Journal of Child Language (2026), 1–13
doi:10.1017/S0305000926100920

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

Young Children Use the Discourse Marker "But" to Provide Clarifications in Everyday Conversations

Luis De la Viña[1], Julia Chunxiao Hu[2], Paul L. Harris[3] and Samuel Ronfard[1]

[1]Department of Psychological & Brain Sciences, University of Toronto Mississauga, Canada; [2]Department of Psychology, University of Toronto, Canada and [3]Graduate School of Education, Harvard University, USA
Corresponding author: Luis De la Viña; Email: luis.delavina@mail.utoronto.ca

(Received 16 June 2025 UTC; revised 22 May 2026 UTC; accepted 06 July 2026 UTC)

Abstract

To communicate effectively, speakers must monitor and refine their utterances based on what their interlocutors know. Young children's use of the contrastive marker *but* offers a window into this ability. Using longitudinal transcripts, we examined how seven children (ages 2–5) used *but* to clarify messages in everyday conversations. We coded 1,446 utterances to identify clarifications addressing gaps in listener knowledge. Children rarely used *but* for clarification at age 2, with marked increases between ages 2 and 3. By age 3, children often used *but* to provide privileged information in response to adults' knowledge-gap cues. These findings raise questions about the role of conversational experience in the development of children's mentalising abilities.

Keywords: clarifications; theory of mind; communication

1. Introduction

Verbal communication requires speakers to quickly translate their thoughts into words, often leading them to omit nuances, use ambiguous expressions, or leave out details that listeners need to understand their communicative intentions. The ability to clarify one's discourse is therefore essential for maintaining effective communication. Clarifications may be prompted by conversational partners through explicit requests or signs of confusion or misunderstanding. In other cases, however, clarifications occur in the absence of such cues, potentially because speakers recognise that their message requires revision or additional detail. Although these two types of clarifications may involve different communicative skills, both require speakers to monitor message adequacy in relation to their audience's presumed knowledge or understanding (Clark & Brennan, 1991). Thus, clarifications in spontaneous conversation provide a naturalistic window into how children's sensitivity to common ground is deployed and refined in everyday interaction.

1.1. Tracking and repairing common ground in early childhood

A large body of research shows that young children communicate differently based on what their interlocutor knows. By 30 months of age, children adapt their vocalisations and

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

pointing depending on whether an adult is knowledgeable or ignorant about an object's location (Behne et al., 2014; Knudsen & Liszkowski, 2012; O'Neill, 1996). By age 3, they offer more specific descriptions when requesting an object that is unfamiliar to the listener (Nayer & Graham, 2006) and more complex expressions when a referent requires disambiguation (Abbot-Smith et al., 2016). Three-year-olds also produce more elaborated narratives with more explicit references when addressing naïve versus well-informed listeners (Menig-Peterson, 1975; Perner & Leekam, 1986).

Young children can also adjust their communicative actions in real time when they realise that their interlocutor lacks knowledge (see Harris, 2017, for a review). When an adult misnames a common object or makes an unusual request, two- and three-year-olds frequently provide corrections, express disagreement, ask for clarification, or express doubt (Wellman et al., 2019), and these detections are often followed by repair attempts, even when there is no benefit at stake (Grosse et al., 2010). With age, children become increasingly competent in using their interlocutors' feedback to detect and repair shortcomings in their own messages. Although preschoolers often produce egocentric and ambiguous referential descriptions (e.g., saying "the blue one" when more than one option fits that description), by ages 4–5 they generally provide adequate repairs when given explicit feedback (Bacso & Nilsen, 2017; Nilsen & Mangal, 2012). However, their ability to repair utterances remains fragile, particularly when feedback is nonverbal (Bacso & Nilsen, 2022) or vague (Matthews et al., 2012). Research suggests that improvements in children's clarification abilities are supported by feedback from communicative partners (Ju et al., 2025; Matthews et al., 2007), as well as by developing theory-of-mind and executive functioning skills (Khu et al., 2020; Matthews et al., 2018; Nilsen & Graham, 2009).

1.2. The current study

In sum, considerable research has examined children's ability to tailor their utterances based on their listeners' informational needs, yet these studies have typically relied on constrained experimental tasks that do not reflect everyday communicative contexts. Moreover, developmental inferences have been drawn primarily from cross-sectional comparisons. We addressed these gaps by using naturalistic, longitudinal transcripts to examine children's use of the discourse marker *but* to elaborate a prior statement with clarifying information. These clarifications are relevant because they may reveal how children adjust their speech to their listeners' informational needs as they arise in everyday conversation.

Although speakers use various discourse markers for clarification (e.g., *because*, *well*, *so*; Schiffrin, 2012), examining multiple markers would have produced an unwieldy and heterogeneous dataset, complicating reliable coding. We focused on *but* for two reasons. First, prior research has documented young children's spontaneous use of *but. Children use it* to elaborate or qualify prior statements and to serve pragmatic functions in conversational narratives (Bloom et al., 1980; Peterson & McCabe, 1991). Second, unlike *because* or *so*, which often introduce explanations, reasons, or consequences, *but* is especially useful for identifying instances in which speakers revisit a prior utterance and add information that constrains its interpretation (e.g., *I'll just go walking around, but not far*). Recent work, however, suggests that young children may not fully grasp all contrastive meanings of *but*, despite producing it in apparently sophisticated ways (Skarabela et al., 2023). For example, in the phrase *James is tall, but he is not good at basketball*, the listener must infer that height creates an expectation of being good at basketball. According to Skarabela et al. (2023), children do not begin to grasp these uses

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

of *but* in an adult-like way until the school years. Our focus was therefore not on whether children possess an adult-like semantic understanding of *but*, but on whether they use it interactionally to elaborate or qualify prior utterances.

Accordingly, we developed a coding scheme to identify cases where children's *but*-statements provided clarifying information in response to a perceived gap in the listener's knowledge or understanding. This included cases where listeners signalled confusion or ignorance, as well as cases where children provided information plausibly unavailable to the listener.

2. Methods

2.1. CHILDES corpus

The current study utilised language samples from the Child Language Exchange System (CHILDES; MacWhinney, 2000). We systematically searched for North American English transcripts that provided dense, longitudinal language samples collected regularly between ages 2 and 5 in comparable naturalistic contexts (i.e., at-home conversations with a parent or researcher). Only seven datasets met these criteria: Abe (Kuczaj, 1977), Emily, Emma, Matt, Roman (Weist & Zevenbergen, 2008), and Adam and Sarah (Brown, 1973). See the Supplementary Materials for additional details about the transcripts.

2.2. Procedure

2.2.1. Search for relevant "but" statements

We searched the transcripts for instances in which children used the word *but*, yielding a total of 1,446 cases for coding. Relevant cases were defined as instances in which *but* was used to elaborate on a previous statement made by the child. Specifically, these instances included: (1) an initial statement, (2) the connective *but*, and (3) a second statement elaborating on the initial one. The two statements needed to be close in proximity (within the same sentence or a few lines apart in the transcript) and to concern the same topic as, or be related to, the child's initial statement. The following interaction is illustrative:

Father: *yeah, what did you do at school today?*
Child: *we made a little snowman.*
Child: *it was a snowman.* (initial statement)
Child: *but it didn't have much snow* (elaboration on initial statement)

All other uses of *but* were coded as *Other*. This included cases in which: (a) *but* was used to change the topic or respond directly to an adult's statement (e.g., expressing a different desire or belief), (b) *but* was not produced spontaneously (e.g., imitation, reading, or singing), or (c) the context was unclear.

To assess reliability, two of the authors independently coded the utterances of four children (Adam, Abe, Matt, and Sarah), representing 40% of the dataset. Inter-rater agreement was excellent (Cohen's $\kappa = .84$), and disagreements were resolved through discussion. The remaining 60% of cases were coded by one author. Nearly half of the utterances (781 of 1,446) met the criteria for relevant *but* statements and were retained for the remaining analyses.

2.2.2. Coding for clarification cases

Once relevant instances were identified, we established two independent criteria to determine which *but* statements constituted clarifications aimed at filling a perceived

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

knowledge gap or resolving confusion on the part of the adult listener. To be counted as a clarification, the child's *but* statement needed to either (a) be preceded by implicit or explicit expressions of ignorance or confusion from the adult, signalling a need for clarification, or (b) include information that was unlikely to be shared by the listener, thereby justifying the addition of details to build common ground.

Based on these criteria, we devised a two-step coding scheme. In Step 1, we coded whether the immediate conversational context preceding the child's *but* statement included knowledge-gap cues from the adult, such as expressions of ignorance, confusion, or misunderstanding related to the child's prior statements (see Table 1). These cases were coded as Prompted. When no such cues were identified, statements were coded as Unprompted. This distinction captured whether the child's elaboration was motivated by a knowledge gap that had either been signalled by the listener or presumed by the child.

Step 2 assessed whether the child's *but* statement involved the provision of privileged information – information unlikely to be known by the adult listener. Although determining what a listener knows can be difficult, we focused on cases where such an inference could be reliably made. Specifically, information was coded as privileged when it involved (a) non-shared personal experiences (e.g., past events the listener did not witness), (b) private internal states (e.g., sensations, intentions, emotions, dreams), or (c) imagined events or pretend scenarios. These three subcategories involve information that is either private or unobservable to the listener, or information to which the child has privileged access based on their own perspective or role in the interaction, even when some shared background knowledge already exists. We limited our analysis to these three subcategories because they involve minimal coder inference about common ground (see Table 1 for detailed descriptions and Table 2 for illustrative examples).

To ensure consistency, two authors independently coded all *but* statements. Inter-rater agreement for Step 1 (Prompted vs. Unprompted) was substantial (Cohen's $\kappa = 0.70$). For Step 2, agreement was assessed separately for each subcategory, yielding $\kappa = 0.82$ for non-current personal experiences, $\kappa = 0.79$ for internal subjective states, and $\kappa = 0.80$ for non-real events. Disagreements were resolved through discussion. Statements meeting the criteria for at least one subcategory were coded as Privileged Information; all others were coded as Other.

Combining the two coding steps yielded four types of *but* statements: (1) Unprompted + Other, (2) Unprompted + Privileged Information, (3) Prompted + Other, and (4) Prompted + Privileged Information. Statements coded as Unprompted + Other were not considered clarifications, as neither the conversational context nor the content justified providing additional information to the listener. The remaining three categories were treated as clarifications.

2.3. Data preparation and analysis

For each identified *but* statement, we used the exact age (years, months, and days) as reported in the CHILDES transcript at the time the utterance was produced. These age values were used in all analyses and, where relevant, were grouped into 12-month age ranges for statistical modelling.

All analyses were conducted in R 4.4.3 (R Core Team, 2025). Interrater agreement was assessed with Cohen's kappa using the irr package (Gamer et al., 2019). Remaining analyses used generalised linear mixed-effects models (GLMMs) fit with lme4 (Bates et al., 2015) using maximum likelihood estimation with the Laplace approximation. Wald chi-square

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

Table 1. Coding scheme

Step	Coding target	Coding categories	Description
Step 1: Type of *but* statement	The adult's prior statements	Prompted	The adult prompts the child (either explicitly or implicitly) by displaying ignorance, confusion, or misunderstanding related to the child's statements prior to the clarification.
		Unprompted	No cues suggesting the adult prompted the child's clarification by indicating ignorance, confusion, or misunderstanding.
Step 2: Conversational content	The child's utterances prior to the *but* statement and the *but* statement itself	Privileged information	The child's conversation involves information that is not readily accessible to the listener (adult), such as narrations or descriptions of: a) events that are not occurring in the "here and now," including past events and assumptions about what could or will happen in the future. b) their own internal subjective states, including desires, intentions, preferences, beliefs, emotions, and physical sensations. c) the content of dreams, make-believe stories, or pretend play.
		Other	The child's utterance involves information that is not specified in the *Privileged information category.*

tests were computed using the Anova function in car (Fox & Weisberg, 2019). Predictor effects were further evaluated with likelihood ratio tests comparing full models to null models that excluded the predictors of interest but retained the same random-effects structure (Dobson, 2002; Forstmeier & Schielzeth, 2011). Estimated marginal means and post hoc pairwise comparisons were obtained using emmeans (Lenth, 2025). Full regression tables are provided in the Supplementary Materials. ChatGPT (GPT-5.5; OpenAI, 2026) was used to assist with constructing and troubleshooting R code for data analysis.

3. Results

3.1. How often did children clarify their utterances using "but"?

As shown in Figure 1, 66% of relevant *but* statements (n = 781) were prompted by knowledge-gap cues, and 85% of these involved privileged information. In total, 90% of relevant *but* statements met our criteria for clarification: *but* statements prompted by knowledge-gap cues, containing privileged information, or both. While two-thirds were prompted, the remaining third were unprompted; of these, 71% still qualified because they contained privileged information. Overall, these patterns suggest that although children sometimes provided clarifying information spontaneously, they more often did so in response to adult knowledge-gap cues.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

Table 2. Examples of conversations coded into each coding category

Conversation excerpt	Coded as	Reason
Adult: *what happened in 4 months? Or 4 months ago? Whatever.* Child: *4 months this happened.* Child: *and it was a bad crash.* Adult: *oh.* Child: ***but no one got hurted.*** Adult: *how lucky*	Step 1: Prompted	The adult explicitly asks about what happened, signalling his ignorance to the child. Notice that the child's *but* statement speaks to this question.
	Step 2: Privileged information	The event the child is describing is make-believe (he is referring to his toy cars). This information is not readily accessible to the listener (adult).
Child: *he never find me.* Adult: *well (.) you have to hide.* Child: *I'm going to hide in there and he's going to look all around for me* ***but he will not see me.***	Step 1: Unprompted	No cues of ignorance, confusion, or misunderstanding from the adult.
	Step 2: Privileged information	The child is describing a belief about what will happen in the future. This information is not readily accessible to the listener (adult).
Adult: *well are you gonna marry Batman &-uh Robin or is Donna going to marry Robin?* Child: *Donna's gonna marry him (.)* ***but I'm gonna marry him faster***	Step 1: Prompted	The adult explicitly asks about what the child plans to do, signalling her ignorance.
	Step 2: Privileged information	The child is describing future plans and intentions. This information is not readily accessible to the listener (adult).
Child: *did you see the clouds?* Adult: *uhhuh.* Child: *they were clouds (.) Mom.* Adult: *that was smoke left over from the firework.* Child: *you thought that.* Child: ***but I thought they were clouds.***	Step 1: Unprompted	No cues of ignorance, confusion, or misunderstanding from the adult.
	Step 2: Privileged information	The information that the child describes is not readily accessible to the listener (adult).

3.2. Did children's tendency to clarify increase with age?

To evaluate whether children's tendency to clarify increased with age, we fitted a linear mixed-effects regression model predicting the proportion of clarifications per 1,000 words spoken from age range (12-month intervals), with a random intercept for participant to account for the nested structure of the data. The full model fit the data significantly better than the null model including only the random intercept, $\chi^2(2) = 9.7$, $p < 0.01$. This model yielded a significant main effect of age range, Wald $\chi^2(2) = 11.99$, $p < 0.01$.

The model's conditional R^2, reflecting both fixed (age) and random (participant) effects, accounted for 64% of the variance, whereas the marginal R^2, reflecting the fixed effects alone, explained 22% of the variance in clarification rates. As shown in Figure 2, children produced increasingly more clarifications with age, with a relatively marked increase between ages 2 and 3. Prior to age 3, children rarely clarified (0.58 clarifications per 1,000 words between 24 and 35 months); this rate increased by about 2.5 times during age 3 (36–47 months), reaching approximately 2.6 clarifications per 1,000 words during age 4 (48–59 months).

3.3. Did the frequency of different types of clarification change with age?

We also examined whether age predicted changes in the three clarification types: Prompted + Privileged Information, Prompted + Other, and Unprompted + Privileged

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

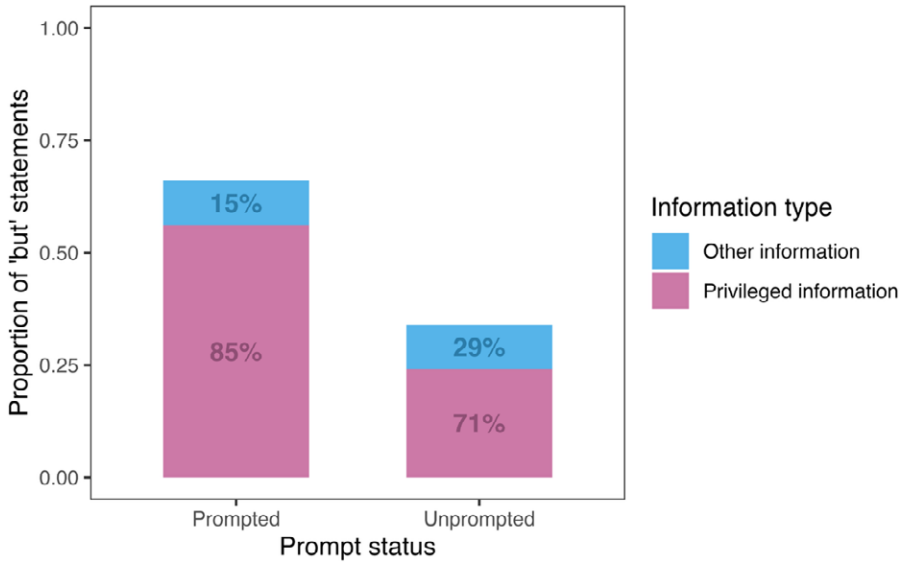


Figure 1. Proportion of relevant "but" statements by prompt status and information type. Percentages inside each bar segment indicate the proportion of privileged versus other information within each prompt-status category.

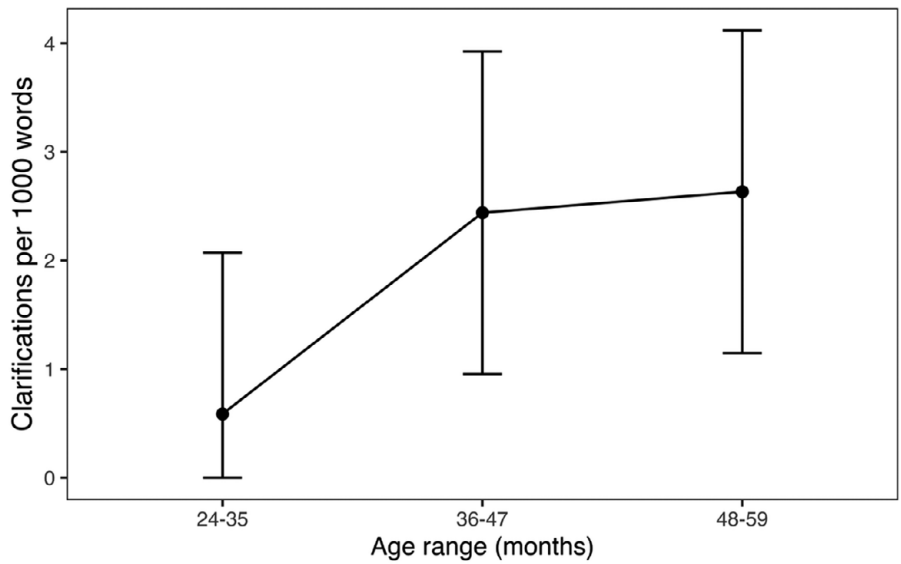


Figure 2. Estimated marginal means of clarification proportions per 1,000 words spoken from ages 2 to 5 years (95% confidence intervals).

Information. We fitted a mixed-effects linear regression predicting the proportion of clarifications per 1,000 words from age range, clarification category, and their interaction, with a random intercept for participant ID. Model diagnostics indicated approximate

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

normality of residuals but mild heteroskedasticity. Given the exploratory nature of these analyses and the robustness of mixed-effects models to mild assumption violations, we retained the linear model for ease of interpretation. The full model fit the data significantly better than the null model with only the random intercept, $\chi^2(11) = 45.48$, $p < 0.001$. The model revealed no significant main effects of Category or Age Range, but a significant Category × Age Range interaction, $\chi^2(6) = 11.03$, $p = 0.03$, indicating that developmental trajectories differed by clarification type. Simple effects analyses showed that only the Prompted + Privileged Information category increased significantly from 24–35 months to 36–47 months ($p < 0.001$), whereas the Unprompted + Privileged Information and Prompted + Other categories remained relatively stable (see Figure 3).

3.4. Did children talk differently about the various types of privileged information?

Among clarifications involving privileged information, internal subjective states were most common (52%), closely followed by non-current events (49%) and non-real events (23%). To examine age-related changes in the type of privileged information children provided, we fitted a mixed-effects linear regression predicting the proportion of privileged-information clarifications per 1,000 words from age range, privileged information subcategory, and their interaction, with a random intercept for participant ID. Because the interaction was not significant, it was removed. Model diagnostics indicated approximate normality of residuals but mild heteroskedasticity. The final model revealed significant main effects of Privileged Information Subcategory, Wald $\chi^2(2) = 7.51$, $p = 0.02$ (see Figure 4), and Age, Wald $\chi^2(2) = 19.46$, $p < 0.001$ (see Figure 5). Pairwise comparisons showed that clarifications involving non-real events were less frequent than those involving non-current events ($p = 0.04$), with no other significant differences between subcategories. Overall, children produced more privileged-information clarifications with age, and the frequency of these clarifications varied by subcategory.

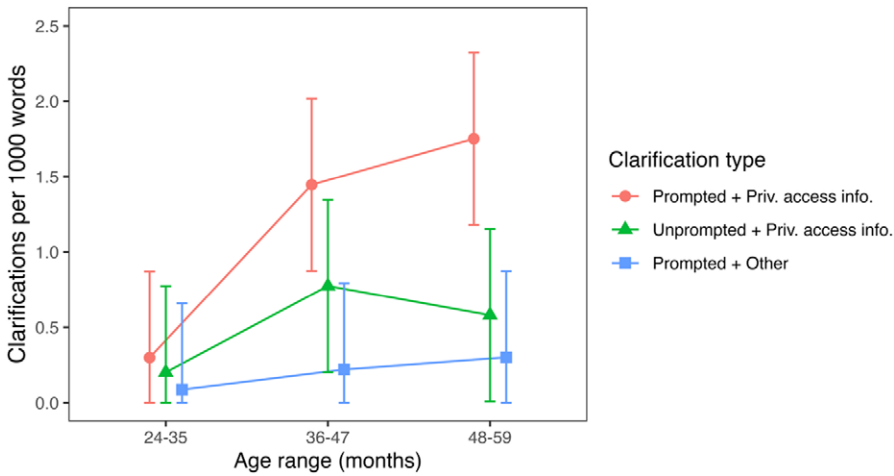


Figure 3. Estimated marginal means of clarification type proportions per 1,000 words spoken from ages 2 to 5 years (95% confidence intervals).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

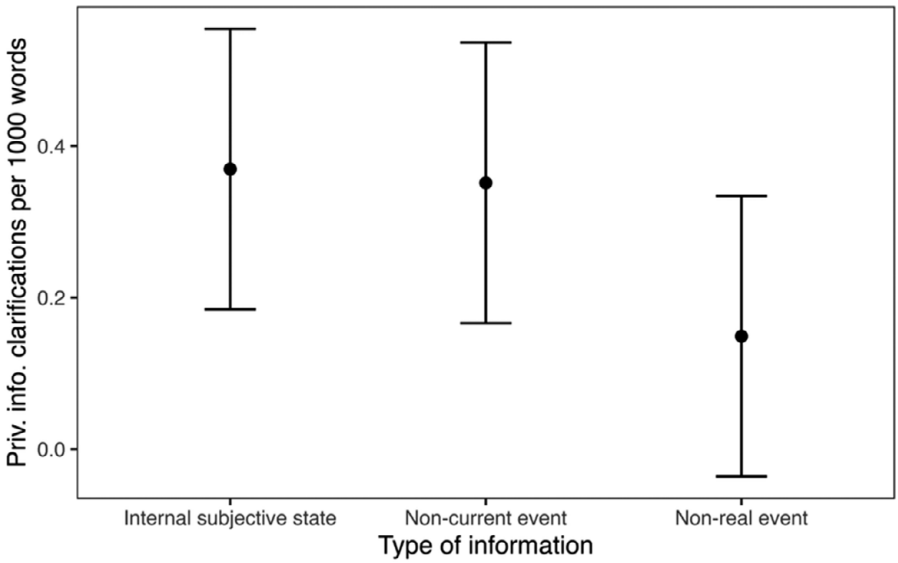


Figure 4. Estimated marginal means of privileged-information clarifications per 1,000 words spoken by subcategory (95% confidence intervals).

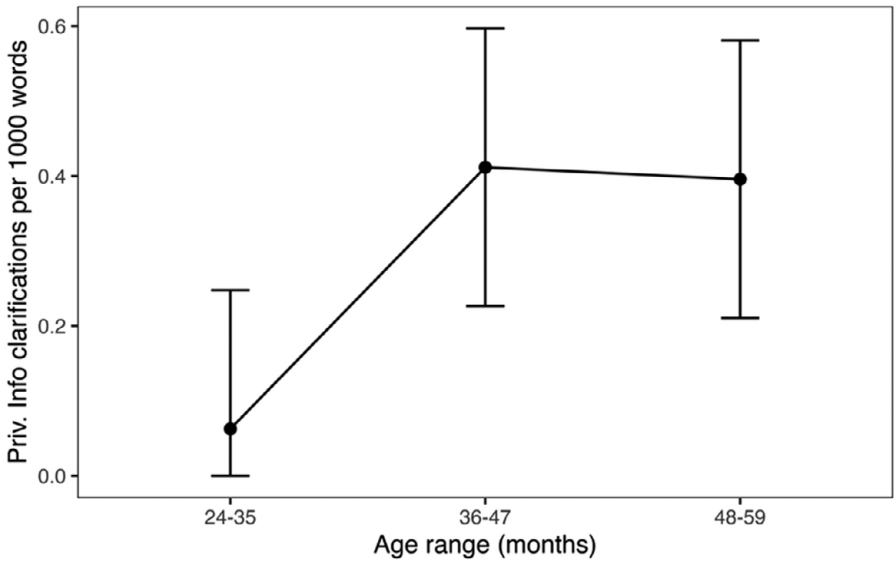


Figure 5. Estimated marginal means of privileged-information clarifications per 1,000 words spoken from ages 2 to 5 years (95% confidence intervals).

4. Discussion

We examined children's use of *but* to clarify messages in everyday conversations with adults, focusing on the development of their ability to monitor and address listeners' informational needs by adding details to build common ground. Using naturalistic,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

longitudinal transcripts, we identified instances in which children used *but* to elaborate on a prior statement. These utterances were coded for: (1) whether the conversational context included a prompt from the adult and (2) whether the child provided information that was not readily accessible to the listener.

Among *but* statements that children used to elaborate a prior utterance, 90% functioned as clarifications. The frequency of these clarifications increased substantially between ages 2 and 4. At age 2, children rarely used *but* to clarify statements. By age 3, they produced nearly 2.5 clarifications per 1,000 words. This increase was primarily driven by cases in which the adult cued a knowledge gap, and the child responded with privileged information.

Prior work has shown that preschoolers are sensitive to the importance of shared knowledge in communication (see Bohn & Köymen, 2018; Harris, 2017, for reviews) and adjust their utterances based on listeners' informational needs to establish common ground (Abbot-Smith et al., 2016; Bacso & Nilsen, 2017, 2022; Matthews et al., 2012; Nayer & Graham, 2006; Nilsen & Mangal, 2012). The current research extends this literature by showing that three- to four-year-olds spontaneously clarify their utterances in conversational contexts where shared knowledge is incomplete. These naturalistic findings suggest that preschoolers' sensitivity to common ground emerges in everyday interaction without the explicit epistemic asymmetries typically used in experimental tasks.

The rapid increase in clarifications between ages 2 and 4 is likely driven by intertwined developmental mechanisms that support children's ability to track and respond to gaps in shared understanding with conversational partners, including theory of mind (ToM) and executive functions (Khu et al., 2020; Matthews et al., 2018; Nilsen & Graham, 2009). Although the present study did not directly measure these capacities, children's ability to tailor clarifications to adult knowledge gaps has direct implications for research on ToM development. Specifically, the age-related changes in privileged-information clarifications suggest that children's performance reflects more sophisticated knowledge-state tracking than standard ToM tasks would predict. Wellman and Liu's (2004) scale assesses differential knowledge access in perceptual, here-and-now scenarios, whereas the privileged-information clarifications in our study often involved non-real content, non-current events, and internal subjective experiences. The latter two in particular showed substantial age-related increases between ages 2 and 4. These aspects of differential knowledge access may be particularly important for everyday conversation, as they require speakers to share information inaccessible to the listener through direct observation.

What specific ToM understanding supports such clarifications? Recent work suggests that children's early recognition of first-person authority – the understanding that speakers have privileged access to their own mental states – may be one key component (Yang et al., 2025). Recognising first-person authority may help children understand that their subjective experiences and memories of past events are unavailable to others, allowing them to identify situations in which clarification is needed.

Research also shows that clarification skills improve with listener feedback (Ju et al., 2025; Matthews et al., 2007), suggesting that repeated exposure to communicative failures and input from others helps children learn when and how to repair their speech. Consistent with this idea, bilingual toddlers – who are more likely to encounter misunderstanding than their monolingual peers – tend to show more advanced clarification skills (Wermelinger et al., 2017). Misunderstandings and clarification requests of the kind observed in our study may create a tutoring context that helps children recognise that others do not always know what they know, and that some types of information are unlikely to be mutually shared. For example, children may encounter more clarification

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

requests in conversations about past experiences compared to conversations about the here and now, because past events often require more contextual details to establish common ground with others. These conversations may offer opportunities for children to practice tracking others' epistemic states, thereby scaffolding their ToM development. This proposal aligns with a large body of work showing that conversation plays a critical role in children's ToM development (see de Rosnay & Hughes, 2006; Harris et al., 2005; Slaughter & Peterson, 2011; Symons, 2004, for reviews). Future research could investigate how individual differences in children's conversational experience and ToM jointly shape the frequency and manner of their clarifications.

The current findings should be interpreted in light of some important limitations. First, our study was limited to a small sample of North American children from predominantly highly educated families. Second, we focused on a particular type of pragmatic move – using the discourse marker *but* – to add detail or nuance to a previous statement. Although *but* is a common means of clarification in English, other discourse markers (e.g., *well*, *so*) can serve related functions and also appear in early speech (e.g., Choi, 2007; Sprott, 1992). Finally, because our analyses were based on verbal transcripts, they lacked information about nonverbal cues (e.g., facial expressions, intonation), and some parent–child interactions may have involved common ground established outside the recorded interaction. Despite these limitations, the data indicate that there are marked gains in children's tendency to offer clarifications, especially when prompted to supply otherwise privileged or opaque information.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0305000926100920.

Disclosure of use of AI tools. This project incorporated the use of generative artificial intelligence. OpenAI's ChatGPT was used for refining language and ensuring clarity of expression (GPT-5.5; OpenAI, 2026). AI-generated input was minimal and limited to wording or technical suggestions; it was reviewed and edited by the authors to ensure its accuracy, alignment with scholarly standards, and suitability for publication.

Competing interests. The authors declare that there is no conflict of interest regarding the publication of this article. Luis De la Viña's research was conducted with the support of the Connaught Scholarship for International students at the University of Toronto.

References

Abbot-Smith, K., **Nurmsoo, E.**, **Croll, R.**, **Ferguson, H.**, & **Forrester, M.** (2016). How children aged 2;6 tailor verbal expressions to interlocutor informational needs. *Journal of Child Language*, **43**(6), 1277–1291. https://doi.org/10.1017/S0305000915000616

Bacso, S. A., & **Nilsen, E. S.** (2017). What's that you're saying? Children with better executive functioning produce and repair communication more effectively. *Journal of Cognition and Development*, **18**(4), 441–464. https://doi.org/10.1080/15248372.2017.1336438

Bacso, S. A., & **Nilsen, E. S.** (2022). Children's use of verbal and nonverbal feedback during communicative repair: Associations with executive functioning and emotion knowledge. *Cognitive Development*, **63**, 101199. https://doi.org/10.1016/j.cogdev.2022.101199

Bates, D., **Maechler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behne, T., **Carpenter, M.**, & **Tomasello, M.** (2014). Young children create iconic gestures to inform others. *Developmental Psychology*, **50**(8), 2049–2060. https://doi.org/10.1037/a0037224

Bloom, L., **Lahey, M.**, **Hood, L.**, **Lifter, K.**, & **Fiess, K.** (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language*, **7**(2), 235–261. https://doi.org/10.1017/S0305000900002610

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

Bohn, M., & **Köymen, B.** (2018). Common ground and development. *Child Development Perspectives*, **12**(2), 104–108. https://doi.org/10.1111/cdep.12269

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Choi, I. (2007). *How and when do children acquire the use of discourse markers?* CamLing, 40–47.

Clark, H., & **Brennan, S. A.** (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

de Rosnay, M., & **Hughes, C.** (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology*, **24**(1), 7–37. https://doi.org/10.1348/026151005X82901

Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed.). Chapman & Hall/CRC.

Forstmeier, W., & **Schielzeth, H.** (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, **65**(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fox, J., & **Weisberg, S.** (2019). *An R companion to applied regression* (3rd ed.). Sage Publications.

Gamer, M., **Lemon, J.**, **Fellows, I.**, & **Singh, P.** (2019). *irr: Various coefficients of interrater reliability and agreement* (R package version 0.84.1). https://CRAN.R-project.org/package=irr

Grosse, G., **Behne, T.**, **Carpenter, M.**, & **Tomasello, M.** (2010). Infants communicate in order to be understood. *Developmental Psychology*, **46**(6), 1710–1722. https://doi.org/10.1037/a0020727

Harris, P. L. (2017). Tell, ask, repair: Early responding to discordant reality. *Motivation Science*, **3**(3), 275–286. https://doi.org/10.1037/mot0000075

Harris, P. L., **de Rosnay, M.**, & **Pons, F.** (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, **14**(2), 69–73. https://doi.org/10.1111/j.0963-7214.2005.00337.x

Ju, N., **San Juan, V.**, **Nilsen, E.**, & **Graham, S. A.** (2025). That's not the one I wanted: Feedback improves 5-year-olds' communicative perspective-taking. *Journal of Cognition and Development*, **26**(5), 793–815. https://doi.org/10.1080/15248372.2025.2492021

Khu, M., **Chambers, C. G.**, & **Graham, S. A.** (2020). Preschoolers flexibly shift between speakers' perspectives during real-time language comprehension. *Child Development*, **91**(3), e619–e634. https://doi.org/10.1111/cdev.13270

Knudsen, B., & **Liszkowski, U.** (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, **15**(1), 113–122. https://doi.org/10.1111/j.1467-7687.2011.01098.x

Kuczaj, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, **16**(5), 589–600. https://doi.org/10.1016/S0022-5371(77)80021-2

Lenth, R. V. (2025). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.11.0) [R package]. https://CRAN.R-project.org/package=emmeans

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: The database* (3rd ed.). Lawrence Erlbaum Associates Publishers.

Matthews, D., **Biney, H.**, & **Abbot-Smith, K.** (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, **14**(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

Matthews, D., **Butcher, J.**, **Lieven, E.**, & **Tomasello, M.** (2012). Two- and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science*, **4**(2), 184–210. https://doi.org/10.1111/j.1756-8765.2012.01181.x

Matthews, D., **Lieven, E.**, & **Tomasello, M.** (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, **78**(6), 1744–1759. https://doi.org/10.1111/j.1467-8624.2007.01098.x

Menig-Peterson, C. L. (1975). The modification of communicative behavior in preschool-aged children as a function of the listener's perspective. *Child Development*, **46**(4), 1015–1018. https://doi.org/10.2307/1128416

Nayer, S. L., & **Graham, S. A.** (2006). Children's communicative strategies in novel and familiar word situations. *First Language*, **26**(4), 403–420. https://doi.org/10.1177/0142723706064834

Nilsen, E. S., & **Graham, S. A.** (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, **58**(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Nilsen, E. S., & **Mangal, L.** (2012). Which is important for preschoolers' production and repair of statements: What the listener knows or what the listener says? *Journal of Child Language*, **39**(5), 1121–1134. https://doi.org/10.1017/S0305000911000432

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, **67**(2), 659–677. https://doi.org/10.1111/j.1467-8624.1996.tb01758.x

OpenAI. (2026). *ChatGPT (GPT-5.5 thinking)* [Large language model]. https://chat.openai.com/

Perner, J., & **Leekam, S. R.** (1986). Belief and quantity: Three-year-olds' adaptation to listener's knowledge. *Journal of Child Language*, **13**(2), 305–315. https://doi.org/10.1017/S0305000900008072

Peterson, C., & **McCabe, A.** (1991). On the threshold of the storyrealm: Semantic versus pragmatic use of connectives in narratives. *Merrill-Palmer Quarterly*, **37**(3), 445–464.

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.4.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Schiffrin, D. (2012). *Discourse markers* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511611841

Skarabela, B., **Cuthbert, N.**, **Rees, A.**, **Rohde, H.**, & **Rabagliati, H.** (2023). Learning dimensions of meaning: Children's acquisition of *but*. *Cognitive Psychology*, **147**, 101597. https://doi.org/10.1016/j.cogpsych.2023.101597

Slaughter, V., & **Peterson, C. C.** (2011). How conversational input shapes theory of mind development in infancy and early childhood. In M. Siegal & L. Surian (Eds.), *Access to language and cognitive development* (pp. 3–26). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199592722.003.0001

Sprott, R. A. (1992). Children's use of discourse markers in disputes: Form-function relations and discourse in child language. *Discourse Processes*, **15**(4), 423–439. https://doi.org/10.1080/01638539209544821

Symons, D. K. (2004). Mental state discourse, theory of mind, and the internalization of self–other understanding. *Developmental Review*, **24**(2), 159–188. https://doi.org/10.1016/j.dr.2004.03.001

Weist, R. M., & **Zevenbergen, A.** (2008). Autobiographical memory and past time reference. *Language Learning and Development*, **4**(4), 291–308. https://doi.org/10.1080/15475440802293490

Wellman, H. M., & **Liu, D.** (2004). Scaling of theory-of-mind tasks. *Child Development*, **75**(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., **Song, J. H.**, & **Peskin-Shepherd, H.** (2019). Children's early awareness of comprehension as evident in their spontaneous corrections of speech errors. *Child Development*, **90**(1), 196–209. https://doi.org/10.1111/cdev.12862

Wermelinger, S., **Gampe, A.**, & **Daum, M. M.** (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, **155**, 84–94. https://doi.org/10.1016/j.jecp.2016.11.005

Yang, Q. T., **Jeffries, I. E.**, **Zheng, I.**, & **Harris, P. L.** (2025). Young children endorse first-person over second-person claims concerning mental states. In *Society for research in child development biennial meeting* [Poster presentation]. Minneapolis, MN.

Cite this article: De la Viña, L., Hu, J.C., Harris, P.L., & Ronfard, S. (2026). Young Children Use the Discourse Marker "But" to Provide Clarifications in Everyday Conversations. *Journal of Child Language* 1–13, https://doi.org/10.1017/S0305000926100920

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:42:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0305000926100920