Journal of Child Language (2023), **50,** 1226–1244
doi:10.1017/S0305000922000344

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Learning conversational dependency: Children's response using *un* in Japanese

Tomoko TATSUMI[1,*] and Giovanni SALA[2]

[1]Kobe University, Kobe, Japan
[2]National Center for Geriatrics and Gerontology, Obu, Japan
*Corresponding author. Tomoko Tatsumi, 1-2-1 Tsurukabuto, Nada-ku, Kobe, Hyogo, JAPAN.
E-mail: tt@port.kobe-u.ac.jp

(Received 14 July 2021; revised 12 May 2022; accepted 28 May 2022)

Abstract

This study investigates how Japanese-speaking children learn interactional dependencies in conversations that determine the use of *un*, a token typically used as a positive response for yes-no questions, backchannel, and acknowledgement. We hypothesise that children learn to produce *un* appropriately by recognising different types of cues occurring in the immediately preceding turns. We built a set of generalised linear models on the longitudinal conversation data from seven children aged 1 to 5 years and their caregivers. Our models revealed that children not only increased their *un* production, but also learned to attend relevant cues in the preceding turns to understand when to respond by producing *un.* Children increasingly produced *un* when their interlocutors asked a yes-no question or signalled the continuation of their own speech. These results illustrate how children learn the probabilistic dependency between adjacent turns, and become able to participate in conversational interactions.

Keywords: child language; adjacency; conversation; statistical modelling; Japanese; corpus

Introduction

Young children start using language in conversational interactions. This means that they need to be able to manage interactions by, for instance, demonstrating their participation, taking turns, and reacting appropriately. One of these challenges is responding by producing specific tokens such as yeah in English. Crucially, children need to produce these tokens at interactionally appropriate moments during a conversation. This study investigates how children learn to use such tokens, specifically the Japanese token un, by focusing on the interactional dependencies that condition its usage.

The Japanese token *un* can be translated as *yes*, *yeah,* or *aha* in English, as it primarily serves as a positive response marker, backchannel, or acknowledgement. It is one of the most frequent linguistic forms used in Japanese daily conversations by both adults and children. Mastery over *un* is essential in Japanese conversations, especially because the communication style implies that listeners assume an active role (e.g., Clancy, 1987; Kita

© The Author(s), 2022. Published by Cambridge University Press.

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

& Ide, 2007). Learning to use this form is thus an important part of both the language acquisition and socialization process.

Responding in Conversations

Responding is essential for managing turn-taking and building up a conversation. It refers to the "second pair part" of an adjacency pair, which comprises two adjacent turns by different speakers (Schegloff & Sacks, 1973). In the second pair part, the speaker replies (e.g., answers, rejects, accepts, agrees, and acknowledges) to their interlocutor's "first pair part", including questions, requests, announcements, and so on. Responses to questions are typically answers. Yes-no questions tend to be followed by responses using tokens such as *yes, no*, or *yeah.* These responses signal agreement, disagreement, or acceptance of the propositional content in the preceding discourse. They contribute to the grounding process in which conversational participants continually seek and provide evidence that they understand each other in communication (Clark & Brennan, 1991). Concurrently, speakers respond on a non-propositional level, signalling their understanding or acknowledgement of their interlocutors' linguistic production or speech act. One of such tokens is *uh-huh* in English, which Schegloff (1982) explained as a form that lets other participants in the conversation know that the speaker is paying attention to or understanding the ongoing discourse. This token also serves for signalling who takes the turn at a given point. Tokens such as *uh-huh* are often referred to as backchannels, reactive tokens (Clancy, Thompson, Suzuki, & Tao, 1996), continuers (Schegloff, 1982), and interjections (Stivers, Sidnell, & Bergen, 2018), depending on the scope and perspective of the researcher. The term backchannel assumes that there are two channels in conversation that operate simultaneously – namely, the "main" channel through which the speaker sends messages and the "back" channel over which the listener provides useful information without claiming the floor; that is, without switching the speaker's and listener's roles (White, 1989; Yngve, 1970). The listener typically uses backchannels to signal that they have heard or understood the speaker or to actively support the speaker's continuation.

How the Japanese *un* is Used in Conversations

Un is a typical token for responding in Japanese and is one of the most frequent linguistic forms in everyday conversations. Angles, Nagatomi, and Nakayama (2000) listed *un*'s functions as a positive response to yes-no questions, backchannel (to support the interlocutor's continuation of their speech), acknowledgement of having heard before answering, self-confirmation (by using *un* at the end of an utterance after a speaker expresses their opinion), and as a response to suggestions and commands or strong requests. Sadanobu (2002) proposed a different account that *un* in conversations can signal acknowledgement at different levels: agreement with speaker's argument (e.g., the positive response to yes-no questions), comprehension of speaker's argument (e.g., backchannels), and recognition of speaker's speech act (e.g., acknowledgement). The token's position within a turn is predictive of these different functions. For example, *un* for backchanneling tends to be turn-initial or turn-final, and *un* for positive responses tends to be turn-initial (Angles et al., 2000; Togashi, 2002). Turn-final *un*s may serve to either emphasise or change turns (Tanaka, 2010). In summary, although classifications differ across researchers, and despite the apparent multifunctionality of *un*, these previous studies have shown that speakers use *un* to respond to its preceding utterance (either their

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

Example 1. Use of *un* after the interlocutor's unfinished sentence (MiiPro corpus, Nanami: 40504.xml, context: toy play)[1]

1	Child	*okyakusan wa koko ni in no* guest TOP here LOC be-NONPAST MODAL "the guest is here"
2	Child	*okyakusan wa koko de* guest TOP here LOC "the guest is here…"
3	Caregiver	***un***
4	Child	*tsukutten no okaasantachi* make-ASP-NONPAST MODAL mothers "mothers are making"
5	Caregiver	***un***

interlocutors' or their own) and to signal their understanding, agreement, and/or acknowledgement.

An essential characteristic of *un* is its listener-centred usage. This token often takes place in the speech of an interviewer who listens to the interviewee's discourse and engages with it actively (Tanaka, 2010). That is, the listener's use of *un* makes the conversation flow smoothly by promoting the speaker's discourse. The listener's production of *un* is one of the expected behaviours when the speakers' discourse appears to continue (Kushida, 2009). Example 1 illustrates a typical backchanneling usage of *un* when the speaker provides a short pause without completing a sentence.

In this example, the caregiver produces *un* when the child utters a sentence without completing its proposition (the turn-final *de* in line 2 signals the lack of completion and/or further continuation). After the caregiver's *un*, the child continues her discourse. This exemplifies the backchanneling function of *un,* which is to indicate that the listener is paying attention to the speaker's discourse, to claim that the listener has heard and understood the preceding discourse without problems, and to present a positive stance on the continuation of the discourse (Nishizaka, 2008). Importantly, Sadanobu (2002) points out that *un* itself has no semantic meaning, and that producing *un* is, in this regard, purely an action. This non-semantic and actional nature of *un* renders it an ideal target for research on how children learn to interact verbally with people in a conversation.

Un has often been studied under the Japanese concept of *aizuchi*, which has attracted a considerable attention in the research on the social and interactive aspects of Japanese conversation. *Aizuchi* is a non-technical term that refers to the behaviour of reacting to the interlocutor either using short lexical tokens (e.g., *un*, *hai* "yes", and *hontoo* "really"), non-lexical tokens (e.g., *nn* and *hun* "hmm"), or head-nods (Iwasaki, 1997; Miyata & Nisisawa, 2007). The Japanese communication style is relatively listener-centric, in that listeners assume an active role in conversations (e.g., Clancy, 1987; Kita & Ide, 2007). This implies that listeners are expected to produce *aizuchi* frequently to signal their state of understanding and promotive attitude towards the speakers' discourse. Furthermore,

[1]Grammatical glosses: ASP: progressive/stative aspect, COND: conditional, CONN: connective, GEN: genitive, HORT: hortative, LOC: locative, MODAL: modal particle, NEG:negation, NOM: nominative, NONPAST: nonpast tense, ONOMA: onomatopoeia, PAST: past tense, QUOT: quotative, TOP: topic

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

researchers argue that this behaviour strengthens the emotional and phatic aspect of communication. According to Iwasaki (1997), *aizuchi* contributes to a culturally encouraged pattern of behaviour through which conversation participants signal their interdependence between themselves. He argues that this culture-specific concept of interdependency escapes the notion of politeness proposed by Brown and Levinson (1987). This is considered one of the reasons why Japanese speakers use *aizuchi* more often than English speakers (Kawamori, Kawabata, & Shimazu, 1998; Kita & Ide, 2007). Likewise, Clancy et al. (1996) showed that backchannels are more frequent in Japanese (target tokens including *un, aa*, and *ee*) than in English (e.g., *hmm*, *huh* and *oh*; 29.9% versus 16.9% of all turns). Children in a Japanese-speaking environment need to adapt themselves to this characteristic pattern of conversation. The importance of the listener's attitude is sometimes emphasised in everyday conversations as well. For example, Clancy (1987) reported that Japanese caregivers consistently demanded responses from young children who had not reacted to the questions addressed to them. This norm of being an attentive and responsive listener highlights the significance of *un* in Japanese language socialization.

Previous Studies on the Development of Children's Response

Although tokens such as *un* are short and simple forms that proficient Japanese speakers use frequently and almost unconsciously, a child's acquisition of these tokens may be difficult because it requires a certain grasp of the nature of conversation and the participants' roles at a given moment in an ongoing interaction.

Responding requires an understanding of the basic structure of conversational interactions, including turn-taking (Casillas, Bobb, & Clark, 2016). Even before their first linguistic production, children are sensitive to temporal contingency in communicative interactions, and can respond to their caregivers as well as appeal to common ground with them by non-linguistic means, such as eye-gaze, pointing, and vocalizations (for a review, see Stephens & Matthews, 2014). Children at 2;6 can also attend to prosodic and lexico-syntactic cues to predict turn structures, as Lammertink, Casillas, Benders, Post, and Fikkert (2015) revealed in their eye-tracking study. However, substantial learning is nevertheless required for responding by linguistic means. Children need to distinguish different kinds of communicative acts, such as questions, imperatives, and statements, and learn what types of responses are expected or allowed for each of them. Learning these distinctions is supposed to take time and can extend beyond the age of 3, with some complex communicative acts, such as indirect requests or ironies, requiring more time to learn (Bucciarelli, Colle, & Bara, 2003; Rakoczy & Tomasello, 2009).

Generally, children fail to respond to questions more often than adults. Casillas et al. (2016) observed that children between 1;8-3;5 took more time before answering questions when compared to adults; however, they could provide quick and simple yes-no answers (including *yeah* and other minimal phrases of assent or denial) to questions from the earliest observed stages. Similarly, Stivers et al. (2018) found that four- and five-year-old English-speaking children failed to respond 33% of the times, while the rate was 6% for adults. Nonetheless, children's broad response patterns are quite similar to those of adults: they respond to most questions by confirming answers (including *yes*, *no*, *yeah*, *uh huh*, and head nods). These studies suggest that responding to yes-no questions by producing simple tokens such as *un* is a strategy available even for very young children. By contrast, it may take more time to acquire backchannel responses. Hess and Johnston (1988) tested English-speaking children aged 7 to 11 years, and found that the frequency of

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

backchannel responses to their interlocutors' instructions increased with age. They discussed that children took a relatively long time to learn a variety of discourse signals and become capable, as listeners, of providing collaborative feedback to the speakers. The ability of backchannel responses may correlate with the listener's skill in general, including asking speakers appropriate questions for successful communication (e.g., Cosgrove & Patterson, 1977). The literature on Japanese *aizuchi* seems to support the relatively late development of backchannels. For instance, Miyata and Nisisawa (2007) analysed conversational data of a boy and his caregiver, and showed that the child's *aizuchi* was much less frequent than his caregiver's during the 1;5 to 3;1 period, despite the caregiver's frequent elicitation of *aizuchi* using final modal particles and verb forms.

The Question on the Learning Process

Although these studies provide valuable information about the development of children's responding behaviours, few have examined the mechanisms underlying children's learning of response tokens. First, any putative mechanism should include the statistical learning of the probability of usage patterns in the input language, as its importance has been widely confirmed in the language acquisition literature (e.g., Ambridge, Kidd, Rowland, & Theakston, 2015; Ellis, 2002). This consideration leads to the prediction that children match their probability of *un* production up to the target probability in the language they experience. Second, children also need to learn when to produce *un*. Our study thus focuses on the moment-to-moment interactional contexts within a conversation, which primarily condition the use of *un*. Although speakers process numerous elements in the prior interaction during a conversation, the immediately preceding utterance from the interlocutor is considered to weigh most for them in deciding what to say next.

The significance of immediately preceding utterance has been studied intensively in conversation analytic studies, which have demonstrated that adjacency pairs are predictable patterns in conversations. Adjacency pairs comprise two consecutive turns from different speakers that are related in the way that the first pair part implies the next pair part, such as question–answer and greeting–greeting (Schegloff & Sacks, 1973). For instance, when someone asks a question, the interlocutor is expected to answer the question in the next turn. Different descriptive studies have confirmed the importance of adjacent pairs for explaining the use of response tokens in both adult conversations (e.g., Tanaka, 2010) and child-caregiver conversations (Montes, 1999).

Based on these previous studies, we assume that children learn the probabilistic dependency between adjacent conversational turns to produce *un* appropriately. In particular, we focus on the formal cues that signal two kinds of interactions: those in which a speaker is asked a yes-no question by their interlocutor, and those in which the interlocutor signals the intention to continue speaking. According to Kushida (2009), Japanese-speaking adults react to certain cues that signal their interlocutor's continuation of their own discourse. These cues typically include: (1) a prolonged pronunciation of the syllable-final sound (often with emphasised contours); (2) final modal particles, such as *ne* and *sa*; (3) conjunction particles, such as the connective conjugational form of verbs, adjectives and auxiliary verbs; and 4) conjunctions, such as *sorede* "then" and *demo* "but" at the end of a turn. Other relevant studies include Kita and Ide (2007) and Miyata and Nisisawa (2007), which examined *aizuchi* in general. Miyata and Nisisawa (2007) examined the final modal particles (*ne* and *sa*) and other particles, including case-markers, a focus marker (*mo*), a topic marker (*wa*), and conjunctive particles, as well

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

as connective and conditional verb-endings (conditional *-tara* and *-ba*, consecutive *nagara*). Kita and Ide (2007) mention that final modal particles *ne* and *yo* are closely related to the use of *aizuchi*. Although these studies group different response tokens under the category of *aizuchi*, we start by studying how speakers learn a specific token. In fact, it is impossible to know a priori whether speakers process *aizuchi* as a category. In addition, to understand whether the above-mentioned forms actually serve as predictive cues for the token *un*, it is necessary to adopt a quantitative approach.

Aim of this study

This study aims to understand how children learn to use *un* in everyday conversations. We hypothesise that children's acquisition of this token is a process in which they learn the predictive interactional cues in the immediately preceding turn by their interlocutor to produce *un* in the following turn. Among potentially numerous formal cues that would condition *un* usage, we first focus on yes-no questions and analyse whether children produce *un* when their interlocutor asks this type of questions. Next, we explore back-channeling or acknowledgement usages. We examine whether children produce *un* when their interlocutor signals continuation of their own speech. Our method involves identifying such interactions by coding potential formal cues in immediately preceding turns, and building statistical models of children's and caregivers' production of *un* following these cues. This allows us to test whether children's production of *un* increases following these cues in the preceding turns. To the best of our knowledge, this study is the first quantitative modelling of how children learn to use a linguistic token at interactionally appropriate moments during a conversation.

Method

Data

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

After downloading the utterance-unit CSV files from the LuCiD Toolkit version of the CHILDES corpora (Chang, 2017), all data were reorganised into a turn-unit dataset using R (R Core Team, 2020). There were 313,214 turns in all in the final dataset, of which 141,758 were derived from the seven target children. The remainder were mostly from their mothers (144007 turns). We used the data from these speakers alone. The age range of the children was from 1;10 to 6;1. Unclear utterances in the original corpora were removed from the analysis. Although the number of speakers was limited because of the availability of corpus data of Japanese child-caregiver conversations, the number of conversational turns is sufficiently large for testing our hypothesis using statistical models.

Coding and Variables

Our variables included the production of *un* as a dependent variable, and children's age in month, speaker type (child or caregiver) and the presence or absence of different formal

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

cues in the immediately preceding turn as independent variables. These variables were coded for each conversational turn (i.e., change of speakers) in the dataset.

We only coded the *un* in the turn-initial position to focus on the typical usage in which the speaker uttered *un* right after the interlocutor's preceding turn (of all 35738 occurrences of *un*, 2294 cases of not turn-initial usage were excluded). We then coded different forms in the immediately preceding turn as potential predictive cues for interlocutors' yes-no questions and continuation. We defined yes-no questions as those turns that ended with the question coding "?" but did not have *wh*-words (e.g., *nani* "what", *dooshite* "why") in the original corpora. The "?" in the original corpora is used for coding yes-no questions, which are overtly marked by the final particles for question and/or by intonations considering contextual information (S. Miyata, personal communication, 16-28 December, 2020). We also coded the final modal particle for question *ka* as well as *wh*-words for additional analyses.

In addition to the yes-no questions, we coded whether the preceding utterances had any potential cues that are considered to signal speakers' further continuation of their speech. These cues include the final modal particle *ne* and predicates in a connective form. Among the many possible cues mentioned earlier, these two types of cues were selected based on the token counts for a reliable quantitative modelling. The final particle *ne* is an utterance-final particle which is characteristically used to establish a common ground between the speaker and the addressee. Cook (1990, p. 31) shows an example of this particle *oimohori shitai ne* "(We) want to go digging up potatoes, don't we?", in which she used the subject "we" and a tag question in the English translation to convey the particle's modal meaning. It can achieve various goals, such as getting another's attention, introducing a new topic, keeping the floor (continuing talking), teaching children, and mitigating face-threatening acts (Cook, 1990). The function of keeping the floor and creating common ground between conversation participants is particularly important for our analysis with continuation cues. The connectives are *-te* or *-de* suffixes for a non-finite inflection of verbs, adjectives, and auxiliary verbs. They typically mark a cumulative and non-contrastive relationship with the next clause, as in *tabete nonda* (eat-CONN drink-PAST) "(someone) ate and drank" but is also used as a turn-final element implying a continuation of further speech, as in *tabete …* (eat-CONN) "(someone) eat (and…)". All these formal cues occurring at the end of a turn were automatically coded by using the information on the morphological tier of the original corpora.

Analysis

All quantitative analyses were performed with Generalized Linear Modelling (GLM), using the GAMLSS R package (Rigby & Stasinopoulos, 2005). GLM allows researchers to implement distributions other than normal distribution. The dependent variable is the production of *un*, which was coded as 0 (no occurrence) or 1 (occurrence). As the dependent variable was binary, the binomial distribution with a logit link function was employed in all models. Independent variables included the presence of children's age, speaker types (children or caregiver), and potential predictive cues in the immediately preceding turn from the interlocutor. The child-caregiver pair was added into the model as a random intercept. The strength of this study lies in modelling a speaker's behaviour as they produce or do not produce *un* at a certain moment in a conversational interaction. They likely produce *un* in their next turn when they recognise a relevant cue in their interlocutor's most recent utterance. Our models will explore the changing relationship

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

with which children associate the different cues and their production of *un* in adjacent turns.

Results

Do Children Learn to Produce un as Frequently as Caregivers?

To test whether children learn to use *un* as frequently as caregivers, we built a GLM of the *un* production using the independent variable of children's age in months and speaker types, along with the two-way interaction.

First, our model showed that the children produced *un* less than the caregivers did (mean = 0.076, SD = 0.265 and mean = 0.157, SD = 0.364, respectively). The difference was significant (estimate = −2.134, SE = 0.044, t = −48.00, $p < .001$), in line with the results of previous studies showing children's overall tendency to respond less than adults. As shown in Figure 1, children and caregivers exhibited different trends (estimate = 0.030, SE = 0.001, t = 32.61, $p < .001$). Children increasingly produced *un* as they grew up (estimate = 0.030, SE = 0.001, t = 32.61, $p < .001$; from a separate model for children's data only) while caregivers' production rate decreased across the observed period (estimate = −0.018, SE = 0.001, t = −24.77, $p < .001$; from a separate model for caregivers' data only). These results reveal that children learn to use *un,* perhaps in an input-driven way that is similar to learning other words. Children assimilate the probability of their production of this form to that of their caregivers, almost reaching the same level when children turn five. The change in caregivers' production is another interesting finding in need of further investigation. It may imply that their use of the simple token *un* decreases

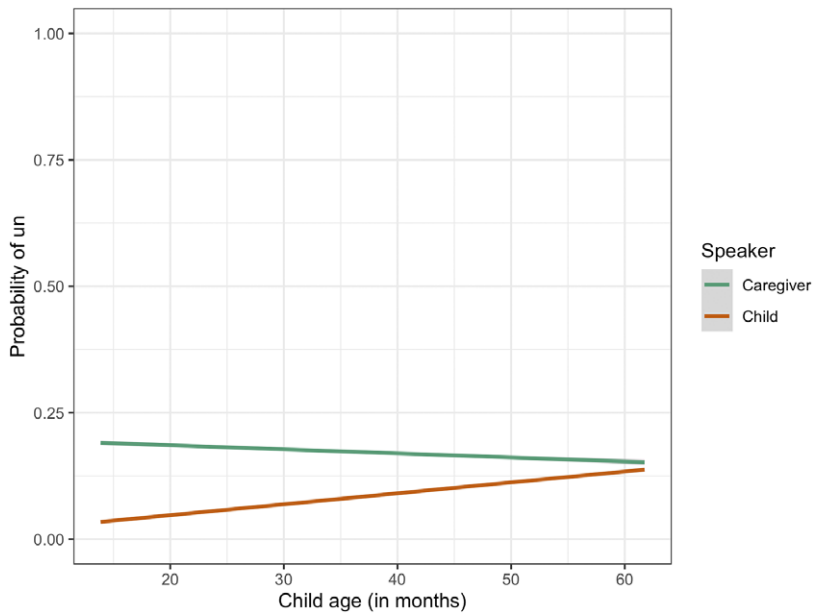


Figure 1. The probability of *un* production by speaker type and child's age.

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

Example 2. Child's use of *un* after the caregiver's question (child's age: 2;2.14, Aki corpus, 20214.xml, context: toy play)

1	Child	*senro* railway "Railway"
2	Caregiver	*senro tsukuroo ka ?* railway make-HORT MODAL "Shall we make a railway track?"
3	Child	*ka* [unintelligible]
4	Caregiver	*senro tsukuroo ka ?* railway make-HORT MODAL "Shall we make a railway?"
5	Child	*duhka* [unintelligible]
6	Caregiver	*tsukuru ?* make-NONPAST "Are we making it?"
7	Child	***un***

relative to other expressions when conversational interactions become more complex and diverse in accordance with children's growth.

Do Children Learn to Produce un at Interactionally Appropriate Moments?

Producing un after interlocutors' yes-no questions

Our data show many instances where speakers used *un* after yes-no questions as in the following example. In Example 2, a child was asked by his caregiver whether he wanted to make a railway track. The caregiver repeated the question as the child did not respond clearly; after the third question, the child uttered *un*. These yes-no questions are marked by intonation, and/or by the final modal particle *ka*. The child initially seemed to respond by repeating a part of the caregiver's question (*ka* and *duhka* are probably from *tsukuroo ka*); however, as the caregiver kept repeating the question, the child changed his response strategy. This sequence seems to exemplify the developmental change wherein children gradually become capable of choosing appropriate linguistic forms (such as *un*).

To understand whether children learn to produce *un* after yes-no questions, we built a model of *un* production by children's age, speaker type and the dummy-coding of whether the immediately preceding turn ended with a yes-no question. The model showed that both caregivers and children produced *un* more after yes-no questions than after other types of utterances (estimate = 0.482, se = 0.055, t = 8.832, $p < .001$), as illustrated in Figure 2. Additionally, children learn to use *un* more sensitively to this interactional context as they grow up (estimate = 0.021, se = 0.002, t = 10.45, $p < .001$; from a separate model for children's data only). This supports our hypothesis that children learn the predictive interactional cues to produce *un* themselves. They are sensitive to yes-no questions in the immediately preceding utterances by their interlocutors, and utter *un* on this interactional condition.

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

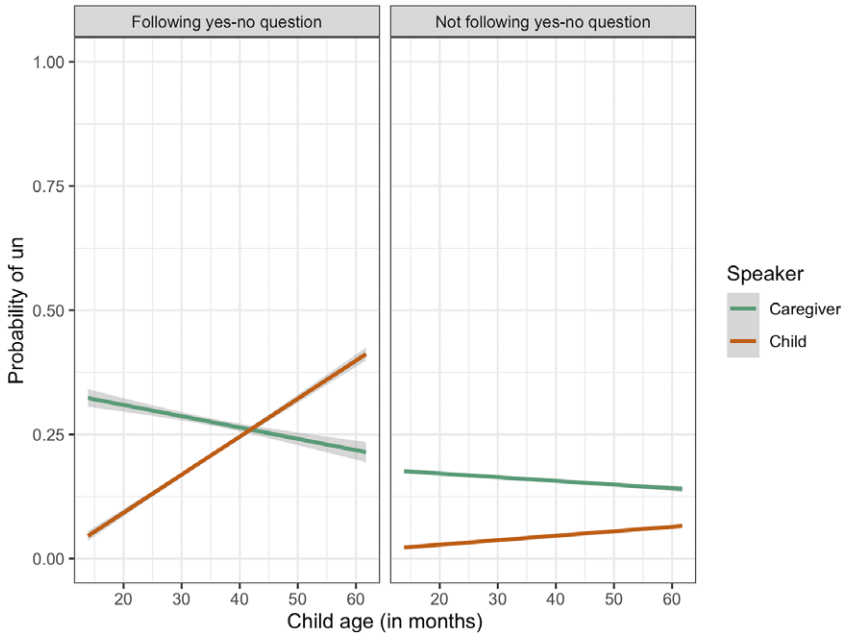


Figure 2. The probability of *un* production by speaker type and child age following or not following the interlocutor's yes-no question.

At the same time, caregivers also changed their language use across the observed period. For instance, they reduced their production of *un* throughout the period (estimate = –0.008, se = 0.002. t = –3.648, p < .001; from a separate model for caregivers' data only). The effect of children's age in both the children and caregivers' *un* production demonstrated not only children's learning, but also the dynamic changes in the way a child and caregiver interacted with each other in their conversations. One possible explanation is that caregivers diversify their response types as children's language develops. The change may also reflect the changes in activities or interactional contexts as children develop.

In addition, we ran an analysis for other relevant cues, the question-marking final modal particle *ka* as well as *wh*-words (that would signal *wh*-questions instead of yes-no questions), to understand how children respond to these individual formal cues that are useful for detecting yes-no questions. We found that the final particle *ka* predicted *un* production in both children and caregivers (estimate = 1.103, SE = 0.175, t = 6.322, p < .001 for both speakers; estimate = 1.017, SE = 0.199, t = 5.102, p < .001 for children) as Figure 3 shows; however, children do not significantly increase their *un* production with regard to this particle (estimate = –0.005, SE = 0.005, t = –1.035, p = .301). Figure 4 illustrates that children learn to respond to turns with *wh*-words distinctly. *Wh*-words negatively predict *un* production in both caregivers and children (estimate = –1.350, SE = 0.155, t = –8.686, p < .001 for both speakers; estimate = –0.505, SE = 0.187, t = –2.705, p = .007 for children). Children's sensitivity to *wh*-words develops with age (estimate = –0.014, SE = 0.005, t = –2.855, p = .004). These results suggest that children learn not to respond to *wh*-questions by using *un*.

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

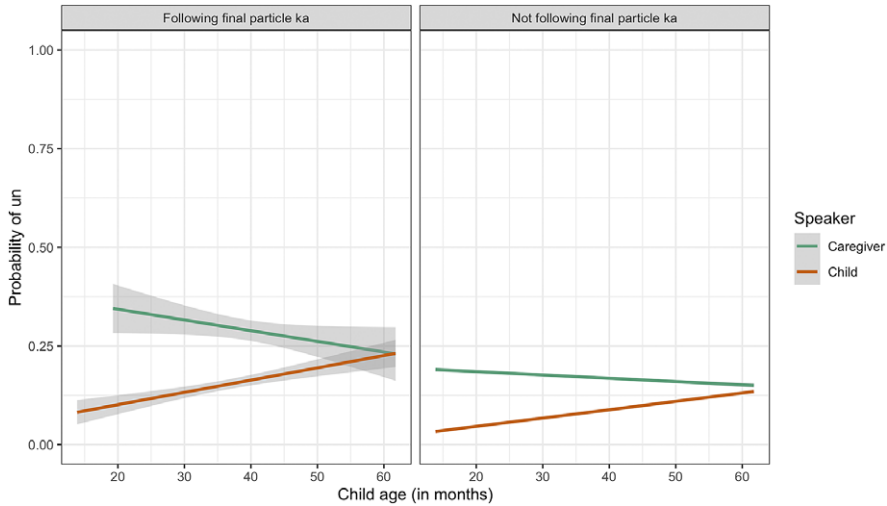


Figure 3. The probability of *un* production following or not following the interlocutor's final modal particle *ka* (question marker) by speaker type and child age.

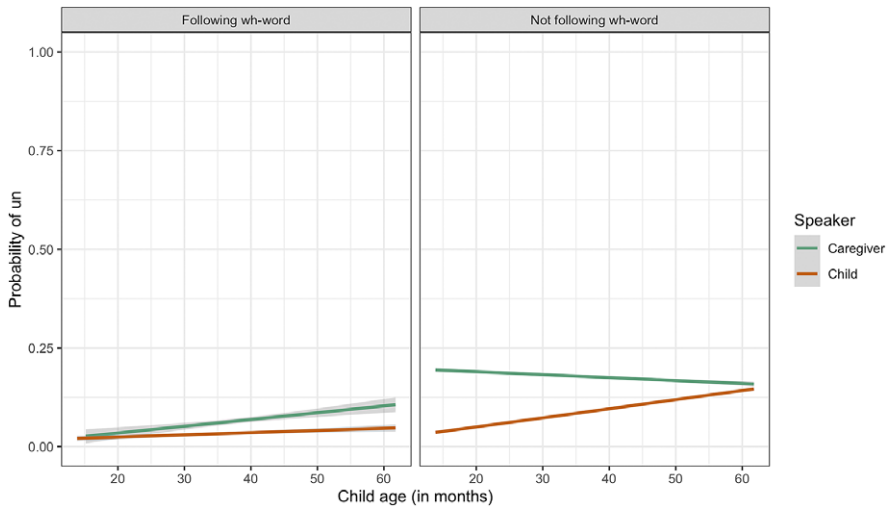


Figure 4. The probability of *un* following or not following a *wh*-words by speaker type and child age.

To sum up, children learn to produce *un* after their interlocutors' yes-no questions. They also show sensitivity to the question-marking final particle *ka* as a cue to produce *un*, and to *wh*-words as a cue to not produce *un*. At the same time, most of the yes-no questions in our data lacked overt question markers such as *ka* (42382 out of 47625). This implies that both caregivers and children are sensitive to intonational cues for questions;

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

this is plausible, given children's sensitivity to prosodic cues (e.g., Lammertink et al., 2015).

Producing un *after the interlocutors' signals for continuation*

This section explores whether children learn to use *un* in another kind of interactional situation where a speaker recognises that their interlocutor would continue talking. Researchers would label the use of *un* in these situations as backchanneling (which actively supports the interlocutors' continuation) or acknowledging (which only acknowledges the interlocutors' utterance or speech act) usage. The preceding model analysis showed that children learn to use *un* after caregivers' yes-no questions, supporting our hypothesis that children can use *un* as a response for yes-no questions. At the same time, children's use of *un* after utterances that are not yes-no questions grows only slightly and does not reach caregivers' rate (see the right plot of Figure 2), seemingly implying that children do not yet use *un* in an adult-like manner, including backchanneling or acknowledgement usages.

Yet, our data have various examples in which children use *un* in a way that could be categorised as backchanneling. In Example 3, the child and caregiver talked about train stations. The caregiver started an utterance with a demonstrative pronoun plus a topic marker (line 2), and did not complete it with a noun argument, but instead used *ne.* The child then responded with *un*, after which the caregiver resumed and finished her sentence. Japanese speakers often talk in a "piece-by-piece" manner, similar to this example, by producing a short and incomplete turn, inviting a backchannel or acknowledgement from the interlocutor before continuing (e.g., Iwasaki, 1997). The final particle *ne*, whose functions include keeping the floor and establishing a common ground (Cook, 1990), is often used for such an interaction. The child in Example 3 probably recognised the particle *ne*, predicted that her interlocutor would keep the floor, and produced *un* to support the interlocutor's continuation.

Note that this modal particle *ne* is often used at the end of a complete sentence as well and may not necessarily signal continuation. In Example 4, the caregiver talked about hippos in a complete sentence with the final particle *ne*, and the child responded with *un.* The final particle in this case does not signal the caregiver's continuation. She did not continue her speech further, but only said *ne,* which closes the sequence on this particular topic by confirming that they achieved a common ground.

Example 3. Child's use of *un* after a final particle *ne* (child's age: 2;8.11, ArikaM corpus 20811.xml, context: toy play)

1	Child	*dok* *kore* *wa* *doko ?* (speech error) this TOP where "Where is this?"
2	Caregiver	*Sore* *wa* *ne* that TOP MODAL "That is …"
3	Child	***un***
4	Child	*sore* *wa* *Niigata* that TOP Niigata "That is Niigata."

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

Example 4. Child's use of *un* after a final particle *ne* (child's age: 2;11.00, Nanami corpus, 21100.xml, context: book reading)

1	Caregiver	*kabasan mo omizu n(o) naka de kurasu no ne* hippo also water GEN inside LOC live-NONPAST MODAL "Hippos also live in the water."
2	Child	***un***
3	Caregiver	*ne:* MODAL "Right"

Example 5. Child's use of *un* after a connective verb (child's age: 2;7.19, Aki corpus 20719.xml, context: toy play)

1	Caregiver	*tamanegi ne* onion MODAL "Onions"
2	Child	*soo* right "Right"
3	Caregiver	*ja onabe ni irete …* then pot LOC put.in-CONN "Then we put them in the pot and …"
4	Child	***un***
5	Caregiver	*tamanegi wa juujuujuu tte yaranaito ne* onion TOP ONOMA QUOT do-NEG-NONPAST-COND MODAL "We've got to fry the onions well."

Example 5 is an instance of the other continuation cue, the connectives. The caregiver talked about how to cook onions, explaining the procedure using an incomplete sentence that included a verb with a connective ending (*ire-te* put.in-CONN) in line 3. The child produced *un* in the next turn, after which the caregiver continued her explanation.

Although there are many potential cues that could signal the speaker's continuation, we focus on the two cues mentioned above: the final modal particle *ne* and connective predicates, whose counts are sufficiently large for a quantitative modelling[2]. By building a GLM, we tested our hypothesis that children learn to produce *un* following these continuation cues from their interlocutors. The model predicted the production of *un* with the independent variables of child's age, speaker types, and the presence or absence of the continuation cues in the interlocutor's immediately preceding turn, as well as all two-way interactions.

The model revealed that the final particle *ne* is a strong and positive predictor of *un* in both caregivers and children (estimate = 1.151, se = 0.096, t = 11.975, p < .001) as Figure 5 shows. Children distinguish this cue to produce *un* (estimate = 0.774, see = 0.115,

[2]We also explored other cues (final particle *yo* and topic marker *wa*) that were frequent enough for model analysis; however neither of these resulted in a significant positive predictor (p = .121 for topic marker *wa*, p = 0.022 for final particle *yo*, both with a negative association with *un* production).

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

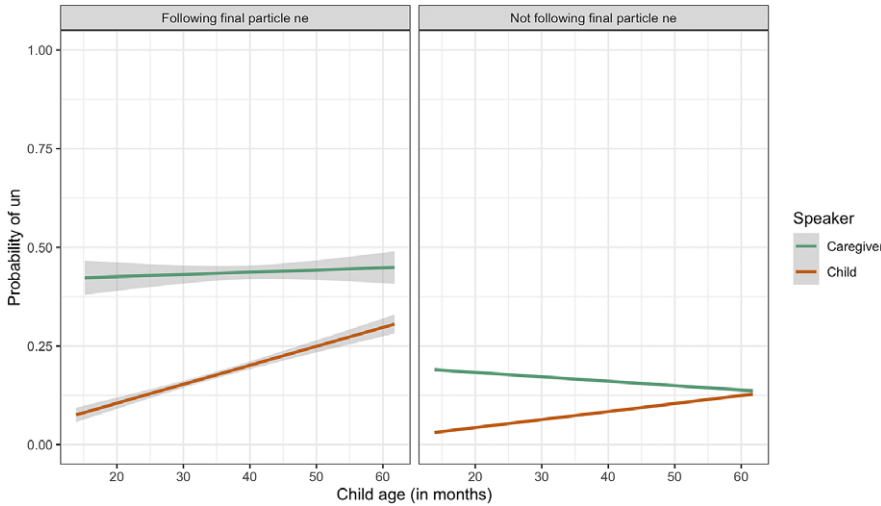


Figure 5. Probability of *un* production following or not following the interlocutor's final particle *ne* by speaker type and child age.

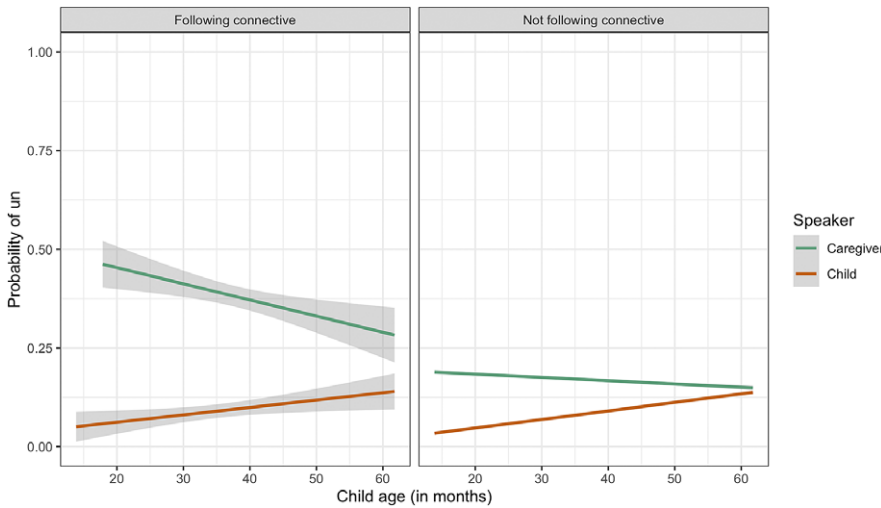


Figure 6. Probability of *un* production following or not following the interlocutor's connective by speaker type and child age.

t = 6.737, p < .001), and marginally increase their production following this cue as they develop (estimate = 0.005, SE = 0.003, t = 1.738, p = .082). As for connective predicates in Figure 6, whereas they clearly predict caregivers' production of *un* (estimate = 1.481, SE = 0.230, t = 6.446, p < .001), they do not do so for children (p = .224). Neither is there a developmental change in this effect (p = .539). This suggests that although children witness the *un* usage that is probabilistically conditioned by the connectives, they have not yet learned to reproduce this usage pattern by themselves.

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

These results suggest that children are in the process of gaining sensitivity to different kinds of cues. The final modal particle *ne* shows a clear predictive pattern in caregivers' production of *un*; this is probably why children are sensitive to this cue from early on. However, they seem to keep adjusting their conditioned usage pattern over development. An important difference between children and caregivers was observed regarding connective predicates. Whereas caregivers tend to produce *un* after the turns that end with a connective predicate, children do not show sensitivity to this kind of cue. Although we can only speculate on the reason for the difference in children's sensitivity between two types of predictive cues, one factor would be the high frequency of use of the final particle compared to the connectives (11,613 vs. 2,584 instances respectively), giving children repeated opportunities to detect the probabilistic adjacency between the cue and *un*.

As mentioned earlier, the final particle *ne* does not always signal the speaker's continuation of their own floor. It can be used at the end of a complete sentence, as in Example 4, probably with a higher frequency. To examine these different kinds of *ne*-ending turns, we analysed the data by coding the likeliness of the speaker's continuation. Specifically, we coded the immediately preceding turns in terms of whether the element before the final particle *ne* was a noun, a case marker (e.g., *ga* NOM, *ni* LOC), or a topic marker (*wa*) to approximately identify the *ne*-ending turns that signal further continuation (e.g., *kore ne*, *kore ga ne*, or *kore wa ne;* all meaning "This is …"). These particles are more likely to signal continuation compared with other elements, such as verbs or adjectives in this position, which would rather make a complete proposition (e.g., *sugoi ne* "it's great") without signalling any lacking element to follow. Consequently, only approximately 10% of the *ne*-ending turns (1,028 out of 10,585) had nouns or these particles before *ne*, thus coded as "likely" to signal continuation. However, the probability of *un* response was clearly higher after this "likely" type of preceding turns in both types of speakers (estimate = 1.471, SE = 0.095, t = 15.56, $p < .001$), as shown in Figure 7. This effect is larger in caregivers than in children (estimate = –0.298, SE = 0.148, t = –2.010, p = .044). Although this interaction was significant, the relatively small effect size (t = –2.010) suggests that the practical significance of this interaction is negligible. This implies that both the caregivers and the children are similarly sensitive not only to the particle *ne* in general, but also to the types of *ne*-ending turns. They can recognise the interactionally important distinctions, which is whether their interlocutor will keep their floor, to respond accordingly in the next turn.

Overall, our results showed that children learn the probabilistic dependency between different cues and the token *un* in adjacent turns. The turns that ended with the final modal particle *ne* and connective predicates were probabilistically associated with the caregivers' production of *un* in the next turn. The children in our data are sensitive to the final modal particle *ne* from early on. This modal particle is a good predictive cue for producing *un* in general. When a speaker seeks a common ground with their interlocutor by marking their utterance with this modal particle *ne*, responding with *un* is likely to be an appropriate behaviour in the next turn. Additionally, both caregivers and children produce *un* with a particularly high probability when they recognise their interlocutor's *ne*-ending turns as incomplete and likely to signal a further continuation. This would be a typical backchanneling usage in which listeners produce *un* to support the speakers' continuation of their own floor. However, we did not find evidence that children are sensitive to the connectives. These cues occur much less frequently and are considered harder to learn, as compared to the final

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

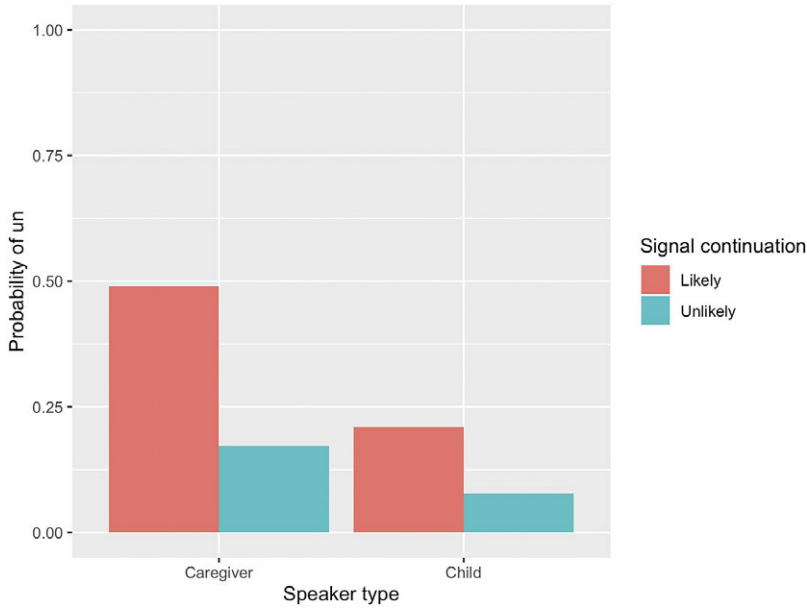


Figure 7. Probability of *un* after the interlocutor's turns that are likely to signal continuation (a noun or case / topic marker + final particle *ne*), or turns that are unlikely to signal continuation (other elements + final particle *ne*).

particle *ne*. Children seem to start with easy cues and gradually become able to process difficult ones as well.

Discussion

This study investigated children's acquisition of a Japanese discourse marker *un*, which is typically used as a positive response for yes-no questions and as a backchannel or acknowledgement. To test whether children learn to produce this token at interactionally appropriate moments, we focused on different cues in the immediately preceding turns from the interlocutor and analysed whether children's production of *un* following these cues changes probabilistically over the course of their development. We built a statistical model for each of the two types of interactional moments: when the interlocutor asked a yes-no question, and when the interlocutor signalled the continuation of their own speech.

Children produce *un* less than caregivers, in line with the findings of previous studies which report that children have more difficulty in responding than adults (Casillas et al., 2016; Stivers et al., 2018). Children gradually increase their production of *un* and reach the caregivers' rate at approximately five years of age. These results are consistent with many studies claiming that children's learning is sensitive to the probabilistic patterns in the input (e.g., Ambridge et al., 2015; Ellis, 2002; Saffran & Kirkham, 2018).

Most importantly, our results supported our hypothesis that children learn the probabilistic dependency between adjacent turns in a conversation to produce *un* appropriately. Children learn different cues to detect relevant interactional distinctions: when their interlocutors ask yes-no questions, and when their interlocutors signal continuation of their speech. Children not only learn the overall frequency distribution

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

of *un*, but also the interactional conditions for using *un*. Children showed sensitivity to yes-no questions in general as well as to formal cues, such as the final modal particle *ka* for interrogation and *wh*-words (which is inversely related with *un* responses). They also exhibited sensitivity to the final modal particle *ne*, especially when it signalled the interlocutors' continuation. These results uphold the idea that children pay attention to different linguistic elements to understand the kind of interaction and accordingly engage during a conversation.

Comparing the two interactional situations, children seem to learn to use *un* as a positive response to yes-no questions earlier than as a backchannel or acknowledgement. While children showed a rapid increase in their use of *un* after yes-no questions, they could not identify some of the predictive cues for interlocutors' continuation during the observed period. Children become sensitive to the final modal particle *ne*, but not to connective predicates. One possible reason could be that they encounter far fewer opportunities to learn connectives as predictive cues, because they occur much less frequently than the final particle *ne*. Learning the adult-like usage of *un* using multiple cues may extend beyond the age of five, longer than the observed period in this study.

As emphasised elsewhere, the most important feature of this study is its focus on the interactional dependency between adjacent turns in a conversation. Our results overall support the importance of the interactional dependency in the context of child language learning. That is, children pay attention to different linguistic elements in an ongoing conversation to detect relevant cues for understanding their interlocutors' communicative acts, and for projecting what to say in the next turn. Focusing on *un*, a token without semantic content, was a way to close in on the effect of interactional dependency. At the same time, any language use in interaction would need an account of how each instance of speech production is situated within an interactional sequence.

To appreciate how children learn language in conversational interactions, we need to understand the kind of challenges they face during an ongoing interaction and study how they learn to meet the challenges by using linguistic means. Conversation includes both children's own turns and their interlocutors' turns, which are sequenced in a certain predictable way. Children not only hear and learn their interlocutors' utterances, but also relate these utterances with their own utterances to learn what to say to engage in a conversation. The concept of input-based learning, which has been central to the usage-based approach, is not very useful for capturing this interactive process because of its unidirectional and static connotation. Instead, children learn how their own utterance affects their interlocutors' subsequent utterances, and vice versa. This constant reaction loop in conversation implies an ample opportunity for language learning.

Finally, quantitative modelling using the data with more than 200,000 conversational turns is an important methodological advantage of this study. The obvious shortcoming of this approach is that we cannot investigate the details of each unique instance as done in qualitative research. However, this approach presents the systematicity in the coding, the exhaustive analysis of the available data, and quantitatively reliable model results. It is also worth noting that our focus on formal cues without attributing any rich interpretation on individual instances is a conservative and justifiable approach, since we do not yet know much about the kind of interpretation children attribute to their language experience through their developing cognition.

Nevertheless, our findings need to be corroborated by future investigations. Research in the field may expand our study by increasing statistical power (i.e., include more subjects) and exploring other interactional settings (e.g., child-child interaction). Moreover, exploring more detailed and complex interactional cues for using *un* or any other

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

linguistic forms, including non-linguistic ones (e.g., gestures and facial expressions), will enhance our understanding of child language development that takes place in everyday conversational interactions.

Acknowledgments. We would like to thank Professor Franklin Chang and Professor Toshihide Nakayama for their helpful comments on the earlier versions of this paper and Professor Suzanne Miyata for her assistance in using the corpus data. We also extend our gratitude to the action editor and anonymous reviewers who helped us improve this paper.

References

Ambridge, B., **Kidd, E.**, **Rowland, C. F.**, & **Theakston, A. L.** (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, **42**(2), 239–273. doi:10.1017/S030500091400049X

Angles, J., **Nagatomi, A.**, & **Nakayama, M.** (2000). Japanese responses hai, ee, and un: yes, no, and beyond. *Language & Communication*, **20**(1), 55–86. doi:10.1016/S0271-5309(99)00018-X

Brown, P., & **Levinson, S. C.** (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Bucciarelli, M., **Colle, L.**, & **Bara, B. G.** (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, **35**(2), 207–241. doi:10.1017/S0305000915000689

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.

Chang, F. (2017). The LuCiD language researcher's toolkit [Computer software]. Retrieved from http://www.lucid.ac.uk/resources/for-researchers/toolkit/ Accessed January 29, 2020.

Clancy, P. M. (1987). 11. The acquisition of communicative style in Japanese. In Schieffelin, B. B., & Ochs, E. (eds.). *Language socialization across cultures*, 213–150. Cambridge: Cambridge University Press.

Clancy, P. M., **Thompson, S. A.**, **Suzuki, R.**, & **Tao, H.** (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, **26**(3), 355–387. doi:10.1016/0378-2166(95)00036-4

Clark, H. H., & **Brennan, S. E.** (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Cook, H. M. (1990). The sentence-final particle ne as a tool for cooperation in Japanese conversation. *Japanese/Korean Linguistics*, **1**, 29–44.

Cosgrove, J. M., & **Patterson, C. J.** (1977). Plans and the development of listener skills. *Developmental Psychology*, **13**(6), 557–564. doi:10.1037/0012-1649.13.6.557

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, **24**(2), 143–188. doi:10.1017/S0272263102002024

Hess, L. J., & **Johnston, J. R.** (1988). Acquisition of back channel listener responses to adequate messages. *Discourse Processes*, **11**(3), 319–335. doi:10.1080/01638538809544706

Iwasaki, S. (1997). The Northridge earthquake conversations: The floor structure and the "loop" sequence in Japanese conversation. *Journal of Pragmatics*, **28**(6), 661–693. doi:10.1016/S0378-2166(97)00070-2

Kawamori, M., **Kawabata, T.**, & **Shimazu, A.** (1998). Discourse markers in spontaneous dialogue. A corpus based study of Japanese and English. In M. Stede, L. Wanner, & E. Hovy (Eds.), *Proceedings of coling/ACL workshop on discourse relations and discourse markers* (pp. 93–99).

Kita, S., & **Ide, S.** (2007). Nodding, aizuchi, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships. *Journal of Pragmatics*, **39**(7), 1242–1254. doi:10.1016/j.pragma.2007.02.009

Kushida, S. (2009). *Kikite ni yoru katari no shinkou sokushin – keizoku shiji, keizoku saisoku, keizoku shikou-* [Recipient's Methods for Facilitating Progress of Telling: Sustaining, Prompting and Promoting Continuation]. *Cognitive studies: Bulletin of the Japanese Cognitive Science Society*, **16**(1), 12–23. doi:10.11225/jcss.16.12

Lammertink, I., **Casillas, M.**, **Benders, T.**, **Post, B.**, & **Fikkert, P.** (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology*, **6**(495), 495. doi:10.3389/fpsyg.2015.00495

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Transcription format and programs* (Vol. **1**). Psychology Press.

Miyata, S. (2004a). *Japanese – Miyata – Aki Corpus*. TalkBank.

Miyata, S. (2004b). *Japanese – Miyata – Ryo Corpus*. TalkBank.

Miyata, S. (2004c). *Japanese – Miyata – Tai Corpus*. TalkBank.

Miyata, S., & **Nisisawa, H. Y.** (2007). The acquisition of Japanese backchanneling behavior: Observing the emergence of aizuchi in a Japanese boy. *Journal of Pragmatics*, **39**(7), 1255–1274. doi:10.1016/j.pragma.2007.02.012

Miyata, S., & **Nisisawa, H. Y.** (2009). *Japanese – MiiPro – Asato Corpus*. TalkBank.

Miyata, S., & **Nisisawa, H. Y.** (2010). *Japanese – MiiPro – Tomito Corpus*. TalkBank.

Montes, R. G. (1999). The development of discourse markers in Spanish: Interjections. *Journal of Pragmatics*, **31**(10), 1289–1319. doi:10.1016/S0378-2166(98)00106-4

Nishizaka, A. (2008). *Hatsugen junban nai ni oite bunsan suru bun: sougo koui no shouten to shite no hannou kikai ba* [A Sentence Dispersed Within a Turn-at-Talk: Response-Opportunity Places as an Interactional Focus]. *Japanese Journal of Sociolinguistic Sciences*, **10**(2), 83–95.

Nisisawa, H. Y., & **Miyata, S.** (2009). *Japanese – MiiPro – Nanami Corpus*. TalkBank.

Nisisawa, H. Y., & **Miyata, S.** (2010). *Japanese – MiiPro – ArikaM Corpus*. TalkBank.

Rakoczy, H., & **Tomasello, M.** (2009). Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition*, **113**(2), 205–212. doi:10.1016/j.cognition.2009.07.013

R Core Team (2020). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/. Vienna, Austria: R Foundation for Statistical Computing.

Rigby, R. A., & **Stasinopoulos, D. M.** (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(3), 507–554.

Sadanobu, T. (2002). *'Un' to 'so' ni imi wa aru ka* [Do *'un'* and *'so'* have meanings?]. In T. Sadanobu (Ed.), *'Un' to 'so' no gengogaku* [Linguistics of *'un'* and *'so'*] (pp. 75–112). Hitujisyobo.

Saffran, J. R., & **Kirkham, N. Z.** (2018). Infant statistical learning. *Annual Review of Psychology*, **69**, 181–203. doi:10.1146/annurev-psych-122216-011805

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Georgetown University Press.

Schegloff, E. A., & **Sacks, H.** (1973). Opening up closings. *Semiotica*, **8**(4), 289–327. doi:10.1515/semi.1973.8.4.289

Stephens, G., & **Matthews, D.** (2014). The communicative infant. In Matthews, D. (Ed.), *Pragmatic development in first langauge acquisition* (pp. 13–36). Amsterdam: John Benjamins.

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. doi: 10.1016/j.pragma.2017.11.013

Tanaka, L. (2010). Is formality relevant? Japanese tokens hai, ee and un. *Pragmatics*, **20**(2), 191–211. doi: 10.1075/prag.20.2.04tan

Togashi, J. (2002). *'Un' to 'hai' no kankei o megutte* [Regarding the relationship between *'Un'* and *'hai'*]. In T. Sadanobu (Ed.), *'Un' to 'so' no gengogaku* [Linguistics of *'un'* and *'so'*] (pp. 127–157). Hitujisyobo.

White, S. (1989). Backchannels across cultures: A study of Americans and Japanese. *Language in Society*, **18** (1), 59–76. doi:10.1017/S0047404500013270

Yngve, V. H. (1970). On getting a word in edgewise. In *Chicago linguistics society, 6th Meeting* (pp. 567–578).

Cite this article: Tatsumi, T., & Sala, G. (2023). Learning conversational dependency: Children's response using *un* in Japanese. *Journal of Child Language* **50,** 1226–1244, https://doi.org/10.1017/S0305000922000344

https://doi.org/10.1017/S0305000922000344 Published online by Cambridge University Press