Journal of Child Language (2025), 1–12
doi:10.1017/S0305000925100184

BRIEF RESEARCH REPORT

Spontaneous Lexical Overlap in Early Conversations: Automated Sequential Coding of Parents and Toddlers

Emily K. Harrington, Pamela A. Hadley and Matthew Rispoli

Department of Speech and Hearing Science, University of Illinois Urbana–Champaign, Champaign, IL, USA
Corresponding author: Emily K. Harrington; Email: emilykh4@illinois.edu

(Received 05 September 2024; revised 19 April 2025; accepted 18 July 2025)

Abstract

This study piloted CHIPUTIL, an automated tool in CLAN for analysing sequential lexical overlap in parent–child conversations. In a sample of 44 dyads (child age M = 1;9), child spontaneous lexical overlap was positively associated with parent imitations and expansions, across the conversation and within sequential turns. Children were more than twice as likely to respond with lexical overlap when parents first produced an imitation or expansion. These findings offer insight into how lexical overlap may unfold in early conversations. We discuss implications of automated coding and future directions in exploring the role of lexical overlap in children's language development.

Keywords: lexical overlap; language sample analysis; automated measures

Spontaneous lexical overlap is a common linguistic phenomenon in conversation, occurring when a speaker repeats lexical content from their communication partner's prior utterance (Che et al., 2018; Sokolov & MacWhinney, 1990; Speidel & Nelson, 1989). Lexical overlap occurs in adult–adult and adult–child conversations (Pickering & Garrod, 2004), across languages (Casla et al., 2022; Che & Brooks, 2021; Chieng et al., 2024; Clark & Bernicot, 2008), and with considerable frequency (Che & Brooks, 2021). In early parent–child conversations, lexical overlap occurs in about 15%–25% of child utterances and 7%–14% of parent utterances (Casla et al., 2022; Che et al., 2018; Che & Brooks, 2021; Masur & Rodemaker, 1999). As such, lexical overlap is a pervasive feature of discourse (Chieng et al., 2024; Pickering & Garrod, 2004; Užgiris et al., 1989).

In addition to being common, lexical overlap has clear benefits within parent–child conversations. It can serve a range of communicative functions, such as confirming, clarifying, or extending a message, which establishes common ground in a transparent way (Clark, 2015; Clark & Bernicot, 2008; Pickering & Garrod, 2004). Lexical overlap allows toddlers to participate in conversation without having to generate novel linguistic content (Kirchner & Prutting, 1987; Snow, 1981). In caregiver utterances, lexical overlap offers children semantically related and temporally adjacent responses (Che et al., 2018), a feature of child-directed speech shown to facilitate language learning (Masek et al., 2021). Taken together, lexical overlap enables toddlers and caregivers to co-construct ideas over successive utterances by repurposing words from one utterance to the next (Pickering & Garrod, 2004). This may promote sustained engagement in conversation (Užgiris et al., 1989) and potentially maximise children's exposure to responsive, child-directed speech. Both characteristics are central to social (Bruner, 1982; Tomasello, 1995) and transactional (Camarata & Yoder, 2002; Sameroff & Chandler, 1975) theories of language development.

Most studies examining lexical overlap have used summary measures, characterising the overall frequency and proportion of parent and child lexical overlap across conversations (Casla et al., 2022; Che et al., 2018; Che & Brooks, 2021; Masur & Rodemaker, 1999). These studies have revealed a positive, concurrent relationship between parent and child lexical overlap (Casla et al., 2022; Masur, 1987; Masur & Eichorst, 2002; Masur & Rodemaker, 1999) and demonstrated that lexical overlap, particularly parent lexical overlap, is associated with children's language outcomes later in development (Casla et al., 2022; Che et al., 2018; Fusaroli et al., 2023; Masur & Eichorst, 2002; Olson & Masur, 2012; Soler et al., 2023; Tamis-LeMonda et al., 2001; Taumoepeau, 2016; Užgiris et al., 1989). However, summary measures offer limited insight into how lexical overlap unfolds within parent–child conversations in real time.

Sequential measures of lexical overlap fill this gap by capturing its turn-by-turn dynamics. Studies employing sequential measures have found that lexical overlap often elicits contingent responses across English-speaking (Farrar, 1987; Masur & Olson, 2008; Olson & Masur, 2012; Scherer & Olswang, 1984) and French-speaking (Clark & Bernicot, 2008) parents and children. For instance, Masur and Olson (2008) found that parents responded to toddler lexical overlap approximately 90% of the time, and toddlers responded to parent lexical overlap approximately 80% of the time. Moreover, speakers respond TO lexical overlap WITH lexical overlap (Clark & Bernicot, 2008; Farrar, 1987; Masur & Olson, 2008; Scherer & Olswang, 1984). Both Farrar (1987) and Scherer and Olswang (1984) examined two specific forms of lexical overlap and found that 2-year-old children were more likely to respond with an imitation if their parents had first produced an expansion. Unfortunately, relatively few studies use sequential measures (Clark & Bernicot, 2008; Farrar, 1987; Masur & Olson, 2008; Scherer & Olswang, 1984; Sokolov, 1993), likely because sequential analysis is time- and labour-intensive. There are compelling theoretical reasons to pursue this work. The transactional model of development emphasises that real-time interactions incrementally shape children's developmental outcomes (Sameroff & Chandler, 1975). Examining how lexical overlap influences caregiver and child participation in early conversations may shed light on how conversations emerge over time and pinpoint precise features of conversation that support children's language development (Camarata & Yoder, 2002; Masur & Olson, 2008). To conduct this work, efficient tools for sequential analysis are needed.

The CHIP and CHIPUTIL programs in Child Language ANalysis (CLAN) were developed to automate the coding of lexical overlap (MacWhinney, 2000). CHIP analyses source–response utterance pairs and codes the response for matches, additions, deletions, and substitutions (Sokolov & MacWhinney, 1990). CHIP has been used in several investigations to efficiently compute summary measures of lexical overlap (Casla et al., 2022; Che et al., 2018; Che & Brooks, 2021; Soler et al., 2023). However, it is not well suited for computing sequential measures. CHIP only codes the response utterance, without linking it back to the source. Thus, automated CHIP measures only encompass the

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

speaker who produced lexical overlap, not the speaker who produced the source utterance. For a sequential analysis, it is necessary to derive measures for both speakers.

To address this limitation, and at the request of the first author, the developers of CLAN created CHIPUTIL. CHIPUTIL extends CHIP's functions by coding the source utterance and linking it to the response utterance coded by CHIP. Additionally, CHIPUTIL can extract and organise these linked source–response pairs into two subsets: those in which the response contains lexical overlap and those in which it does not. Researchers can then apply any CLAN program to these subsets to compute a range of linguistic and discourse measures. Measures computed on the source utterances offer information about the first turn in a conversational exchange, and the linked response utterance indicates whether lexical overlap was produced on the subsequent turn or not. Thus, automated measures can be computed on source utterances and compared relative to the presence or absence of lexical overlap in the response utterances.

In this study, we examine whether certain types of parent utterances increase the likelihood of children's lexical overlap in their subsequent turns. We focus on parent imitations and expansions because these are associated with child lexical overlap in studies employing summary measures (Casla et al., 2022; Masur, 1987; Masur & Eichorst, 2002; Masur & Rodemaker, 1999) and in the few small-scale studies employing sequential analyses (Farrar, 1987; Scherer & Olswang, 1984). To illustrate the differences in these approaches, we ask the following research questions:

1. Is there an association between a summary measure of child lexical overlap and a combined summary measure of parent imitations and expansions?
2. Do parent imitations and expansions increase the likelihood of lexical overlap in the child's subsequent turn?

We hypothesise positive associations for both, with sequential measures offering more insight into how lexical overlap contributes to parent–child conversations.

1. Method

1.1. CHILDES corpus

The current study utilised the Champaign longitudinal corpus (Rispoli & Hadley, n.d.) in the Child Language Data Exchange System (MacWhinney, 2000). The corpus contains transcripts of parent–child language samples recorded when children were 1;9, 2;0, 2;3, 2;6, 2;9, and 3;0 years. All 44 children in this database spoke General American English, were developing language typically, and had no significant medical history. Language samples were collected in a lab playroom where researchers provided a standard set of toys (i.e., baby doll with accessories, play kitchen set, bubbles, blocks, windup toys, and bowling set) and instructed parents to "play as they would at home" for 30 minutes (Hadley et al., 2014). Language samples were video recorded and transcribed by research assistants (for full procedures, see Hadley et al., 2014).

We analysed 44 parent–child language samples collected at child age 1;9 (SD = 0.38 months). Per parent report, children were 86.4% White (n = 38), 9.1% Black (n = 4), and 4.6% Mixed race (n = 2). No children were reported to be Hispanic. Parents included 42 mothers and 2 fathers, with education levels ranging from high school diploma (n = 2), some college or an associate's degree (n = 7), and a bachelor's degree (n = 21), to an advanced degree (n = 14).

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

1.2. Procedures

Transcripts were analysed using CLAN (MacWhinney, 2000). For detailed procedures, refer to Appendix B of Harrington (2025). We first standardised the spelling of lexical items using the FREQ and CHSTRING commands to ensure that CHIP could accurately identify matches among child-like variations of lexical items (e.g., duck and ducky). We used the CHIP program to code for parent and child lexical overlap (Figure 1). Child utterances were coded for spontaneous lexical overlap, which was adapted from the CLAN definition (MacWhinney, 2000) and operationalised as the IMMEDIATE overlap of MEANINGFUL lexical content in a SPONTANEOUS manner. To achieve this, we added unique operators to the CHIP command line. The +q3 operator narrowed CHIP's analysis window from six to three utterances, so only child utterances within two

```
*MOT: puppy dog .
      %mor:  n|puppy n|dog .
      %adu:  $NO_REP
      %chU:  $SOURCE
*CHI: puppy dog .
      %mor:  n|puppy n|dog .
      %chi:  $EXA:puppy-dog $EXACT $DIST = 1
*CHI: bunny rabbit .
      %mor:  n|bunny n|rabbit .
      %chi:  $NO_REP
*MOT: a blue rabbit .
      %mor:  det:art|a n|blue n|rabbit .
      %adu:  $EXA:rabbit $ADD:a-blue $DEL:bunny $DIST = 1
      %chU:  $NON
*CHI: carrots .
      %mor:  n|carrot-PL .
      %chi:  $NO_REP
*MOT: carrots .
      %mor:  n|carrot-PL .
      %adu:  $EXA: carrots $EXACT $DIST = 1
      %chU:  $NON
*MOT: you see carrots ?
      %mor:  pro:per|you v|see n|carrot-PL ?
      %adu:  $EXA:carrots $ADD:you-see $EXPAN $MEXA:-PL $DIST = 2
      %chU:  $NON
*CHI: xxx a waffles .
      %mor:  det:art|a n|waffle-PL .
      %chi:  $NO_REP
*MOT: waffles?
      %mor:  n|waffle-PL ?
      %chi:  $EXA:waffles $DEL:a $REDUC $DIST = 1
      %chU:  $NON
```

The $SOURCE indicates that utterance is linked to a response utterance containing lexical overlap based on CHIP codes.

The $DIST is present on every instance of lexical overlap. It indicates how far a lexical overlap response is away from the source.

The $NON indicates there is no response within the analysis window, or the response does not contain lexical overlap based on CHIP codes.

The $NO_REP indicates there is no overlap present between the source and response

The $EXACT code is an utterance-level code and indicates there are only exact matches between the source and response utterances.

The $EXPAN is an utterance-level code and indicates there are only exact matches and additions between the source and response

The $REDUC is an utterance-level code and indicates there are only exact matches and deletions between the source and response

Figure 1. CHAT transcript excerpt and study-relevant code descriptions. *Note*: This CHAT transcript is annotated to define CHIP and CHIPUTIL codes mentioned throughout this paper. Refer to the CLAN manual (MacWhinney, 2000) for the description of other codes.

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

utterances of the parent were coded. The -h operator excluded conversational devices (e.g., *yeah* and *okay*) and high-frequency function words (e.g., *the*, *is*, and *you*), an approach adapted from Sokolov and Moreton (1994) to direct CHIP to ignore less meaningful instances of child lexical overlap (e.g., CHIP ignored the overlap of *yeah* in M: Yeah; C: Yeah). Child lexical overlap codes were placed on a %chi tier. After running CHIP, we used the FREQ command to search for the words *say* and *tell* on the parent speaker tier to find instances where parents explicitly prompted for lexical overlap (e.g., *Say dog*) and the children responded with prompted lexical overlap (e.g., *Dog*). We identified 156 parent prompts (n = 30 parents) to which children (n = 15) responded 44 times with prompted lexical overlap. The %chi tiers were removed from these 44 utterances. Thus, the remaining CHIP codes reflected spontaneous lexical overlap.

We ran the CHIP program again to code for parent imitations and expansions. A parent imitation was defined as an exact replication of all or some of the child's words, whereas a parent expansion was defined as an exact replication of all the child's words with new additions (MacWhinney, 2000). Again, the +q3 operator was added to the base CHIP command to set the CHIP analysis window to three utterances, and the -h operator was added to direct CHIP to ignore conversational devices. Unlike the child CHIP coding pass, high-frequency function words were not excluded when coding parent imitations and expansions because adding function words to make a child's utterance more grammatically complete is a recognised form of expansion (Baker & Nelson, 1984). Parent lexical overlap codes were placed on an %adu tier.

The first author verified the accuracy of CHIP coding by checking for errors of commission (i.e., CHIP coded overlap in error) and errors of omission (i.e., CHIP missed an instance of overlap) across every child and parent utterance in nine randomly selected transcripts (20%). For child lexical overlap, four errors of commission and three errors of omission were found across 1,144 child utterances (99.4% accuracy). For parent imitation and expansion, two errors of omission were found across 3,248 parent utterances (99.9% accuracy).

Finally, we used the CHIPUTIL program to link parent source utterances with child responses based on the existing CHIP codes in the transcript (Figure 1). A $SOURCE code was placed on a %chU tier of the parent utterance if the child response had a CHIP overlap code on the %chi tier. A $NON code was placed on the %chU tier of the parent utterance if there was no child utterance within the next two utterances or if the child response did not have a CHIP overlap code on the %chi tier of the child response. The +s operator extracted all pairs of parent $SOURCE utterances and child responses into a new file. The -s operator extracted all parent $NON utterances and any corresponding child responses into a new file. This allowed measures to be automatically computed on different subsets of parent utterances: those followed by child lexical overlap responses ($SOURCE) and those that were not ($NON).

1.3. Variables

We used the MLU and FREQ commands to compute the mean length of utterance in morphemes, the number of different words, and the frequency of communication acts and analysis set utterances. Communication acts were defined as any verbal or nonverbal utterance in the transcript, including fully unintelligible utterances. Analysis set utterances were a subset of communication acts, excluding utterances that were (a) nonverbal, (b) fully unintelligible, (c) made up exclusively of words on the conversational device and

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

high-frequency function word lists, and (d) prompts for lexical overlap (for parents) or responses to these prompts (for children).

The FREQ command was used to compute the frequency of child spontaneous lexical overlap, parent imitation, and parent expansion. Refer to the CHIP section in the CLAN manual for the definitions of the codes (MacWhinney, 2000). Child spontaneous lexical overlap was computed by summing the $DIST codes on the %chi tier, which was chosen because it is present on every instance of lexical overlap, regardless of type. The frequency of parent imitations was computed by totalling the $EXACT and $REDUC codes on the % adu tier, and the frequency of parent expansions was computed by totalling the $EXPAN codes on the %adu tier. Parent imitation and expansion variables were combined for a summary measure of parent imitation and expansion.

To compute sequential variables, FREQ summed the frequency of parent imitations and expansions in the $SOURCE transcripts and the $NON transcripts, separately. This resulted in four variables reflecting sequential parent→child turns:

(a) parent imitation→child spontaneous lexical overlap,
(b) parent imitation→no child spontaneous lexical overlap,
(c) parent expansion→child spontaneous lexical overlap,
(d) parent expansion→no child spontaneous lexical overlap.

We computed the frequency of all other parent utterances by subtracting the frequency of parent imitations and expansions from the frequency of parent analysis set utterances in the $SOURCE and $NON transcripts, separately. This resulted in the final two sequential variables:

(a) frequency of other parent utterance→child spontaneous lexical overlap,
(b) frequency of other parent utterance→no child spontaneous lexical overlap.

Together, these six variables characterised every parent analysis set utterance by the type of utterance it was (i.e., parent imitation, parent expansion, and other) and whether children responded to it with spontaneous lexical overlap or not (i.e., $SOURCE or $NON).

1.4. Data analysis

Descriptive statistics were used to summarise children's spontaneous lexical overlap, parent imitation, and expansion. We analysed the association between the summary measures of child spontaneous lexical overlap and parent imitation and expansion using a bivariate Spearman correlation, which was appropriate for the non-normally distributed data.

Next, we restructured the dataset from a participant-level format to a long-form utterance-level format using the *dplyr* (Wickham et al., 2023) and *tidyr* (Wickham et al., 2024) packages in R. This dataset included 17,884 rows, each reflecting one parent analysis set utterance. Columns represented categorical variables of participant ID, parent utterance type, and child lexical overlap. Each parent contributed multiple utterances. This nested data structure justified the use of mixed effects logistic regression (MELR) model, which was fit using the *lme4* package (Bates et al., 2015) in R via maximum likelihood estimation with Laplace approximation. The binary outcome variable was the presence or absence of spontaneous lexical overlap in the child's response. Parent

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

utterance type was modelled as the fixed effect with three levels: parent imitation, parent expansion, and other. Parent "other" utterances were set as the reference level. Participant ID was modelled as the random effect to account for the nested data structure and subject-level variability. An analysis of variance test was used to compare the Akaike information criterion between the base model (with only the random effect) and the full model (with both random and fixed effects) and evaluate the final model fit. Odds ratios were computed from the fixed effects coefficient estimates.

2. Results

Table 1 summarises descriptive data for summary measures of parent and child lexical overlap and related child language variables. Our first research question examined the association between summary measures of child's spontaneous lexical overlap and combined parent imitations and expansions. A bivariate Spearman correlation revealed a moderately strong and positive association ($r(42) = 0.656$, $p < .001$).

Our second research question examined the sequential association between parent utterance type and child spontaneous lexical overlap. Table 2 collapses data across all participants to display the frequency of child spontaneous lexical overlap responses by parent utterance type. Two MELR models were fit to test whether parent imitation and expansion increased the likelihood of child lexical overlap. Relative to a base model with only Participant ID as a random effect, adding parent utterance type as a fixed effect significantly improved model fit ($\chi^2(2) = 95.42$, $p < .0001$). Table 3 presents model estimates and fit statistics. Fixed effects indicated that both parent imitation and

Table 1. Descriptives of child and parent variables

Child variables	Mean	Standard deviation	Min	Max
CDI words produced	204.16	142.84	37	621
Mean length of utterance in morphemes	1.42	0.49	1	3.33
Number of different words	54.05	40.19	7	196
Communication acts	250.95	100.77	112	559
Analysis set utterances	110.75	74.35	9	315
Frequency of spontaneous lexical overlap	22.23	17.26	0	66
Proportion of spontaneous lexical overlap[a]	20.10%	12.43%	0	67.24%
Parent variables	**Mean**	**Standard deviation**	**Min**	**Max**
Communication acts	485.50	140.86	220	837
Analysis set utterances	406.18	118.83	181	691
Frequency of imitation + expansion	44.45	27.07	2	110
Proportion of imitation + expansion[b]	11.14%	6.10%	0.49%	25.11%
Imitation	17.59	12.42	1	46
Expansion	26.86	18.13	0	73

Note: CDI = MacArthur Bates Communicative Development Inventory (Fenson et al., 2007).
[a]Computed by dividing spontaneous lexical overlap by analysis set utterances.
[b]Computed by dividing imitation + expansion by analysis set utterances.

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

Table 2. Frequency and proportion of child spontaneous lexical overlap relative to parent utterance type

	Child spontaneous lexical overlap		
Utterance type	Yes	No	Proportion[a]
Other	671	15,257	4.397%
Parent imitation	102	672	15.179%
Parent expansion	132	1,050	12.571%

[a]The number of utterances in the "Yes" column divided by the total number of parent utterances in that row.

Table 3. Model estimates for the effect of parent utterance type on child spontaneous lexical overlap

Fixed effects				
Variable	Estimate	Std. error	*z*-score	*p*-value
(Intercept)	−3.35	0.15	−22.01	<.0001
Parent imitation	0.94	0.12	7.99	<.0001
Parent expansion	0.80	0.10	7.70	<.0001
Random effect				
Variable	Variance	SD		
Participant ID	0.90	0.95		
Model fit statistics: AIC = 6,588; Deviance = 6,580, Residual degrees of freedom = 17,880				

Note: AIC = Akaike information criterion; SD = standard deviation.

expansion significantly increased the log-odds of child lexical overlap compared to the reference category of other parent utterances (all $p < .0001$). Odds ratios indicated that child spontaneous lexical overlap was more than twice as likely to follow parent imitation (OR = 2.55, 95% CI [2.05, 3.18]) and parent expansion (OR = 2.23, 95% CI [1.82, 2.74]) than other parent utterance types.

3. Discussion

This study piloted the CHIPUTIL program, demonstrating its utility in efficiently analysing spontaneous lexical overlap in a sequential manner. This novel automated approach allowed us to include a larger sample of dyads and longer language samples compared to previous sequential analyses of lexical overlap (Clark & Bernicot, 2008; Farrar, 1987; Masur & Olson, 2008; Scherer & Olswang, 1984). Moreover, by conducting analyses with both summary and sequential measures, we revealed the added value of sequential analyses. Consistent with prior findings (Casla et al., 2022; Masur, 1987; Masur & Eichorst, 2002; Masur & Rodemaker, 1999), we observed a positive, moderately strong association between summary measures of child lexical overlap and parent imitations and expansions. Our sequential findings built on this, revealing that child lexical overlap was twice as likely to follow a parent imitation or expansion than other parent utterance types.

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

These results suggest that the association observed between parent and child summary measures may be partially driven by sequences of parent–child lexical overlap.

Parent imitations and expansions appeared to invite children to participate in multi-turn exchanges. Although our analysis focused on the association between a parent utterance and an immediate child response, these sequences involved three turns: a child utterance, a contingent parent imitation or expansion, and a child response containing spontaneous lexical overlap. These sequences reflected moments of cohesive discourse, where children and parents built on each other's contributions in semantically related ways (Pickering & Garrod, 2004). Parent imitations and expansions helped maintain an established topic and modelled lexical overlap as a natural and meaningful way to sustain conversation (Clark & Bernicot, 2008; Kirchner & Prutting, 1987; Pickering & Garrod, 2004; Snow, 1981; Taumoepeau, 2016; Užgiris et al., 1989). In turn, children repurposed their parents' words in their response. For these children transitioning from single words to word combinations, it is possible these sequences of parent–child lexical overlap allowed them to communicate messages they would not have been able to generate independently (Kirchner & Prutting, 1987; Snow, 1981). In this way, lexical overlap may play a key role in scaffolding early conversations that resemble the multi-turn structure and cohesion typical of adult–adult discourse. Not only does this offer opportunities for children to actively participate in and sustain conversations early in language development, but it also creates conditions well suited for language learning (Casla et al., 2022; Che et al., 2018; Fusaroli et al., 2023; Hirsh-Pasek et al., 2015; Levickis et al., 2014; Masek et al., 2021; Tamis-LeMonda et al., 2001).

The findings point to promising future directions for understanding children's language development. Grounded in the transactional model of development (Camarata & Yoder, 2002; Sameroff & Chandler, 1975), this work highlights the value of studying proximal, real-time transactions to inform the mechanisms underlying development. Our findings suggest that parent imitations and expansions promote child lexical overlap, and together, these sequences of parent–child lexical overlap create contingent, multi-turn exchanges. Similarly, an extensive body of research has demonstrated that parent input that is temporally adjacent, semantically related, and embedded in multi-turn interactions is associated with children's language development (Hirsh-Pasek et al., 2015; Levickis et al., 2014; Masek et al., 2021; Tamis-LeMonda et al., 2001). Taken together, it seems possible that child lexical overlap may serve as an entry point into multi-turn conversations that set in motion the kinds of input and interactions that contribute to later language development. Our future work will explore how these parent–child lexical overlap sequences evolve over time, how early lexical overlap contributes to broader measures of multi-turn, contingent exchanges, and whether this transactional pathway explains language growth over time. CHIPUTIL will allow for efficient replication of sequential analyses, offering a feasible approach to test these hypotheses.

We see promise in the CHIPUTIL program for the broader child language community. To the best of our knowledge, this is the first study to publish on this novel automated coding approach. A major benefit of automated programs is the ability to replicate work across labs in a scalable way. Moreover, CHIPUTIL is flexible in what it allows researchers to analyse. While our work is focused on parent and child lexical overlap, CHIPUTIL allows researchers to explore other features as well. For example, researchers might be interested in understanding whether child lexical overlap is dependent on parent utterance length. This could be easily adapted from our methodology, replacing measures of parent imitation and expansion with a measure of parent utterance length. In this way, CHIPUTIL is a broadly applicable resource for the field.

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

While this study holds promise for future work, several limitations should be considered. The sample lacked racial, ethnic, and parental educational diversity, limiting generalisability. In addition, all children spoke General American English, which is common across the lexical overlap literature. This further motivates the need for efficient and replicable methodologies to extend work on lexical overlap to more diverse samples. While the utterance-level sequential analysis enabled fine-grained analysis of conversational sequences, we recognise our model was simple and did not include specific participant-level predictors. Future work should examine how parent and child characteristics influence sequential patterns. Moreover, this analysis was restricted to three-turn sequences. In future work, we plan to explore automated approaches that characterise longer sequences. Finally, our focus on a single time point limits our ability to speak directly to later language outcomes. However, we believe that the present findings are promising, offering new directions to understand how lexical overlap may contribute to child language development and an efficient and replicable approach to pursue this work.

Acknowledgements. The CHIPUTIL program was created by Brian MacWhinney and Leonid Spektor at the request of E.K.H. We thank them for their efforts.

Disclosure of the use of AI. The authors wrote all original text. ChatGPT was used to provide editorial suggestions to improve clarity and reduce word count forthis Brief Report. The first author reviewed suggested changes, revising some and rejecting others, and the second and third authors reviewed and edited the final draft. The authors take full responsibility for the content of this publication.

Funding statement. Data collection for the archival data used in this study was supported by National Science Foundation Grant BCS-08-22513, awarded to M.R. and P.A.H. Secondary analysis was partially supported by an Illinois Distinguished Fellowship awarded to E.K.H. by the Graduate College at the University of Illinois Urbana–Champaign.

Competing interests. The authors declare no competing interests.

References

Baker, N. D., & **Nelson, K. E.** (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. *First Language*, **5**(13), 3–21. https://doi.org/10.1177/014272378400501301

Bates, D., **Mächler, 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

Bruner, J. (1982). The formats of language acquisition. *American Journal of Semiotics*, **1**(3), 1–16.

Camarata, S., & **Yoder, P.** (2002). Language transactions during development and intervention: Theoretical implications for developmental neuroscience. *International Journal of Developmental Neuroscience*, **20**(3–5), 459–465. https://doi.org/10.1016/S0736-5748(02)00044-8

Casla, M., **Méndez-Cabezas, C.**, **Montero, I.**, **Murillo, E.**, **Nieva, S.**, & **Rodríguez, J.** (2022). Spontaneous verbal repetition in toddler–adult conversations: A longitudinal study with Spanish-speaking two-year-olds. *Journal of Child Language*, **49**(2), 266–301. https://doi.org/10.1017/S0305000921000015

Che, E. S., & **Brooks, P. J.** (2021). Maternal repetition and expansion of child utterances at the outset of combinatorial speech promote growth in MLU. In D. Dionne & L. V. Covas (Eds.), *Proceedings of the 45th annual Boston University conference on language development* (Vol. **45**(1), pp. 120–133). Cascadilla Press. http://www.lingref.com/bucld/45/BUCLD45-10.pdf

Che, E. S., **Brooks, P. J.**, **Alarcon, M. F.**, **Yannaco, F. D.**, & **Donnelly, S.** (2018). Assessing the impact of conversational overlap in content on child language growth. *Journal of Child Language*, **45**(1), 72–96. https://doi.org/10.1017/S0305000917000083

Chieng, A. C. J., **Wynn, C. J.**, **Wong, T. P.**, **Barrett, T. S.**, & **Borrie, S. A.** (2024). Lexical alignment is pervasive across contexts in non-weird adult–child interactions. *Cognitive Science*, **48**(3), Article e13417. https://doi.org/10.1111/cogs.13417

Clark, E. V. (2015). Common ground. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 328–353). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118346136.ch15

Clark, E. V., & **Bernicot, J.** (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of Child Language*, **35**(2), 349–371. https://doi.org/10.1017/S0305000907008537

Farrar, M. J. (1987). *Adult recasts as conditions for imitation in children* [Paper presentation]. The Society for Research in Child Development, Baltimore, MD, United States.

Fenson, L., **Marchman, V. A.**, **Thal, D. J.**, **Dale, P. S.**, **Reznick, J. S.**, & **Bates, E.** (2007). *MacArthur–Bates communicative developmental inventories* (2nd ed.). Singular.

Fusaroli, R., **Weed, E.**, **Rocca, R.**, **Fein, D.**, & **Naigles, L.** (2023). Caregiver linguistic alignment to autistic and typically developing children: A natural language processing approach illuminates the interactive components of language development. *Cognition*, **236**, Article e105422. https://doi.org/10.1016/j.cognition.2023.105422

Hadley, P. A., **Rispoli, M.**, **Holt, J. K.**, **Fitzgerald, C.**, & **Bahnsen, A.** (2014). Growth of finiteness in the third year of life: Replication and predictive validity. *Journal of Speech, Language, and Hearing Research*, **57**(3), 887–900. https://doi.org/10.1044/2013_JSLHR-L-13-0008

Harrington, E. K. (2025). *Reciprocal influences of caregiver language facilitation strategies and child characteristics on caregiver–child interactions* [Doctoral dissertation, University of Illinois Urbana–Champaign]. To be published in Illinois Digital Environment for Access to Learning and Scholarship.

Hirsh-Pasek, K., **Adamson, L. B.**, **Bakeman, R.**, **Owen, M. T.**, **Golinkoff, R. M.**, **Pace, A.**, **Yust, P. K. S.**, & **Suma, K.** (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, **26**(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Kirchner, D. M., & **Prutting, C. A.** (1987). Spontaneous verbal repetition: A performance-based strategy for language acquisition. *Clinical Linguistics & Phonetics*, **1**(2), 147–169. https://doi.org/10.3109/02699208708985009

Levickis, P., **Reilly, S.**, **Girolametto, L.**, **Ukoumunne, O. C.**, & **Wake, M.** (2014). Maternal behaviors promoting language acquisition in slow-to-talk toddlers: Prospective community-based study. *Behavioral Pediatrics*, **35**(4), 274–281. https://doi.org/10.1097/DBP.0000000000000056

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315805672

Masek, L. R., **McMillan, B. T. M.**, **Paterson, S. J.**, **Tamis-LeMonda, C. S.**, **Golinkoff, R. M.**, & **Hirsh-Pasek, K.** (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, **60**, Article 100961. https://doi.org/10.1016/j.dr.2021.100961

Masur, E. F. (1987). Imitative interchanges in a social context: Mother–infant matching behavior at the beginning of the second year. *Merrill-Palmer Quarterly*, **33**(4), 453–472. https://www.jstor.org/stable/23086404

Masur, E. F., & **Eichorst, D. L.** (2002). Infants' spontaneous imitation of novel versus familiar words: Relations to observational and maternal report measures of their lexicons. *Merrill-Palmer Quarterly*, **48**(4), 405–426. https://www.jstor.org/stable/23093778

Masur, E. F., & **Olson, J.** (2008). Mothers' and infants' responses to their partners' spontaneous action and vocal/verbal imitation. *Infant Behavior and Development*, **31**(4), 704–715. https://doi.org/10.1016/j.infbeh.2008.04.005

Masur, E. F., & **Rodemaker, J. E.** (1999). Mothers' and infants' spontaneous vocal, verbal, and action imitation during the second year. *Merrill-Palmer Quarterly*, **45**(3), 392–412. https://www.jstor.org/stable/23092579

Olson, J., & **Masur, E. F.** (2012). Mothers respond differently to infants' familiar versus non-familiar verbal imitations. *Journal of Child Language*, **39**(4), 731–752. https://doi.org/10.1017/S0305000911000262

Pickering, M. J., & **Garrod, S.** (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, **27**(2), 169–225. https://doi.org/10.1017/S0140525X04000056

Rispoli, M., & **Hadley, P. A.** (n.d.) *CHILDES English Champaign corpus* [Data set]. TalkBank. https://www.doi.org/10.21415/RE7A-2P42

https://doi.org/10.1017/S0305000925100184 Published online by Cambridge University Press

Sameroff, A., & **Chandler, M.** (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz, S. Hetherington, S. Scarr-Salapatek, & G. Siegal (Eds.), *Review of child development research* (Vol. **4**, pp. 187–244). University of Chicago Press.

Scherer, N. J., & **Olswang, L. B.** (1984). Role of mothers' expansions in stimulating children's language production. *Journal of Speech, Language, and Hearing Research*, **27**(3), 387–396. https://doi.org/10.1044/jshr.2703.387

Snow, C. E. (1981). The uses of imitation. *Journal of Child Language*, **8**(1), 205–212. https://doi.org/10.1017/S0305000900003111

Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental Psychology*, **29**(6), 1008–1023. https://doi.org/10.1037/0012-1649.29.6.1008

Sokolov, J. L., & **MacWhinney, B.** (1990). The CHIP framework: Automatic coding and analysis of parent–child conversational interaction. *Behavior Research Methods, Instruments, & Computers*, **22**(2), 151–161. https://doi.org/10.3758/BF03203138

Sokolov, J. L., & **Moreton, J.** (1994). Individual differences in linguistic imitativeness. In J. L. Sokolov & C. E. Snow (Eds.), *Handbook of research in language evelopment using CHILDES* (pp. 174–209). Lawrence Erlbaum Associates.

Soler, M. C., **Murillo, E.**, **Nieva, S.**, **Rodríguez, J.**, **Mendez-Cabezas, C.**, & **Rujas, I.** (2023). Verbal and more: Multimodality in adults' and toddlers' spontaneous repetitions. *Language Learning and Development*, **19**(1), 16–33. https://doi.org/10.1080/15475441.2021.2008939

Speidel, G. E., & **Nelson, K. E.** (1989). A fresh look at imitation in language learning. In G. E. Speidel & K. E. Nelson (Eds.), *The many faces of imitation in language learning* (Vol. **24**, pp. 1–22). Springer Science + Business Media, LLC. https://doi.org/10.1007/978-1-4612-1011-5

Tamis-LeMonda, C. S., **Bornstein, M. H.**, & **Baumwell, L.** (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, **72**(3), 748–767. https://doi.org/10.1111/1467-8624.00313

Taumoepeau, M. (2016). Maternal expansions of child language relate to growth in children's vocabulary. *Language Learning and Development*, **12**(4), 429–446. https://doi.org/10.1080/15475441.2016.1158112

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Routledge.

Užgiris, I. C., **Broome, S.**, & **Kruper, J. C.** (1989). Imitation in mother–child conversations: A focus on the mother. In G. E. Speidel & K. E. Nelson (Eds.), *The many faces of imitation in language learning* (Vol. **24**, pp. 91–120). Springer Science + Business Media, LLC. https://doi.org/10.1007/978-1-4612-1011-5

Wickham, H., **François, R.**, **Henry, L.**, **Müller, K.**, & **Vaughan, D.** (2023). *dplyr: A grammar of data manipulation* (Version 1.1.4) [Computer software]. https://CRAN.R-project.org/package=dplyr

Wickham, H., **Vaughan, D.**, & **Girlich, M.** (2024). *tidyr: Tidy messy data* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tidyr

Cite this article: Harrington, E.K., Hadley, P.A., & Rispoli, M. (2025). Spontaneous Lexical Overlap in Early Conversations: Automated Sequential Coding of Parents and Toddlers. *Journal of Child Language* 1–12, https://doi.org/10.1017/S0305000925100184