2019. Multimodal language processing in human communication. ***Trends in Cognitive Sciences*** 23.639–52.

HOLZRICHTER, AMANDA S., and RICHARD P. MEIER. 2000. Child-directed signing in American Sign Language. ***Language Acquisition by Eye***, ed. by Charlene Chamberlain, Jill P. Morford, and Rachel I. Mayberry, 25–40. Mahweh, NJ: Lawrence Erlbaum Associates.

HOSEMANN, JANA; ANNIKA HERRMANN; MARKUS STEINBACH; INA BORNKESSEL-SCHLESEWSKY; and MATTHIAS SCHLESEWSKY. 2013. Lexical prediction via forward models: N400 evidence from German Sign Language. ***Neuropsychologia*** 51.2224–37.

HUTCHBY, IAN, and ROBIN WOOFFITT. 1998. ***Conversation analysis: Principles, practices and applications***. Cambridge: Polity Press.

JANZEN, TERRY, and BARBARA SCHAFFER. 2002. Gesture as the substrate in the process of ASL grammaticization. ***Modality and structure in signed and spoken languages***, ed. by Richard P. Meier, Kearsy Cormier, and David Quinto-Pozos, 199–223. Cambridge: Cambridge University Press.

KEITEL, ANNE, and MORITZ M. DAUM. 2015. The use of intonation for turn anticipation in observed conversations without visual signals as source of information. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00108.

KENDRICK, KOBIN H., and JUDITH HOLLER. 2017. Gaze direction signals response preference in conversation. ***Research on Language and Social Interaction*** 50.12–32.

KENDRICK, KOBIN H., and FRANCISCO TORREIRA. 2015. The timing and construction of preference: A quantitative study. ***Discourse Processes*** 52.255–89.

KOOIJ, ELS VAN DER. 2002. ***Phonological categories in Sign Language of the Netherlands: The role of phonetic implementation and iconicity***. Utrecht: Landelijke Onderzoekschool Taalwetenschap dissertation.

KUSTERS, ANNELIES; MASSIMILIANO SPOTTI; RUTH SWANWICK; and ELINA TAPIO. 2017. Beyond languages, beyond modalities: Transforming the study of semiotic repertoires. ***International Journal of Multilingualism*** 14.219–32.

KYLE, JIM G., and BENCIE WOLL. 1985. ***Sign language: The study of deaf people and their language***. Cambridge: Cambridge University Press.

LAMMERTINK, IMME; MARISA CASILLAS; TITIA BENDERS; BRECHTJE POST; and PAULA FIKKERT. 2015. Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00495.

LATIF, NIDA; AGNÈS ALSIUS; and K. G. MUNHALL. (2018). Knowing when to respond: the role of visual information in conversational turn exchanges. ***Attention, Perception, & Psychophysics*** 80.27–41.

LE GUEN, OLIVER. 2012. An exploration in the domain of time: from Yucatec Maya time gestures to Yucatec Maya Sign Language time signs. ***Endangered sign languages in village communities: anthropological and linguistic insights***, ed. by Ulrike Zeshan and Connie de Vos, 209–50. Berlin: De Gruyter Mouton.

LEVINSON, STEPHEN C. 2006. On the human "interaction engine". ***Roots of human sociality: Culture, cognition and*** *interaction*, ed. by N. J. Enfield and Stephen C. Levinson, 39–69. Oxford: Berg.

LEVINSON, STEPHEN C. 2019. Interactional foundations of language: The interaction engine hypothesis. ***Human language: From genes and brain to behavior***, ed. by Peter Hagoort, 189–200. Cambridge, MA: MIT Press.

LEVINSON, STEPHEN C., and FRANSCISCO TORREIRA. 2015. Timing in turn-taking and its implications for processing models of language. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.00731.

MAGYARI, LILLA, and J. P. DE RUITER. (2012). Prediction of turn-ends based on anticipation of upcoming words. ***Frontiers in Psychology*** 3. Online: https://doi.org/10.3389/fpsyg.2012.00376.

MANRIQUE, ELIZABETH, and N. J. ENFIELD. 2015. Suspending the next turn as a form of repair initiation: Evidence from Argentine Sign Language. ***Frontiers in Psychology*** 6. Online: https://doi.org/10.3389/fpsyg.2015.01326.

MCCLEARY, LELAND EMERSON, and TARCÍSIO DE ARANTES LEITE. 2013. Turn-taking in Brazilian Sign Language: Evidence from overlap. ***Journal of Interactional Research in Communication Disorders*** 4.123–54.

MONDADA, LORENZA. 2007. Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. ***Discourse Studies*** 9.194–225.

NESPOR, MARINA, and WENDY SANDLER. 1999. Prosody in Israeli Sign Language. ***Language and Speech*** 42.143–76.

ORMEL, ELLEN, and ONNO CRASBORN. 2012. Prosodic correlates of sentences in signed languages: A literature review and suggestions for new types of studies. ***Sign Language Studies*** 12.279–315.

PFAU, ROLAND, and MARKUS STEINBACH. 2006. ***Modality-independent and modality-specific aspects of grammaticalization in sign languages***. Potsdam: Universitäts-Verlag.

PIKA, SIMONE; RAY WILKINSON; KOBIN H. KENDRICK; SONJA C. VERNES. 2018. Taking turns: bridging the gap between human and animal communication. ***Proceedings of the Royal Society B: Biological Sciences*** 285. Online: https://doi.org/10.1098/rspb.2018.0598.

POMERANTZ, ANITA, and JOHN HERITAGE. 2012. Preference. ***The handbook of conversation analysis***, ed. by Jack Sidnell and Tanya Stivers, 210–228. Malden, MA: Wiley-Blackwell.

R CORE TEAM. 2017. ***R: A language and environment for statistical computing***. Vienna: R Foundation for Statistical Computing. Online: https://www.R-project.org.

ROSSANO, FEDERICO; PENELOPE BROWN; and STEPHEN C. LEVINSON. 2009. Gaze, questioning and culture. ***Conversation analysis: Comparative perspectives***, ed. by Jack Sidnell, 187–249. Cambridge: Cambridge University Press.

SACKS, HARVEY; EMANUEL A. SCHEGLOFF; and GAIL JEFFERSON. 1974. A simplest systematics for the organization of turn-taking for conversation. ***Language*** 50.696–735.

SANDLER, WENDY. 1999. Prosody in two natural language modalities. ***Language and Speech*** 42.127–42.

SANDLER, WENDY, and DIANE LILLO-MARTIN. 2006. ***Sign language and linguistic universals***. Cambridge: Cambridge University Press.

SCHEIN, JEROME D., and MARCUS T. DELK, JR. 1974. ***The Deaf population of the United States***. Silver Spring, MD: National Association of the Deaf.

SEHYR, ZED SEVCIKOVA; KATHERINE J. MIDGLEY; PHILLIP J. HOLCOMB; KAREN EMMOREY; DAVID C. PLAUT; and MARLENE BEHRMANN. 2020. Unique N170 signatures to words and faces in deaf ASL signers reflect experience-specific adaptations during early visual processing. ***Neuropsychologia*** 141. Online: https://doi.org/10.1016/j.neuropsychologia.2020.107414.

SENGHAS, ANN; and MARIE COPPOLA. 2001. Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. ***Psychological Science*** 12.323–8.

STIVERS, TANYA; N. J. ENFIELD; PENELOPE BROWN; CHRISTINA ENGLERT; MAKOTO HAYASHI; TRINE HEINEMANN; GERTIE HOYMANN; FEDERICO ROSSANO; JAN PETER DE RUITER; KYUNG-EUN YOON; and STEPHEN C. LEVINSON. 2009. Universals and cultural variation in turn-taking in conversation. ***Proceedings of the National Academy of Sciences of the United States of America*** 106.10587–92.

STIVERS, TANYA, and FEDERICO ROSSANO. 2010. Mobilizing response. ***Research on Language and Social Interaction*** 43.3–31.

STOKOE, WILLIAM C. (1960). ***Sign language structure: An outline of the visual communication systems of the American Deaf***. Buffalo, NY: Department of Anthropology and Linguistics, University of Buffalo.

TANO, ANGOUA, and VICTORIA NYST. 2018. Comparing body-part size and shape constructions in village sign languages with co-speech gesture. ***Sign Language Studies***, 18.517–45.

TIJSSELING, CORRIE. 2014. ***'School, waar?': Een onderzoek naar de betekenis van het Nederlandse dovenonderwijs voor de Nederlandse dovengemeenschap***. Utrecht: Utrecht University dissertation.

VAN DER KOOIJ, ELS; ONNO CRASBORN; and WIM EMMERIK. 2006. Explaining prosodic body leans in Sign Language of the Netherlands: Pragmatics required. ***Journal of Pragmatics*** 38.1598–614.

WICKHAM, HADLEY. 2016. ***ggplot2: Elegant graphics for data analysis***. New York, NY: Springer-Verlag.

WILBUR, RONNIE B. 1987. ***American Sign Language: Linguistic and applied dimensions*** (2nd ed.). San Diego, CA: College-Hill Press.

WOLFE, ROSALEE; PETER COOK; JOHN C. MCDONALD; and JERRY SCHNEPP. 2011. Linguistics as structure in computer animation: Toward a more effective synthesis of brow motion in American Sign Language. ***Sign Language & Linguistics*** 14.179–99.

ZESHAN, ULRIKE. 2004. Interrogative constructions in signed languages: Crosslinguistic perspectives. ***Language*** 80.7–39.

ZESHAN, ULRIKE, and NICK PALFREYMAN. 2017. Sign language typology. ***The Cambridge Handbook of Linguistic Typology***, ed. by Alexandra Y. Aikhenvald and R. M. W. Dixon, 178–216. Cambridge: Cambridge University Press.

NOTES

[1] Example of a single-TCU turn: "One phrase is enough." vs. a multi-TCU turn: "One phrase is enough, but I can do two."

[2] It is possible that these participants were treating the stimulus-final freeze as holding of the turn-final sign, an attested cue for prosodic boundaries and turn transitions in sign (Fenlon et al. 2007; Brentari et al. 2011; Groeber & Pochon-Berger 2014; Girard-Groeber 2015).

[3] glmer(Anticipation ~ (1|Participant) + (1|Item), data = All.responses.from.a.participant.group, family = binomial)

[4] glmer(Anticipation ~ Group * Question.status + Turn.duration + Trial.number + Dyad + (1|Participant) + (1|Item), data = Single.endpoint.turns, family = binomial). Note that participant group is treated as a factor in our analysis. As reported, we therefore run one version of this model with early learners as the reference group (i.e., early vs. late and early vs. non-signer) and one with late learners as the reference group (i.e., late vs. early and late vs. non-signer) to fully examine pairwise participant group effects (see Appendices for full model outputs).

[5] This subset approach is fundamental for our understanding of how individual cues may relate to prediction and participant groups with such a naturalistic, varied collection of stimuli (see Appendix Figure 3). If we instead created a model using all the items from analysis 2, we would be comparing performance on turns with a specific cue of interest (lexicalized cues to questionhood) with the total grab bag of all other turns, some of which may facilitate anticipation and some of which may impede it due to a variety of other cues and their combinations that we do not track in the present analysis.

[6] glmer(Anticipation ~ Group + Turn.duration + Trial.number + Dyad + (1|Participant) + (1|Item), data = Single.endpoint.turns.with.global.and.or.signspecific.cues, family = binomial). Note that as before (see Analysis 2) we implement two versions of this model with different reference levels to examine all pairwise participant group effects.

TABLE 1.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 1.7739 | 0.3138 | 5.653 | 1.58e-08 *** |

TABLE 2.

| Fixed effect | Estimate | Standard error | z-value | p-value |
| --- | --- | --- | --- | --- |
| (Intercept) | 1.4623 | 0.2446 | 5.98 | 2.24e-09 *** |

TABLE 3.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
| --- | --- | --- | --- | --- |
| (Intercept) | 1.1448 | 0.2505 | 4.571 | 4.86e-06 *** |

TABLE 4.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 0.522160 | 0.370508 | 1.409 | 0.15874 |
| Group=LateLearners | -0.232030 | 0.372181 | -0.623 | 0.53300 |
| Group=NonSigners | -0.603115 | 0.352376 | -1.712 | 0.08698 . |
| IsQuestion | 0.373459 | 0.140230 | 2.663 | 0.00774 ** |
| DurationSec | 0.476042 | 0.091807 | 5.185 | 2.16e-07 *** |
| Order | -0.159351 | 0.148224 | -1.075 | 0.28234 |
| SignDyad | 0.062194 | 0.175269 | 0.355 | 0.72270 |
| Group=LateLearners*IsQuestion | -0.009457 | 0.118496 | -0.080 | 0.93639 |
| Group=NonSigners*IsQuestion | -0.258092 | 0.110847 | -2.328 | 0.01989 * |

TABLE 5.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 0.290130 | 0.352098 | 0.824 | 0.40994 |
| Group=EarlyLearners | 0.232037 | 0.372233 | 0.623 | 0.53304 |
| Group=NonSigners | -0.371083 | 0.333437 | -1.113 | 0.26575 |
| IsQuestion | 0.364002 | 0.134329 | 2.710 | 0.00673 ** |
| DurationSec | 0.476042 | 0.091809 | 5.185 | 2.16e-07 *** |
| Order | -0.159351 | 0.148228 | -1.075 | 0.28235 |
| SignDyad | 0.062194 | 0.175271 | 0.355 | 0.72271 |
| Group=EarlyLearners*IsQuestion | 0.009457 | 0.118498 | 0.080 | 0.93639 |
| Group=NonSigners*IsQuestion | -0.248636 | 0.103512 | -2.402 | 0.01631 * |

TABLE 6.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 1.16587 | 0.53408 | 2.183 | 0.0290 * |
| Group=LateLearners | -0.29160 | 0.41050 | -0.710 | 0.4775 |
| Group=NonSigners | -0.76805 | 0.38952 | -1.972 | 0.0486 * |
| DurationSec | 0.34756 | 0.14903 | 2.332 | 0.0197 * |
| Order | -0.10993 | 0.23212 | -0.474 | 0.6358 |
| SignDyad | -0.07622 | 0.26318 | -0.290 | 0.7721 |

TABLE 7.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 0.87427 | 0.51262 | 1.705 | 0.0881 |
| Group=EarlyLearners | 0.29160 | 0.41050 | 0.710 | 0.4775 |
| Group=NonSigners | -0.47645 | 0.36372 | -1.310 | 0.1902 |
| DurationSec | 0.34756 | 0.14903 | 2.332 | 0.0197 * |
| Order | -0.10993 | 0.23212 | -0.474 | 0.6358 |
| SignDyad | -0.07622 | 0.26318 | -0.290 | 0.7721 |

TABLE 8.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 0.52271 | 0.37207 | 1.405 | 0.160 |
| Group=LateLearners | -0.21938 | 0.36689 | -0.598 | 0.550 |
| Group=NonSigners | -0.53366 | 0.34744 | -1.536 | 0.125 |
| DurationSec | 0.42637 | 0.09114 | 4.678 | 2.9e-06 *** |
| Order | -0.16598 | 0.14970 | -1.109 | 0.268 |
| SignDyad | 0.09770 | 0.17691 | 0.552 | 0.581 |

TABLE 9.

| Fixed effect | Estimate | Standard error | *z*-value | *p*-value |
|---|---|---|---|---|
| (Intercept) | 0.30333 | 0.35439 | 0.856 | 0.392 |
| Group=EarlyLearners | 0.21938 | 0.36693 | 0.598 | 0.550 |
| Group=NonSigners | -0.31428 | 0.32900 | -0.955 | 0.339 |
| DurationSec | 0.42637 | 0.09114 | 4.678 | 2.9e-06 *** |
| Order | -0.16598 | 0.14970 | -1.109 | 0.268 |
| SignDyad | 0.09770 | 0.17692 | 0.552 | 0.581 |

FIGURE 1.


YOU
YOU-DUAL
PALM-UP
QUESTION-MARK
HOW
HOW-LONG
HOW-MANY
WHO
WHY

FIGURE 2.

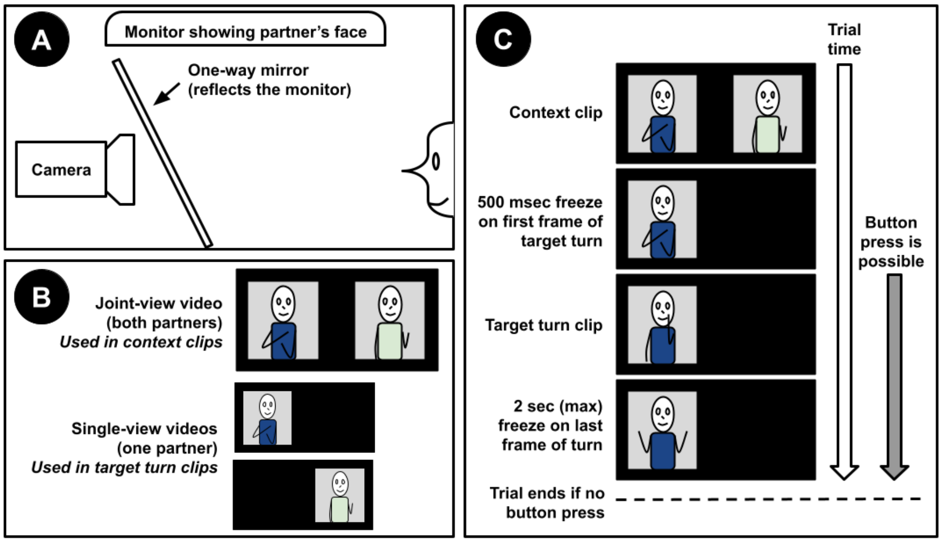
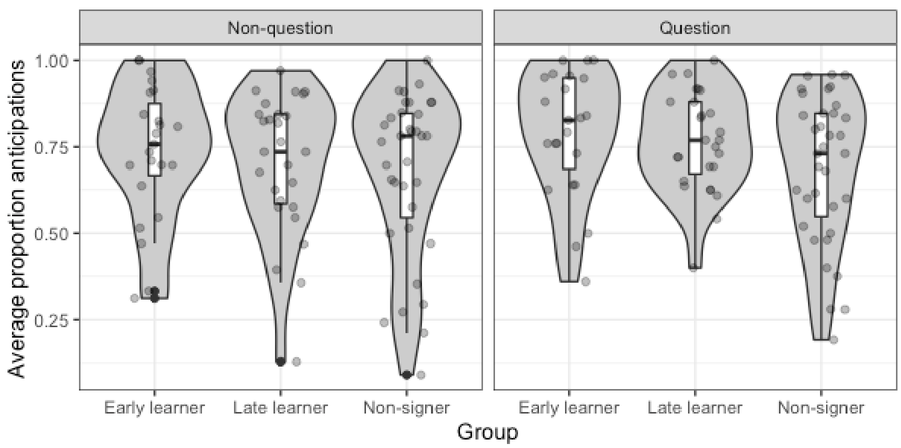
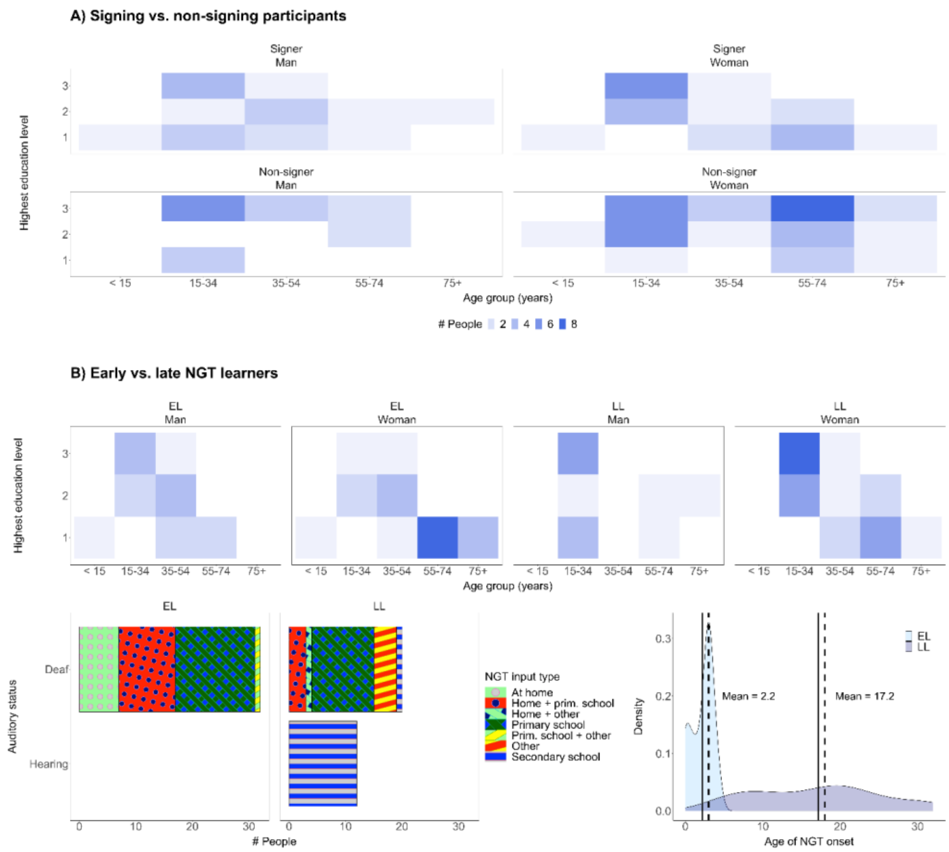

A
Monitor showing partner's face
One-way mirror
(reflects the monitor)
Camera
B
Joint-view video
(both partners)
Used in context clips
Single-view videos
(one partner)
Used in target turn clips
C
Trial
time
Context clip
500 msec freeze
on first frame of
target turn
Target turn clip
2 sec (max)
freeze on last
frame of turn
Trial ends if no
button press
Button
press is
possible

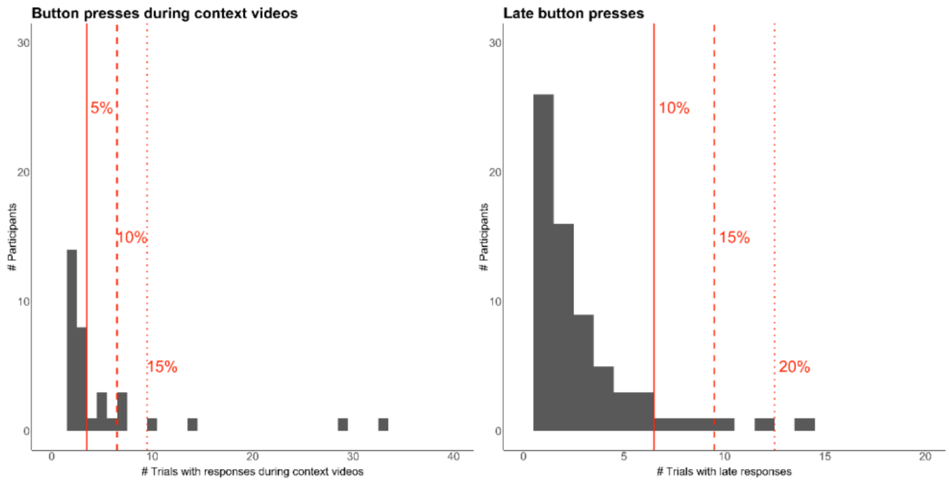
FIGURE 3.


Non-question
Question
Average proportion anticipations
1.00
0.75
0.50
0.25
Early learner
Late learner
Non-signer
Early learner
Late learner
Non-signer
Group

APPENDIX FIGURE 1.



APPENDIX FIGURE 2.



APPENDIX FIGURE 3.