Modeling speech act development in early childhood: the role of frequency and linguistic cues

Mitja Nikolaus, Juliette Maes, Abdellah Fourtassi

Modeling speech act development in early childhood: the role of frequency and linguistic cues

Mitja Nikolaus[1,2] (mitja.nikolaus@univ-amu.fr)
Juliette Maes[2] (juliette.maes@etu.univ-amu.fr)
Abdellah Fourtassi[1] (abdellah.fourtassi@gmail.com)

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

Abstract

A crucial step in children's language development is the mastery of how to use language in context. This involves the ability to recognize and use major categories of speech acts (e.g., learning that a "question" is different from a "request"). The current work provides a quantitative account of speech acts' emergence in the wild. Using a longitudinal corpus of child-caregiver conversations annotated for speech acts (Snow et al., 1996), we introduced two complementary measures of learning based on both children's production and comprehension. We also tested two predictors of learning based on input frequency and the speech acts' quality of linguistic cues. We found that children's developmental trajectory differed largely between production and comprehension. In addition, development in both of these dimensions was not explained with the same predictors (e.g., frequency in child-directed speech was predictive of production, but not of comprehension). The broader impact of this work is to provide a computational framework for the study of communicative development where both measures and predictors of children's pragmatic development can be tested and compared.

Keywords: first language acquisition; pragmatics; speech acts; computational modeling

Introduction

Language development requires not only the acquisition of linguistic structures that allows one to construct grammatically sound utterances (e.g., phonology, words, and syntax) but also the mastery of how to put this structure to use in social communication. This mastery involves both learning to pick utterances that best conveys one's communicative intents (or speech act) and understanding other's communicative intents based on their linguistic utterances (e.g., Grice (1975))

An important line of work in language use has been the study of how categories of speech acts (e.g., "question", "request") emerge in the natural context of child-parent social interactions (for reviews, see Cameron-Faulkner (2014); Casillas & Hilbrink (2020)). Children's learning of speech acts is crucial for their ability to engage in coherent conversations. For example, it is important to recognize that an utterance is a "question" requiring an "answer", or that it is a "request" requiring "acceptance" or "refusal", instead.

Previous studies about children's acquisition of speech act categories have focused on developing coding schemes that best capture children's emerging repertoire of communicative intents. The most comprehensive scheme has been the Inventory of Communicative Acts and its abridged version, INCA-A (Ninio et al., 1994). Using the INCA-A scheme, Snow et al. (1996) analyzed a longitudinal corpus of child-caregiver spontaneous conversations for children aged 14 to 32 months old. They found that children produce a rich set of speech acts from an early age. However, they also observed that some speech acts take longer to emerge than others.

The current study

Following this observation, the current study aims at investigating, in a quantitative fashion, what makes speech acts easy or hard to learn. To answer this question, we define both accurate measures of emergence (the explanandum) and plausible predictors of learning (the explanans).

Concerning the measures, Snow et al. (1996) focused on children's production. While it is true that production provides a rather tangible evidence of learning, it tends to underestimate children's knowledge. In many cases, children may understand the speech act without necessarily attempting to produce it, especially in contexts such the child-caregiver interactions where there is a clear asymmetry in social roles. In fact, social asymmetry could translate into an asymmetry in terms of the speech acts used. Take the case of why-questions: In some small-scale traditional societies, children do not typically ask such questions to caregivers (Gauvain & Munroe, 2020) though they may be perfectly able to answer them. Thus, the production-based measure is not enough, it should be complemented with a comprehension-based measure. The latter can be operationalized in many ways, e.g., in terms of whether or not the child is able to respond to the target speech act in a contingent fashion.

Concerning the predictors of learning, a variety of factors may influence children's learning. Here we focus on testing the role of two different predictors. The first one is frequency: We can imagine that children learn first the speech acts that are used more frequently by the caregiver. Another factor that could influence learning is the difficulty with which children can infer the identity of the speech act from its linguistic expressions. Using examples from the INCA-A scheme, the speech act that consists in "asking for permission" typically involves distinctive words such as "can I" and "please", which could make it easier to learn and understand than, say, the speech act that consists in "giving reason" which could be expressed in a much larger number of ways and does not have linguistic cues that are as distinctive as in the case of asking permission. The goal is thus to test how such linguistic fac-

tors predict the learning trajectory of speech acts.

In what follows, we will first provide a brief description of the data we use as well as the INCA-A coding scheme used to annotate these data for speech acts. Then, we explain how we defined the measures of emergence (in both production and comprehension) and how we characterized the predictors of learning, especially regarding the quantification of the linguistic cues associated with speech acts. Next, we present the results of the analyses that aim at 1) comparing the developmental trajectories of speech acts across production and comprehension, and 2) testing how frequency and the quality of linguistic cues predict the order of emergence of speech acts. Finally, we discuss the findings in the lights of the literature on speech act acquisition.

Data and Methods

Data

We used the data that Snow et al. (1996) used for their longitudinal study examining speech act development of 52 American English speaking children aged 14, 20 and 32 months old. Child-caregiver dyads were invited for three sessions which included a warm-up and a semi-structured free play period. All conversations were recorded, transcribed, and annotated using the INCA-A coding scheme.

INCA-A coding scheme

While a wide range of taxonomies has been developed to study children's emerging speech acts (Cameron-Faulkner, 2014), INCA-A is the most comprehensive to date (Ninio et al., 1994). The coding scheme has two levels: 1) the interchange level that characterizes the topic of the conversation (e.g., "discussing a recent event") and may span multiple utterances, and 2) the illocutionary force (e.g., "Ask a yes/no question") which is determined at the utterance level. Here, we focus on the illocutionary force, more commonly known as the speech act. INCA-A has 67 different speech act types, which are grouped into high-level categories such as directives, speech elicitations, commitments, declarations, markings, statements, questions, performances, evaluations, demands for clarification, text editing, and other vocalizations.

INCA - Abridged Again: INCA-A^2

Our preliminary investigation of INCA-A has revealed the presence of several couples of speech acts that were either very similar or hierarchically related (see Cameron-Faulkner & Hickey (2011) for a similar observation). We found that these shortcomings add noise to the our measures of speech act emergence and spuriously inflate the error rate of our models. Thus, we created an abridged version of the already abridged INCA-A (henceforth called INCA-A^2) in a systematic fashion where we collapsed 1) couples of speech acts categories that overlap to a high degree *both* conceptually and linguistically (e.g., "Criticize or point out error in nonverbal act" (`CR`) overlaps with "Disapprove scold protest disruptive behavior" (`DS`)), 2) couples of speech acts where the meaning of one act was included in the other (coarser-grained) act (e.g., the speech act "Ask a limited-alternative yes/no question" (`TQ`) is part of the higher-level category "Ask a yes/no question" (`YQ`).) (A full list of speech acts and how they were collapsed is given in the appendix). The resulting coding scheme reduced the number of speech act types from 67 to 46. The results in the main text are obtained using INCA-A^2. However, we also report the results using the original INCA-A version in the appendix. For a transparent reading, the labels for collapsed speech act categories (e.g. `YQTQ`) were simply obtained by concatenating the labels for the individual categories (e.g., `YQ` and `TQ`).

Measures of speech act emergence

Here we introduce measures of speech acts' age of emergence both at the level of children's production and comprehension. This will allow us to rank speech acts by order/difficulty of emergence in development and, later, test which factors predict this order.

Production By analogy to work in word learning (Braginsky et al., 2019), we define the age of acquisition of a speech act in production as the month by which at least 50% of the observed children produce it.[1] More precisely, for each speech act `S`, we proceed as follows:

1. For each age in the dataset (i.e., 14, 20 and 32 months), calculate the fraction of children who are producing `S` at least twice.
2. Perform a logistic regression over these fractions
3. Measure the age of first production as the age where the logistic regression curve surpasses the value 0.5.

Comprehension As pointed out in the introduction, studying speech act emergence only from a production point of view may underestimate children's pragmatic competence. Thus, we additionally introduce a measure for children's comprehension which we define as the ability of children to respond to a target speech act in a contingent fashion (e.g., responding to a "yes/no question" with "yes" or "no"). More precisely, for each speech act `S`, we proceed as follows:

1. Find all utterances produced by the caregivers labelled as `S`.
2. Find all cases where these utterances are followed by an utterance of the child.
3. For each occurring follow-up utterance, annotate whether its speech act is contingent as a response to `S`.[2]
4. For each age (14, 20 and 32 months), calculate the fraction of contingent follow-up utterances.
5. Perform a logistic regression over the fractions.[3]
6. Measure the age of comprehension as the age where the logistic regression curve surpasses the value 0.5.

[1] In line with (Snow et al., 1996), we consider that a child acquired a speech act if it is produced at least twice at a certain age.

[2] Annotating contingency was done using a binary scale, indicating whether the speech act was *possibly* contingent (1) or clearly non contingent (0).

[3] We only regard datapoints where the fraction was calculated over at least 2 examples, i.e. where there were at least two utterances with follow-ups.

In both production and comprehension, we only used in the analyses the speech acts for which we could successfully perform the logistic regression (i.e. we excluded speech acts where we had less than two datapoints at two different ages.) This left us with a set of 23 speech acts for production and 29 for comprehension. [4]

Predictors of speech act emergence

We test two predictors of speech acts' development: the frequency of use by the caregiver and the quality of its linguistic cues. While measuring frequency required a mere count of occurrences in the input, the characterization of the linguistic cues required the use of sophisticated tools we borrowed from the field of Natural Language Processing (NLP).

The intuition behind this second factor is explained in the introduction and can be summed up as follows: The easier it is to map a given speech act to its linguistic instances, the more these linguistic instances contain rich and consistent cues pointing to the speech act, and — as the hypothesis goes — the easier it is for children to learn it.

We quantify the quality of the linguistic cues of a given speech act by the accuracy of classification by an automatic classification model (as measured in per-label F_1-score). We proceed as follows. We use a Conditional Random Field (CRF, Lafferty et al., 2001), a simple probabilistic model that is typically used in speech act recognition in adult dialogues as it takes into account the context of the conversation (i.e., how preceding labels are sequentially organized).[5]

We train the model to automatically classify speech acts given a set of linguistic features.[6] Next, we evaluate the model on a held-out test set (20% of the data). The accuracy of this model reached 75.3% in our INCA-A^2 coding scheme (increasing from 72.3% obtained with the original, but noisier scheme: INCA-A), while inter-annotator agreement for the corpus is reportedly ranging from 81% to 89% (Snow et al., 1996). Given these high accuracy scores (close to state-of-the-art scores for models in adult dialogs), our model can be understood as successfully learning the linguistic cues that characterize each speech act.

Finally, for each speech act, we define the quality of linguistic cues as its F1-score on the held-out test set when classifying adults' utterances.[7]

[4] While the resulting sets of speech acts may appear small compared to the original size, it is due to the fact that the original frequency distribution was highly skewed: A small set of speech acts were used very frequently while many have very few instances, and therefore, did not provide enough data to fit a logistic regressor.

[5] In addition to CRF, we tested both simpler models (random forests and linear support vector machine) as well as state-of-the-art neural network based models (using a hierarchical LSTMs encoder in combination with a CRF decoder). The CRF model was performing the best in terms of accuracy. We ascribe the poor performance of the neural network model to the lack of large-scale training data.

[6] These features are: speaker (caregiver/child), unigrams and bigrams of the target utterance, repetitions (number of words that are repeated from the previous utterance) and part of speech tags. We also experimented with other features such as words from previous utterances but found no performance improvements.

[7] We only test on adults' utterances as we assume these utterances represent the input children are learning from. The quality of linguistic cues in adults' utterances is what may predict their learning by children.

Results

We present two sets of results. The first concerns the analysis of age of emergence of speech acts as quantified by our measures of production and comprehension. The second set of results concerns the analysis of how our hypothesized predictors (i.e., frequency and quality of linguistic cues) correlate with the age of emergence of speech acts both in production and comprehension.

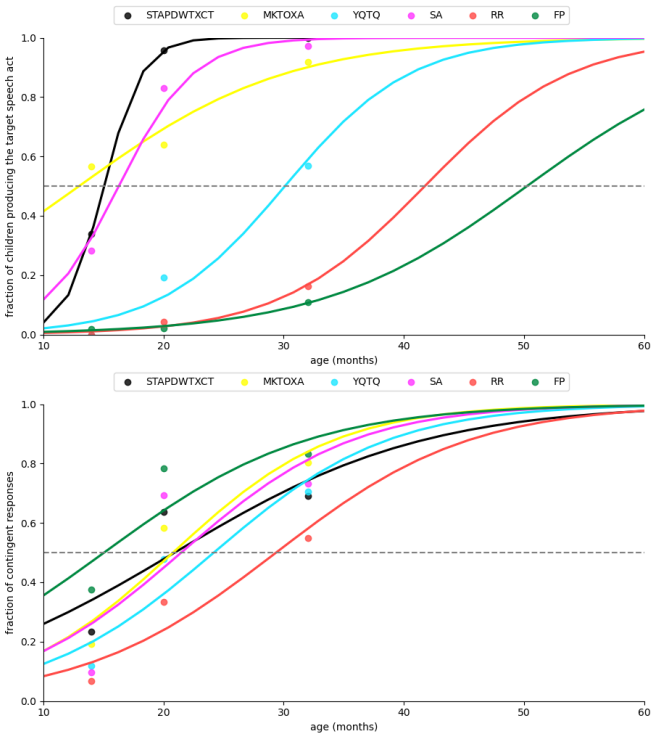

Figure 1: Quantification of the age of acquisition in terms of production (top) and comprehension (bottom) of 6 example speech acts.

Trajectories of speech act emergence

Concerning the measure of production, Figure 1 (top) is an illustration of the proportion of children who use speech acts across time as well as the best logistic fits we used to predict their precise age of emergence (we only selected a few examples of speech acts for ease of visibility and to illustrate the range of variance). We can observe clear variance in terms of when these speech acts emerge, in line with the qualitative observations made by Snow et al. (1996). For example, the category of Statements (`STAPDWTXCT`), markings (`MKTOXA`) and answers to wh-questions (`SA`) are produced early, while polar questions (`YQTQ`), demands for clarification (`RR`) and demands for permission (`FP`) are produced later.

To illustrate the emergence of speech acts in terms of comprehension, we first show observed adjacency pairs for

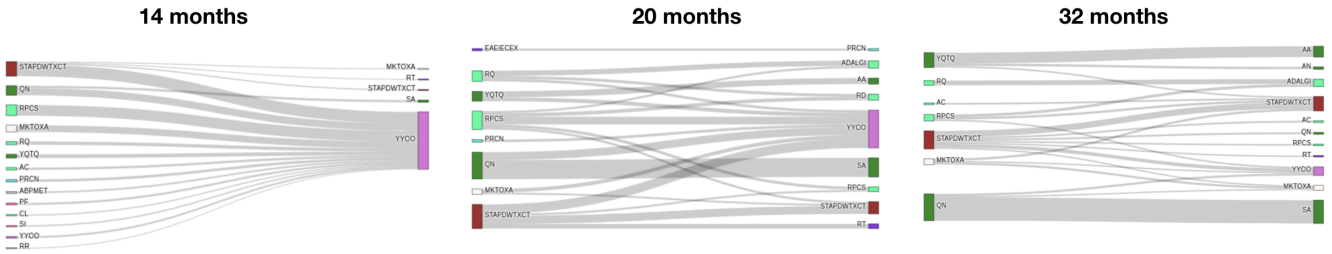

Figure 2: Adjacency pairs of speech acts for children of 14, 20, and 32 months. Utterances by the caregiver are on the left, responses by the children on the right. Filtered to display speech acts that occur in at least 0.01% of the data for better visibility.

adult-child turns for different ages in Figure 2. The younger children respond with unintelligible utterances or utterances without clear function (YYOO) in most of the cases displayed. Children at 20 months of age show some consistent patterns in their response behavior: Polar and product questions (YQTQ and QN) are answered with adequate responses (AA and SA). Polite requests (RQ) are either accepted (ADALGI) or refused (RD). Requests or suggestions (RPCS) are also usually accepted or refused, although in some cases children answer with a statement (STAPDWTXCT), which is not contingent. Additionally, there is still a large amount of utterances without clear function (YYOO). Only by the age of 32 months, most of the parents' utterances are addressed with contingent responses.

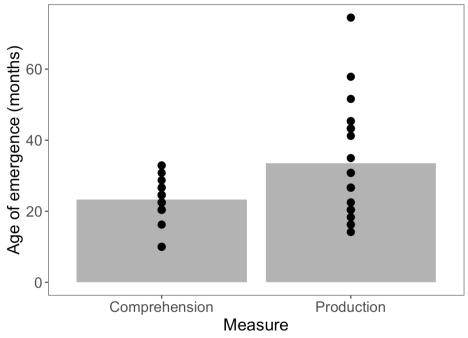

Figure 3: The distribution of the speech acts' age of emergence in comprehension and production.

Using this data on child-caregiver adjacency pairs, we computed the age of emergence in comprehension as the month at which the proportion of contingent responses surpasses 50% of children's total responses. Figure 1 (bottom) illustrates the proportion of contingent responses made by children across time as well as the best logistic fits used to predict the speech acts' precise age of emergence. We show the same examples of speech acts as in production for comparison. While there are similar trajectories in production and comprehension for some speech acts (e.g. RR), we also observed some striking differences in other cases. For example, "demands for permission" (FP) is produced very late (around 52 months), but they are already understood a lot earlier (around 16 months). Figure 3 shows the full distribution of age of emergence in both production and comprehension. It shows that, overall, comprehension of speech acts precedes their production. Indeed, a paired t-test shows a mean difference of 9.61 months ($p < 0.01$).

Finally, we ask how the trajectory of emergence in comprehension compares to that of production. For instance, does production follow the same pattern/order of comprehension, only delayed? Pearson's correlation between the two developmental trajectories is $r = 0.3$ ($p = 0.19$), indicating that speech acts emerge differently in production and comprehension, and suggesting that these two dimensions of development may be explained by different factors.

Predicting the emergence of speech acts

What makes a speech act easy or hard to acquire? Here we investigate the extent to which frequency and quality of linguistic cues predict the order of emergence both in production and comprehension. The results are shown in Figures 4 and 5, respectively.

For production, we found that frequency (but not the quality of the linguistic cues) predicts the speech acts' order of emergence ($r \approx -0.47$, $p < 0.03$). As for comprehension, we found the opposite pattern: While frequency showed no correlation whatsoever with age of emergence, the quality of linguistic cues led to a small correlation in the right direction, although this effect is not statistically significant ($r = -0.17$, $p = 0.37$) (probably due to low statistical power for this small sample size). Finally, the predictors themselves are highly correlated ($r = 0.77$, $p < 0.001$). [8]

Discussion

This work had two major goals: 1) provide a quantitative account of the developmental trajectory of speech acts in early childhood and 2) test some hypotheses about what could explain/predict this trajectory. For the first goal, we introduced two complementary measures that quantify the age of emergence of speech acts both in children's production and comprehension. We found that these two measures did not correlate, i.e., showing that speech acts may develop differently

[8] This high collinearity made it inadequate to run a multiple regression.

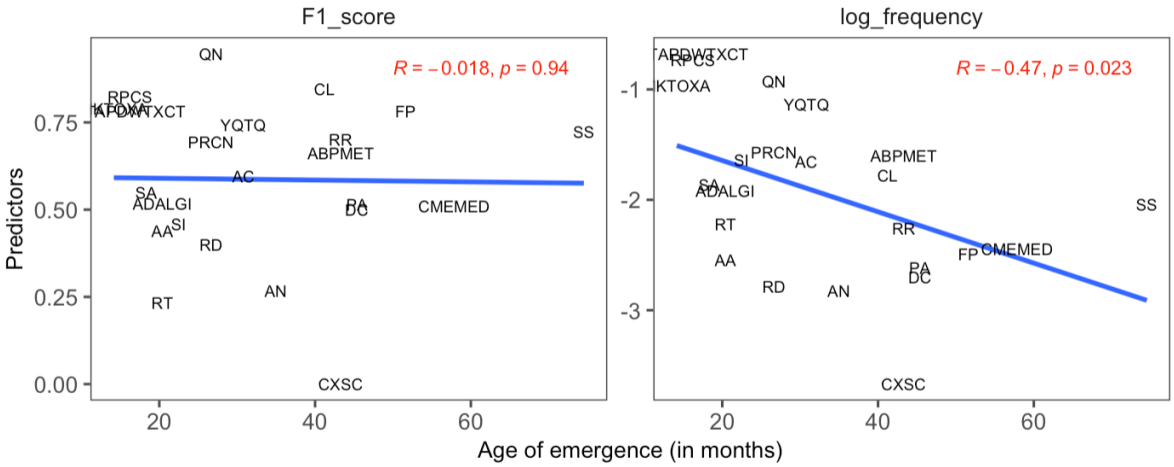

Figure 4: Predictors of age of emergence in production.

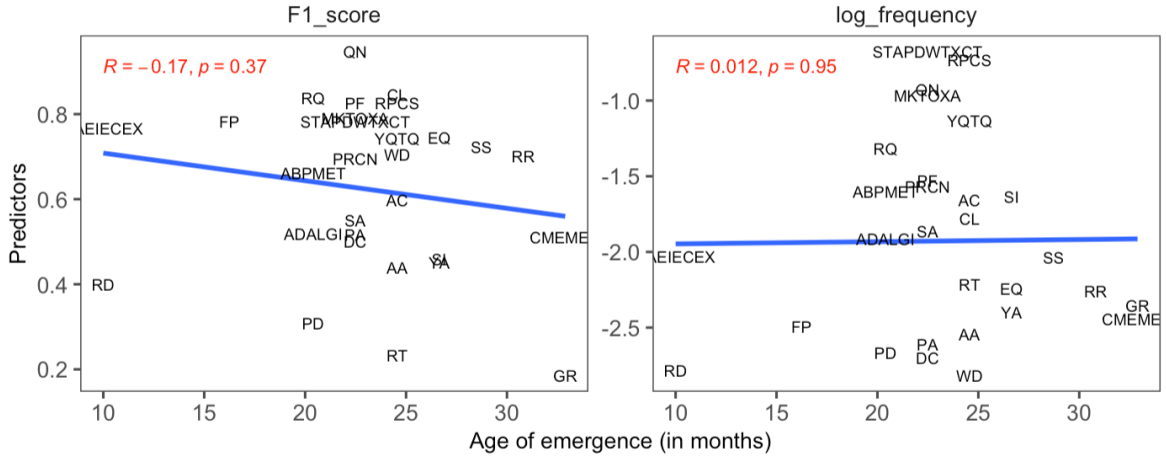

Figure 5: Predictors of age of emergence in comprehension.

in production and comprehension, and suggesting that factors that would be relevant for learning in production may not necessarily be the same in comprehension.

As for the second goal, and in order to explain what makes some speech acts emerge before others, we tested two factors: the frequency in the caregiver's input and the quality of the linguistic features that cue the speech act. Although these factors were highly correlated, we found that they provided, overall, non-redundant information. Generally speaking, frequency tended to predict production more whereas the quality of linguistic cues tended to predict comprehension more.[9]

The non-redundancy in the information provided by linguistic cues — beyond and above frequency — can be illustrated more clearly with some special cases. For example, the quality of linguistic cues for "giving reason" (GR) is very low compared to "requests to repeat" (RR) or "eliciting question" (EQ), while all have a similar (low) frequency. Indeed, there is a variety of ways one can express the act of "give reasons" in linguistic terms, which makes it harder to recognize this speech act based only on the linguistic features of its instances. In comparison, the set of linguistic terms typically used to express the act of requesting repetition or eliciting question is much more constrained, making its recognition easier. Indeed, this difference did predict emergence in comprehension: GR emerges later than RR and EQ

Take also the case of "stating intent" (SI) and "prohibiting" (PF). Both of these speech acts are similarly frequent, but the linguistic cues for PF are better/more consistent. Indeed, learning wise, we found that PF was understood earlier than SI. Finally, an interesting example is that of "asking for permission to carry out an act" FP which has high consistent linguistic cues while being very infrequent in caregivers' talk (caregivers do not frequently ask permission from children). Nonetheless, we found that this speech act is acquired very

[9] That said, statistical tests in the correlations are to be taken with a grain of salt given the relatively small sample size.

early in terms of comprehension, highlighting the predictive power of linguistic cues beyond frequency.

Findings in the current work allow us to make some links with literature on the development of communicative intents. On prominent example is that of Wh-questions vs. yes/no questions (also known as polar questions). Snow et al. (1996) found that children produce Wh-questions before polar questions and Moradlou et al. (2020) shows that the same order is found in comprehension. Our work confirms both of these findings (cf. Figure 5; `QN` is acquired earlier than `YQTQ` in both production and comprehension). This order was predicted both by the quality of linguistic cues which was much higher for Wh-questions than for polar questions and, to a lesser extent, by frequency.

Another interesting case is that of "Yes/no requests" vs. "yes/no questions for information." In production, we replicated Snow et al. (1996)'s finding that children produce yes/no questions as requests later than yes/no questions for information (very few children produced the first act and only at 32 months). This fact is also in line with the literature on politeness suggesting that children produce polite requests quite late (Axia & Baroni, 1985). Interestingly however, in comprehension we found the opposite pattern: Children responded more contingently to the yes/no requests earlier than they did to yes/no questions for information. In line with the general trend found in Figures 4 and 5, the order of production was predicted by frequency but not by linguistic cues, and the order of comprehension was predicted by the linguistic cues but not by frequency.

Limitations and future work

Finally, this current work has introduced both novel measures and research methods that we hope will pave the way to a more quantitative approach to the study of children's speech act development in the wild. That said, there is still room for improvement in future work.

Concerning the measures, while it is easier to quantify acquisition through production, it is trickier to have a perfect measure of comprehension. Here we provided a contingency-based measure. Such an operationalization has allowed us to uncover new interesting phenomena (namely that children understand some speech act before they produce them), however, measuring contingency can be difficult because responses can be contingent in various ways, e.g., asking a yes-no question like "Do you want a banana?" can be followed by many speech acts that can all be contingent such as "Yes!", "I just ate one", or "now?". In this work, we used a broad binary annotation that judges whether a response is possibly contingent (like the three previous examples) or totally inappropriate (e.g., a "greeting" after a "yes-no question"). In addition to this theoretical difficulty, there was a practical difficulty related to the fact that children (especially the younger ones) do not always respond (leading to more data exclusion), and sometimes they respond in an unintelligible fashion, a case which we had to classify as non-contingent, possibly under-estimating children's early age of comprehension. Second, concerning the research methods, here we introduced an NLP-based method (CRF model) that allowed us to provide a quantification of the linguistic cues to speech acts despite the variability and complexity that characterize natural conversations (as opposed to controlled lab designs). While this method was enough to investigate the question at hand, i.e., whether the quality of linguistic cues play a role in facilitating the learning of speech act, it only provided a partial response to the larger question of how speech acts emerge. Indeed, a more comprehensive answer would involve a diversity environmental factors. For example, several multimodal cues — besides language — likely play a role in signaling communicative intents such as vocal and visual cues. Indeed, such cues are picked up on by adults and children and are integrated to optimize language understanding and learning (e.g., Fourtassi & Frank, 2020; Fourtassi et al., 2021). In order for such an account to capture development with naturalistic data, efforts should continue to develop and combine methods in both NLP and Computer Vision.

Acknowledgements

We thank the anonymous reviewers for their comments and feedback.

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)

References

Axia, G., & Baroni, M. R. (1985). Linguistic politeness at different age levels. *Child Development*, 918–927.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Cameron-Faulkner, T. (2014). The development of speech acts. *Pragmatic development in first language acquisition*, 37–52.

Cameron-Faulkner, T., & Hickey, T. (2011). Form and function in irish child directed speech. *Cognitive Linguistics*, *22*(3), 569–594.

Casillas, M., & Hilbrink, E. (2020). 3. communicative act development. *Developmental and Clinical Pragmatics*, *13*, 61.

Fourtassi, A., & Frank, M. C. (2020). How optimal is word recognition under multimodal uncertainty? *Cognition*, *199*.

Fourtassi, A., Regan, S., & Frank, M. C. (2021). Continuous developmental change explains discontinuities in word learning. *Developmental Science*, *24*(2), e13018.

Gauvain, M., & Munroe, R. L. (2020). Children's questions in social and cultural perspective. *The Questioning Child: Insights from Psychology and Education*, *183*.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Moradlou, S., Zheng, X., Ye, T., & Ginzburg, J. (2020). Wh-questions are understood before polar-questions: Evidence from english, german, and chinese. *Journal of Child Language*, 1–27.

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.

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.

Appendix

The appendix can be downloaded from the following OSF project: `https://osf.io/m53jt/` .

Source code of the model and experimentation scripts can be found here: `https://github.com/mitjanikolaus/childes-speech-acts`.