

ARTICLE

# Shifting toward progressive and balanced interaction: A longitudinal corpus study of children's responses to Who-questions in Japanese

Tomoko Tatsumi<sup>1,2,3</sup>  and Julian Pine<sup>2,4</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; <sup>2</sup>University of Liverpool, Department of Psychology, Liverpool, United Kingdom; <sup>3</sup>Kobe University, Graduate School of Intercultural Studies, Kobe, Japan and <sup>4</sup>ESRC International Centre for Language and Communicative Development (LuCiD), United Kingdom

**Corresponding author:** Tomoko Tatsumi; Email: [tomoko.tatsumi2@liverpool.ac.uk](mailto:tomoko.tatsumi2@liverpool.ac.uk)

(Received 11 November 2023; revised 19 November 2024; accepted 27 December 2024)

## Abstract

Children's speech becomes longer and more complex as they develop, but the reasons for this have been insufficiently studied. This study examines how changing linguistic choices in children are linked to interactive factors by analysing Who-question sequences in Japanese child–caregiver conversations. The interactive factors in focus are progressivity and balanced joint activity, which are core aspects of conversational interaction. Our analysis reveals that as children respond to Who-questions, their responses grow in length and multifunctionality. This growth is positively associated with progressivity, namely a quicker completion of the question sequence, and reduced functional load in the interlocutor's contributions, resulting in more balanced joint activity. These findings suggest that children adapt their linguistic choices by observing and aligning them with their interactive goals in conversational sequences.

**Keywords:** conversation; response; sequence

## 要旨

発達に伴って一般に子供の発話は長く複雑になるが、なぜこの変化が起きるのかは十分に研究されていない。本研究は、子どもと保護者の会話における「誰」疑問文のシークエンスを分析することで、子供の言語的選択に、会話のやりとりの進行性と均衡性が関わるかどうかを探った。量的分析により、子供の疑問文への返答が長く複雑になる変化が、以上の2要因と結びつくことが明らかになった。長い返答は、質問のシークエンスの素早い完了に寄与する。またシークエンス内の後続ターンにおいて相手による負荷を減らし、参加者間でより均衡のとれたやりとりを生み出す。これらの研究結果により、子供がやりとりのシークエンスにおける目的に沿うように自身の言語的選択を調整、変化させることが示唆される。

**关键词:** 会话; 应答; 第一语言习得; 日本語; 連鎖

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

## 1. Introduction

Language acquisition occurs through everyday linguistic interactions, which often follow a sequentially structured pattern in which different participants take turns and respond to each other. This dynamic, unpredictable, and rapid interaction poses a challenge for children who strive to participate despite their limited linguistic knowledge (Bates et al., 1975; Clark, 2018; Ninio & Snow, 1988). Initially, children's responses may be non-verbal or consist of simple words, which may or may not be considered appropriate by their conversational partners. Over time, children show the developmental changes in their linguistic choices, eventually exhibiting more adult-like behaviour. However, the factors that drive this transformation in children's response behaviour remain insufficiently explored. One of the possible factors is social interaction. The primary goal of social interaction is to manage relationships (Enfield, 2014), which predicts that children, as well as adults, assess and alter their own linguistic behaviour with respect to this goal. This study investigates the developmental changes in children's responses to Who-questions and the impact of these changes on conversational sequences by analysing corpora of naturalistic Japanese conversation. Our aim is to shed light on the relationship between the local dynamics of question interaction and developmental changes in children's response behaviour.

### 1.1. Children's responses to questions as cooperative actions

Engaging in conversation is primarily cooperative behaviour. Important conversation analytic studies such as Atkinson and Heritage (1984) have pointed out speakers' tendency to maximize cooperation and affiliation while minimising conflict during a conversation. One way to achieve cooperation is by accomplishing an activity intended by the speaker of the preceding turn. In his argument on adjacency pairs, Schegloff (2007: 59) states, that '(s)equences are the vehicle for getting some activity accomplished, and that response to the first pair part which embodies or favours furthering or the accomplishment of the activity is favoured – or, as we shall term it, preferred – second pair part'. In the case of a Question-Answer sequence, the question is typically followed by an answer in the next turn, thus accomplishing the activity that is initiated by the question. Progressivity, as discussed by Goffman (1983), Heritage (2013), Lerner (2006), Robinson (2006), Sacks (1987), and Stivers and Robinson (2006), refers to the basic feature of interaction that each component of the organisation of interaction should 'progress' to the next relevant component immediately after, or contiguously to, the prior component. Progressivity is the norm or default; therefore, behaviour that halts progressivity is accountable (Garfinkel, 1967) and thus examined by participants for its interactional import.

Numerous previous studies have documented developmental changes in how children gradually become able to respond to questions in an adult-like manner. Infants typically use their first words to direct others' attention and negotiate immediate activities (Dore, 1974; Snow et al., 1996), rather than to respond to questions. Around the age of two to three, children are generally responsive to questions in both caregiver-child and peer interaction (Ervin-Tripp, 1979; Gallagher, 1981; Snow, 1977; Wellman & Lempers, 1977), with a gradual increase in responsiveness in later years (Garvey & Hogan, 1973; Mueller, 1972; Van Hekken & Roelofsen, 1982). Children's

understanding of the question-response structure in conversation allows them not only to respond to others but also to ask questions of others. Mueller (1972) investigated peer interactions among children aged 3;6 to 5;6 and found that ~85% of their utterances successfully received replies from others. At these ages, children can produce well-formed utterances, tailored to their listener's perspective, and timed to coincide with the listener's attention. In terms of response types, children generally seem to show adult-like preferences. They prefer answer responses to non-answer responses from as early as 2–3 years (Dore, 1977), and informative responses to non-informative responses from 5 to 11 years (Van Hekken & Roelofsen, 1982). These studies suggest that children learn to respond to their interlocutors' questions in a cooperative manner. Children's turn-taking latency also changes. It is initially slower than adults, but becomes quicker as they get older (Casillas et al., 2016; Ervin-Tripp, 1979). Casillas et al. (2016) also observed that turn-taking latency was influenced by the increasing complexity of conversations during children's development. Additionally, the type of interlocutor (e.g., adult or peer) and the number of interactants play a crucial role in children's successful and smooth participation in conversations, with multi-party interactions presenting greater challenges than dialogic ones (Ervin-Tripp, 1979). Children gradually become skilled at joining varied conversational interactions by mastering interactive and linguistic strategies. The first part of this study tests whether these developmental changes in question-answering, mainly reported for English, also occur in Japanese. Our prediction is that children's responses will become more cooperative, showing an increase in the provision of requested information and their relevance to the interaction.

### 1.2. Linguistic and functional complexity in response utterances

Turning our attention to the linguistic expressions used in interactions, speakers' responses to questions are far more complex and varied than the mere provision of a word corresponding to the requested information. A response can assume a virtually infinite range of linguistic forms. While speakers can produce a minimally informative answer, they can also produce a more complex answer by expressing their attitude toward the content or the interlocutor. They can also adjust the turn-taking or steer the ongoing interaction in a particular direction. For instance, responding to a question 'what is this?' with 'cat' or 'it's a cat, isn't it?' might seem inconsequential, as both are accepted as correct and cooperative answers. However, linguistic details such as the tag question or copula verb construction in this specific example, represent the speaker's linguistic choices in a given social interaction, reflecting distinct functions like expressing certitude or engaging the interlocutor. The presence of such linguistic variation in actual conversation underscores the fact that question-answering cannot be reduced to a simple exchange of information between participants.

The general developmental trajectory in child language involves the use of longer utterances (Ervin-Tripp, 1978; Ochs et al., 1979) and a wider variety of constructions, which also applies to responses to questions (Mueller, 1972). Responses from younger children are often characterised by norm-violating expressions, like a simple 'No' in response to a request (Stivers et al., 2018). This elementary example illustrates the extent to which adults' responses, whether they seek to soften a negative response or to account for it (e.g., 'uh sorry I don't have time'), entail social and interactive calculation. Previous

studies on child language have investigated question-answering by focusing on grammatical well-formedness, interactional aspects such as responsiveness and turn-taking, or informativeness (see Stivers et al. (2018) for a summary), but have overlooked how children develop in terms of their strategic choice of linguistic expressions within actual social interactions. The importance of focusing on linguistic expressions in interaction has been pointed out in a number of important qualitative studies on social interaction (Ervin-Tripp et al., 1990; Goodwin & Goodwin, 1987; Kyratzis & Ervin-Tripp, 1999). In addition, connecting the functional analysis of complex speech in interaction with other levels of analysis, such as the analysis of developmental change, has been a challenge. One way to do this systematically is to restrict the range of interactions and turn designs and to focus on the primary functional properties of utterances in these interactions in a way that allows quantitative evaluation of development change.

This study focuses specifically on Who-questions and answers in Japanese. Question sequences are particularly well-suited to analysing the coherence within sequenced turns as the ‘strongest demand for relevance in a conversation arises when a question is asked’ (Ervin-Tripp, 1978: 359). As such, a question creates a face-threatening situation that conversational participants need to jointly manage, which explains why its sequence is organized via a very robust set of social norms (Dale & Spivey, 2006; Sidnell, 2016). We chose Who-questions as the target for this study because these questions focus on a relatively controlled range of interactions, primarily those revolving around information requests for animate references. Another reason is that who-questions emerge relatively early, along with what-questions and where-questions, compared with other wh-questions about time, activities, reasons, and manners in different languages such as English, Korean, and Japanese (Bloom et al., 1982; Cairns & Hsu, 1978; Clancy, 1989; Uno, 2017). In child–caregiver interaction, questions are often not a genuine search for information, as their referents tend to be given information and also present at the time of speaking (Uno, 2017). Nonetheless, answering these questions with an animate reference is generally a valid choice. In sum, Who-question sequences allow us to evaluate children’s responses in terms of their provision of a requested piece of information and also of adult-like ‘normal’ linguistic choices in social interaction.

Concerning the target language of this study, Japanese presents an intriguing case for exploring speech in social interaction because of its diverse array of modal and politeness-related expressions (Burdelski, 2011; Chang et al., 2021; Clancy, 1985; Cook, 1990; Nakamura, 2014). These expressions add layers of social and interactional dynamics that wrap around the propositional content of the sentence, as posited by Japanese grammarians (e.g., Masuoka, 1991; Kenkyūkai, 2003). For example, a Japanese speaker in a situated interaction would rather produce utterances such as *zettai ame furu yo* (‘(I believe) it will definitely rain’, in plain register) or *kitto ame ga furimasu* (‘It is likely to rain’, in polite register) rather than a simple *ame ga furu* ‘it will rain’. The former two expressions encompass a broader dimension such as the speaker’s judgement on the propositional content, epistemic stance as well as social attitude. These functions make the utterances appropriate and natural in a given social interaction. Since such functionally complex expressions are customary in a normal conversation, the child’s task lies in eventually assimilating these norms and incorporating such expressions into their responses to questions. Accordingly, we hypothesise that, as children develop, their responses will become longer and assume a broader array of functions.

### 1.3. Children's linguistic choice and interaction

As the literature review indicates, children gradually acquire the ability to produce more adult-like speech. Nevertheless, the reasons for this developmental change remain relatively unexplored within the research field. To address this question, and understand how children, as active participants, shape interaction with their own speech, various studies have explored the significance of conversational sequences in child language acquisition. Maternal responses that were contingent upon infant vocalisations have been shown to enhance both the quantity and quality of such vocalisations (Bloom et al., 1987; Goldstein & West, 1999; Topping et al., 2013). Warlaumont et al. (2014) demonstrated that the relevance of children's utterances increases conversational interaction. Adults were observed to respond frequently when children's linguistic production was related rather than unrelated to speech. Furthermore, these adult responses had a positive impact on the likelihood of the children's speech-related vocalisation in subsequent turns. Investigating a longitudinal dataset, Nikolaus et al. (2022) found that caregivers' temporally contingent responses to children's utterances positively influenced the speech-relatedness and intelligibility of the children's subsequent utterances. These studies illustrate how semantic relatedness or contingency within neighbouring turns promotes more interaction. Building on this line of research, our focus is on exploring how children's linguistic choices affect the dynamics of conversational interaction. If children learn to be proficient conversationalists, their behaviour should contribute to the defining factors of a normal conversation: progressive interaction and joint activity.

As conversation-analytic studies contend, progressivity constitutes a fundamental aspect of conversational interactions (Heritage, 2013; Sacks, 1987; Stivers & Robinson, 2006). For a progressive interaction, children need to infer the intentions underlying their interlocutor's action and find a way to meet those intentions. When confronted with an information-requesting question, children must recognize the expectation to supply the requested information or, at minimum, offer something relevant to the interlocutor's intent. This allows them to contribute to the completion of the ongoing sequence, facilitating the transition to a new one. Insights from developmental psychology suggest that humans possess inherent prosocial and cooperative tendencies (Dunfield et al., 2011; Warneken & Tomasello, 2006, 2007). Even preverbal children exhibit the capacity to engage in joint activities, grasp others' intentions, understand goals, and show helping behaviours in various social contexts (Ashley & Tomasello, 1998; Liszkowski et al., 2006; Smiley, 2001). These cognitive abilities form the foundation for a goal-oriented and cooperative, thus progressive, conversation.

Another fundamental structure is joint activity. Conversational participants take turns and jointly make contributions to establish a common ground, and to cope with any breakdowns in communication. Child-caregiver interactions indeed shift toward more joint activity as children mature (Dale & Spivey, 2006; Sokolov, 1993). Initially, caregivers tend to take the lead, constantly encouraging the child to engage to sustain the interaction. As the child develops, conversations tend to become more balanced, with both parties actively engaging in the exchange. This equilibrium entails children adopting behaviours similar to those of adults and obtaining a degree of control over the interaction.

The current study focuses on progressive and balanced interaction as potential factors underlying children's changes in their linguistic behaviour. It postulates that these factors play an essential role in influencing how children monitor and adjust their own speech to achieve more effective social interaction in the future.

1.4. Goal of this study

The current study aims to explore how children’s linguistic behaviour changes over development, and how these changes are related to interactional dynamics. To this end, we test the following hypotheses by running quantitative analyses on a corpus-based dataset of Who-question sequences in Japanese.

- 1. Children’s responses become more cooperative, showing an increase in the provision of requested information and relevance.
- 2. Children’s responses become longer and serve more functions as they develop.
- 3. Children’s longer and multifunctional responses contribute to a quicker completion of the interactive sequence (more progressivity) and to more joint and balanced interaction within the sequence.

2. Method

2.1. Data

In this study, we used the seven Japanese longitudinal corpora available in the CHILDES database (MacWhinney, 2000). These corpora consist of naturalistic conversations, mostly between target children and their caregivers, who are all monolingual Japanese speakers. They include data from three children (Aki, Ryo, and Tai) that comprise the Miyata corpus (Miyata, 2004a, 2004b, 2004c), and four children (ArikaM, Asato, Nanami, and Tomito) that comprise the MiiPro corpus (Miyata & Nisisawa, 2009, 2010; Nisisawa & Miyata, 2009, 2010).

All data were reorganised into a turn-unit dataset by using R (R Core Team, 2024). We extracted all Who-question sequences that started with a Who-question by a caregiver, followed by the target child’s and the caregiver’s turns (815 sequences in total). We excluded the sequences that were preceded by another question in the immediately preceding turn, and those that included participants other than the target child and caregiver in the following 10 turns. We also removed overlaps of sequences within 5 turns (on the basis of 5.177, the mean + standard deviation of the sequence completion score, as detailed in the coding section), to avoid including follow-up or repeated who-questions within a sequence as a new target sequence. The age range of the children was from 1;1 (years; months) to 5;2.

Table 1 shows an example of a Who-question sequence. For the sake of systematicity, we used the turn as the unit of analysis. Turns were defined as single utterances or

Table 1. Labels for utterances in a sequence (Nanami, 3;2.20)

Order of utterances	Speaker	Utterances	English translation	Turn label
1	Caregiver	<i>fuun dare ga tsukuru no ?</i>	ok, who makes this?	A1
2	Child	<i>u</i>	<i>uh</i>	C1
3	Child	<i>Arichan</i>	<i>Arichan</i>	C1
4	Caregiver	<i>Arichan ga tsukuru no ?</i>	Does Arichan make this?	A2
5	Child	<i>un</i>	yeah	C2

sequences of utterances in the original transcripts bounded by speaker changes. The only exceptions were instances where the child produced no verbal response immediately after the Who-question at the beginning of a sequence, in which case, we coded no response for the C1 turn, although there was no speaker change between the question and the caregiver's succeeding utterance. Turns could consist of a single short utterance, such as a label or a backchannel, or of multiple utterances varying in length. Overlaps between speakers were common in the corpora, but less common in our Who-question dataset, where none of the Who-question turns overlapped with neighbouring turns. Where there were overlaps between the turns following Who-question turns, we systematically followed the order of utterances in the transcripts and treated the first overlapping utterance as belonging to a different preceding turn from the second overlapping utterance, which was taken to be the beginning of a different succeeding turn. We sometimes refer to different turns within a sequence by the turn labels shown below.

## 2.2. Coding

In addition to the information available in the original corpora, we coded several other variables for our analyses.

The first analysis looks at how children's responses become more cooperative. To this end, we classified children's responses in the C1 turn into basic response types; Cooperative and Uncooperative. Note that here the terms 'Cooperative' and 'Uncooperative' do not refer to children's intentions or attitudes. Instead, they mean that the verbal production or the lack of it can be interpreted as cooperative or uncooperative by the interlocutor in interaction or by an observer. Cooperative responses are further classified as either Expected or Unexpected to distinguish responses that include an animate reference from those that do not. The animate reference need not be the right answer but can be any word that refers to an animate entity that is not included in the preceding Who-question. Unexpected responses are related to the question but lack the animate reference. These responses consist of function words, metalinguistic expressions (acknowledging a problem in answering the question such as 'I don't know', and repairs like 'huh?') or repetitions of any word in the preceding who-question (e.g., producing 'clock' following the question 'who has this clock?'). These response types were automatically classified by defining categories in terms of the formal features in the transcription and morphological coding in the original corpora. After identifying the lack of a child response as well as unintelligible and nonverbal responses, our R script coded the animate references, and the relatedness of response to classify all the C1 turns into two categories with subcategories: Cooperative (Expected and Unexpected responses), Uncooperative (Unrelated, No response, and unintelligible responses).

The second and third sets of analyses required the coding of the functions of linguistic expressions. We defined five functional categories, namely Reference, three Modalities (Evaluation, Action, and Social), and Interaction, each of which is associated with a group of different linguistic forms listed in Table 2. Reference points to entities. When a speaker produces a noun, noun phrase, or personal pronoun (e.g., 'dinosaur', 'grandma's train', 'me'), this is coded as a Reference. By definition, all Expected responses are referential. While some responses minimally refer to an object by using only its label, other responses are longer and include additional linguistic expressions. Japanese grammarians have characterized Japanese sentences as having a proposition at the core, with the core accompanied by different modality elements (Masuoka, 1991; Nihongo Kijutsu Bunpō



**Table 2.** Definition and corresponding linguistic forms for function categories (note that these are not general or exhaustive, but are intended only to cover the linguistic forms in our response dataset)

Function categories	Definition	Linguistic forms
Reference	Objective expression of things or events	Referential expressions including nouns, noun phrases, and pronouns (e.g. <i>kaiju</i> ‘dinosaur’, <i>baaba no densha</i> ‘grandma’s train’)
Evaluation Modality	Speaker’s evaluation of the proposition (e.g., assertion, negation, epistemy, and evidentiality)	Finite predicates with verbs, adjectives, and nouns (e.g. <i>tabeta</i> ‘(I) ate’, <i>hebi da</i> ‘it is a snake’), verbal particles for epistemic and evidential functions (e.g., <i>tabeta kamo</i> ‘might have eaten’), and evidential adverbs ( <i>kitto</i> ‘perhaps’)
Action Modality	Expressions of initiating speech acts (e.g., questions, requests, and commands)	Verbs and modal final particles for marking intentions, questions, proposals, requests, and commands (e.g. <i>hoshii?</i> ‘do you want it?’, <i>kite</i> ‘Come!’)
Social Modality	Expressions for the speaker’s communicative attitude toward the interlocutor	Utterance-final modal particles for intersubjectivity (e.g., <i>ne</i> for establishing common ground), politeness marker (e.g., <i>-mas</i> , <i>-des</i> for polite register), and polite social formulas (e.g., <i>doozo</i> for offer)
Interaction	Forms related to turn-taking	Response tokens (e.g., <i>un</i> ‘yes’, <i>hee</i> for showing surprise), attention-getting tokens (e.g., <i>hora</i> ‘look!’), vocatives (e.g., <i>anta</i> ‘you’), and fillers (e.g., <i>eeto</i> , <i>aa</i> , etc.)

Kenkyūkai, 2003). Although the definition of modality varies considerably from one researcher to another and also across languages, we posit three categories on the basis of important overlaps in Japanese studies on modality; Evaluation Modality, Action Modality, and Social Modality. Evaluation Modality refers to epistemic and evidential expressions. Epistemic modality expresses the degree of speaker’s certainty that what they are saying is true. Evidential modality expresses the source of evidence a speaker has for their statement (de Haan, 2006). Action Modality covers a range of speech act categories such as questions, requests, proposals, invitations, and commands. It includes not only speaker-oriented modalities for cases in which the speaker gives someone an order or permission (Bybee, 1985; de Haan, 2006), but also other cases in which the speaker is in need and asks someone for help. These expressions tend to initiate a new sequence other than the who-question sequence in our dataset. Social Modality expresses the speaker’s engagement with the interlocutor. It includes speakers’ assumptions about the degree to which their attention or knowledge is shared by the addressee (Evans et al., 2018), as well as politeness and formulaic addresses. In addition to these is the Interaction category. It refers to the speech elements or vocalisations for negotiating turns at the interface with other turns. We thus have five function categories, each of which may or may not be expressed by linguistic means in an utterance or turn. Using our coding scheme, the turn *kaiju da yo* ‘that’s dinosaur’, for example, refers to an object (*kaijuu* ‘dinosaur’), encodes



judgement modality (*da yo* is assertive) as well as social modality (*yo* marks intersubjectivity). It does not contain linguistic expressions that signal action modality or interaction functions. Another example is *anoo kore kore ne otoosan* 'uhm, this, this is, father'. This turn has a reference (*otoosan* 'father'), encodes social modality (*ne* for establishing common ground), and interaction (*anoo* is a filler to start a turn). Table 2 explains our working definitions and coding scheme.

We coded all children's response turns as well as caregivers' turns into these functional categories using R. Each of the categories was defined by linguistic forms that were found in our data following the classification in Table 2. This allowed us to ensure consistency by coding the data automatically. A turn can have some or all of the functions in Table 2, ordered in different ways, and can also use multiple linguistic elements to express a function. We dealt with the variation and complexity of naturalistic speech data by coding the presence and absence of these five functions regardless of order and redundancy. This means that the number of functions in a turn varies between 1 and 5. Note that we do not assume a one-to-one or clear-cut mapping between function and form, but treat linguistic forms as cues for functions. Despite being very broad, we believe that our five functional categories are useful in elucidating developmental changes in child speech, and may also be useful for comparing different languages in future studies.

Sequence completion is a measure of progressivity for the third set of analyses. It refers to the turn at which participants start a new sequence (e.g., asking a different question, introducing a new topic, etc.). In other words, it is the turn immediately after the last turn of a who-question sequence. This factor was also coded automatically using R commands to ensure consistency. The beginning of a new sequence was identified as the first turn that introduces a new topic in a given who-question sequence. In our coding scheme, a new topic means a word that is not included in any of the following categories that constitute turns within a who-question sequence: (1) animate references within the 10 turns from the who-question, (2) grammatical words (existential/copula verbs, case particles, modal particles, formal nouns, demonstrative pronouns and adverbs), (3) tokens for acknowledgement, denial, interjections and fillers (e.g., *un* yes, *chigau* not right, *hee* ah, *anoo* uhm), and (4) metalinguistic expressions for perception and understanding including repair request (e.g., *wakannai* 'don't understand', *dore* 'which?') as well as part-words that are mostly considered as false-starts or hesitation. Another way to identify the beginning of a sequence is by detecting expressions that initiate a move, which basically corresponds to the first pair part of an adjacency pair. These expressions include wh-questions and follow-up questions (e.g., *kore/kotchi wa?* 'and this?', and *dore ga ii* 'which do you like better?'), commands (e.g., *kudasai* 'give me', *matte* 'wait' *mite* 'look'), and offer (*doozo* 'here it is'). The sequence completion for a sequence was coded by detecting these forms, searching from C1 to later turns for each sequence. We acknowledge that this method, like any other, can only approximate sequence completion in spontaneous interaction. Still, this automated method is an improvement in hand-coding in terms of systematicity and reproducibility. The following example illustrates this coding process conducted automatically in R. C1 turn has a new word compared with A1, which is *obaasan* 'grandmother'. This turn is considered within the sequence because it is an animate reference. The next A2 turn adds two new animate reference nouns, one of which (*obaachan*) repeats the referent of the C1 noun *obaasan*, and the other is a new animate referent, *oneechan* 'older sister'. It also adds *chigau* 'not right', and *deshoo* 'isn't it'. A2 turn is also a part of the who-question sequence because none of these words add a new topic or

initiate a new move. This turn is the last turn in the who-question sequence because the following C2 turn adds a new topic *hambaaga* ‘hamburger’, which is not included anywhere in the preceding turns A1–A2.

Example of the coding of sequence completion (Ryo, 2;1.18)

	Utterances	All new words compared with preceding turn(s)	Words within the categories defining who-question sequences	New topic/ initiation of a new sequence
A1	<i>dare ga motte kite kaetta no kore?</i> 'Who brought this home?'			
C1	<i>Obaasan</i> 'Grandmother'	<i>obaasan</i>	<i>obaasan</i> (animate reference)	
A2	<i>Obaachan chigau oneechan deshoo ?</i> 'Not grandmother It was your sister, right?'	<i>Obaachan chigau oneechan deshoo</i>	<i>obaachan, oneechan</i> (animate reference) <i>chigau</i> (metalinguistic expression) <i>deshoo</i> (grammatical particle)	
C2	<i>Hambaaaga nai</i> 'There's no hamburger'	<i>hambaaaga nai</i>	<i>nai</i> (metalinguistic expression)	<i>hambaaaga</i> is a word that introduces a new topic = sequence completion

2.3. Analysis

We tested our hypotheses by using generalized mixed-effect linear models (lme4 packages, Bates et al., 2015) on R (R Core Team, 2024), with corpora, which corresponds to speaker/dyad distinctions, as a random effect.

In addition to these quantitative analyses, we report qualitative analyses on a range of examples of question sequences from our conversational dataset. These examples not only complement our arguments but also highlight the variation and complexity that we find in real social interactions. The dataset and analysis script are published online on the OSF website ([https://osf.io/v3xhu/?view\\_only=f48936b74c2549a8ad61a086a492ef2f](https://osf.io/v3xhu/?view_only=f48936b74c2549a8ad61a086a492ef2f)).

3. Results

3.1. Do children’s responses become more cooperative as they get older?

To test the prediction that children’s responses become more cooperative as they get older, we first classified children’s responses to Who-questions into Cooperative, Uncooperative, and Other responses. The cooperative category is further subdivided into Expected (with an animate reference) and Unexpected (without an animate reference). Example 1 shows a typical example of an Expected response, in which C1 is a good answer to the question in A1. Following C1, A2 turn confirms the answer, with a slight modification of the name in question, and moves on by asking a new question. Unexpected responses are related to the question but lack the animate reference. They

include utterances that focus on the child's lack of the knowledge required to answer the question, repair utterances, and so on. Examples 2 and 3 are two such instances, in which the child indicates their inability to answer the question, with different linguistic means. Note how the child and adult try to understand each other, which results in an extended sequence in Example 3. Uncooperative responses are not related to the question, including instances of no verbal response. Example 4 is an instance of no verbal response from the child. In this case, the adult closes the sequence by giving the answer to her own question. These different examples show not only the different types of child responses to Who-questions, but also the impact of these responses on the interactional dynamics.

Example 1. (Nanami, 2;8.19)

A1	<i>kore dare ?</i>	Who is this?
C1	<i>Batakochan</i>	<i>Batakochan</i> (character name)
A2	<i>Batakosan. kore wa nan da ?</i>	<i>Batakosan</i> . What is this?
C2	<i>Dokinchan</i>	<i>Dokinchan</i> (character name)
A3	<i>Dokinchan ka.</i>	<i>Dokinchan</i>

Example 2. (Tai, 2;5.19)

A1	<i>kondo dare ?</i>	Who is it this time?
C1	<i>kondo ne wakaranai</i>	I don't know this time
A2	<i>un. fuun. Paaman tte</i>	Hm, all right. It's <i>Paaman</i> (character name).
C2	<i>Paaman</i>	<i>Paaman</i>

Example 3. (Ryo, 2;4.22)

A1	<i>kinoo dare to asonda?</i>	Who did you play with yesterday?
C1	<i>n ?</i>	Hm ?
A2	<i>kinoo dare to asonda?</i>	Who did you play with yesterday?
C2	<i>n ?</i>	Hm ?
A3	<i>dare to asonda?</i>	Who did you play with?
C3	<i>n?</i>	Hm ?
A4	<i>wakattenai ?</i>	You are not getting it?
C4	<i>n ?</i>	Hm ?
A5	<i>wakannai ?</i>	You don't get it?
C5	<i>un</i>	Yes

Example 4. (ArikaM, 3;1.4)

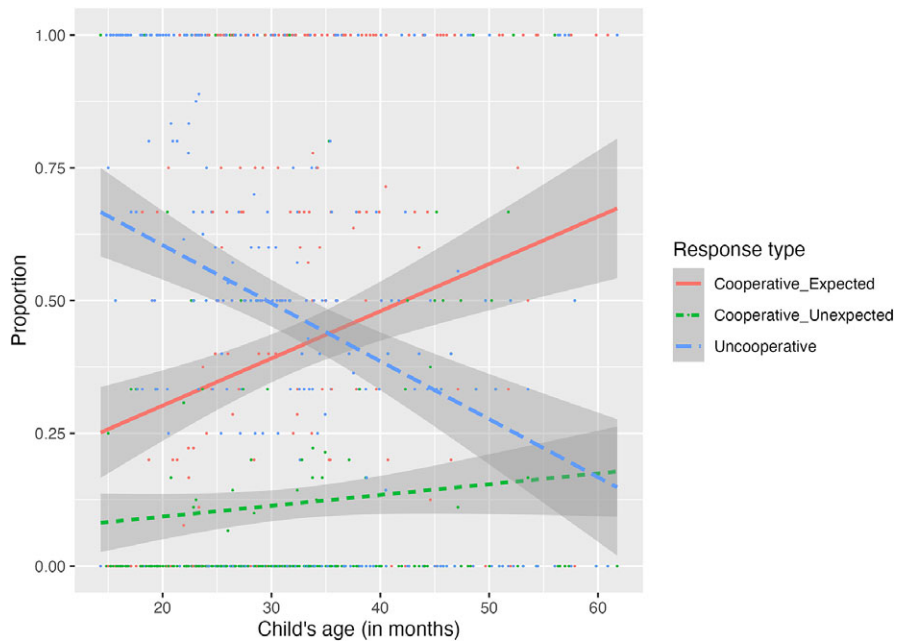
A1	<i>oyatsu wa dare ga taberu no ?</i>	Who is eating the snacks?
C1	[No Response]	[No Response]
A2	<i>Akko to Arichan ?</i>	<i>Akko</i> and <i>Arichan</i> ?
C2	<i>un</i>	Yeah

Table 3 shows the breakdown of children's responses to Who-questions. Cooperative and Uncooperative responses each account for approximately half of the data.

More importantly, as shown in Figure 1, the proportion of these different types of responses changes across development. Our results from a generalized mixed-effect model show that children's responses become more cooperative as they get older. Expected responses, which pick out an animate referent in answer to a Who-question, increase over development (Estimate = 0.044, SE = 0.008,  $z = 5.321$ ,  $p = 1.03e-07$ ).

**Table 3.** Counts of different response types

Response types		Criterion	Count	
Cooperative	Expected	with an animate reference that addresses the Who-question	320	
	Unexpected	without animate reference but related to the question	98	
Uncooperative	Unrelated	not related to the question	98	
	No response	no verbal response	271	397
	Unintelligible	unintelligible utterance	28	



**Figure 1.** Proportion of children's response types by their age.

Cooperative but unexpected responses also show a significant increase (Estimate = 0.030, SE = 0.010,  $z = 2.863$ ,  $p = 0.004$ ). This means that children become cooperative not only in their provision of information, but also in their meta-interactive behaviours that signal, for example, non-understanding, non-hearing or their lack of knowledge for answering the question. In contrast, uncooperative responses decrease over time (Estimate =  $-0.059$ , SE = 0.008,  $z = -7.046$ ,  $p = 1.84e-12$ ). These age effects complement the previous studies' findings that children's preference for cooperative responses is adult-like (Dore, 1977; van Hekken & Roelofsen, 1982), by supporting our hypothesis that children's responses become more cooperative as they get older.

Our analyses in this first section have shown that children's responses become more cooperative as they develop. Not only do children's Expected Cooperative responses (i.e., with an animate reference) increase, but also the Unexpected Cooperative responses (i.e., without animate reference but relevant) increase. The increase in this latter type of response suggests that children exhibit increasing meta-interactional awareness, indicating their improved ability to navigate interactions by addressing the risk of communication breakdowns. Given these developmental changes in children's responses, an obvious question is what these changing linguistic choices bring to the interaction. Before directly addressing this question in the last section, we will now narrow our focus to children's answer utterances to gain insight into what children try to achieve in interaction.

### 3.2. What are children's response utterances designed to do?

This section attempts to capture the developmental changes in children's linguistic expressions, particularly in terms of utterance length and multifunctionality. Children's responses become longer in terms of word count. Notably, not only does the proportion of children's no responses diminish over time, but the length of Expected responses also exhibits an increase. Figure 2 visually represents the effect of children's age on the length of their Expected Cooperative responses (Estimate = 1.250, SE = 0.178,  $z = 7.007$ ,  $p = 2.44e-12$ ). The earliest stages are characterized by child responses consisting of a

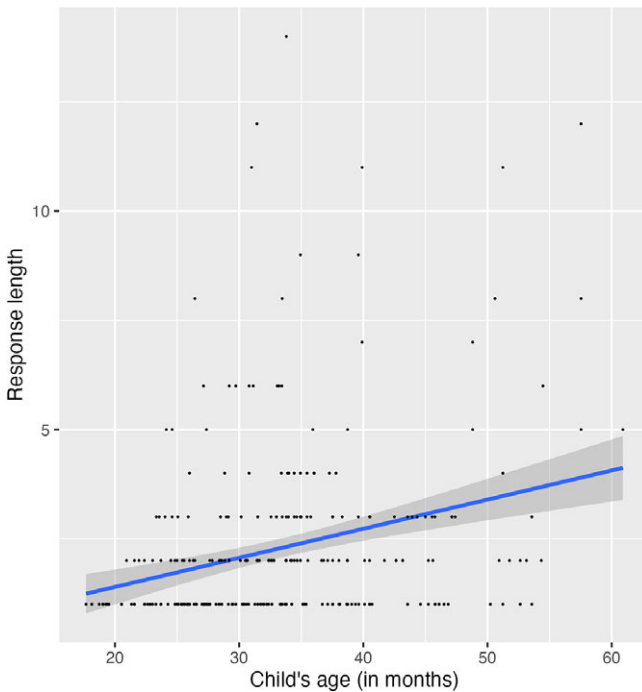


Figure 2. Utterance length (in word count) of children's C1 turn responses by age.

single word, yet such responses become less frequent in the later stages. Typically, a single word is sufficient for providing the information sought by a Who-question. For instance, Example 5 shows a child's simple reply of *kakka* 'Mom'. In contrast, a longer answer in Example 6, exhibits greater complexity. It involves a *kore wa X* 'this is X' construction, as well as the use of the modal particle *ne* whose important function is to establish an affective common ground with the interlocutor (Cook, 1990). This response also uses a genitive modification ('sister rabbit's') to give detailed information. As these contrastive examples illustrate, the developmental increase in the length of the response utterance seems to imply that children not only furnish the requested information but also encode additional functions in their utterances. Note that we do not assume that children have an adult-like functional or semantic understanding of all these linguistic expressions, but that these expressions shape the interaction since they are interpreted and responded to by the interlocutor.

Example 5. (Tai, 1;7.8)

A1	<i>kore dare ga noru no koko ni ?</i>	Who will ride this here?
C1	<i>Kakka</i>	Mom
A2	<i>n ?</i>	Hm?
C2	<i>Kakka</i>	Mom
A3	<i>Okaasan noru no?</i>	Will your mother ride?

Example 6. (Aki, 2;10.28)

A1	<i>kore dare no uchi ?</i>	Whose house is it?
C1	<i>kore wa ne usagi no oneesan no ne kodomo</i>	This is the sister rabbit's child
A2	<i>kodomo?</i>	Child?

We now turn to the functions of children's linguistic choices. As explained in the method section, we used five functional categories; Reference, Evaluation Modality, Action Modality, Social Modality, and Interaction. The multifunctionality is coded in terms of number of function categories. Most early verbal responses are only referential, and simply provide the piece of information requested by the interlocutor's Who-question (one function in Figure 3 and Figure 4). The earlier Example 1 is such an instance (C1 only provides a character's name as their response). The proportion of two-function turns also reduces over development. In contrast, 3- to 5-function turns increase, showing that children's responses become functionally more loaded. Example 7 below shows one such multifunctional response. The C1 includes discursive fillers (*anoo* 'uhm'), a modal particle (*ne*), which is typically used for establishing common ground with the interlocutor, in addition to the target reference 'father'. These items are coded into Interaction, Social, and Reference functions, respectively.

Example 7. (Nanami, 1;11.10)

A1	<i>dare ni oshiete moratta no kondo joozu ni dekite ?</i>	Who taught you this, you did it well this time?
C1	<u><i>anoo kore kore ne otoosan</i></u>	Uhm, this, this is, father
A2	<i>un un un</i>	Yeah yeah yeah

Example 8 shows a case in which the child uses Evaluation and Social modality as well as Reference (*kaijuu* 'dinosaur'). What follows the noun is a finite copula verb *da* that expresses an assertive stance toward the proposition. The modal particle *yo* reflects both

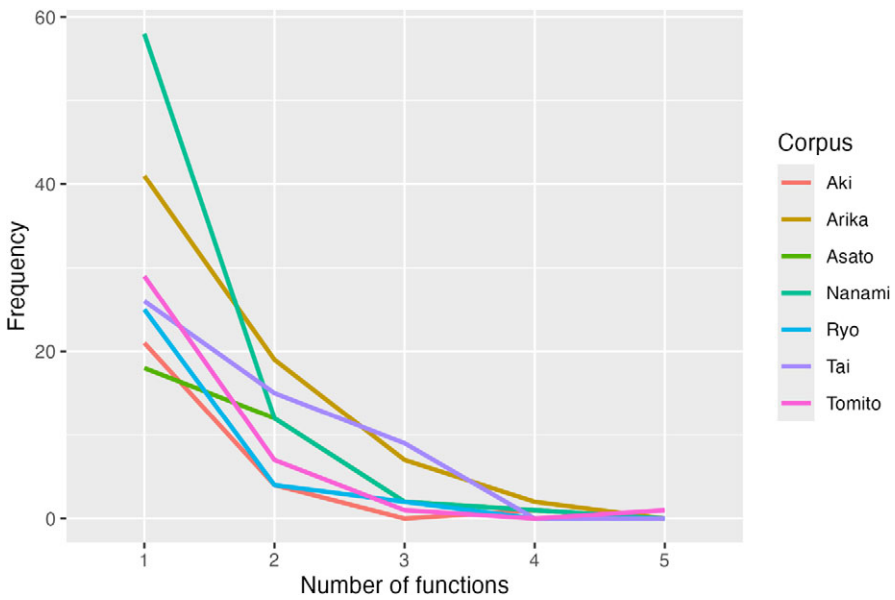


Figure 3. Distribution of number of functions by corpus (child).

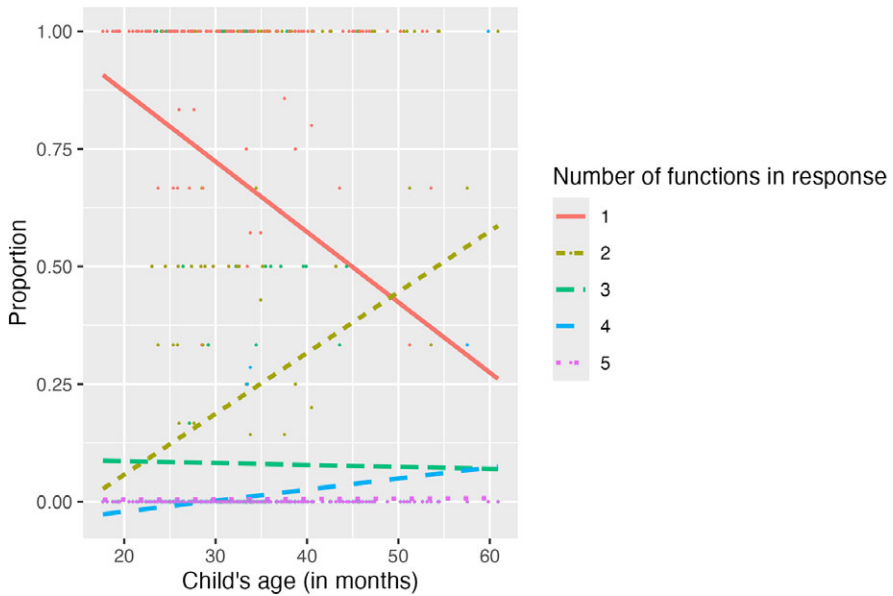


Figure 4. The number of functions in children's C1 response turn by their age.



Evaluation and Social modalities, as it is assertive, and marks the stance that the propositional information belongs to the speaker rather than the interlocutor.

Example 9 shows another kind of response, in which the child answers the question and then asks the interlocutor to play with them by verbally expressing Action modality. Action modality is less frequent than other categories in our data. This modality, which includes requests and questions, characteristically initiates a new sequence, and is therefore not the most typical modality for the turn immediately following a question. Yet children sometimes actively manipulate the interaction by throwing in a question or command of their own.

All these verbal elements specify and manage different aspects of the ongoing interaction. Children do not simply answer with the requested information, but encode, for example, their stance toward the information or the interlocutor. Such functions influence the way in which the interlocutor responds in the next turn. The construction *da yo* in Example 8 marks the fact that the child speaker is assertive about the fact that it is a dinosaur, and that this information belongs to the child himself rather than to the interlocutor. This turn is followed by the caregiver's questioning in A2. The hortative verb construction ('let's...') in Example 9 creates a new sequence that invites the caregiver to respond to the child's proposal. These examples show how the linguistic choices in a turn shape the next turn. Children, by choosing what to say, manipulate the interaction.

Example 8. (Asato, 3;3.18)

A1	<i>dare o ?</i>	Who?
C1	<i>kaijuu da yo</i>	That's dinosaur
A2	<i>kaijuu o?</i>	Dinosaur?
C2	<i>un</i>	Yeah

Example 9. (Asato, 3;3.18)

A1	<i>dare ga ?</i>	Who?
C1	<i>Baaba no shinkansen to densha issho ni oikakekko shiyoo yo</i>	Grandma's bullet train and train. Let's play tag together!

We analysed how children's Expected responses changed in terms of the number of functions they expressed. Figure 3 shows how this variable differs among the seven children in our dataset. The general distribution is similar across all children, responses with only one function are the most common. The more functions the less responses.

Figure 4 shows the general tendency for responses to include more functions with age. The proportion of responses with one function decreases (Estimate =  $-0.072$ , SE =  $0.017$ ,  $z = -4.232$ ,  $p = 2.32e-05$ ), and conversely those with multiple functions increase. Note that the responses with only one function are those with only Reference. Reference is the defining element of Expected responses, but the proportion of answers with only Reference decreases sharply over development. We do not discuss every observed combination of the five functions, but instead add that the probability of all other categories (Evaluation, Action, Social, and Interaction) increases, and that children code more functions in their C1 turns over development.

This second set of analyses looked at the developmental changes that occur at the linguistic level by focusing on the length and multifunctionality of children's answer responses. Children produce longer responses as they develop, which implies children's coding of more functions than a minimal provision of requested information. Our qualitative analysis has in fact illustrated that children generate responses that encompass

diverse functions. More importantly, children increasingly produced multifunctional responses over development. Given these results, our next question is what serves as the driving force behind these changes in children's linguistic choices. The last section thus examines the hypothesised relationship between children's responses and the subsequent interaction within a question sequence.

### 3.3. How do children's responses shape subsequent interaction?

To investigate the underlying factors behind the observed shifts in children's linguistic choices, we posit that these changes benefit interaction. Our first prediction is that children's longer and more multifunctional responses contribute to increased progressivity in interaction, – resulting in a quicker completion of question sequences. Our second prediction holds that such responses on the part of the child create more balanced joint activity within the sequence.

The first prediction refers to the progressivity of conversational sequences, which is considered as one of the general factors behind adults' conversational structures. Conversation-analytic studies have demonstrated that we tend to conclude a sequence as quickly as possible (Stivers & Robinson, 2006), to minimize the collaborative effort of conversational participants (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). In a question-answering sequence, we typically answer a question in its immediately following turn. The quick completion of a given question sequence, rather than a slow or delayed completion, is the normal and preferred pattern of conversational interaction. Furthermore, it yields various advantages, including heightened cooperation and information acquisition, as a swift sequence completion creates space for new sequences to follow. In light of these observations, we posit progressivity as one of the driving forces propelling the changes in children's linguistic behaviours. These changes refer to, as we have shown in the prior section, children's response utterances becoming longer and marking more functions as they develop. If indeed the long and multifunctional responses facilitate progressive interaction, then this can serve as a motivation for children to produce such responses. To empirically assess this idea, we examined the relationship between the number of turns leading to the completion of a question sequence and the number of functions in C1 turns. Our prediction is that a multifunctional C1 turn tends to close the Who-question sequence quickly, without needing additional turns, either by child or adult, to complement or sustain the interaction.

To investigate the relationship between C1 turn and the characteristics of the sequence, we created the variable of sequence completion. Figure 5 shows its distribution by individual child. Sequence completions at turn 1 (C1), 2 (A2), 3 (C2), and 4 (A3) are common. A score of 10 groups all sequences that extend over C5. Despite individual differences, sequence completion occurs most often at A2, meaning that the second turn from the caregiver (following the child's response to the who-question) tends to initiate a new sequence, in all corpora.

Figure 6 shows the negative relationship between the length of Who-question sequences (number of turns) and the length of C1 turns (number of words). Our model analysis confirms that the longer the C1 turn the quicker the sequence completion (Estimate =  $-0.103$ , SE =  $0.019$ ,  $z = -5.344$ ,  $p = 9.11\text{e}-08$ ). Figure 7 also shows a negative association with the number of functions encoded in C1 turns (Estimate =  $-0.245$ , SE =  $0.052$ ,  $z = -4.729$ ,  $p = 2.26\text{e}-06$ ). As detailed in the method section, the number of

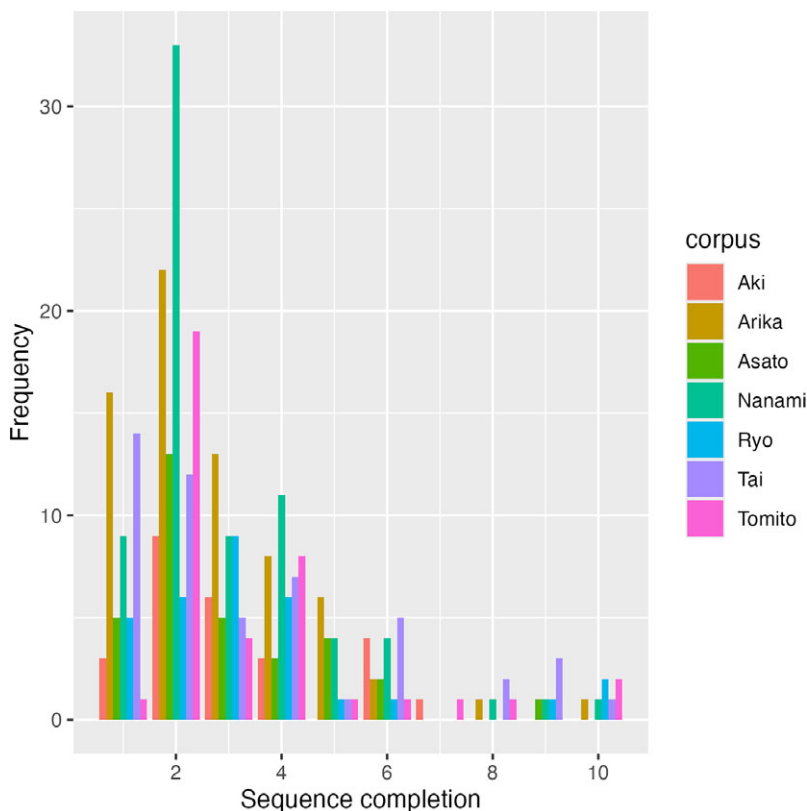
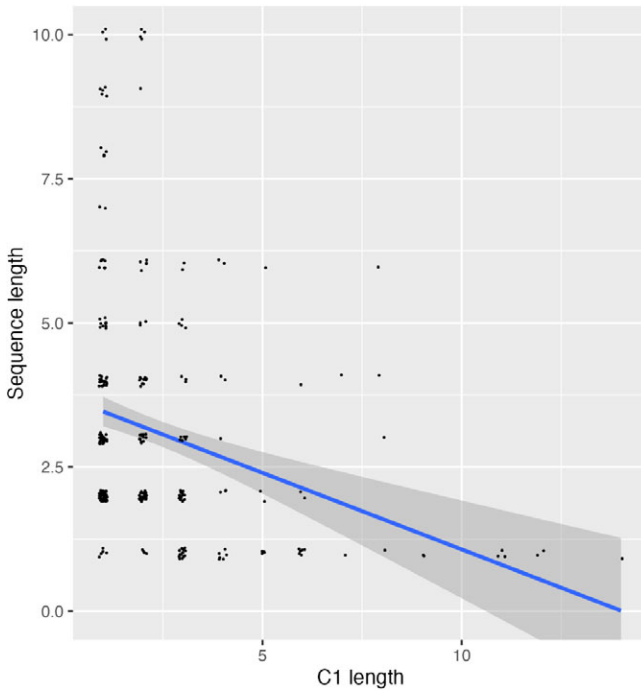


Figure 5. Distribution of sequence completion by corpus (child).

functions ranges between 1 (only Reference function) and 5 (Reference, Evaluation, Action, Social, and Interaction functions). These results are in line with our prediction that longer or multifunctional responses help complete the question sequence quickly. Despite its greater cognitive effort, a longer or multifunctional response has an interactional advantage, namely greater progressivity.

Next, we investigate our second prediction that children's multifunctional responses create more balanced joint activity within a sequence. Conversation is a cooperative activity between different participants, who constantly seek to establish normal interaction through a range of resources such as repair, back-channelling and so on, to cope with or hedge difficulties. When a speaker's turn lacks a certain element, it is possible that another participant attempts to provide it in subsequent turns. Example 10 can be seen as such an example. The child in the C1 turn gives a simple answer *nena* 'sis'. Although this is the right reference for answering the Who-question, the caregiver in A2 turn follows this response up by asking a fully-fledged question, including an affirmative finite verb and modal particle, for confirmation. The sequence ends in A3, with the caregiver's acknowledgement token *un*, after the child repeats the same word three times in C2. This example suggests not only that producing the target answer word alone sometimes invites more interaction in a sequence, but also that the interlocutor may be prompted to provide the



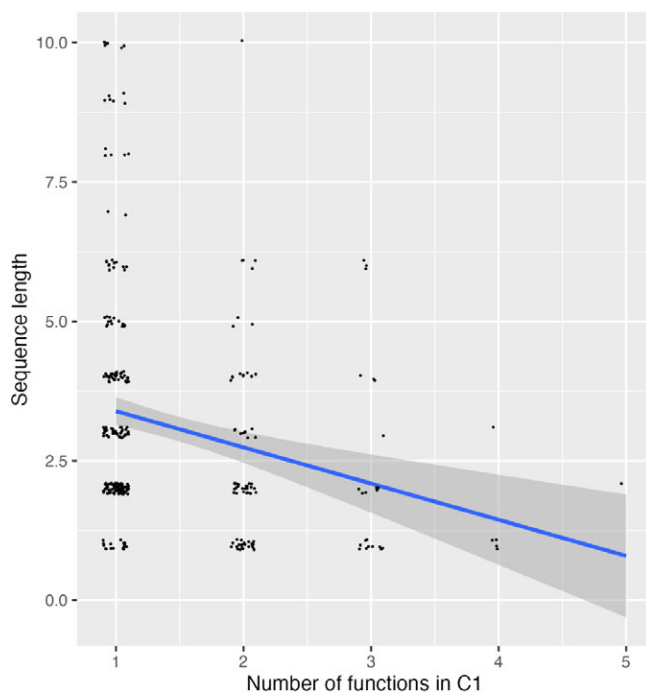
**Figure 6.** Relationship between the length of Who-question sequences (number of turns) and the length of C1 turns (number of words).

absent functions. In other words, participants complement each other in pursuit of normal interaction. When a young child interacts with an adult, the child's short and simple response will require the adult to make more effort to manage the interaction. As the child grows older and produces responses with more words and functions, the interaction will become more balanced, with both the child and adult contributing to the interaction.

Example 10. (Ryo, 1;10.12)

- |    |                        |                            |
|----|------------------------|----------------------------|
| A1 | dare tsukutta no ?     | Who made this?             |
| C1 | <u>nena</u>            | Sis                        |
| A2 | Oneechan tsukutta no ? | Did your sister make this? |
| C2 | Nena nena nena         | Sis sis sis                |
| A3 | un                     | Yeah                       |

We therefore hypothesise that the functions in children's C1 turn and those in their interlocutor's (A's) subsequent turns within a Who-question sequence are in a predicted trade-off relationship. To test this idea, we investigated whether the number of functions in the child response in C1 turn is negatively associated with the number of functions in caregivers' turns that follow children's responses within the question sequence. The model results support our prediction that child's multifunctionality is a negative predictor of caregiver's multifunctionality (Estimate =  $-0.203$ , SE =  $0.064$ ,  $z = -3.167$ ,  $p = 0.002$ ) as



**Figure 7.** Relationship between the length of Who-question sequences (number of turns) and the number of functions in C1 turns.

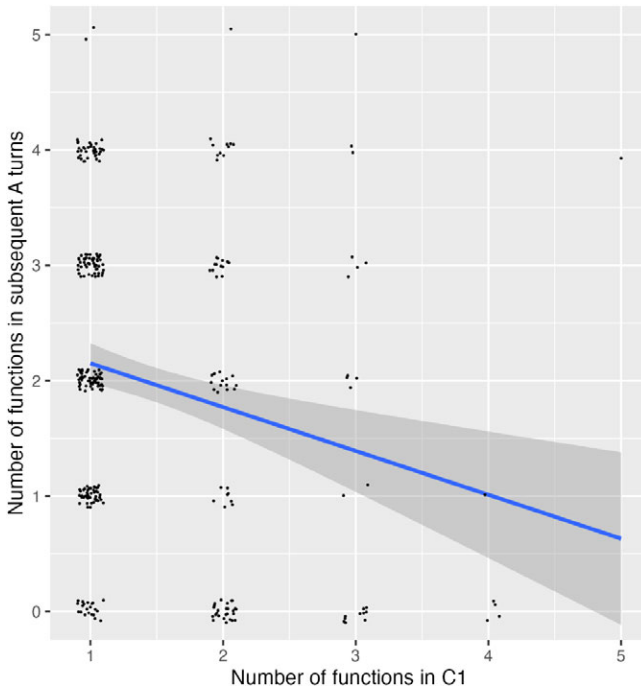
shown in Figure 8. This result implies a trade-off relationship between the child and caregiver regarding functional load in their conversational turns.

This section has demonstrated the association between children's long and multifunctional responses and enhanced progressivity and more balanced interaction. In conjunction with the developmental changes that children produce longer and more functionally-loaded utterances, we argue that children's linguistic behaviour evolves in a way that facilitates progressive and joint interaction.

#### 4. Discussion

This study has examined Who-question sequences focusing on the interactive factors of progressivity and joint activity, which potentially link the children's immediate linguistic choices and their developmental adaptations. Our corpus-based quantitative analyses revealed that (1) children's response becomes more cooperative throughout development, (2) their response utterances become longer and include more functions over time, and (3) their longer and multifunctional responses correlate with more progressivity and balanced interaction within question sequences.

As previous studies have reported, children's responses shift toward greater cooperativity over development. In our dataset of Japanese Who-question sequences, an increase was observed both in children's provision of animate references in the C1 turn and in their non-answer responses that maintain relevance to the question, such as saying *wakaranai*



**Figure 8.** Relationship between the number of functions of A turns (interlocutor's turns until the end of sequence) and the number of functions of C1 turn.

'I don't know'. These cooperative responses indicate children's awareness and linguistic articulation of normative conversational structure. Conversely, non-cooperative responses – namely irrelevant or absent responses – diminish as children get older. With regard to the existing claim of an adult-like preferences for answer responses over non-answer responses or for informativeness in children (Dore, 1977; Van Hekken & Roelofsen, 1982), our results suggest a gradual longitudinal shift toward information-providing and cooperative responses. This implies a learning process whereby children adjust their behaviour choices through experience.

Another important developmental change, illuminated by our second set of analyses, reveals that children's answers to Who-questions become longer and more multifunctional. While a Who-question can be succinctly answered with a single word, children often craft longer utterances to express evaluative, social and interactive stance in C1 turns. They become able to attend linguistically to these varied communicative aspects, which make them competent conversational participants. Who-questions necessitate not merely an informational exchange, but an interaction wherein participants mutually invest in sharing their belief and attitude toward the propositional content, saving face, and modulating turn-taking. Such demands increase the proficiency of children's linguistic production, requiring the right choice of words and constructions that effectively communicate their intentions.

Our third set of analyses explored the impact of C1 turns' length and multifunctionality on the dynamics of question sequences. These factors showed a positive correlation with a quicker completion (progressivity) of question sequences, implying that children's

long or multifunctional utterances in C1 turns generally contribute to a quicker completion of the ongoing sequence. Consequently, they are able to achieve a local interactional goal, namely, satisfying the interlocutor's request expressed in the form of a question. Furthermore, these utterances are associated with balanced interaction amongst participants. Children, by adding varied functions to their utterances, actively engage themselves in communication. We identify this as a pivotal phenomenon that is necessary for explaining the longitudinal transition in the nature of interaction from being adult-led to being a more balanced joint activity (Dale & Spivey, 2006; Sokolov, 1993). From a broader perspective, these findings suggest a probabilistic contingency between children's C1 turn and its subsequent sequence. Experiencing this contingency between their own linguistic behaviour and its consequence in interaction is an important learning opportunity if children wish to adapt to normal conversational interaction.

Contemplating the relationship between children's linguistic behaviour and subsequent interaction invites different possible interpretations. One crucial question is the level of goal-directedness that we attribute to children's linguistic behaviour. On the one hand, imitation or copying behaviour is a basic learning strategy especially until around two years of age. A number of studies have shown that infants copy others' behaviour with or without an understanding of the goal or intention. For example, 2-year-olds imitate model's inefficient actions for reaching an object with a tool without revising them according to the goal (Nagell et al., 1993). This means that children focused more on reproducing the observed behaviour than attaining the goal. Early linguistic choices are also heavily dependent on copying and repeating of the most recent or frequent forms in the perceived input (Bloom et al., 1974). These behaviours are motivated not only by learning needs but also by the basic pro-social tendency for children to try to behave like the people around them (Uzgiris, 1981). On the basis of these studies, one might assume that children, at least at the earliest stages, are not capable of accessing or paying attention to the goals of the interaction, let alone of evaluating conversational sequences against these goals.

On the other hand, children are known to be teleological, showing certain intentional and goal-directed actions already by the end of the first year (Carpenter et al., 1998; Gergely & Csibra, 2003). For example, children copy adults' adjectives more when these adjectives serve descriptive or contrastive functions in the input, but not when they are produced by a slip-of-the-tongue (Bannard et al., 2013; Nielsen, 2006). Although we cannot assume that children recognise goals in a completely adult-like manner, children's developmental change in their response behaviour in the current study seems to suggest that children pay attention to basic goal-related intentions. If children perceive the goal of answering a question and thereby finishing a sequence of interaction, they probably (at least attempt to) plan and execute behaviour to attain this goal. They will thus choose their linguistic expressions to achieve this goal and assess these expressions against observed outcomes – namely, whether the goal was in fact achieved. If their use of functionally complex utterances tends to lead to a more successful subsequent interaction, children may learn to increase their use of this kind of utterance in future interactions. Similarly, balanced joint activity may also serve as an interactive goal for children in a way that children try to match the level of contribution with that of adults'. As both progressivity and joint activity inherently define normal cooperative conversations, they may also motivate children to fine-tune their own linguistic behaviour choices over time.

Another pivotal factor in elucidating children's linguistic behaviour is children's active moves to manipulate interactions. Children's long and multifunctional utterances encompass an array of actional modalities such as proposals, commands, and requests. They may also extend the existing topic or introduce a new one. They sometimes



combine these new spontaneous moves with their question-answering (e.g., provide the requested information first and then propose a new activity in the same turn). Expressing these different demands verbally in interaction often necessitates diverse and complex utterances.

Our study underlines the significance of interactional goals within conversational interaction. Given our prevailing knowledge regarding children's abilities to comprehend goals and intentions (Carpenter et al., 1998; Gergely & Csibra, 2003), self-monitor their behaviour (e.g., self-repair, Forrester, 2008), and construct social understanding (Carpendale & Lewis, 2004), children's linguistic behaviour must also be understood in the light of their conversational goals. Progressivity, or satisfying the interlocutor's request quickly, could represent such a goal, as it is fundamental in explaining the structure of question sequences. Joint activity is also essential in conversation and, more broadly, in social interaction, in that it encapsulates turn-taking and cooperation. Our results reveal that children participate in sequenced interactions where specific facets of their linguistic behaviour in a turn invite quicker completion and more balanced interaction in immediately succeeding turns. The mechanism through which children associate their linguistic behaviour with subsequent interactions merits further, more comprehensive investigation.

As a study on naturalistic corpus data, the current study investigated sequences initiated by Who-questions, irrespective of preceding interactions and the context of their speech and activities. Certain sequences are unmistakably identified as information requests, while others are not. A typical discourse function of Wh-questions is to introduce a topic, allocate a turn and elicit conversation, with the latter two elements being particularly essential in child-caregiver conversations (Snow, 1986). Children may experience less pressure to deliver an informative response in these cases, which is a potential source of bias in our dataset. There are also likely to be individual differences between children, which are difficult to evaluate in this study because our dataset consists of data from seven child-caregiver pairs which extend over different developmental periods and include interactions over a range of different activities. A future study could explore naturalistic interactions within a controlled task to discern more directly the relationship between children's goal-oriented linguistic behaviours and their interactive repercussions, as well as individual differences.

Children participate in social interactions regardless of their linguistic capability, and their speech significantly shapes interactions. Their participation creates opportunities to learn from contingent interactions and refine future behaviour choices. Investigating this process constitutes a meaningful challenge for any ecological approach to language acquisition that regards social interaction as both the site and goal of language use.

**Acknowledgement.** Julian Pine is a Professor in the International Centre for Language and Communicative Development (LuCiD) at The University of Liverpool. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

## References

- Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development*, 7(2), 143–163. <https://doi.org/10.1111/1467-9507.00059>
- Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. Cambridge University Press; Editions de la Maison des sciences de l'homme.

- Bannard, C., Klinger, J., & Tomasello, M. (2013). How selective are 3-year-olds in imitating novel linguistic material?. *Developmental Psychology*, 49(12), 2344–2356. <https://doi.org/10.1037/a0032062>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 21(3), 205–226.
- Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, 14(2), 211–227. <https://doi.org/10.1017/S0305000900012897>
- Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, 6(3), 380–420. [https://doi.org/10.1016/0010-0285\(74\)90018-8](https://doi.org/10.1016/0010-0285(74)90018-8)
- Bloom, L., Merkin, S., & Wootten, J. (1982). ‘Wh’-questions: Linguistic factors that contribute to the sequence of acquisition. *Child Development*, 53(4), 1084. <https://doi.org/10.2307/1129150>
- Burdelski, M. (2011). Language socialization and politeness routines. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (1st ed.). Wiley. <https://doi.org/10.1002/9781444342901>
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.9>
- Cairns, H. S., & Hsu, J. R. (1978). Who, why, when, and how: A development study. *Journal of Child Language*, 5(3), 477–488. <https://doi.org/10.1017/S0305000900002105>
- Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children’s social understanding within social interaction. *Behavioral and Brain Sciences*, 27(01). <https://doi.org/10.1017/S0140525X04000032>
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i. <https://doi.org/10.2307/1166214>
- Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language*, 43(6), 1310–1337. <https://doi.org/10.1017/S0305000915000689>
- Chang, F., Tatsumi, T., Hayakawa, H., Yoshizaki, M., & Oka, N. (2021). The role of parental input in the early acquisition of Japanese politeness distinctions. *Collabra: Psychology*, 7(1), 18989. <https://doi.org/10.1525/collabra.18989>
- Clancy, P. M. (1985). The acquisition of Japanese. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: The data* (1st ed.). Psychology Press. <https://doi.org/10.4324/9781315802541>
- Clancy, P. M. (1989). Form and function in the acquisition of Korean wh-questions. *Journal of Child Language*, 16(2), 323–347. <https://doi.org/10.1017/S0305000900010448>
- Clark, E. V. (2018). Conversation and language acquisition: A pragmatic approach. *Language Learning and Development*, 14(3), 170–185. <https://doi.org/10.1080/15475441.2017.1340843>
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13(2), 259–294. [https://doi.org/10.1207/s15516709cog1302\\_7](https://doi.org/10.1207/s15516709cog1302_7)
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Cook, H. M. (1990). The sentence-final particle *ne* as a tool for cooperation in Japanese conversation. In H. Hoji, P. M. Clancy, Stanford Linguistics Association, & Center for the Study of Language and Information (U.S.) (Eds.), *Japanese/Korean linguistics* (pp. 29–44). Southern California Japanese/Korean Linguistics Conference, Stanford, Calif. Published for the Stanford Linguistics Association by the Center for the Study of Language and Information.
- Dale, R., & Spivey, M. J. (2006). Unraveling the Dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56(3), 391–430. <https://doi.org/10.1111/j.1467-9922.2006.00372.x>
- de Haan, F. (2006). Typological approaches to modality. In W. Frawley, E. Eschenroede, S. Mills, & T. Nguyen (Eds.), *The Expression of modality* (pp. 27–70). Mouton de Gruyter. <https://doi.org/10.1515/9783110197570>
- Dore, J. (1974). A pragmatic description of early language development. *Journal of Psycholinguistic Research*, 3(4), 343–350. <https://doi.org/10.1007/BF01068169>
- Dore, J. (1977). ‘Oh Them Sheriff: A Pragmatic Analysis of Children’s Responses to Questions11The laboratory facility and staff, the data collection, and some of the data analyses for the research reported in

- this chapter were supported by Grant No. 5-20284 from The Grant Foundation to the Rockefeller University. This research was also supported in part by a grant from the Carnegie Corporation of New York to the Rockefeller University. In *Child Discourse* (pp. 139–163). Elsevier. <https://doi.org/10.1016/B978-0-12-241950-8.50014-9>
- Dunfield, K., Kuhlmeier, V. A., O'Connell, L., & Kelley, E. (2011). Examining the diversity of prosocial behavior: helping, sharing, and comforting in infancy. *Infancy*, **16**(3), 227–247. <https://doi.org/10.1111/j.1532-7078.2010.00041.x>
- Enfield, N. J. (2014). *Relationship thinking: Agency, enchrony, and human sociality*. Oxford University Press.
- Ervin-Tripp, S. (1978). Some features of early child–adult dialogues. *Language in Society*, **7**(3), 357–373. <https://doi.org/10.1017/S0047404500005777>
- Ervin-Tripp, S. (1979). Children's verbal turn-taking. In E. Ochs (Ed.), *Developmental pragmatics* (Nachdr., pp. 391–414). Academic Press.
- Ervin-Tripp, S., Guo, J., & Lampert, M. (1990). Politeness and persuasion in children's control acts. *Journal of Pragmatics*, **14**(2), 307–331. [https://doi.org/10.1016/0378-2166\(90\)90085-R](https://doi.org/10.1016/0378-2166(90)90085-R)
- Evans, N., Bergqvist, H., & San Roque, L. (2018). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition*, **10**(1), 110–140. <https://doi.org/10.1017/langcog.2017.21>
- Forrester, M. A. (2008). The emergence of self-Repair: A Case Study of One Child During the Early Preschool Years. *Research on Language & Social Interaction*, **41**(1), 99–128. <https://doi.org/10.1080/08351810701691206>
- Gallagher, T. M. (1981). Contingent query sequences within adult–child discourse. *Journal of Child Language*, **8**(1), 51–62. <https://doi.org/10.1017/S0305000900003007>
- Garfinkel, H. (1967). *Studies in ethnomethodology* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall. <https://www.taylorfrancis.com/books/9781003320609>
- Garvey, C., & Hogan, R. (1973). Social speech and social interaction. Egocentrism Revisited. *Child Development*, **44**(3), 562. <https://doi.org/10.2307/1128013>
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, **7**(7), 287–292. [https://doi.org/10.1016/S1364-6613\(03\)00128-1](https://doi.org/10.1016/S1364-6613(03)00128-1)
- Goffman, E. (1983). The interaction order: American Sociological Association, 1982 Presidential Address. *American Sociological Review*, **48**(1), 1. <https://doi.org/10.2307/2095141>
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, **113**(1), 52–58. <https://doi.org/10.1037/0735-7036.113.1.52>
- Goodwin, C., & Goodwin, M. H. (1987). Children's arguing. In S. U. Philips, S. Steele, & C. Tanz (Eds.), *Language, gender, and sex in comparative perspective* (pp. 200–248). Cambridge University Press.
- Heritage, J. (2013). *Garfinkel and Ethnomethodology* (1st ed.). Wiley.
- Kyrtzidis, A., & Ervin-Tripp, S. (1999). The development of discourse markers in peer interaction. *Journal of Pragmatics*, **31**(10), 1321–1338. [https://doi.org/10.1016/S0378-2166\(98\)00107-6](https://doi.org/10.1016/S0378-2166(98)00107-6)
- Lerner, G. H. (2006). On the 'semi-permeable' character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs (Ed.), *Interaction and grammar* (Digitally print. hardback version (pp. 238–276). Cambridge Univ. Press.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-olds point to provide Information for others. *Journal of Cognition and Development*, **7**(2), 173–187. [https://doi.org/10.1207/s15327647jcd0702\\_2](https://doi.org/10.1207/s15327647jcd0702_2)
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition* (0 ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315805672>
- Masuoka, T. (1991). *Modariti no Bunpō*. Kuroshio Shuppan.
- Miyata, S. (2004a). *Japanese – Miyata – Aki Corpus* [Dataset]. TalkBank.
- Miyata, S. (2004b). *Japanese – Miyata – Ryo Corpus* [Dataset]. TalkBank.
- Miyata, S. (2004c). *Japanese – Miyata – Tai Corpus* [Dataset]. TalkBank.
- Miyata, S., & Nisisawa, H. Y. (2009). *Japanese – MiiPro – Asato Corpus* [Dataset]. TalkBank.
- Miyata, S., & Nisisawa, H. Y. (2010). *Japanese – MiiPro – Tomito Corpus* [Dataset]. TalkBank.
- Mueller, E. (1972). The maintenance of verbal exchanges between young children. *Child Development*, **43**(3), 930. <https://doi.org/10.2307/1127643>
- Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology*, **107**(2), 174–186. <https://doi.org/10.1037/0735-7036.107.2.174>

- Nakamura, K. (2014). The acquisition of polite language by Japanese children. In K. E. Nelson, A. Aksu-Koç, C. E. Johnson, & A. Aksu-Koc (Eds.), *Children's Language: Volume 10: Developing Narrative and Discourse Competence* (0 ed.). Psychology Press. <https://doi.org/10.4324/9781410605153>
- Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, 42(3), 555–565. <https://doi.org/10.1037/0012-1649.42.3.555>
- Nihongo Kijutsu Bunpō Kenkyūkai (Ed.). (2003). *Gendai Nihongo bunpō. 4: Modariti / Nihongo Kijutsu Bunpō Kenkyūkai* (Hrsg.) (2. Auflage). Kuroshio shuppan.
- Nikolaus, M., Prévot, L., & Fourtassi, A. (2022). *Communicative feedback as a mechanism supporting the production of intelligible speech in early childhood. Proceedings of the annual meeting of the cognitive science society*, 44.
- Ninio, A., & Snow, C. E. (1988). Language acquisition through language use: The functional sources of children's early utterances. Categories and processes in language acquisition, 11–30. In Y. Levy (Ed.), *Categories and processes in language acquisition* (1. [print.], pp. 11–30). Erlbaum.
- Nisisawa, H. Y., & Miyata, S. (2009). *Japanese – MiiPro – Nanami Corpus* [Dataset]. TalkBank.
- Nisisawa, H. Y., & Miyata, S. (2010). *Japanese – MiiPro – Nanami Corpus* [Dataset]. TalkBank.
- Ochs, E., Schieffelin, B. B., & Platt, M. (1979). Propositions across utterances and speakers. In E. Ochs & B. B. Schieffelin, *Developmental pragmatics* (pp. 251–268). Academic Press.
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robinson, J. D. (2006). Managing trouble responsibility and relationships during conversational repair. *Communication Monographs*, 73(2), 137–161. <https://doi.org/10.1080/03637750600581206>
- Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 54–69). Multilingual Matters.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.
- Sidnell, J. (2016). A conversation analytic approach to research in early childhood. In A. Farrell, S. L. Kagan, & E. K. M. Tisdall (Eds.), *The SAGE handbook of early childhood research* (pp. 255–276). SAGE.
- Smiley, P. A. (2001). Intention understanding and partner-sensitive behaviors in young children's peer interactions. *Social Development*, 10(3), 330–354. <https://doi.org/10.1111/1467-9507.00169>
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4(1), 1–22. <https://doi.org/10.1017/S0305000900000453>
- Snow, C. E. (1986). Conversations with children. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (2nd ed., pp. 69–89). Cambridge University Press.
- Snow, C. E., Pan, B. A., Imbens-Bailey, A., & Herman, J. (1996). Learning how to say what one means: A longitudinal study of children's speech act use\*. *Social Development*, 5(1), 56–84. <https://doi.org/10.1111/j.1467-9507.1996.tb00072.x>
- Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental Psychology*, 29(6), 1008–1023. <https://doi.org/10.1037/0012-1649.29.6.1008>
- Stivers, T., & Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society*, 35(03). <https://doi.org/10.1017/S0047404506060179>
- Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics*, 124, 14–30. <https://doi.org/10.1016/j.pragma.2017.11.013>
- Topping, K., Dekhinet, R., & Zeedyk, S. (2013). Parent–infant interaction and children's language development. *Educational Psychology*, 33(4), 391–426. <https://doi.org/10.1080/01443410.2012.744159>
- Uno, M. (with Ortega, L.). (2017). *Developing question constructions in Japanese as a first language: The roles of type of referent and parental input*.
- Uzgiris, I. C. (1981). Two Functions of Imitation During Infancy. *International Journal of Behavioral Development*, 4(1), 1–12. <https://doi.org/10.1177/016502548100400101>
- Van Hekken, S. M. J., & Roelofsen, W. (1982). More questions than answers: A study of question–answer sequences in a naturalistic setting. *Journal of Child Language*, 9(2), 445–460. <https://doi.org/10.1017/S0305000900004803>
- Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, 25(7), 1314–1324. <https://doi.org/10.1177/0956797614531023>

- Warneken, F., & Tomasello, M.** (2006). Altruistic helping in human infants and young chimpanzees. *Science*, **311**(5765), 1301–1303. <https://doi.org/10.1126/science.1121448>
- Warneken, F., & Tomasello, M.** (2007). Helping and cooperation at 14 months of age. *Infancy*, **11**(3), 271–294. <https://doi.org/10.1111/j.1532-7078.2007.tb00227.x>
- Wellman, H. M., & Lempers, J. D.** (1977). The naturalistic communicative abilities of two-year-olds. *Child Development*, **48**(3), 1052. <https://doi.org/10.2307/1128359>

---

**Cite this article:** Tatsumi, T., & Pine, J. (2025). Shifting toward progressive and balanced interaction: A longitudinal corpus study of children's responses to Who-questions in Japanese. *Journal of Child Language* 1–27, <https://doi.org/10.1017/S0305000925000029>