Title
Analysing Communicative Intent Coordination in Child-Caregiver Interactions

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Agrawal, Abhishek
Favre, Benoit
Fourtassi, Abdellah

Publication Date
2024

Peer reviewed

Communicative Intent Coordination in Child-Caregiver Interactions

Abhishek Agrawal (abhishek-amit.agrawal@univ-amu.fr)
Benoit Favre (benoit.favre@univ-amu.fr)
Abdellah Fourtassi (abdellah.fourtassi@univ-amu.fr)

Aix-Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France

Abstract

Social interaction plays a key role in children's development of language structure and use. In particular, children must successfully navigate the complex task of coordinating their communicative intents with people around them in early conversations. This study leveraged advanced NLP techniques to analyze a large corpus of child-caregiver conversations in the wild, combining methods for communicative intent inference and for turn contingency evaluation. Key findings include the prevalence of classic adjacency pairs like question-response; caregivers initiated the overwhelming majority of these sequences. We also document new developmental shifts in intent expression and an interesting dissociation between frequency vs. well-coordinated use across the early years of development. This framework offers a new approach to studying language development in its naturalistic, social context.

Keywords: communicative intent; adjacency pairs; contingency; social coordination; language acquisition; large-scale investigation

Introduction

Breakthroughs in language acquisition can emerge from investigating children's early social environment (Snow, 1972). This investigation has recently been made easier thanks to technological advancements in recording and the advent of multi-lab collaborations (MacWhinney, 2000; Roy, Frank, DeCamp, Miller, & Roy, 2015; Bergelson et al., 2023). These developments facilitate large-scale, dense, and diverse data collection, leading to more generalizable and robust findings.

Nevertheless, progress in large-scale data collection is of limited impact if not accompanied by a similar effort to develop computational tools that allow effective and automatic information extraction and analysis. While several methods exist that enable the processing of children's overall linguistic input (e.g., the quantity of speech heard), we have fewer options when navigating the early *interactive* context. This aspect is crucial, however, as learning is deeply rooted in social interaction (Tomasello, 2003; Kuhl, 2007; Clark, 2018; Song, Spier, & Tamis-Lemonda, 2014; Bruner, 1983; Nelson, 2007; Nikolaus & Fourtassi, 2021, 2023).

Theories in language development (especially the ones that fall under the social-pragmatic account) highlight the fact that language acquisition occurs in a communicative context where children use words to express intents and interpret others' words as intentional (Tomasello, 2001). Thus, a comprehensive understanding of language learning in natural contexts depends on our ability to characterize child-caregiver coordination of communicative intents. Such characterization has been notoriously elusive, given that intent is not directly observed; it can only be inferred, and that the context from which this inference can be made is high-dimensional, involving – among other things – the interlocutor's communicative sequence, their shared knowledge, and the environment in which communication occurs.

While the full characterization of intent coordination in natural contexts may still be unachievable, many aspects can be studied, especially the ones that can be inferred from the structure and content of conversations. We will address this aspect of the problem in the current work. In the remainder of the introduction, we briefly summarize how intent coordination has been studied from a developmental point of view in early child-caregiver conversations. Second, we argue that advancements in Natural Language Processing (NLP) show promise for automatizing such studies. We end the introduction by specifying the contribution of the current study, using NLP methods to investigate early child-caregiver coordination at scale.

Intent coordination in early child-caregiver conversations

Developmental researchers have investigated the nature of children's communicative intents in natural interactions with caregivers. Some have focused on the emergence of broad distinctions (e.g., declarative vs. imperative intents) in the pre-verbal phase via gestures (e.g., Dore, 1973; Bates, Camaioni, & Volterra, 1975). Others have investigated more specific, finer-grained categories such as "disagree with proposition," "promise," and "justify request," which develop in tandem with children's expressive language (Ninio, Snow, Pan, & Rollins, 1994; Snow, Pan, Imbens-Bailey, & Herman, 1996; Bergey, Marshall, DeDeo, & Yurovsky, 2022; Nikolaus, Maes, Auguste, Prévot, & Fourtassi, 2022).

By identifying children's communicative intent "inventory", this literature provides one crucial building block. Nevertheless, since our goal is to characterize coordination, we need to *also* investigate children's intent expression in relation to the interlocutor in the context of ongoing communication. Theories in Conversation Analysis have identified several common sequences in natural dialog (known as adjacency pairs) that can help us detect coordination at the broad intentional level, such as question-response, request-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

acceptance/refusal, and greeting-greeting (Schegloff, 1986). They are especially helpful in our case given that some of these sequences, e.g., question-response, are commonplace in early child-caregiver conversations, and they have been used as a window into conversational development in general (Stivers, Sidnell, & Bergen, 2018; Chouinard, Harris, & Maratsos, 2007; Peirolo, Xu, & Fourtassi, 2024).

That said, adjacency pairs are not enough to fully characterize coordination. For instance, many categories of communicative intents (as documented in the developmental literature) do not fall within a clear adjacency pair, such as "make a statement," "express a wish," and "promise" (Ninio et al., 1994). For an accurate analysis, it is important to delve into the verbal context to determine whether the child's intent is well-coordinated. One way researchers have operationalized this step in a conversation is by evaluating the extent to which the child's *expression* of intent is contingent on the caregiver's previous message and the context of the conversation, e.g., by being on-topic (Bloom, Rocissano, & Hood, 1976; Abbot-Smith, Dockrell, Sturrock, Matthews, & Wilson, 2023).

How NLP can help

Natural Language Processing (NLP) has been providing – increasingly accurate – automated solutions to the two (complementary) labeling problems we mentioned above that are necessary to characterize intent coordination in child-caregiver conversation. In particular, techniques from dialog act labeling (Mezza, Cervone, Stepanov, Tortoreto, & Riccardi, 2018; Kumar, Agarwal, Dasgupta, Joshi, & Kumar, 2017; Nikolaus et al., 2022) can help classify child and caregiver's utterances in terms of the communicative intent they express. Further, techniques from dialog evaluation (Dziri, Kamalloo, Mathewson, & Zaiane, 2019; Yi et al., 2019; Yeh, Eskenazi, & Mehri, 2021) can help determine whether the child's utterance is contingent on the caregiver's prior utterance (and vice versa). Thus, given a linguistic utterance and the conversational context, we can automatically a) infer what type of intent is being communicated and b) judge whether this intent is compatible with the interlocutor's based on the conversational context. Combined, these two measures provide reasonable automatic measures of intents coordinated in early child-caregiver conversations.

The current study

The main contribution of the study is the analysis of child-caregiver intent coordination at a large scale, thanks to the use of automatic labeling methods.

The paper is organized as follows. First, we introduce the corpora and the NLP models we used to generate communicative intents and response contingency labels. In the Analysis section, we document major patterns in child-caregiver intent coordination, discussing how the results based on manual annotation in a relatively small corpus generalize to results based on automatic annotation in a large corpus.

Methods

Corpus

The main corpus we use for manual annotation (as well as for models' training) is the New England corpus (Snow et al., 1996), publicly available via the CHILDES repository (MacWhinney, 2000). The corpus contains longitudinal recordings of N=52 children aged 14, 20, and 32 months interacting with their caregivers. The corpus contains around 56k utterances (for both children and adults).

Manual annotation

Communicative intents The authors in Snow et al. (1996) labeled each utterance in their corpus for communicative intents, based on INCA-A (Ninio et al., 1994), the most comprehensive coding scheme to date. It contains both easy and challenging intent types (67 categories), allowing the study of development over infancy and preschool.

Response contingency In a previous study (Agrawal, Nikolaus, Favre, & Fourtassi, 2024), we manually annotated a subset of the New England corpus for response contingency. For each change of turns (hereafter turn-switch), we annotated if the response was on-topic, given the conversational context. We used a simple 3-point scale, where a response was labeled as contingent, non-contingent, or uninterpretable. Data from 20 and 32 months were annotated, but not from 14 months (as most utterances at this young age were made of unintelligible speech). Around 4k turn-switches were annotated (out of a total of around 13k in the corpus).

Automatic annotation

Communicative intents The model we use for annotating communicative intents is the one introduced by Nikolaus et al. (2022). The authors used manual annotation of Snow et al. (1996) to train a variety of state-of-the-art models, of which a simple Conditional Random Field (CRF) model proved to be the most effective. It reached 72.33% accuracy, approaching 81% accuracy of human inter-annotation agreement.

We used an identical training procedure to replicate the results by Nikolaus et al. (2022). Then, we used the trained model to automatically annotate all the conversational turn-switches in the English-language CHILDES corpora of children aged 20 to 32 months (excluding data from the New England corpus) (around 345k instances).

Response contingency The model we used for annotating response contingency is the one we introduced in Agrawal et al. (2024). From a variety of modeling techniques, the best-performing approach was fine-tuning a pre-trained Language Model (DeBERTaV3, He, Gao, and Chen (2022)) on the manual annotation. More specifically, the model was fine-tuned to predict the contingency of a turn given the conversational context made of five previous turns. The model reached an F1 score of 74% for children's responses, approaching the 82% F1 score of human inter-annotation agreement. Using the trained model, and similar to what we did with the communicative intent model, we annotated all conversational turn-switches in the English-language CHILDES corpora of children aged 20 to 32 months (around 345k instances).

Analysis

To analyze child-caregiver intent coordination, we can either start with the caregiver's intent and observe how the child responds or vice versa. In both cases, we will analyze, for each major intent in the Initiator, what intent categories are provided by the Responder and how contingent they are. For example, given a question by the caregiver, we observe i) the distribution of the responses given by the child, such as a direct answer, a statement, or another question, and ii) the extent to which each of these attested pairs (i.e., question-answer, question-statement, and question-question) are contingent based on their conversational context.

For a detailed comparison of small-scale manual-based annotation vs. large-scale automatic-based annotation, we first restrict the analysis to data of children aged 32 months old, both in the Responder role (Part 1) and Initiator role (Part 2). Then, in Part 3, we study developmental patterns between 20 and 32 months old.

Part 1: Children as Responders

Figure 1 displays the river plots indicating which categories of intents in the Initiator lead to which categories in the Responder (using manual and automatic data). The thickness of the bands provides a visual representation of the probability with which each response is given. For visual clarity, the figures include only the most frequent categories in the Initiator and Responder.

While the river plots provide the frequency distribution of the attested sequences, the contingencies of each sequence are provided as heat maps in Figure 2 (for data based on manual and automatic annotations).

We identify the following major sequences:[1]:

Question (YQ, QN) → Answer (AA, AN, SA) Question-initiated sequences are, by far, the most common. We distinguish two types: Yes/No-questions (YQ) and Wh-questions (QN).

For Yes/No-questions (YQ), the river plot of manual annotation (the plot on the left in Figure 1) shows that, in most cases, they are followed appropriately with Yes/No-answers (AA or AN). When looking at the corresponding heat map (the plot on the left in Figure 2), we see that almost all Yes/No-answers are contingent. This finding is replicated and generalized – in the automatic data – to other corpora/children (The plots on the right in Figures 1 and 2).

Note, however, that a minority of responses to Yes/No-questions (YQ) are Statements (ST). While this is not the expected category (at least from an adjacency-pair perspective), looking at the heat maps, we can see that not all of these instances (i.e., YQ → ST) are inappropriate when considering the conversational context. In fact, around half of them (60% in manual data and 46% in automatic data) are contingent. Here is an excerpt of a contingent Yes/No question (YQ) → Statement (ST):

- got it?
- too heavy.

And a non-contingent YQ → ST:

-is that good enough?
-he's getting angry too.

As for Wh-questions (QN), according to the river plots (both manual and automatic, Figure 1), the overwhelming majority are followed by a Wh-answer (SA), a sequence that is highly contingent (Figure 2). A minority of Wh-questions (QN) are followed by a Statement (ST) or even by another question (QN). Again, while these responses are not expected from a categorical point of view, the heat maps show that this is not necessarily inappropriate or ill-coordinated. Here is an excerpt of a contingent Wh-question (QN) → Wh-question (QN), where the response is a clarification question:

-what it is?
-what?

But there are also non-contingent QN → QN, as in:

-what color is that?
-where the other crayon?

Request (RQ, RP) → Approval/Refusal (AD/RD) We distinguish two types of requests: The one that takes the form of a Yes/No-suggestion (RQ) (e.g., "should we get another box?") and a more general one that is not restricted by a specific form (RP) (e.g., "let's play house.").

The Yes/No-suggestions (RQ) are mapped correctly to Acceptance (AD) and Refusal (RD).[2] As one would expect, both responses are perfectly contingent according to the heatmaps, in both manual and automatic annotations. Here is an excerpt of Request (RQ) → Approval (AD):

-should we get another box?
-yeah.

The more general form of Request (RP) appears slightly more complex; it leads partly to expected categories from an adjacency-pair point of view (i.e., Approval/Refusal, AD/RD); both are highly contingent according to the heatmaps. Here is an excerpt of RP → AD:

-let's play house.
-okay.

[1] All the example excerpts in the following subsections are taken from the New England corpus.

[2] The Refusal RD is not as frequent as the Acceptance AD and, thus, does not make it to the river plots, but we still show its contingency numbers in the heat maps.

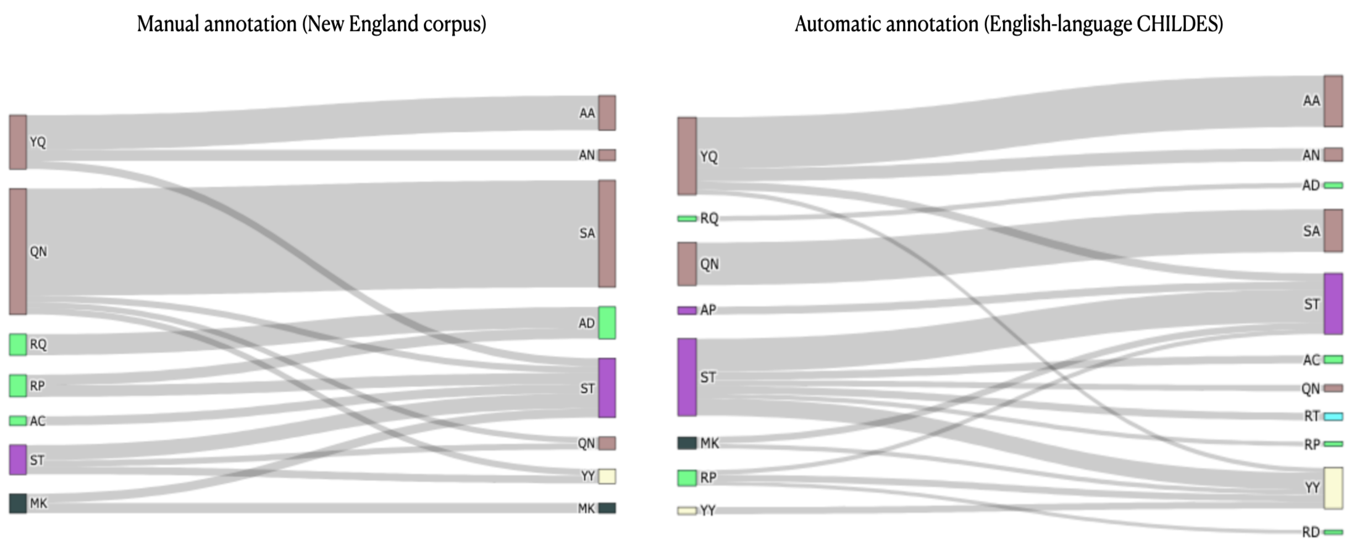

Figure 1: Adjacent pairs for 32-month-old children responding to caregivers. Each plot should be read from left to right: The initiating intents on the left (the parent) and the responding intents on the right (the child). Communicative intents occurring less than 1% of the time were filtered out for a clear representation.

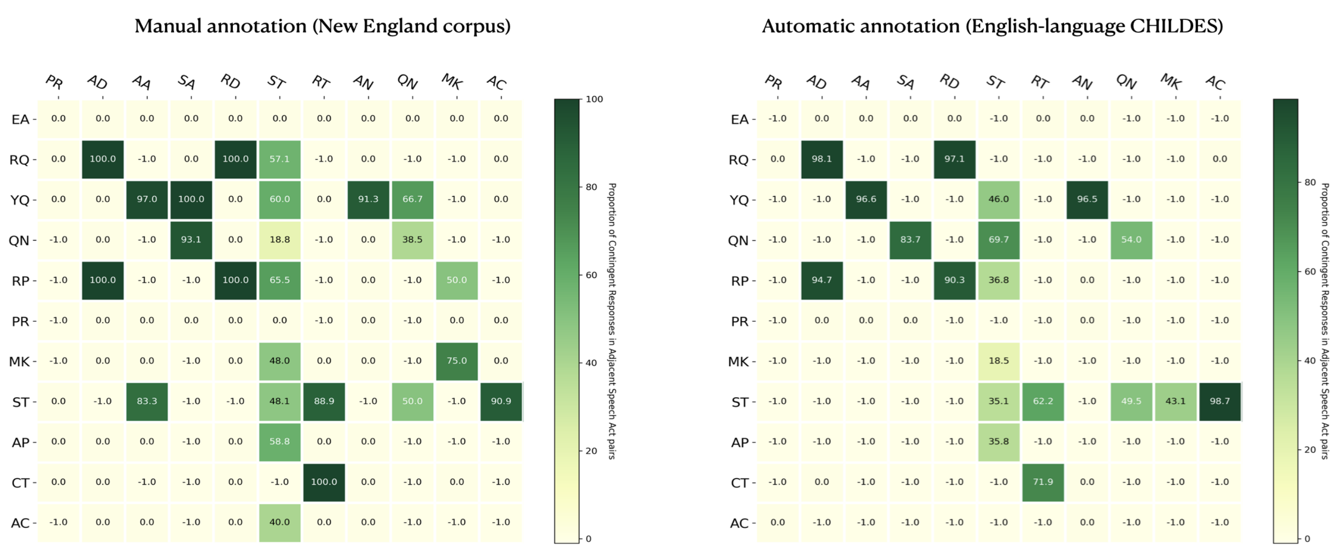

Figure 2: The proportion of contingent responses by the 32-month-old child to the caregiver for an adjacent communicative intent pair. The rows represent the caregiver's communicative intent, and the columns represent the child's. Contingent responses occurring less than approximately 0.4% of the times in frequency are marked as -1.0 in the figure.

Intent label	Short Description	Intent label	Short Description	Intent label	Short Description
EA	Elicit sound	RQ	Yes/No (Suggestion)	YQ	Yes/No (Question)
QN	Wh-question	RP	Request/Suggest	PR	Peform game move
MK	Social norm	ST	Statement	AP	Agree (proposition)
CT	Correct wrong form	AC	Show attentiveness	AD	Agree (act)
AA	Yes (Y/N question)	SA	Wh-answer	RD	Refuse (act)
RT	Imitate/Repeat	AN	No (Y/N question)	YY	Non-sensical utterance

Table 1: A list and short description of all the communicative intent labels displayed in the river plots and heat maps above.

However, this request category (RP) can trigger other response types, mostly Statements (ST). Heatmaps indicate that the contingency of Request (RP) → Statement (ST) sequences should be considered case by case, requiring the investigation of the conversational context. Here is an excerpt of a contingent RP → ST:

-well let's see what's in the next box.
-toys in it.

and a non-contingent RP → ST:

-look at all the furniture.
-other boy.

Statement (ST) → Statement (ST) Statements (ST) are frequent in both the Initiator and the Responder. However, unlike questions and requests, they do not necessarily elicit a specific response. Instead, we see a variety of response types in both river plots. Crucially, because they do not define an adjacency pair, studying the coordinated use of ST necessitates investigating their contingency in the conversational context.

Most Statements (ST) are followed by other Statements (ST) (Figures 1). The heatmaps indicate that their contingency is to be examined on a case-by-case basis. Here is an excerpt of a contingent, well-coordinated ST → ST:

-two beds.
-I don't see see a other bed.

And an excerpt of a non-contingent, ill-coordinated ST → ST:

-I thought you just fell down.
-I want this Cookie Monster.

Social norms (MK) → Social norms (MK) This category includes a variety of social norms, such as greetings and thanks. One interesting sequence is MK → MK, which is generally highly contingent according to the heat map. Excerpt:

-Bye.
-good night.

Part 2: Children as Initiators

When we look at the data from the perspective of children as Initiators and caregivers as Responders, we see a different picture. Figure 3 shows the river plots, both manual (left) and automatic (right) for caregivers responding to 32 months old children. The major sequences do not look like standard adjacency pairs (unlike what we observed when caregivers were the Initiators). Only a few sequences start with genuine elicitation, like questions or requests. Here is an excerpt of a Wh-question (QN) → Wh-answer(SA):

-why?
-Because she's hungry.

Most sequences exist only because caregivers follow up on the child's previous answers and statements, especially using follow-up questions. Here is an excerpt of a Wh-answer (SA) → Wh-question(QA):

-this one.
-who's that one?

Other responses are caregivers agreeing with or acknowledging the children's answers to earlier questions. An excerpt of a Wh-answer (SA) → Agreeing (AP):

-I have Cookie Monster.
-you sure do.

As for the contingency, it is no surprise that most caregivers' responses and follow-ups are highly contingent regardless of the sequence.[3] All these patterns were observed in both manual and automatic data.

Part 3: Developmental patterns

In this section of the results, we showcase another important way automatic annotation is helpful: The *dense* study of developmental trajectories.

Indeed, as mentioned in the Corpus sub-section, the New England corpus has usable data from children aged 20 and 32 months. In all analyses above, we used automation to replicate findings with other children of the *same* age (focusing on 32 months); here, we use automation to evaluate children's coordination at all months between 20 and 32 in CHILDES,[4] providing a denser estimation of the developmental trajectories. Such a study enables a richer, more continuous understanding of development than comparing only two data points: 20 and 32 months. The latter is more likely to be confounded by noise and idiosyncratic variability.

Figure 4 shows children's development in terms of the communicative intents used in the Responder role (Figure 1), focusing on the top five most frequent communicative intents. Manual and automatic annotations show convergent evidence regarding the frequency increase in children's use of two communicative intents: Yes-answers (to Yes/No questions) and Wh-questions. The first is used in a highly contingent fashion (as one would expect). However, Wh-questions start low on contingency and witness only minor improvement over the developmental period under study. Further, manual and automatic annotations agree that Imitations and Statements slightly decrease in relative frequency, although both of these communicative intents continue improving in terms of contingency. Finally, Wh-answers is the only category that shows some differences, increasing (in relative frequency) in manual and (slightly) decreasing in automatic annotations.

[3] Data not shown here but provided in the online supplementary details which can be found here: `https://osf.io/q96jk/?view_only=bfd57508b53343cd9c65233890de603a`

[4] Remember that the models are trained on data from both 20 and 32 months old children, so the ages in between are not quite out-of-distribution; see also experiments in Agrawal et al. (2024).

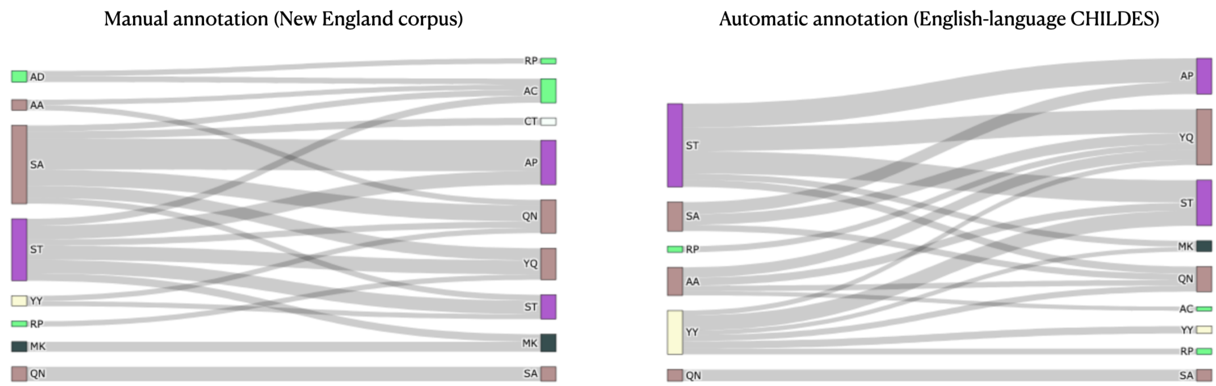

Figure 3: Adjacent pairs for caregivers responding to 32-month-old children. Each plot should be read from left to right: The initiating intents on the left (the child) and the responding intents on the right (the caregiver). Communicative intents occurring less than 1% of the time were filtered out for a clear representation.

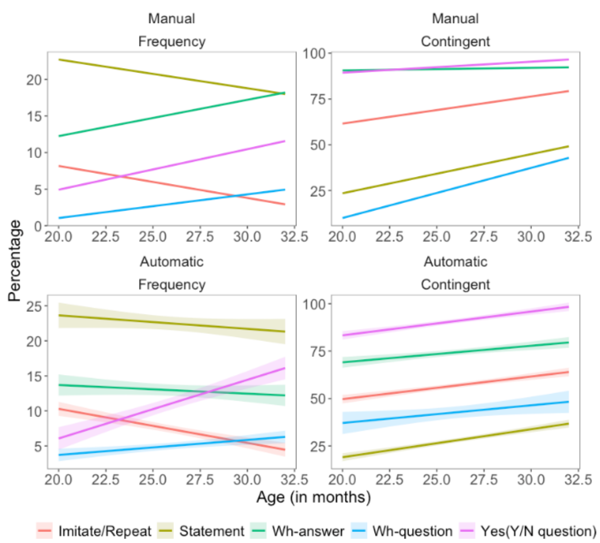

Figure 4: The development of children's top frequent communicative intent categories between 20 and 32 months. We show – side by side – the development of their (relative) frequency and overall contingency (overall contingency is averaged over all caregivers' Initiators). The lines represent linear fits, and the envelopes (for automatic, dense data) indicate the confidence intervals.

Conclusions

This study analyzed intent coordination in early child-caregiver conversations at a large scale. This was enabled by our use of modern NLP techniques to automatically label two key characteristics of intent coordination in conversations: classification of utterances in terms of their communication intent type and judging their level of coordination/contingency based on the conversational context.

This investigation led to several findings. For instance, we documented several robust and early-emergent patterns of intent coordination. The most frequent, we found, fell within classic adjacency pairs such as Question $\rightarrow$ Response and Request $\rightarrow$ Acceptance/Refusal. When investigating the conversational context, these patterns were highly contingent. This is a not a trivial finding: One could imagine situations where children would, e.g., respond to a Wh-question with a Wh-answer that is still off-topic and, therefore, ill-coordinated. This was not the case in the overwhelming majority of cases.

Further, we observed sequences that did not fall within a classic adjacency pair. Most of these sequences tend to involve a declarative statement. The degree of coordination in such cases cannot be predicted based on the intent category. Here, looking at the conversational context is crucial. Indeed, we did not observe any systematic contingency pattern, with many instances being contingent and others being non-contingent.

There were major differences between children and adults, as children initiated much fewer sequences than parents did. This difference, however, may be due to the properties of the specific context of the recorded interactions in CHILDES (Dideriksen, Christiansen, Tylén, Dingemanse, & Fusaroli, 2023; Bodur, Nikolaus, Prévot, & Fourtassi, 2023; Jiang, Frank, Kulkarni, & Fourtassi, 2022).

Development-wise, we found that the frequency of communicative intents used by children undergoes major changes in the second and third years of life. While many of these developmental shifts in terms of frequency have already been reported in previous work (Snow et al., 1996), one new contribution here is testing their robustness at a large scale. More important, we showed that frequency of use cannot be the sole indicator of development: Some of the most frequent intent types (e.g., statements) are also the least coordinated/contingent in the conversational context (Figure 4).

To sum up, the current study provides a new framework to study the development of intent coordination in a naturalistic context. The next step is to investigate how such coordination may vary across cultures and how it predicts language learning in the wild (Misiek & Fourtassi, 2022; Chieng, Wynn, Wong, Barrett, & Borrie, 2024).

Acknowledgements

This work, carried out within the Institute of Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government (France 2030), managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). Furthermore, this study was also supported by the ANR MACOMIC (ANR-21-CE28-0005-01) grant. This work was performed using HPC resources from GENCI–IDRIS (Grant 2022-AD011013886).

References

Abbot-Smith, K., Dockrell, J., Sturrock, A., Matthews, D., & Wilson, C. (2023, December). Topic maintenance in social conversation: What children need to learn and evidence this can be taught. *First Language*, *43*(6), 614–642. (Publisher: SAGE Publications Ltd) doi: 10.1177/01427237231172652

Agrawal, A., Nikolaus, M., Favre, B., & Fourtassi, A. (2024, May). Automatic coding of contingency in child-caregiver conversations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation.* Torino, Italy: European Language Resources Association - International Committee on Computational Linguistics.

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226. (Publisher: Wayne State University Press)

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., R. Hamrick, L., … Cristia, A. (2023, December). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, *120*(52). doi: 10.1073/pnas.2300671120

Bergey, C., Marshall, Z., DeDeo, S., & Yurovsky, D. (2022). Learning communicative acts in children's conversations: A hidden topic markov model analysis of the childes corpora. *Topics in Cognitive Science*, *14*(2), 388–399.

Bloom, L., Rocissano, L., & Hood, L. (1976, October). Adult-child discourse: Developmental interaction between information processing and linguistic knowledge. *Cognitive Psychology*, *8*(4), 521–552. doi: 10.1016/0010-0285(76)90017-7

Bodur, K., Nikolaus, M., Prévot, L., & Fourtassi, A. (2023). Using video calls to study children's conversational development: The case of backchannel signaling. *Frontiers in Computer Science*, *5*. doi: https://doi.org/10.3389/fcomp.2023.1088752

Bruner, J. (1983). *Child's talk: Learning to use language*. W.W. Norton.

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), e13417. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13417` doi: https://doi.org/10.1111/cogs.13417

Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's Questions: A Mechanism for Cognitive Development. *Monographs of the Society for Research in Child Development*, *72*(1), i–129. (Publisher: [Society for Research in Child Development, Wiley])

Clark, E. V. (2018, July). Conversation and Language Acquisition: A Pragmatic Approach. *Language Learning and Development*, *14*(3), 170–185. doi: 10.1080/15475441.2017.1340843

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, *152*(3), 864.

Dore, J. (1973). A Developmental Theory of Speech Act Production*. *Transactions of the New York Academy of Sciences*, *35*(8 Series II), 623–630. doi: 10.1111/j.2164-0947.1973.tb01535.x

Dziri, N., Kamalloo, E., Mathewson, K., & Zaiane, O. (2019, June). Evaluating coherence in dialogue systems using entailment. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 3806–3812). Minneapolis, Minnesota: Association for Computational Linguistics. doi: 10.18653/v1/N19-1381

He, P., Gao, J., & Chen, W. (2022, September). DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing..

Jiang, H., Frank, M. C., Kulkarni, V., & Fourtassi, A. (2022). Exploring patterns of stability and change in caregivers' word usage across early childhood. *Cognitive Science*, *46*(7), e13177.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental science*, *10*(1), 110–120.

Kumar, H., Agarwal, A., Dasgupta, R., Joshi, S., & Kumar, A. (2017). *Dialogue act sequence labeling using hierarchical encoder with crf.*

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs, Vol. 1, 3rd ed.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. (Pages: xi, 366)

Mezza, S., Cervone, A., Stepanov, E., Tortoreto, G., & Riccardi, G. (2018, August). ISO-standard domain-independent dialogue act tagging for conversational agents. In E. M. Bender, L. Derczynski, & P. Isabelle (Eds.), *Proceedings of the 27th international conference on computational linguistics* (pp. 3539–3551). Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Misiek, T., & Fourtassi, A. (2022, August). Caregivers exaggerate their lexical alignment to young children across several cultures. In *Proceedings of the 26th workshop on the semantics and pragmatics of dialogue - full papers.* Retrieved from http://semdial.org/anthology/Z22-Misiek_semdial_0005.pdf

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory*. Harvard University Press.

Nikolaus, M., & Fourtassi, A. (2021). Modeling the interaction between perception-based and production-based learning in children's early acquisition of semantic knowledge. In *Proceedings of the 25th conference on computational natural language learning.* Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.conll-1.31

Nikolaus, M., & Fourtassi, A. (2023). Communicative feedback in language acquisition. *New Ideas in Psychology*, *68*, 100985.

Nikolaus, M., Maes, E., Auguste, J., Prévot, L., & Fourtassi, A. (2022). Large-scale study of speech acts' development in early childhood. *Language Development Research*, *2*(1), 268–304.

Ninio, A., Snow, C. E., Pan, B. A., & Rollins, P. R. (1994, June). Classifying communicative acts in children's interactions. *Journal of Communication Disorders*, *27*(2), 157–187. doi: 10.1016/0021-9924(94)90039-6

Peirolo, M., Xu, Z., & Fourtassi, A. (2024). Development of flexible role-taking in conversations across preschool. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015, October). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. doi: 10.1073/pnas.1419773112

Schegloff, E. A. (1986). The routine as achievement. *Human studies*, *9*(2-3), 111–151.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565. (Place: United Kingdom Publisher: Blackwell Publishing) doi: 10.2307/1127555

Snow, C. E., Pan, B. A., Imbens-Bailey, A., & Herman, J. (1996). Learning How to Say What One Means: A Longitudinal Study of Children's Speech Act Use*. *Social Development*, *5*(1), 56–84. doi: 10.1111/j.1467-9507.1996.tb00072.x

Song, L., Spier, E. T., & Tamis-Lemonda, C. S. (2014, March). Reciprocal influences between maternal language and children's language and cognitive development in low-income families. *Journal of Child Language*, *41*(2), 305–326. doi: 10.1017/S0305000912000700

Stivers, T., Sidnell, J., & Bergen, C. (2018, January). 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

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language Acquisition and Conceptual Development* (pp. 132–158). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620669.007

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press. doi: 10.2307/j.ctv26070v8

Yeh, Y.-T., Eskenazi, M., & Mehri, S. (2021, November). A Comprehensive Assessment of Dialog Evaluation Metrics. In *The First Workshop on Evaluations and Assessments of Neural Conversation Systems* (pp. 15–33). Online: Association for Computational Linguistics. doi: 10.18653/v1/2021.eancs-1.3

Yi, S., Goel, R., Khatri, C., Cervone, A., Chung, T., Hedayatnia, B., ... Hakkani-Tur, D. (2019, October). Towards Coherent and Engaging Spoken Dialog Response Generation Using Automatic Conversation Evaluators. In *Proceedings of the 12th International Conference on Natural Language Generation* (pp. 65–75). Tokyo, Japan: Association for Computational Linguistics. doi: 10.18653/v1/W19-8608