Combinatorial Capacity of English Negation in Early Child Language

Anonymous CogSci submission

Abstract

How does negation develop in early child language? Previous work has hypothesized the concept of negation arises via conmbining negative morphemes with different communicative functions such as rejection. However, these specific functions have been challenging to assess given their abstract concepts. Here we examine the developmental trajectories of seven communicative functions served by negation in children's production. Going beyond previous studies that have mainly relied on manual annotations and small sample sizes, we used large-scale corpora of child speech in English. We automatically identified each function based on their distinct syntactic characteristics. Our analyses demonstrated frequent usage of negation in all seven functions between the age of 24 - 36 months; yet there are notable differences in production variability of negation depending on the specifc function examined.

Keywords: negation; combinatorial capacity; corpus anlysis.

Introduction

Negation is an abstract concept 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 inform us to 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? What are the specific communicative functions that negation combines with in early language development?

Regarding these two questions, there have been several influential hypotheses. Starting a century and a half ago, Darwin (1872) thought that negation has roots in the expression of human emotions and desires. He hypothesized 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 this negation function is the first to emerge in children's early speech.

On the other hand, 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", 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.

Additionally, Choi (1988) introduced "prohibition" as one function of 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.

A function similar to prohibition is "inability", in that both involve conceptualizing actions and negating them, possibly interacting with early development of motor control. As discussed in Choi (1988), children tend to communicate their own inability to perform an action with different negative morphemes (e.g. *I can't reach* / *I cannot zip it*). It was suggested that utterances as such emerge after the first phase (i.e. non-existence, rejection, prohibition).

Another 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 instance a child may say: "It's not sharp". Later researchers

formulated 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. A particular sub-function of denial is "labeling", which is often realized as the negation of nominal or adjectival predicates such as "this is not a bunny" or "not red". These utterances are often used to introduce new linguistic labels and in turn would possibly facilitate word learning. Therefore, it is possible that labeling helps the development of abstract negation.

However, previous research on the origins and functions of negation has faced two major issues. First, it mainly relied on human annotation and identification of negative utterances in corpora which is costly and difficult. As a result, most studies focused on only a handful of children and a relatively small sample of their speech. This led to the second caveat, which is that the limited data sizes in previous experiments possibly influenced their claimed findings. 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 differences in how early a negative function is attested. It awaits to be seen whether similar results would hold in much larger sample sizes.

In this study, we investigate the development of negation in child speech of English, focusing on the utterance level. To address the issues raised above, we used of large-scale corpora of child utterances (MacWhinney, 2000). Specifically, we asked: (1) what communicative functions are negation combined with and to what extent? (2) within the same communicative function, does the developmental trajectory differ depending on particular lexical items that negation modifies (e.g. *like* or *want* for rejection)? (3) taking all functions into account, do they share similar developmental characteristics, or would there be function-specific differences?

Experiments

Data and preprocessing

For developmental data of child language in English, we turned to the CHILDES database (MacWhinney, 2000).[1]. We focused on speech produced by children with typical development within the age range of 12 - 72 months. Negative structures were first identified based on whether a structure contains any of the three negative morphemes: *no*, *not* and *n't*. Since the matters of interest are individual utterances and what negation *combines* with, cases consisting of one negative morpheme were excluded (e.g. *no*!). Preprocessing led to a data set of 365,260 negative utterances from a total of 811 children across 56 corpora.

Negation functions

Besides the communicative function of rejection, non-existence, prohibition, inability and denial (labeling), we expanded with two other functions: epistemic negation (Choi, 1988) and possession (see Table 1). For each function, we first characterized the syntactic features of the linguistic structures that the negative morphemes are combined with. Then negative utterances of different communicative functions were automatically extracted in a rule-based fashion, i.e. based on the syntactic characteristics. In order to do this, we resorted to the available morphosyntactic information provided from CHILDES (Sagae, Davis, Lavie, MacWhinney, & Wintner, 2010), such as part-of-speech (POS) tags as well as grammatical or syntactic dependency relations. After extracting negative utterances, the developmental trajectories of different constructions of interest were analyzed.

In what follows we introduce each construction and present the results. While our focus is child utterances, we used parents' speech as references and therefore our plots often contrast the relative frequency (or ratio) of these constructions between children's and parents' production at the corresponding age of the child.

Rejection For the function of rejection, we examined cases where the lemma 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*
(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

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

Function	Linguistic Composition	Examples
Rejection	with *like* or *want*	*I not like it*, *not want it*
Non-existence	expletives; with a nominal; *no more*	*there is no soup*; *no juice*; *no more milk*
Prohibition	with imperative subjectless *do*	*do not spill milk*
Inability	with modal *can*	*I cannot zip it*
Labeling	modifying nominal or adjectival predicatives	*that's not a crocodile*; *it's no interesting*
Epistemic negation	with *know*, *think*, *remember*	*I not know*
Possession	with *have*; or possesive pronouns	*not have the toy*; *not mine*

Table 1: Communicative functions of negation in early child language of English.

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.

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

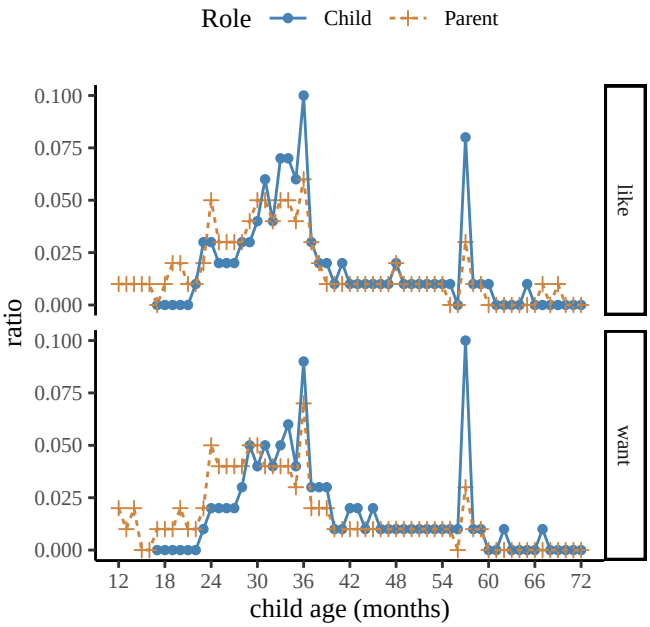

Figure 1: Rejection

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 1, 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.

Non-existence For the function of non-existence, we extracted utterances that either have expletives marked by *there* (e.g. (5)), 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. (6)). With utterances such as (6) in particular, in order to not confuse with the function of labeling (see below), 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).

(5) *there's no water*

(6) *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. As shown in Figure 2, children began increasing their use of the negative morphemes to express non-existence around the age of 22 months, which is similar to that for the function of rejection and in opposition to what was initially proposed by Bloom (1970).

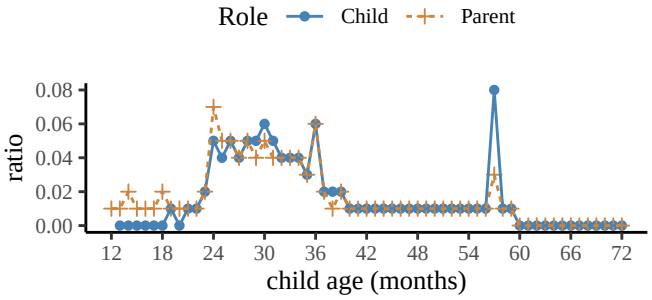

Figure 2: Non-existence

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. (7)). In certain cases the negative morphemes and the auxiliary together modifies a head verb. For instances as such, in order to not overlap with rejection, non-existence and epistemic nega-

tion 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. Based on Figure 3, children started more regular usage of negative morphemes to articulate prohibition around the age of 23 months. This finding also corresponds to the proposal from (Choi, 1988), which suggested that prohibition emerges around similar times compared to rejection and non-existence.

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

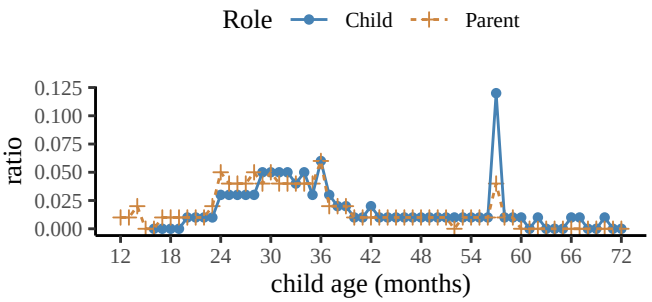

Figure 3: Prohibition

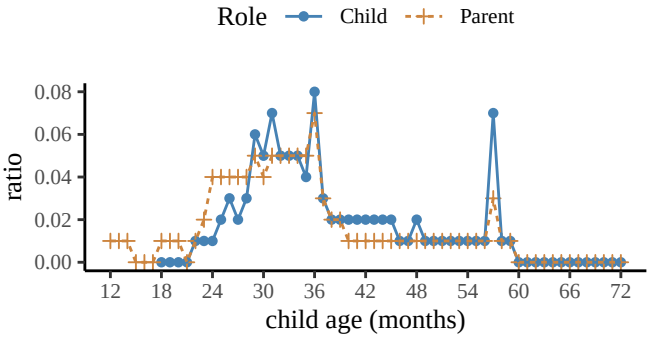

Figure 4: Inability

Inability We analyzed instances where the negative morphemes co-occur with the auxiliary *can* (*can* and *could*; e.g. (8)) for the function of inability. 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 (e.g. *can't play*) or do not contain a subject other than I (e.g. *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).

(8) *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 4, the developmental trajectory of this function is similar to that for prohibition (Choi, 1988), 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.

Labeling To capture the function of denials, in parituclar the instances of labeling, we concentrated on cases where negative morphemes are adopted to indicate the identity (e.g. (9)), and/or characteristics (e.g. (10)) of a predicative nominal. In addition, we also included instances where the negative morphemes are used to modify a predicative adjective (e.g. (11)). Following these criteria, utterances where the negative morpheme is modifying 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).

(9) *that's not a farmer*

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

(11) *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 speech. Based on results from Figure 5, 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 more frequently around the age of 24 months.

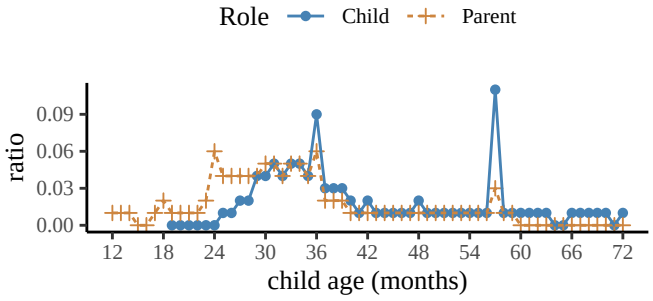

Figure 5: Language learning via labeling

Epistemic Negation The function of epistemic negation was originally discussed by Choi (1988). Although there has been no proposal for negation originating in children's understanding of their own or others' epistemic/mental states, previous studies have reported many instances where the negative morphemes are combined with mental state verbs such as *know*, *think*, and *remember*. Here we focused on these

three verbs and analyzed utterances that articulate the concept of unknowing (e.g. (12)) or uncertainty (e.g. (13)). The verbs in these cases are 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 state of another speaker (e.g. (14)) were also selected. This led to a subset of 32,793 utterances in total (Child: 10,389; Parent: 22,404).

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

(13) *I don't think so*

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

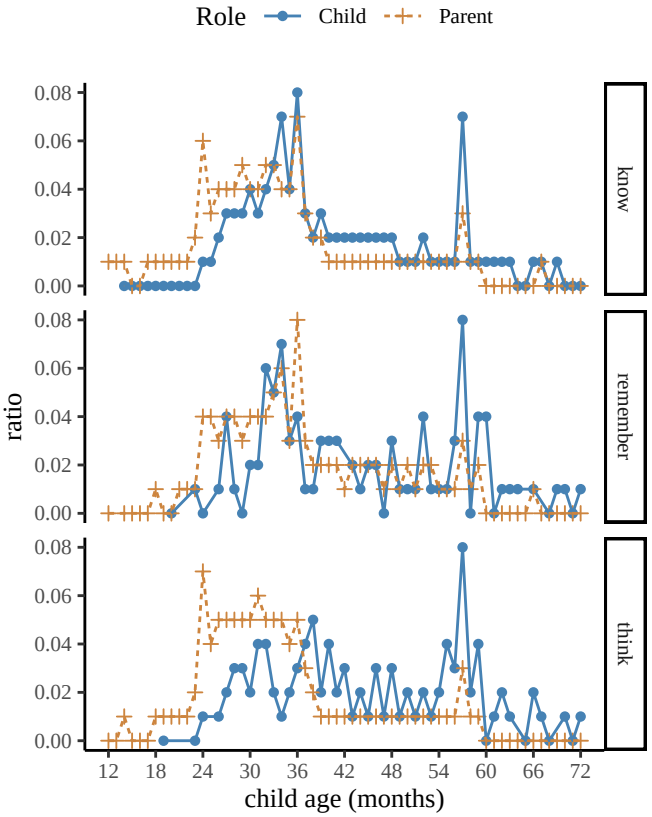

Figure 6: Epistemic

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 6, 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.

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*). The number 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 7, is comparable to the other functions, where children started to combine negative morphemes around the age of 22 months..

(15) *not mine*

(16) *I don't have it*

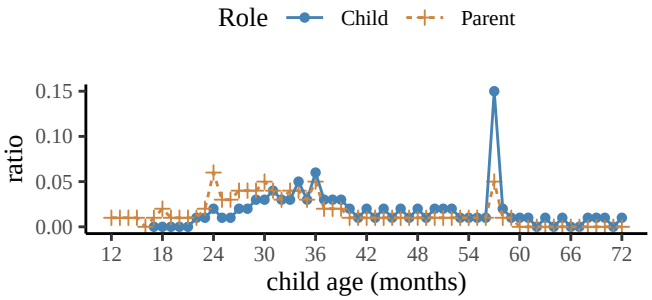

Figure 7: Possession

Overall Figure 8 shows the developmental trajectory for all analyzed negative functions. The y-axis is the relative frequency of each construction relative to the frequency of all constructions within a monthly age period. Children's negative utterances 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 with respects to the production variability of individual function. Comparing across functions, the usage of negative morphemes for non-existence is the most variable or the least consistent, with there being more variability in child speech than parent speech (Child: 1.61; Parent: 1.08). On the other hand, the opposite patterns are observed for prohibitions (Child: 0.62; Parent: 0.92) and labeling (Child: 0.49; Parent: 0.94).

Discussion

Previous research on the development of negation have used human-annotated small-scale corpora to study early communicative functions of negation in children's speech. By contrast, this study presented an automatic and large-scale approach using POS and syntactic dependency relations to define and extract relevant structures. In particular, we presented analyses on utterances combined with negation to convey rejection, non-existence, prohibition, inability, labeling,

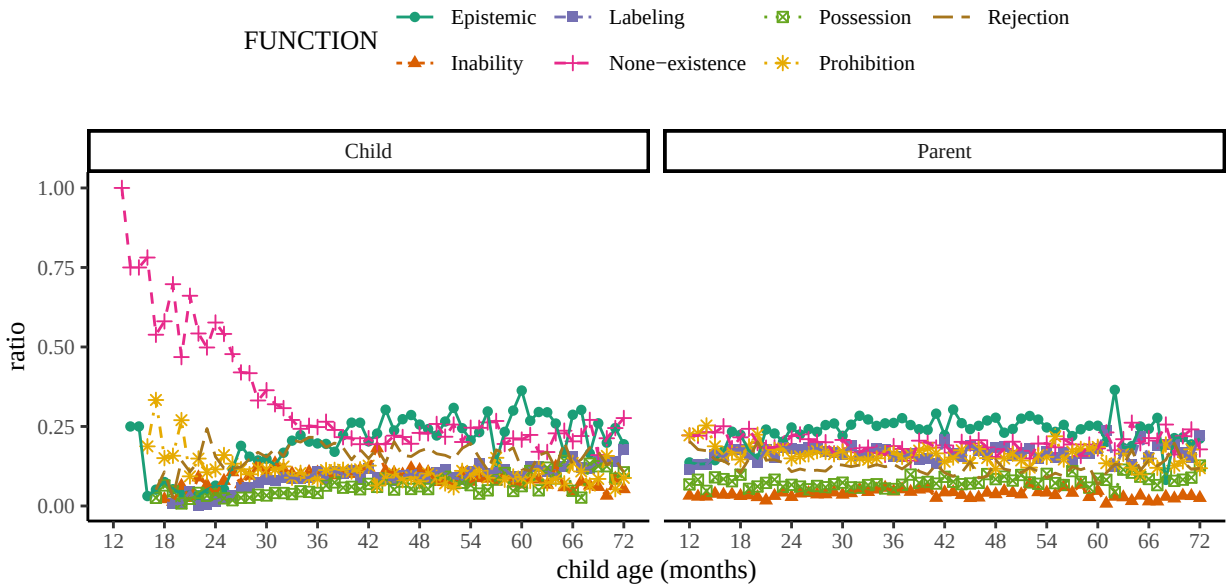

Figure 8: All functions.

epistemic states and possession. Overall our results provided evidence that negation is frequently combined with all these functions between the age of 24 to 36 months, and that there are quantifiable differences in the production variability of the negative morphemes depending on the specific function.

One important limitation of our approach is that we have used data from children's production to assess the development of negation as a concept. While it is possible that patterns in children's production reflect their comprehension and semantic development as well, this is not guaranteed. Most importantly, there are production-specific effects (e.g. length of utterance, ease of pronunciation) that we have not taken into account yet. Thus, results from production data alone do not suffice to conclude that early development of negation is necessarily conceptual. Further experiments testing children's comprehension of negative utterances that serve different communicative functions are necessary to better understand the origin of negation more thoroughly.

In future work, we plan to investigate the production trajectory of positive counterparts to our negative structures (e.g. *I don't know* vs. *I know*). We intend to also apply our analyses to individual children in order to assess individual differences. Lastly, our experiments thus far have concentrated on multi-word syntactic structures at the utterance level, therefore cases where negations are used as discourse markers in single-word or few-word expressions were excluded. However, discourse markers as responses to a previous utterance(s) 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*). Inclusions of negative structures at the discourse level, along with more carefully manual inspections, would allow us to paint a more clear picture about the development 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.