Title

Communicative Feedback as a Mechanism Supporting the Production of Intelligible Speech in Early Childhood

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Nikolaus, Mitja
Prévot, Laurent
Fourtassi, Abdellah

Publication Date

2022

Peer reviewed

Communicative Feedback as a Mechanism Supporting the Production of Intelligible Speech in Early Childhood

Mitja Nikolaus[1,2] (mitja.nikolaus@univ-amu.fr)
Laurent Prévot[2] (laurent.prevot@univ-amu.fr)
Abdellah Fourtassi[1] (abdellah.fourtassi@univ-amu.fr)

[1]Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[2]Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France

Abstract

Children start to communicate and use language in social interactions from the very early stages in development. This allows them to experiment with their current linguistic knowledge and receive valuable feedback from their interlocutors. We conducted a large-scale corpus study to examine the quality of positive and negative Communicative Feedback signals that caregivers provide in terms of time-contingent responses and clarification requests. We found evidence for the effect of such feedback in supporting children's production of intelligible speech.
The broad impact of this paper is to highlight how general social feedback mechanisms that govern human communication can also support child language acquisition.

Introduction

Much of computational research in language acquisition has traditionally focused on investigating learnability from the linguistic input. While such an approach has been insightful about the role of the input in language development, it tends to consider – whether implicitly or explicitly – that children only passively absorb the information they are exposed to. However, children start to actively interact with people around them very early in development and use their growing linguistic knowledge to establish some form of rudimentary communication. This early social interaction also plays a role in the acquisition of language (Bruner, 1985; Tomasello, 2005; Kuhl, 2007; E. V. Clark, 2018).

Currently, the dominant line of research on the role of social interaction focuses on children's ability to make pragmatic inferences about caregiver's communicative intents, taking into account the context of language use, common ground, as well as social cues such as gaze and pointing (Tomasello et al., 2005; Senju & Csibra, 2008; Yurovsky & Frank, 2017; Bohn & Frank, 2019; Tsuji et al., 2020).

Another dimension of social interaction is that it offers opportunities for caregivers to provide *feedback* on children's linguistic productions. Children start communicating long before their linguistic skills are mature (i.e., intelligible and grammatically sound). Their vocalizations start from being non-speech (e.g., crying, laughing) and become increasingly speech-related (e.g., babbling), but still largely unintelligible. Then, their linguistic productions become increasingly intelligible[1] although not always grammatical (e.g., "Want play!"). Finally, children's productions become grammatical/well-formed (e.g., "I want to play!"). As children move from one stage to the next, the range of available social feedback mechanisms and their specificness increases (i.e. their ambiguity decreases, see Figure 1).

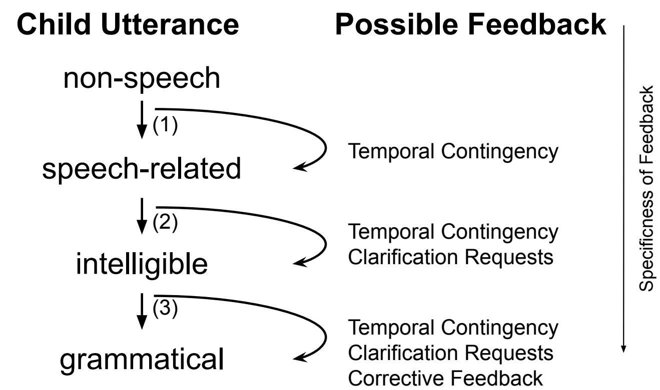

Figure 1: Developmental steps in children's linguistic productions before they reach the final grammatical/well-formed stage. As children move from one stage to the next, the range of available feedback mechanisms and their specificness increases (as the communicative intent of the child becomes increasingly easier to decode). Communicative Feedback (the subject of the current paper) includes contingency and clarification requests, but excludes corrective feedback.

To support the transition from non-speech to speech-related vocalization (Transition (1) in Figure 1), the caregiver can provide feedback in the form of *temporal contingency* (e.g., by responding more often and faster to speech-related than to speech-unrelated vocalization, thus "reinforcing" the former), a mechanism that has been studied for example by Warlaumont et al. (2014).[2] Once children become able to produce speech-related utterances (that can be either intelligible or unintelligible), additional feedback mechanisms become possible, e.g., *clarification requests* (Purver, 2004) can

[1]We define intelligibility by contrast to unintelligible utterances whose *communicative intent* is difficult to infer (e.g., babbling). An utterance is intelligible if a communicate intent can be decoded from it even though it is not necessarily grammatically correct.

[2]This mechanism has also been referred to as "responsiveness". Here we call it temporal contingency in order to distinguish it from other kinds of contingency, namely input-contingency or content-contingency (see also McGillion et al., 2013).

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

be given following unintelligible vocalizations, encouraging the child to produce more intelligible speech (Transition (2) in Figure 1; see e.g., E. V. Clark (2020)). Finally, *corrective feedback* (e.g., a correct reformulation of an incorrect utterance by the child) has been studied more extensively in the development literature (e.g., Brown, 1973; Saxton, 2000; Chouinard & Clark, 2003; Hiller & Fernández, 2016), but this form of feedback can only be provided at the last stage (Transition (3) in Figure 1) where children are capable of producing intelligible utterances. Indeed, if an utterance is unintelligible, the caregiver cannot infer the child's communicative intent and, thus, will not be able to *correct* or reformulate its linguistic expression.

Communicative Feedback Here, we focus on a subset of the social feedback mechanisms reviewed above which we call *Communicative Feedback* (CF). We define CF as the form of social feedback whose goal is to signal *communicative success or failure*, rather than to correct the linguistic *content* of a child's production. Therefore, CF includes rather non-specific mechanisms such as time-contingent responses and clarification requests (e.g., "What?", "Which one?").

We are interested in this communication-focused feedback because it has been understudied compared to content-focused feedback (especially corrective feedback), although it is arguably a better candidate for a universal mechanism. Indeed, contrary to corrective feedback, Communicative Feedback is not specific to language acquisition, as it is a fundamental mechanism for communication in general (see "communicative grounding"; Stalnaker (1978); H. H. Clark (1996)). Further, both communicative repair mechanisms and caregiver's temporal contingency have been observed in a diversity of cultures (Dingemanse et al., 2015; Richman et al., 1992; Bornstein et al., 1992). Finally, CF is the only mechanism that can a priori be used across all stages of child utterance development, as explained in Figure 1.

Communicative Feedback and language acquisition Though CF is more about communication management than about correcting or refining the content of the child's utterance, it can still help with language learning, indirectly, via a reinforcement-like mechanism. CF can be positive or negative, signaling communicative success or failure, respectively. The hypothesis is that the child would seek positive signals and avoid negative signals, motivated by the desire to be understood. Utterances that receive positive feedback are more likely to be correct and will be repeated, whereas utterances that receive negative signals are more likely to be incorrect and will be avoided or revised in future interactions.

Negative CF signals include a) lack of contingency (e.g., the caregiver providing a non-contingent response, or no response at all) and b) clarification requests, which could be verbal or non-verbal (e.g., "What?" or a puzzled face). Positive CF signals include a) high contingency (e.g., fast and on-topic verbal responses, successful shared attention) and b) explicit verbal and non-verbal signs of understanding (backchannel responses, repeating the child's utterance, a cheering face and a head nod, etc.).

We find in the development literature evidence that supports CF as a mechanism for language learning, although most research has investigated only the early stages of utterance development described in Figure 1. For example, it has been shown that caregivers are more responsive to child speech-related utterances than to non-speech utterances and that, critically, children also react differently to positive and negative CF, producing more speech-related utterances following high temporal contingency (Bloom, 1988; Goldstein et al., 2003; Warlaumont et al., 2014).

Very few studies examined CF for later stages of development. E. V. Clark (2020) documented caregiver's use of clarification requests and Gallagher (1977); Saxton et al. (2005) tested how children revise their utterances when they receive such requests. However, though very insightful to our understanding of the phenomenon, this previous work remains incomplete as it has either relied on qualitative/anecdotal reporting or on experimentally controlled settings (as opposed to systematic and quantitative study of natural/spontaneous child-caregiver conversations).

The current study and novelty of our work We conduct a quantitative large-scale corpus study on the role of CF in children's language development. The paper makes two main contributions. First, we test the extent to which previous work by Warlaumont et al. (2014) — on how temporal contingency can help children transition towards speech-related utterances (i.e., Transition (1) in Figure 1) — can be replicated with different datasets of child-caregiver interactions. Second, we investigate how CF (both temporal contingency and clarification requests) can help children transition towards more intelligible utterances (i.e., Transition (2) in Figure 1).

To ensure reproducibility, we make the source code of all analyses publicly available: `https://github.com/mitjanikolaus/childes-communicative-feedback`.

Methods

Unit of analysis: U-R-F

To study CF in child-caregiver conversations, we use as unit of analysis a 3-part micro-structure sequence consisting of 1) child's utterance, 2) caregiver response (or lack thereof), and 3) the child follow-up (following previous work like Warlaumont et al. (2014)).[3] Hereafter, we will call this sequence U-R-F (Utterance, Response, Follow-up). Table 1 shows some examples of U-R-F from the dataset we used.

Data

We used transcribed conversations from an English subset of the CHILDES corpus (MacWhinney, 2014). The subset involves children aged 10 to 48 months[4] for which timing in-

[3] We disregard case where the follow-up occurs more than 60s after the response.

[4] We chose 10 months as a minimum age because at this age children typically start to produce their first intelligible utterances. As

Table 1: Examples of U-R-F sequences taken from the Thomas corpus (Lieven et al., 2009).

Child utt.	Caregiver resp.	Child follow-up
Moon	A big moon.	And a firework.
Uh no big smoke .	<no response>	xxx .
[=! babble]	what?	put in there please.

formation (start and end time of each utterance) is available. We converted the children's ages into equidistant bins of 6 months for plotting and analyses. In total, our data consists of 21 corpora[5] with 1787 transcripts from 326 children. We extracted and analyzed a total of 367,774 U-R-F sequences.

Annotations

For each U-R-F, we annotate the **speech-relatedness** and **intelligibility** of all child utterances and follow-ups, as well as whether there was a caregiver **response** and whether the response was a **clarification request.**

Speech-relatedness All corpora in CHILDES follow the CHAT transcription format (MacWhinney, 2017) which includes so called "paralinguistic events".[6] All utterances that contain at least one transcribed word or one speech-related event were annotated as speech-related, others as non-speech.

Intelligibility We labelled all utterances as either intelligible or unintelligible using a rule-based approach on the transcriptions. Not all corpora transcribed unintelligible speech exactly the same way, so we manually verified which conventions were used in each corpus.

In CHAT, phrases are either explicitly labelled as unintelligible ("xxx") or labelled as phonological fragments (e.g., "&baba", "baba@p"). The latter case is used for "vocalizations that cannot be mapped to words" (and are therefore coded phonetically instead, see MacWhinney, 2017). As they cannot be mapped to words, we assume they are also unintelligible to the interlocutor.[7] Further, there are some event codes that refer to unintelligible utterances and babbling (e.g., "&=vocalize", "&=babble", "baba@b"). Utterances that contain at least one unintelligible word were labelled as unintelligible, all others as intelligible.

Temporal contingency While the contingency of caregiver response behavior can by measured in many ways (McGillion et al., 2013), here we focus on one instantiation of contingency known as temporal contingency. Using timing information available in the transcripts, we annotate for each child's utterance whether the caregiver response is given or not. Following Warlaumont et al. (2014), we considered all cases in which a caregiver's utterance follows the child's utterance within a response latency of one second as **response**. All other cases (a caregiver's utterance that follows with a greater delay, or no utterance at all) are considered as **no response**.[8]

Clarification requests To detect clarification requests, we used a model that was recently developed for automatic annotation of speech acts in child-caregiver conversations (Nikolaus et al., 2021). This model uses the INCA-A coding scheme, which was specifically designed for the study of child-caregiver conversations (Ninio et al., 1994).

All utterances that were labelled as "Eliciting questions (e.g., hmm?)" (`EQ`) or "Requests to repeat utterance" (`RR`) were treated as clarification requests. The most common utterances falling into these categories are open clarification requests, e.g., "what?", "hm?", "what, darling?", "huh?". Less frequently, there are also restricted ones such as "what about backside?" or "some what?".

Analyses

Our analyses are organized into three main parts: 1) replicating work by Warlaumont et al. (2014) on the development of speech-relatedness via temporal contingency, 2) investigating the development of speech intelligibility via temporal contingency, and 3) investigating the development of speech intelligibility via clarification requests.

General developmental trajectories Since we are both replicating work on the development of speech-related speech and investigating the development of intelligible speech, we start our analysis by providing an overview of the developmental trajectories of both phenomena in our dataset (Figure 2). As expected, children's utterances become increasingly speech-related as well as increasingly intelligible. The proportion of speech-related utterances converges clearly before the proportion of intelligible utterances.

Development of Speech-related Vocalizations (Replication of Warlaumont et al. (2014))

Following Warlaumont et al. (2014), we first calculated a measure for caregiver's temporal contingency on child speech-relatedness. This measure was defined as the ratio of caregiver responses to speech-related child utterances subtracted by the ratio of caregiver responses to non-speech utterances. When applied to our dataset, we obtain:

$$\frac{\#(U_{speech} \wedge R_{response})}{\#U_{speech}} - \frac{\#(U_{non-speech} \wedge R_{response})}{\#U_{non-speech}} \approx 0.13 \quad (1)$$

A one-sided t-test indicated that the value was significantly greater than 0 ($SE = 0.008, p < 0.001$), replicating the orig-

a maximum age, we chose 48 months because data in CHILDES becomes sparse after this age.

[5]Namely: Bernstein, Bloom, Braunwald, Brent, Edinburgh, Gleason, MPI-EVA-Manchester, MacWhinney, McCune, McMillan, Nelson, NewmanRatner, Peters, Rollins, Sachs, Snow, Soderstrom, Thomas, Tommerdahl, VanHouten, Weist

[6]Events in CHILDES can be transcribed either as "paralinguistic events" ("[=! crying]"), or "simple events" ("&=crying").

[7]There may however be exceptions in which the utterance remains intelligible, we will return to this case in the discussion.

[8]We also ran all experiments with a more conservative response latency threshold (2 seconds) and this higher threshold did not change the conclusions of the paper.

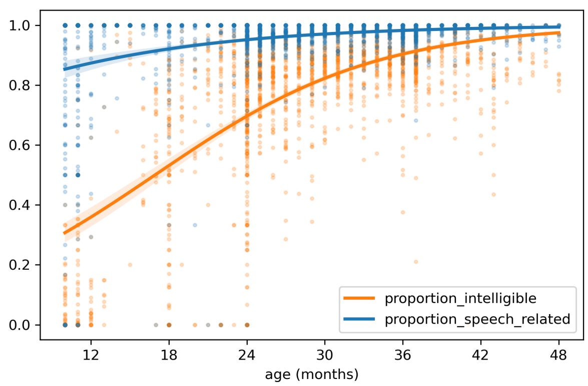

Figure 2: Proportion of speech-related utterances and intelligible utterances. Each data point represents a transcript. The plot shows fitted logistic regression curves and their 95% confidence intervals.

inal results, and confirming that temporal contingency contains useful information for learning speech-related vocalizations.

Second, we calculated a measure for the child's follow-up depending on whether there was a caregiver's response to the child's utterance. This measure was defined (in the previous study) as the ratio of speech-related child follow-ups to speech-related utterances that received a response subtracted by the ratio of speech-related child follow-ups to speech-related utterances that did not receive a response:

$$\frac{\#(U_{speech} \wedge R_{resp} \wedge F_{speech})}{\#(U_{speech} \wedge R_{resp})} - \frac{\#(U_{speech} \wedge R_{no\_resp} \wedge F_{speech})}{\#(U_{speech} \wedge R_{no\_resp})} \approx 0.01 \quad (2)$$

This value was also significantly positive ($SE = 0.003, p < 0.05$), again replicating the results of the original study and confirming that children are sensitive to temporal contingency when learning speech-related vocalizations.[9]

Note that, for comparison with Warlaumont et al. (2014), our replication study used the same measures. However, for our next analyses, we will use mixed-effect models instead, because they allow more rigorous statistical testing as well as the ability to control for other variables. We also ran the equivalent mixed-effects models for this replication, and the results confirmed the significance of both effects, even when controlling for age.

[9] Warlaumont et al. (2014) obtained the following values: 0.065 for caregiver contingency and 0.036 child contingency. The difference in effect sizes could arise from differences in the properties of the datasets used in the original vs. the replication, e.g., different conversational contexts or because in our data (which relies on transcriptions instead of automatic speech classification) probably not all non-speech utterances were transcribed exhaustively.

Development of Intelligibility via Temporal Contingency

Following the general reasoning in Warlaumont et al. (2014), we first study the extent to which caregiver's time-contingent response behavior depends on the intelligibility of the child's utterance. Second, we study the extent to which the child's follow-up show improved intelligibility when following responsive behavior from the caregiver.

Caregiver's temporal contingency Figure 3 shows the results of how caregivers' response behavior depends on the intelligibility of the children's utterances.

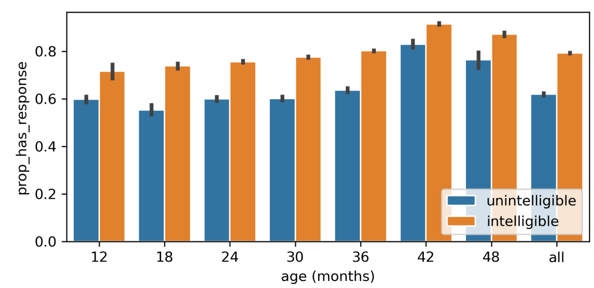

Figure 3: Comparison of proportion of caregiver responses for intelligible and unintelligible child utterances.

To quantify this effect, we used the following mixed-effects GLM that predicts whether a caregiver response was given as a function of whether the child utterance was intelligible:

$$\texttt{has\_resp} \sim \texttt{utt\_is\_intelligible} * \texttt{age} + (1|\texttt{child}) \quad (3)$$

The estimated fixed effects were: utt_is_intelligible $= 0.81$ $(SE = 0.01, p < 0.001)$; age $= 1.23$ $(SE = 0.03, p < 0.001)$; utt_is_intelligible:age $= -0.29$ $(SE = 0.05, p < 0.001)$.

These results confirm the qualitative observations in Figure 3, that is, utt_is_intelligible is a predictor of caregiver's response contingency (more intelligible utterances leads to more contingency and vice versa). This effect was significant even controlling for age.

Child sensitivity to temporal contingency In Figure 4, we show how the intelligibility of the child's follow-up depends on whether there was a caregiver's response to an intelligible child's utterance.

To quantify this effect, we similarly used a GLM that predicts whether a child follow-up is intelligible as a function of whether there was a caregiver's response or not, taking into account only U-R-Fs for which the initial utterance by the child was intelligible:

$$\texttt{follow\_up\_is\_intelligible} \sim \texttt{has\_resp} * \texttt{age} + (1|\texttt{child}) \quad (4)$$

The estimated fixed effects were: has_resp $= 0.38$ $(SE =$

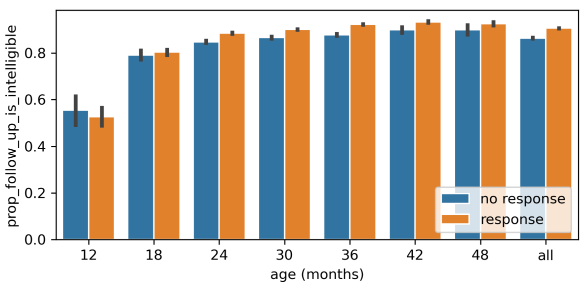

Figure 4: Comparison of the proportion of intelligible follow-ups depending on whether the child's previous intelligible utterance received a response from the caregiver or not.

$0.01, p < 0.001$); `age` $= 0.77$ ($SE = 0.04, p < 0.001$); `has_resp:age` $= 0.12$ ($SE = 0.06, p = 0.06$).

Here again, the statistical analysis confirms the qualitative observations in Figure 4, that is, there was a positive impact of caregiver's responses (`has_resp`) on children's follow-up being more intelligible. This effect was significant above and beyond the child's age.

Development of Intelligibility via Clarification Requests

We first study the extent to which caregivers' clarification requests are dependent of the intelligibility of the child's utterance. Second, we study the extent to which children's follow-ups increase in intelligibility after a clarification request.

Caregiver's clarification requests Figure 5 shows how caregiver's clarification requests depend on the intelligibility of the preceding child utterance across development.

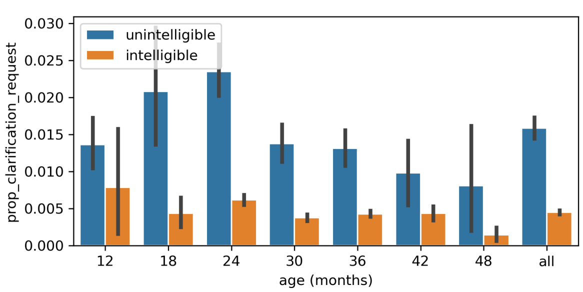

Figure 5: Comparison of proportion of clarification requests for intelligible and unintelligible child utterances.

We used the following GLM predicting whether the caregiver's response is a clarification request, as a function of whether the child's utterance was intelligible:

$$\texttt{resp\_is\_clar\_req} \sim \texttt{utt\_is\_intelligible} * \texttt{age} + (1|\texttt{child}) \quad (5)$$

The estimates were: `utt_is_intelligible` $= -1.14$ ($SE = 0.06, p < 0.001$); `age` $= -0.59$ ($SE = 0.15, p < 0.001$); `utt_is_intelligible:age` $= 0.02$ ($SE = 0.21, p = 0.9$).

As expected, we found a negative effect of `utt_is_intelligible` showing that clarification requests are more likely to be made by the caregiver after unintelligible utterances from children. This effect was also significant above and beyond age.

Child sensitivity to clarification requests Here, we investigate if caregiver's clarification requests lead to more intelligible follow-ups from children. For this analysis, we do not compare the intelligibility of child follow-ups as a function of the presence vs. absence of clarification request (similarly to the analysis on sensitivity to temporal contingency, where we compare the distinction response vs. no response), because the temporal-contingency-based mechanism creates a confound. More precisely, when caregivers do not give a clarification request, this is usually because the child utterance was already intelligible and is thus likely to receive a response from the caregiver, leading to the continuation of intelligibility in child follow-up. In both cases (presence vs. absence of clarification request), we can predict high follow-up intelligibility, hence the confound we need to avoid.

Thus, to be able to test the specific effect of clarification request without interference from the temporal contingency mechanism, we compare the intelligibility of the follow-up to that of the child's previous utterance within the *same* U-R-F.

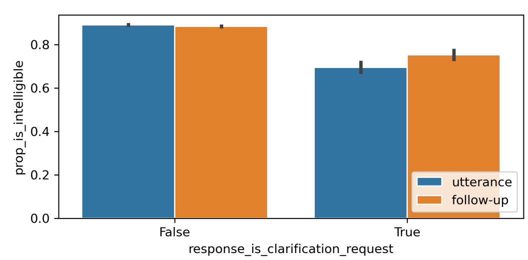

Figure 6: Comparison of proportion of intelligible utterances before (utterance) and after (follow up) clarification requests and other responses.

Figure 6 compares the effect of the caregiver's response (presence vs. absence of clarification requests) on the intelligibility of utterances before and after the response. We observe that in the absence of a clarification request, *both* the child's follow-up and her previous utterance are more intelligible. This observation illustrates the confound previously mentioned: Comparing intelligibility of follow-ups alone would be misleading. However, when we compare intelligibility before and after a response, we observe that intelligibility improved more when the response is a clarification request (right side of Figure 6). We quantify this effect through testing an interaction in the following model, demonstrating that children are sensitive to this kind of CF and that they can use it to improve their intelligibility:

$$\texttt{is\_intelligible} \sim \texttt{resp\_is\_clar\_req} * \texttt{before\_after} + (1|\texttt{age}) + (1|\texttt{child}) + (1|\texttt{urf\_id}) \quad (6)$$

The estimates were as follows: resp_is_clar_req $= -1.02$ ($SE = 0.04, p < 0.001$); before_after $= 0.14$ ($SE = 0.03, p < 0.001$); and more importantly, the interaction term: resp_clarification_req*before_after $= 0.46$ ($SE = 0.05, p < 0.001$). The positive interaction term demonstrates that the difference between before and after is larger in the case of clarification request responses, than it is in other responses.

Next, we zoom in on the case of clarification requests (bars on the right in Figure 6, and we test whether the observed effect holds over development (Figure 7).

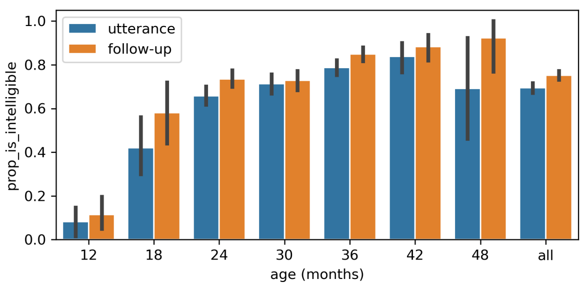

Figure 7: Comparison of proportion of intelligible utterances before (utterance) and after (follow up) a clarification request.

We used the following model, taking into account only the subset of U-R-Fs where the response is a clarification request:

$$\texttt{is\_intelligible} \sim \texttt{before\_after} * \texttt{age} + (1|\texttt{child}) + (1|\texttt{urf\_id}) \quad (7)$$

The estimates were: before_after $= 0.44$ ($SE = 0.1, p < 0.001$); age $= 2.74$ ($SE = 0.43, p < 0.001$); before_after:age $= 0.06$ ($SE = 0.52, p = 0.9$), indicating that the effect remains significant above and beyond age. The intelligibility of children's utterances increased after receiving negative feedback in the form of a clarification request from the caregiver.

Discussion

The broad goal of this paper was to investigate how some general social feedback mechanisms that are part of human communication (H. H. Clark, 1996) can help children learn language. As a case study, we explored how this feedback can support children to produce intelligible speech.

What is special about this feedback (which we call communicative feedback, CF) is that it does not aim to correct or refine the content of the child's utterances (as in the case of corrective feedback). It only seeks to establish/repair communication via positive or negative signals of understanding. We argued that CF can help with language development indirectly: As children seek to be understood, they are sensitive to social signals that indicate whether their communicative intent (e.g., requesting an object or seeking attention) was successfully achieved, and they revise/adjust their expression if necessary, aligning with the correct linguistic conventions.

We provided evidence that CF is useful for learning how to produce intelligible speech. To this end, we investigated not only *positive* CF (temporal contingency) but also one kind of *negative* CF (clarification requests).

Our results indicated that both are contingent on the intelligibility of child utterances across all observed ages, thus providing a useful feedback signal. Critically, we also found evidence that children are sensitive to these signals and produce more intelligible utterances if their caregivers are responsive and improve the intelligibility of utterances if the caregiver asks for clarification.

Limitations and future research directions This work relied on publicly available transcriptions of child-caregiver interactions in naturalistic environments. The recordings could be of varying quality and there might be cases in which utterances are transcribed as "unintelligible" because of poor audio quality, or noise. These cases are a confound to our analyses, since we considered all such cases as unintelligible to the caregiver, while they might just be unintelligible to the transcribing person. That said, manual verification of several examples suggested that in most cases the utterances were most likely also unintelligible to the interlocutor. Further, we observed a continuous increase of the intelligibility of children's utterances in our corpora (see Figure 2, which indicates that the intelligibility is not (only) a phenomenon of the transcription but an indicator of the children's linguistic development.

We quantified children's sensitivity to CF by measuring its effect on immediate child follow-ups and observed a significant influence. This operationalization could *overestimate* actual learning effects, which could be forgotten in the long term. However, it could also *underestimate* learning effects which may become visible only at a later point in time. Future research is required to explore the long-term effects of CF.

Regarding negative CF signals, we studied the role of clarification requests. While we were able to demonstrate their usefulness, their presence in the observed conversations was rather scarce (We analyzed a total of 2235 clarification requests, which formed 0.5% of all U-R-F sequences).

In future research, many other positive and negative CF signals could be quantified, including facial expressions (e.g., frowning as negative feedback), actions (e.g., providing requested objects as positive feedback), and content contingency (e.g., responding on-topic as positive feedback). This effort will require collecting more multimodal data of child-caregiver conversations where such cues can be captured, as well as the development of machine learning methods that can perform annotation at scale.

Acknowledgements

We would like to thank Juliette Millet, Marvin Lavechin, Maureen de Seyssel, and the anonymous reviewers for their insightful feedback on this work.

This work, carried out within the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille - A*MIDEX (Archimedes Institute AMX-19-IET-009), a French "Investissements d'Avenir" Programme.

References

Bloom, K. (1988). Quality of adult vocalizations affects the quality of infant vocalizations. *Journal of Child Language*, *15*(3), 469–480.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*, 223–249.

Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., ... Vardi, D. (1992). Maternal Responsiveness to Infants in Three Societies: The United States, France, and Japan. *Child Development*, *63*(4), 808–821. doi: 10.1111/j.1467-8624.1992.tb01663.x

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

Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, *1*(1), 111–114.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of child language*, *30*(3), 637–670.

Clark, E. V. (2018). Conversation and language acquisition: A pragmatic approach. *Language Learning and Development*, *14*(3), 170–185.

Clark, E. V. (2020). Conversational repair and the acquisition of language. *Discourse Processes*, *57*(5-6), 441–459.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., ... others (2015). Universal principles in the repair of communication problems. *PloS one*, *10*(9), e0136100.

Gallagher, T. M. (1977). Revision behaviors in the speech of normal children developing language. *Journal of Speech and Hearing Research*, *20*(2), 303–318.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Hiller, S., & Fernández, R. (2016). A data-driven investigation of corrective feedback on subject omission errors in first language acquisition. In *Proceedings of the 20th signll conference on computational natural language learning* (pp. 105–114).

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

Lieven, E., Salomo, D., & Tomasello, M. (2009, August). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3), 481–507. (Publisher: De Gruyter Mouton Section: Cognitive Linguistics) doi: 10.1515/COGL.2009.022

MacWhinney, B. (2014). *The childes project: Tools for analyzing talk, volume i: Transcription format and programs*. Psychology Press.

MacWhinney, B. (2017). Tools for analyzing talk part 1: The chat transcription format. *Carnegie.*.

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development*, *5*(3), 240–248.

Nikolaus, M., Maes, J., Auguste, J., Prevot, L., & Fourtassi, A. (2021). Large-scale study of speech acts' development using automatic labelling. In *Proceedings of the 43rd annual meeting of the cognitive science society.*

Ninio, A., Snow, C. E., Pan, B. A., & Rollins, P. R. (1994). Classifying communicative acts in children's interactions. *Journal of communication disorders*, *27*(2), 157–187.

Purver, M. R. J. (2004). *The theory and use of clarification requests in dialogue* (Unpublished doctoral dissertation).

Richman, A. L., Miller, P. M., & LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology*, *28*(4), 614–621. (Place: US Publisher: American Psychological Association) doi: 10.1037/0012-1649.28.4.614

Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language*, *20*(60), 221–252.

Saxton, M., Houston–Price, C., & Dawson, N. (2005, July). The prompt hypothesis: Clarification requests as corrective input for grammatical errors. *Applied Psycholinguistics*, *26*(3), 393–414. doi: 10.1017/S0142716405050228

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current biology*, *18*(9), 668–671.

Stalnaker, R. C. (1978, December). Assertion. *Pragmatics*, 315–332. doi: 10.1163/9789004368873_013

Tomasello, M. (2005). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–691.

Tsuji, S., Jincho, N., Mazuka, R., & Cristia, A. (2020). Communicative cues in the absence of a human interaction partner enhance 12-month-old infants' word learning. *Journal of experimental child psychology*, *191*, 104740.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological science*, *25*(7), 1314–1324.

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349.