Combinatorial Capacity of English Negation in Early Child Language

Anonymous CogSci submission

Abstract

Negation is very important for langauge and thought. How does it develop in the language of children? There has been many guessses like rejection, non-existence, denail, etc. but it has been hard to assess because these concepts are vaguge. Here we assess the combinatorial capacity of early negation in children's productions, and use words negation combines with as a proxy for early concepts expressed by it. We show some important stuff.

Keywords: negation; combinatorial capacity; corpus anlysis.

Introduction

Negation is an abstract concept, lexicalized in all previously studied human languages, and crucial to everyday communication. It can help a coffee shop divide its menu into "coffee" and "not coffee" sections, with the "not coffee" section bringing together diverse items with no common label. It can help us regulate each others' actions in a sign like "no mask, no entry". It can also communicate our deepest wants and dislikes, for example when we say "I don't like Mondays". But how does this crucial abstract concept emerge in the human mind? Does it emerge in its abstract and general form from the beginning or does it develop from limited and context-specific communicative functions?

There has been several influential hypotheses on the conceptual origins of negation and the functions it plays in early communication. Starting a century and a half ago, Darwin (1872) thought that negation has roots in the expression of human emotions and desires. He hypothesized that the earliest manifestation of negation and affirmation in infants is when they refuse food from parents, by withdrawing their heads laterally, or when they accept the food, by inclining their heads forward. He suggested that head shaking and nodding as common gestures for negation and affirmation have developed from this early habit. Similarly, many researchers studying early functions of negative morphemes like *no* proposed that children use them to "reject" or "refuse" (Bloom, 1970; Choi, 1988; Pea, 1978). For example, when they are asked "do you want juice?", they may say "no", "not want it", or "don't like it". Pea (1978) proposed that this function of negation is the first to emerge in children's early language.

Bloom (1970) suggested that the use of negation to expresses "non-existence" emerges before rejection or refusal. For example, when an object that children expect to be present is not present, children may say: "no window", "no fish in the bathroom", "Kathrine have no socks on" or "I do not have underpants". Two close concepts to non-existence discussed in the literature are "disappearance" and "non-occurrence" (Pea, 1978; Villiers & Villiers, 1979). Disappearance refers to situations where an object disappears and children use negation to express it such as "no food. all gone" or "no more noise". Non-occurrence refers to cases when an expected action or event does not occur as in "not working" or "doggie not barking". Some researchers referred to these cases as "failures" and included examples like "no fit in da box" or "it don't fit" (Cameron-Faulkner, Lieven, & Theakston, 2007; Choi, 1988). Non-existence can also be expressed by negation of locative prepositional phrases (e.g. "no in there" or "daddy was not on the phone"). While rejection was hypothesized to interact with human emotions and desires, non-existence (broadly construed to include "disappearance" and "non-occurrence") likely interacts with human perception. Choi (1988) proposed that children's early linguistic negation is used to express both rejection and non-existence.

Choi (1988) introduced "prohibition" as a function of early negation and suggested that it emerges as early as rejection and non-existence. In cases of prohibition, children use linguistic negation to stop others from performing some action. For example they may say: "don't go" or "do not spill milk". A special case of prohibition is "self-prohibition". For example, a child may approach prohibited food but immediately say "no, don't eat" to stop themselves from doing the prohibited action. Choi (1988) also discussed instances of negation in which children communicate their own "inability" to perform an action, for example "I can't reach" or "I cannot zip it". She suggested that these instances emerge after the first phase (i.e. non-existence, rejection, prohibition). Instances of prohibition and inability are similar in that both involve conceptualizing actions and negating them, possibly interacting with early development of motor control.

A fourth function of negation discussed in the literature is "denial". Bloom (1970) originally defined it as asserting that "an actual or supposed predication was not the case". For example a child may say: "It's not sharp". Later researchers labeled it as "truth-functional negation" and suggested that it is used to negate the truth of a proposition (Cameron-Faulkner et al., 2007; Pea, 1978). This definition depends on the assumed logical system and its assumptions on what type of propositions receive truth values. In this study, we focus on a

sub-function of traditional "denials", namely "labeling". This is often realized as the negation of nominal or adjectival predicates such as "this is not a bunny", "not red", or "this isn't a reptile". Parents often use such sentences to introduce novel linguistic labels and facilitate word learning. Therefore, it is possible that word learning helps the development of abstract negation. A fifth function discussed by Choi (1988) is "epistemic negation". There has been no proposal for negation originating in children's understanding of their own or others' mental states. However, previous studies have reported many instances of negation modifying mental state verbs such as *know*, *think*, and *remember* (e.g. "I not know").

Previous research on the origins and functions of negation has faced two major issues. First, it has had to rely on human annotation and classification of negative utterances in corpora which is costly, time-consuming, and difficult. As a result, most studies have had to focus on a handful of children and a relatively small sample of their speech. Second, the results of previous studies have shown great variability among children with respect to the emergence of different functions in children's speech. For example, Nordmeyer & Frank (2018) looked at the speech of five children in the Providence corpus Demuth, Culbertson, & Alter (2006) and found a great deal of individual variation in how early a negative function is attested, likely due to the prevalence of different types of activities children engage in such as book reading with parents vs. eating food. In this paper, we address these issues by focusing on the combinatorial and compositional capacity of negation in children's speech. We start with the widely accepted assumption that negation is a higher order operator, taking other concepts as its argument. The question we ask is: what type of concepts does linguistic negation operate on in early child language? Do we find negation composing with a limited set of lexical items? Or do we find it operating on a variety of lexical items and syntactic constructions?

Experiments

Data and preprocessing

For developmental data of child language in English, we turned to the CHILDES database (MacWhinney, 2000), which provides child-parent conversational interactions.[1]. We focused on speech produced by children with typical development within the age range of 12 - 72 months. As this study focuses on negation constructions at the utterance-level, we extracted individual utterances with any of the three negation markers to be investigated in this study: *no*, *not* and *n't*. Utterances with only one lexical token (e.g. *no !*) were not considered as here we aim to address particularly the question of what negation markers could *combine* with. In addition, cases where the negation markers only serve as a discourse marker (e.g. *no I like this one /no no no Mommy*) were excluded as well. Preprocessing led to a data set of 365,260 utterances with negation structures from a total of 811 children across 56 corpora.

[1]Code and data are in quarantine at https://somewhereonearth.

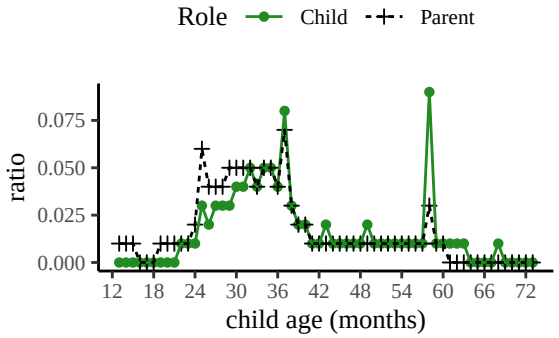

Figure 1: Distribution of the number of utterances with negative morphemes in child and parent speech.

Negation functions

The current English data from CHILDES contains morphosyntactic information (Sagae, Davis, Lavie, MacWhinney, & Wintner, 2010) such as part-of-speech (POS) information as well as grammatical or syntactic dependency relations.[2] We used POS tags of *neg* and *qn* to select negative utterances and morphemes, the latter of which was mainly for cases with *no* as a quantifier. We further excluded cases of enumeration (e.g. *no no no*), communicators, or discourse markers. After extracting negative utterances, the developmental trajectories of different constructions of interest were analyzed. Each construction was defined to roughly match one of the communicative functions previously discussed in the literature. In what follows we introduce each construction and present the results. Our plots often contrast the frequency of these constructions in children's speech as well as parents' speech at the corresponding age of the child.

Rejection For the function of "rejection", we examined cases where the lemma form of the head verb of the phrase is either *like* or *want*, and the head verb is modified by one of the three negative morphemes. Each of the utterances either takes a subject or has no subject at all. And the existence of a subject was determined via searching for a word in the utterance that has the *SUBJ* dependency relation with the head verb.

Additionally, other than expressions that the speaker used to describe their own emotion (e.g. (1)) or their (in)ability to do so (e.g. (2)), we also included cases that express rhetorical inquiries of emotions from one interlocutor addressed to another (e.g. (3)) as well as instances where the speaker is describing the emotion of somebody else (e.g. (4)). Overall our data extraction resulted in a total of 21,034 utterances (Child: 9,608; Parent: 11,426).

(1) *I no like sea / don't wanna go*

(2) *I can't like that*

(3) *don't you wanna try it*

[2]Besides using the provided POS and syntactic dependency information in CHILDES, we also experimented with the state-of-the-art parser from Stanza [@qi-etal-2020-stanza], an open-source natural language processing library. There were no notable differences in the analyses for the negation constructions.

(4) *Sarah doesn't like that either*

To compare the patterns between child and parent speech, we measured the following four metrics. The first one is the relative ratio of each of the three negative morphemes overall. For instance, given the 9,608 utterances from child speech that serve as rejections, there are 8,531 cases with the negative morpheme *no*; then the ratio of these utterances was calculated as 8,531 / 9,608 = 0.41. The second one is the relative ratio of negative morphemes within different head verbs (e.g. *like* vs. *want* for rejection). For example, again with child speech that express rejections, utterances where the negative morphemes modify the head verb *like* occur for 3,268 times; then the ratio of these cases was computed as 3,268 / 9,608 = 0.16. The third one is the relative ratio of the negative utterances at different ages of the child. For instance, for rejection, at the age of 36 months, the total number of instances with the negative morphemes in child speech is 888; then their ratio was calculated as 888 / 9,608 = 0.04.

The last one is the amount of variability in the production of the specific function across the age span of the child, which was measured with entropy (Cover & Thomas, 1991) For example, after computing the relative ratio ($P(x\_i)$) of the negative utterances at a number of N ages of the child for the specific function, the production variability is calculated using the equation below.

$$H(X) = -\sum_{i=1}^{N} P(x_i) log_2 P(x_i)$$

In both child and parent speech, when articulating desires or emotions with either of the two head verbs *like* and *want*, the most frequently used negative morpheme is *n't* combined with an auxiliary verb. Comparing the two different head verbs, overall the negative morphemes co-occur with *want* more frequently. With that being said, the amount of variability in both child and parent production is similar, a pattern that holds for both head verbs (Child *like*: 0.12; Child *want* 0.12; Parent *like*: 0.12; Parent *want*: 0.12).

On the other hand, when looking at the developmental trajectory, as presented in Figure @ref(fig:emotion), children's usage of negative morphemes is comparable regardless of the particular head verb. In general, children start applying the negative morphemes for the function of rejection more regularly around the age of 22 months. Within the context of the corpus data that we analyzed, their usage of these morphemes is the most frequent during the age range of 25 - 36 months.

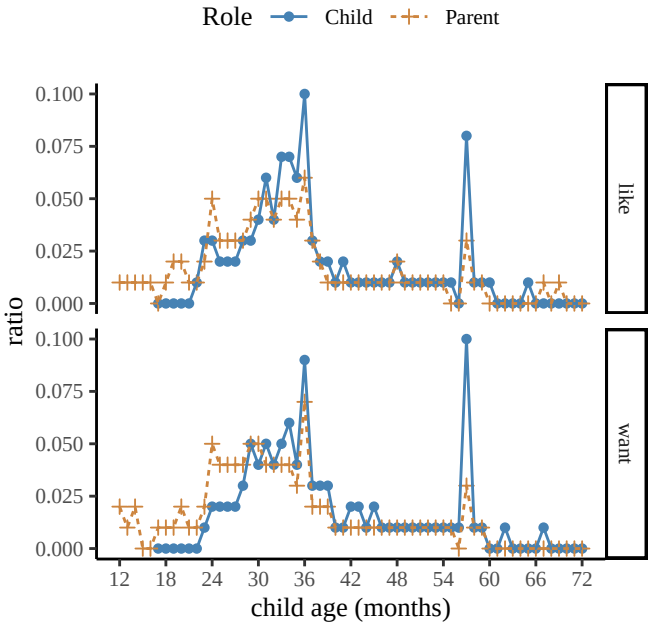

Figure 2: Rejection

Epistemic Negation To find epistemic uses of negation, we focused on utterances that articulate the concept of unknowing (e.g. (5)) or uncertainty (e.g. (6)). using the mental state verbs *know*, *remember* or *think* as the head verb, modified by the negative morphemes or the combination of negation with auxiliaries. By these search criteria, instances where the speaker inquires about or describes the negative epistemic position of another speaker (e.g. (7)) were also selected. This led to a subset of 32,793 utterances in total (Child: 10,389; Parent: 22,404).

(5) *I not know / I didn't remember*

(6) *I don't think so*

(7) *don't you remember / She doesn't know this*

In both child and parent speech, the most frequently used negative morpheme that modifies epistemic state is *n't*, a pattern that is consistent across the three different head verbs. And the negative morphemes tend to co-occur more often in cases that describe the state of unknowing, which is indicated mainly by the verb *know*. Based on results from Figure @ref(fig:epistemic), the production of *know* for expressions of epistemic state starts earlier in comparison to *remember* and *think*. On the other hand, the production variability for each of the head verb (˜0.12) is comparable to each other regardless of the particular speaker. Overall, children began to apply the negative morphemes to articulate this function in a more regular fashion around the age of 25 months.

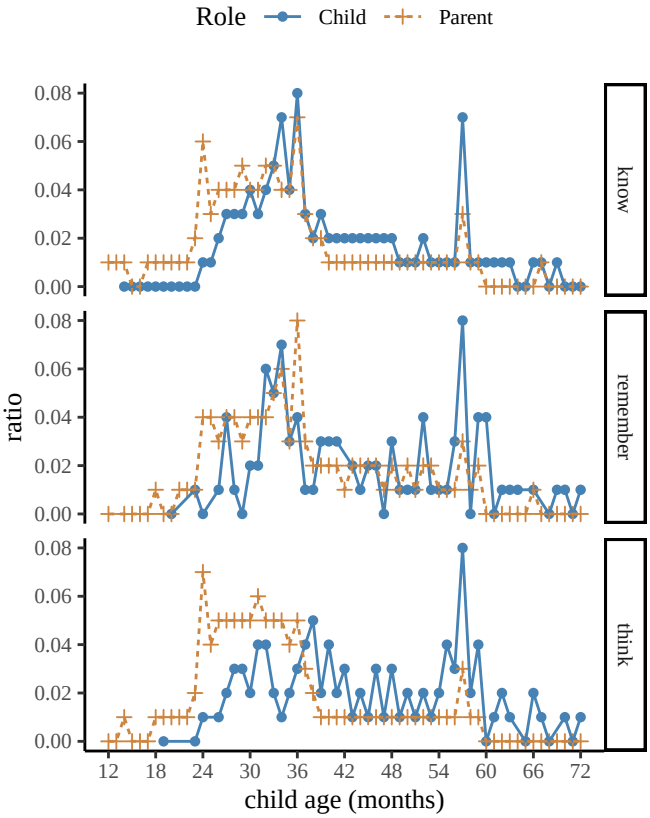

Figure 3: Epistemic

Prohibition For utterances that articulate the function of prohibition, we focused on cases where the negative morphemes are combined with the auxiliary verb *do* (*do*, *does*, *did*) and the auxiliary does not take any subject (e.g. (8)). In certain cases the negative morphemes and the auxiliary together modifies a head verb. For instances as such, in order to not overlap with the function of rejection, epistemic state, non-existence and possession (see below), our search excluded cases where the head verb has any of the following lemma forms: *like*, *want*, *know*, *think*, *remember*, *have*. This resulted in a total of 21,197 utterances (Child: 6,140; Parent: 15,057).

After applying our metrics, overall the most frequently used negative morpheme is *n't* when articulating prohibition. The amount of production variability for this function is comparable in both child and parent speech, with an approximate value of 0.12. The developmental trajectory of using the negative morphemes to serve this function (Figure @ref(fig:prohibition)) is comparable to that of the previous ones, where children started more regular usage of negative morphemes around the age of 23 months.

(8) *don't blame Charlotte* / *don't*

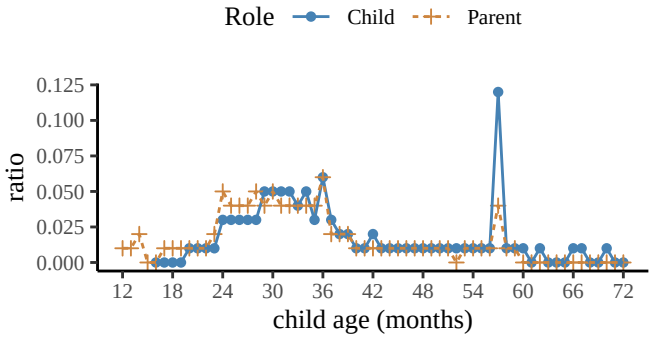

Figure 4: Prohibition

Inability We analyzed instances where the negative morphemes co-occur with the auxiliary *can* (*can* and *could*; e.g. (9)). Again, for instances with a head verb modified by the negative morphemes and the auxiliary, we filtered out cases where the head verbs are the focus for other functions. Cases that do not have a subject (*can't play*) or do not contain a subject other than I (*you can't do that*) could yield ambiguous readings without taking a larger discourse context into account; they could be a rhetorical question or also express the concept of prohibition. Therefore to avoid potential ambiguity, we restricted our analyses only to cases with a subject *I*. This led to a subset of 9,150 utterances (Child: 5,410; Parent: 3,740).

(9) *I can't see* / *I can't*

Comparing child and parent production, the negative morpheme that is used most frequently is also *n't*. As shown in Figure @ref{fig:inaibility} The developmental trajectory of this function is similar to that for prohibition, and the negative morphemes are applied more regularly starting around the age of 23 months. The amount of production variability in both child and parent speech is approximately 0.12.

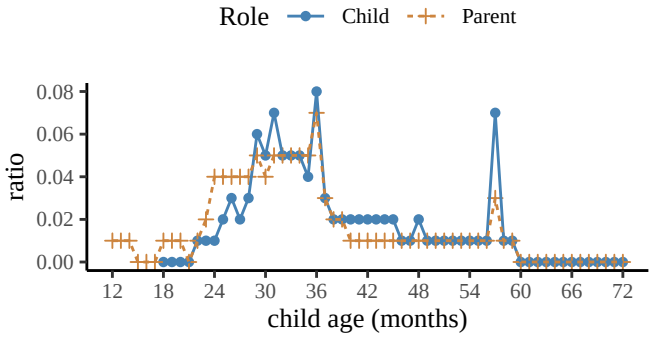

Figure 5: Inability

Labeling To capture labeling instances of denials, we concentrated on cases where negative morphemes are adopted to indicate the identity (e.g. (10)), and/or characteristics (e.g. (11)) of a predicative nominal. In addition, we also included instances where the negative morphemes are used to modify a predicative adjective (e.g. (12)). Following these criteria, utterances where the negative morpheme is modify-

ing a nominal or adjectival predicate of a copula verb were extracted. None of the utterances contained expletives (*there is no book*). The existence of a predicate was identified with the help of POS information and dependency relation. The POS of the predicate has to be either noun (*n*) or adjective (*adj*), and its dependency relation with the copula has to be *PRED*. This resulted in a total of 20,329 utterances (Child: 4,793; Parent: 15,536).

(10) *that's not a farmer*

(11) *I'm not a heavy baby Mum*

(12) *It's no good*

Comparing the three negative morphemes, the most frequently used is *not* regardless of the specific speaker, and the amount of production variability is comparable (˜0.12) between child and parent speehc. Based on results from Figure @ref(fig:learning), the developmental trajectory of using the negative morphemes in the domain of language learning is comparable to previous domains. Children started using the negative morphemes for the function of labeling nominal objects more frequently around the age of 24 months.

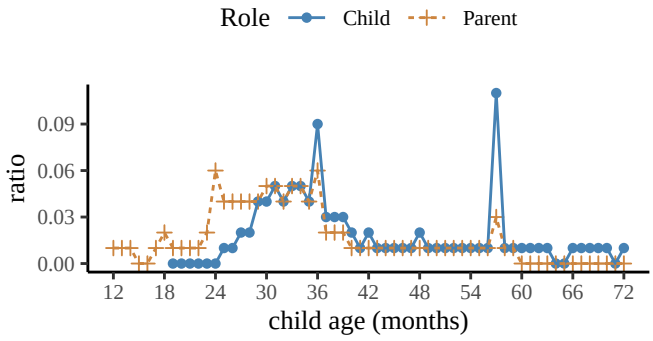

Figure 6: Language learning via labeling

Non-existence For the function of non-existence, we extracted utterances that either have expletives marked by *there* (e.g. (13)), or cases where the negative morphemes are modifying a nominal (i.e. its syntactic head based on the CHILDES annotation is a nominal; e.g. (14)). With utterances such as (14) in particular, in order to not confuse with the function of labeling, we did not include any cases where the syntactic head of the negative morphemes is a predicate of a copula verb (e.g. *this is not candy*). This led to a total of 34,672 utterances (Child: 16,866; Parent: 17,806).

(13) *there's no water*

(14) *no (more) candy / not your mouth*

In both child and parent speech, the most frequently occurred negative morphemes to indicate non-existence is *no*. The amount of production variability approximates 0.12 regardless of the specific speaker. Again for comparison of child and parent speech, we calculated the relative ratio of (i) each of the three negative morphemes overall; (ii) usage of negative morphemes with the two different communicative functions; (iii) utterances expression motor control with the three negative morphemes at different ages of the child. Overall the most frequently used negative morpheme is *n't* when applied in the domain of motor control. Comparing the two communicative functions, the negative morphemes tend to co-occur more often when expressing inability. As shown in Figure @ref(fig:existence), children began increasing their use of the negative morphemes to express non-existence around the age of 22 months.

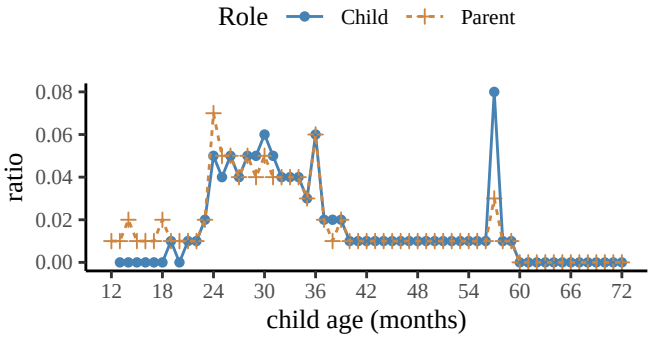

Figure 7: Non-existence

Possession The last function that we investigated includes utterances that are combined with negative morphemes to denote possession. Specifically, we selected cases where the negative morphemes are modifying a possessive pronoun (e.g. (15)), as well as instances where the negative morphemes are combined with auxiliary verbs to modify a head verb with the lemma form *have* (e.g. (16)). Again similarly to our search for utterances that express non-existence, we excluded cases in which the syntactic head of the negative morphemes is a predicate of a copula verb (e.g. *this is not mine*). As a result, the total of utterances that were subjected to analysis for this function is 9,265 (Child: 2,899; Parent: 6,366). The developmental trajectory for this function, as shown in Figure @ref(fig:possession), is comparable to that for the function of non-existence.

(15) *not mine*

(16) *I don't have it*

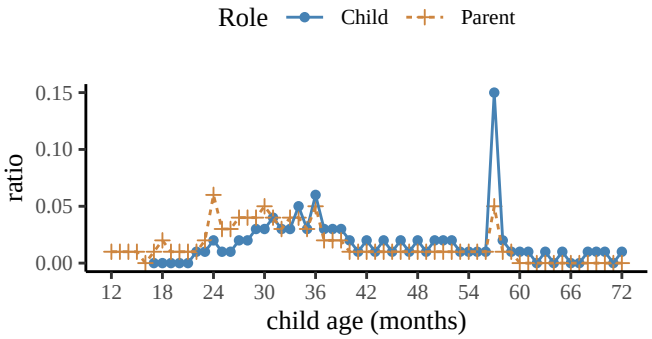

Figure 8: Possession

Overall Figure @ref(fig:all) shows the developmental trajectory for all previous negative constructions. The y-axis is the relative frequency of each construction relative to the

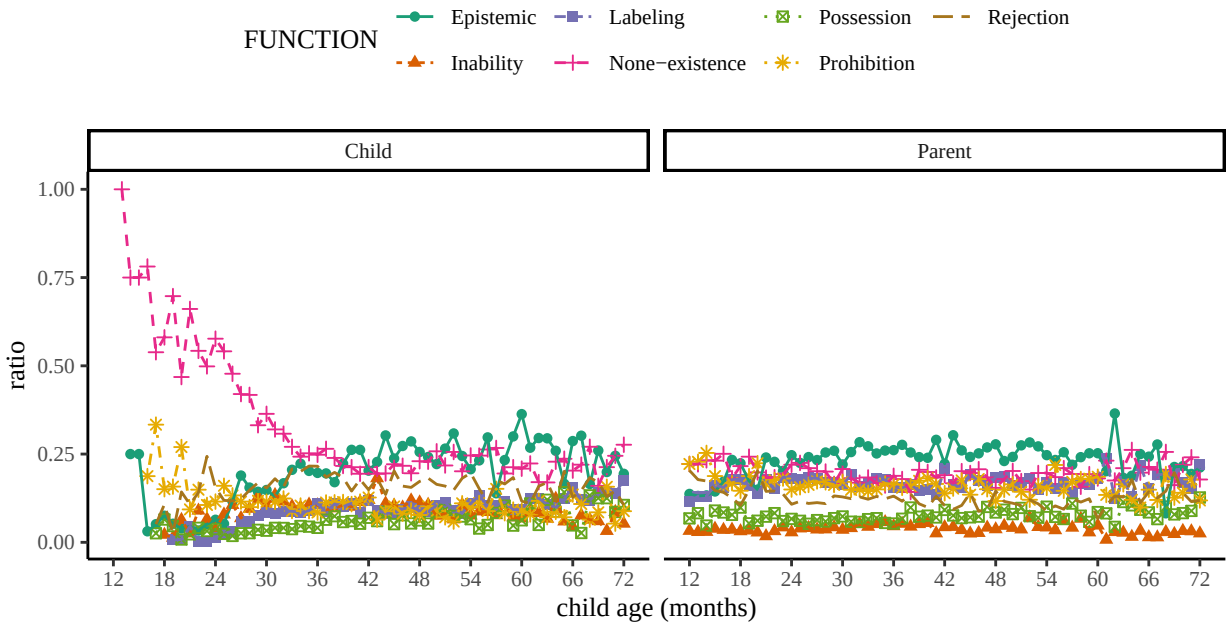

Figure 9: All functions.

frequency of all constructions within a monthly age period. Children's negative constructions bear considerable resemblance to parent speech in terms of the overall production frequency. Early on, the most frequently applied function is non-existence, while the functions with relatively smaller number of occurrences include possession and inability. With that being said, there are a number of observable differences between child and parent utterances with respects to production variability. In both child and parent production, the function that has the highest amount of variability is non-existence (Child: ; Parent:). There is also a considerable difference in variability with prohibitions (Child: ; Parent:) and labeling (Child: ; Parent:).

Discussion

Previous research on the development of negation had used human-annotated small-scale corpus data to study early functions of negation in children's speech. This study presented an automatic and large-scale approach using part of speech tagging and syntactic dependency relations to define and extract relevant constructions for different functions of negation. We presented data on constructions conveying rejection, prohibition, inability, none-existence, Possession, Labeling, and epistemic states in the speech of children and adults. Our results provide preliminary evidence for frequent use of negation in all these constructions between 24 to 36 months of age.

We should add two important limitations of the approach presented here. First we have used data from children's productions to assess the development of negation as a concept. While it is possible that patterns in children's productions reflect their comprehension and semantic development as well, this is not guaranteed. Most importantly, there are production-specific effects (length of utterance, ease of pronunciation, ...) that we have not taken into account yet. Therefore, we can't conclude that early emergence of some functions such as non-existence or prohibition is necessarily conceptual. Second, given this approach's reliance on multiword syntactic constructions to express different communicative functions, we miss early gestural, single-word, or few-word expressions of these functions. To capture these cases, traditional studies with human annoation and classification would be more suitable.

In future work on this project, we plan to investigate the emergence of positive counterparts to our constructions (e.g. *I don't know* vs. *I know*). This would allow us to compare the production trajectory of the negative constructions relative to their positive counterpart. We plan to also focus on the developmental trajectory of individual children to assess individual differences in the development of negation using the methods developed in this study. Lastly, our experiments thus far have concentrated on the utterance level, therefore cases where negations are used as discourse markers were excluded. However, discourse markers operating on a previous turn have important semantic and conceptual roles in the communication between children and parents (e.g. Parent: *do you want some bread?*; Child: *no no no*). In future work, we plan to also include such discourse level negation to our analyses to paint a more clear and thorough picture about the production of negation.

References

Bloom, L. M. (1970). *Language development: Form and function in emerging grammars* (PhD thesis). Columbia University.

Cameron-Faulkner, T., Lieven, E., & Theakston, A. (2007). What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language*, *34*(2), 251.

Choi, S. (1988). The semantic development of negation: A cross-linguistic longitudinal study. *Journal of Child Language*, *15*(3), 517–531.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–173.

MacWhinney, B. (2000). *The childes project: Tools for analyzing talk. Transcription format and programs* (Vol. 1). Psychology Press.

Nordmeyer, A., & Frank, M. C. (2018). Individual variation in children's early production of negation. In *CogSci*.

Pea, R. (1978). *The development of negation in early child language* (PhD thesis). University of Oxford.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of childes transcripts. *Journal of Child Language*, *37*(3), 705–729.

Villiers, P. A. de, & Villiers, J. G. de. (1979). Form and function in the development of sentence negation. *Papers and Reports on Child Language Development*, *17*, 57–64.