

Combinatorial Capacity of English Negation in Child Language

Anonymous CogSci submission

Abstract

Negation is very important for language and thought. How does it develop in the language of children? There has been many guesses like rejection, non-existence, denial, etc. but it has been hard to assess because these concepts are vague. 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: Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

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 that otherwise cannot be labeled. It can help us regulate others' actions in a sign like "no mask, no entry". It can also communicate our deepest wants and dislikes. But how does this crucial abstract concept emerge in humans? Does language play a role in its emergence or does language simply adopt it for communication?

There has been several influential hypotheses on the conceptual origin of negation.

In this paper, we address the same issues with a slightly different approach. We start with the widely accepted assumption that negation is a higher order operator or function, operating on lower level concepts. The question we ask is: what type of concepts does linguistic negation operate on in early child language? Do we find negation starting in a limited conceptual domain and then expanding to others? Or do we find it operating across different conceptual domains as early as we can attest it?

Darwin (1998) 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. Considering early functions of negative morphemes like *no*, many researchers proposed that children use them to "reject" or "refuse" (Bloom, 1970; Choi, 1988; Pea, 1978). For example, they may say

"no" when asked "do you want juice?", say "not want it", or say "don't like it". Pea (1978) proposed that this function of negation is the first to emerge in children.

Motor control: prohibition (do not spill milk), inability (I cannot zip it)

Bloom (1970) suggested that the first function of negation in children's speech is to express non-existence. Relates to children's development of object permanence.

Perceptual: non-existence (no juice, no more milk, no fish in the bathroom, I do not have underpants), failure, Locatives (no in there, daddy was not on the phone), non-events (the dog not barking)

A third possible domain and path to the acquisition of negation is language itself. Word learning places its own constraints on the conceptual space. One possibility is that negation develops, and is aided by the act of labeling and categorizing objects and actions for linguistic communication. This function would manifest itself in labeling acts with nominal predicates such as "this is not a bunny", "not red", or "this isn't a reptile".

There has been no proposal for negation originating in the child's understanding of her own or other's epistemic states. In fact, most development theory of mind accounts assume that this ability emerges later in children. However, many corpus instances of negation modify mental state verbs such as *know*, *think*, and *remember* (e.g. "I not know"). Therefore, we also report the prevalence and emergence of such cases.

Caveat on production vs comprehension.

Focusing on the utterance level in this study

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 utterances produced by children with typical development within the age range of 12 - 72 months, then extracted cases with any of the three negation markers that are of interest 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

¹Code and data for our study are in quarantine at <https://somedwhereonearth>.

markers could *combine* with. Preprocessing led to a data set of 365,260 utterances with negation structures from a total of 811 children across 56 corpora.

Figure @ref(fig:speaker_stats)

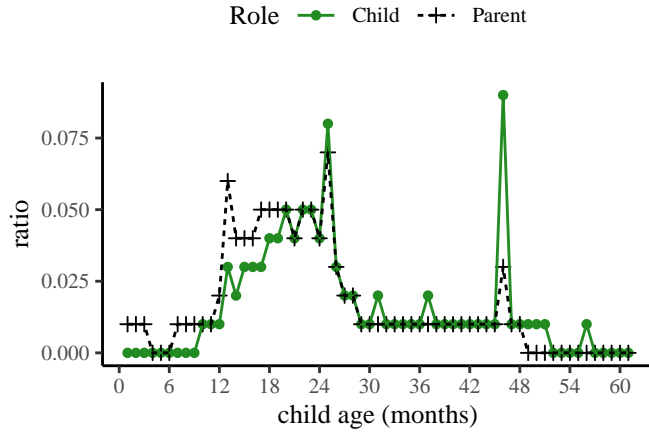


Figure 1: Distribution of the number of utterances with negative morphemes in child and parent speech.

Negation domains

In this section, we describe in details our automatic extraction of syntactic structures that express different types of negation concepts. 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 take advantage of information as such when automatically identifying our constructions of interest. An utterance with negative morpheme(s) was only considered when the negative morphemes has either a POS of *neg* or *qn*, the latter of which was mainly for cases with *no* as a quantifier. Furthermore, the syntactic functions and relations of the negative morphemes should not be enumeration (*no no no*), communicators or discourse markers

After extracting all instances with negative morphemes, the developmental trajectory of each construction type as described in the previous section was analyzed. While the matter of interest here is child speech, we also compared patterns in child production to those in parent speech as references at the corresponding age of the child. Then we combined the development of all construction types for analysis as a whole

Emotion In order to investigate utterances that express emotions, particularly the concept of rejections, we focused on specific 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

²Besides using the provided POS and syntactic dependency information in CHILDES, we also experimented with the state-of-the-art parser from Stanza [atq-etal-2020-stanza]. There were no worth noting differences in the analyses for the negation constructions.

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 emotions (e.g. (1)) or their (in)ability to do so, 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 emotions 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, for each domain, we calculated the relative ratio of (i) each of the three negative morphemes overall; (ii) usage of negative morphemes with the two different head verbs; (iii) utterances expression emotion with the three negative morphemes at different ages of the child. In both child and parent speech, when expressing emotion 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 are co-occur with *want* more frequently.

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 did not start using negative morphemes in this domain until the age of 7 months; their usage of these morphemes for the concept of rejection is the most frequent during the age range of 13 - 21 months, then gradually decreases as they age.

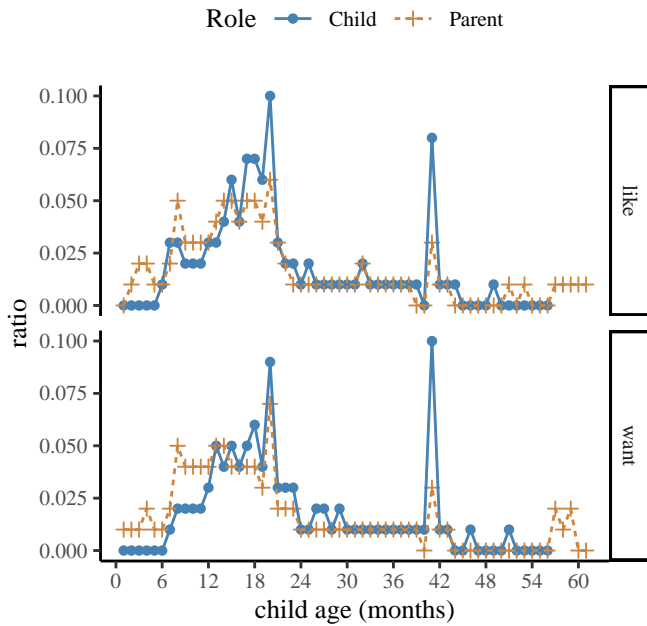


Figure 2: Emotion

Theory of mind With regards to the domain of theory of mind, we attended to cases that express epistemic position. Specifically, we focused on utterances that articulate the concept of unknowing (e.g. (5)) or uncertainty (e.g. (6)). The cases that were subject to analyses here included either *know*, *remember* or *think* as the head verb, modified by the negative morphemes or the combination of the negative morphemes with auxiliaries. By these search criteria, instances where the speaker inquires about or describes the 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 in expressing epistemic position 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 describes the state of unknowing, which is indicated mainly by the verb *know*. Given results from Figure @ref{fig:epistemic}, there does not seem to be a quite consistent developmental trajectory of child speech in the domain of theory of mind. While regardless of the head verb, children appeared to start applying the negative morphemes in this domain around the age of 4. Nevertheless, there seems to be more variability in expressing an epistemic position with *remember* and *think* as the children age. By contrast, the pattern for instances with *know* gradually decreases after the age of 16 months.

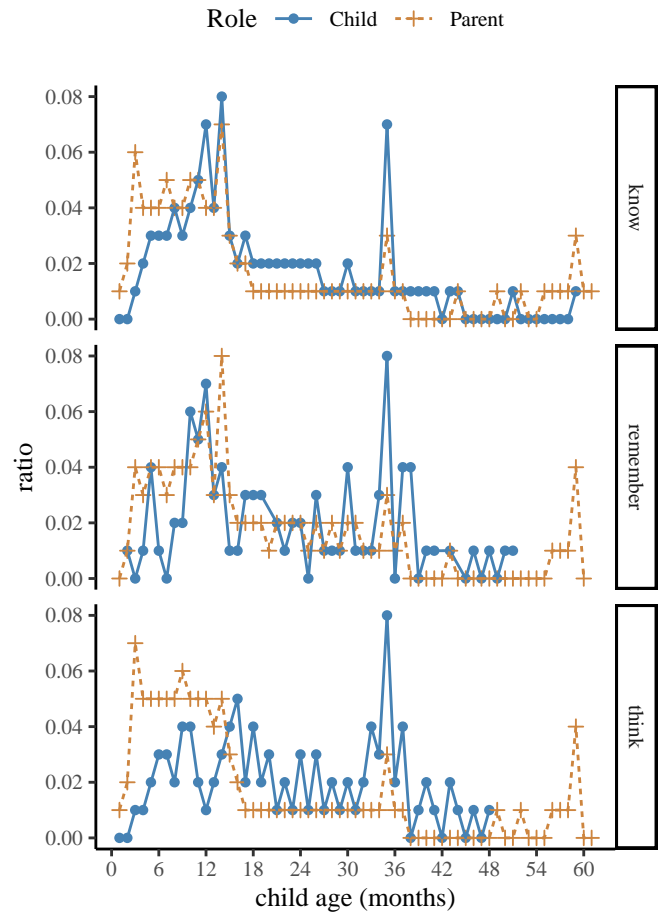


Figure 3: Theory of mind

Motor control For utterances that articulate the concept of motor control, we focused on two individual aspects with different communicative functions. The first one includes cases that indicate prohibition (e.g. (8)), and the second one contain cases that articulate inability (e.g. (9)). For the former, we analyzed cases where the negative morphemes are combined with the auxiliary verb *do* (*do*, *does*, *did*) and the auxiliary does not take any subject; whereas for the latter, we analyzed cases where the negative morphemes co-occur with the auxiliary *can* (*can* and *could*). For both functions, the negative morphemes and the auxiliary together modifies a head verb. In order to not overlap with the domain of Emotion, Theory of Mind and Perception (see below), our search excluded cases where the head verb has any of the following lemma forms: *like*, *want*, *know*, *think*, *remember*, *have*.

For inability in particular, cases that do not have a subject (*can't play*) or 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 excluded instances as such. In other words, when searching for utterances that articulate inability, we restricted our analyses only to cases with a subject *I*. This led to a subset of 235 utterances (Child: 113;

Parent: 122).

(8) *do: don't blame Charlotte*

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

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:motor_control}, the developmental trajectory of using the negative morphemes in the domain of motor control is similar to that in the domain of emotion. Children started combining negative morphemes in syntactic structures that express prohibition and inability around the age of 6 months, then gradually increases until the age of 22 months.

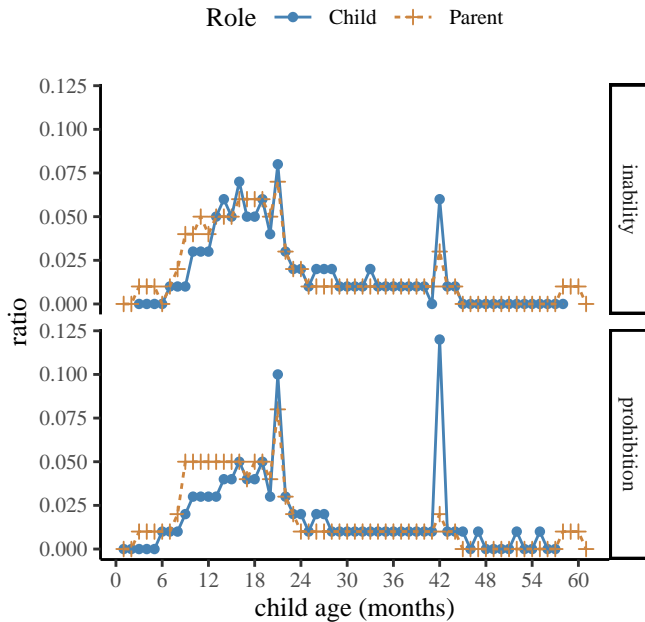


Figure 4: Motor control

Language learning Within the domain of language learning, we concentrated on cases where negative morphemes are adopted to label the identity (e.g. (10)), and/or characteristics (e.g. (11)) of a predicative nominal. In addition, we also included instances where the negators are used to modify a predicative adjective (e.g. (12)). 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 contain 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 27,689 utterances (Child: 55 utterances; Parent: 61 utterances).

(10) *that's not a farmer*

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

(12) *It's no good*

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.

Comparing the three negative morphemes, the most frequently used is *not* in both child and parent speech. 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 around the age of 7 months and their usage as such became more regular around the age of 10 months. However, the frequency of applying the negative morphemes in the language learning domain began to decrease around the age of 18 months.

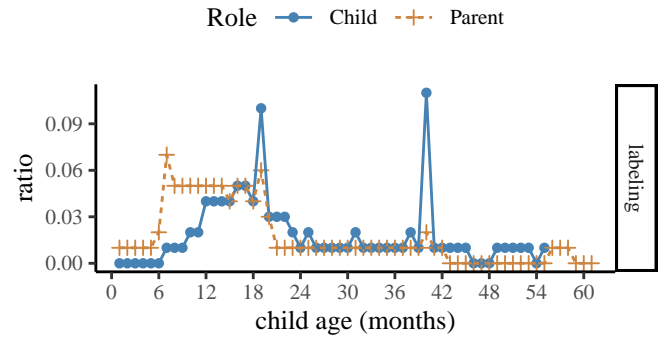


Figure 5: Language learning

Perception The last domain to be analyzed here is Perception, with two particular functions. First one is indication of non-existence, while the second one denotes possession. For expressions 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)). For expressions of possession, we selected utterances where the negative morphemes are modifying a possessive pronoun, 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. (15)). With utterances such (14) and (15), in order to not confuse with the domain of Language Learning, we did not include any cases

where the syntactic head of the negative morphemes is a predicate of a copula verb. As a result, our search did not find any utterances that express the function of possession in this case, and the total of utterances that were subjected to analysis in this domain is 23,408 (Child: 55; Parent: 61).

- (13) *there's no water*
- (14) *no (more) candy / not in your mouth*
- (15) *not mine / I don't have it*

In both child and parent speech, the most frequently occurred negative morphemes in the domain of Perception is *no*. 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:perception}, the developmental trajectory of using the negative morphemes in the domain of Perception is also comparable to previous domains. Children began applying the negative morphemes to this domain around the age of 8 months and their usage gradually increased after the age of 11 months. Similarly to the domain of Language Learning, the frequency of using the negative morphemes in indicating non-existence started to decrease around the age of 19 months.

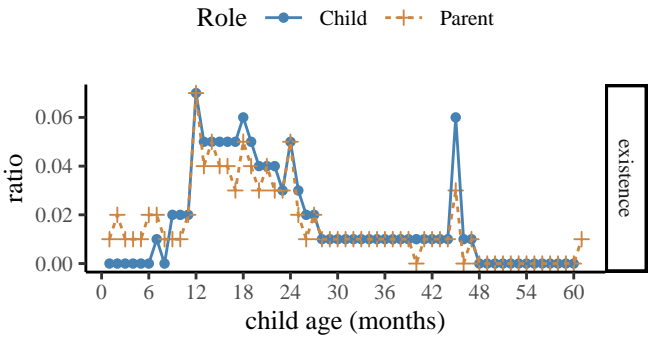


Figure 6: Perception

An overall look at all domains

Discussion

References

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

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

Darwin, C. (1998). *The expression of the emotions in man and animals*. John Murray.

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

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 chldes transcripts. *Journal of Child Language*, 37(3), 705–729.

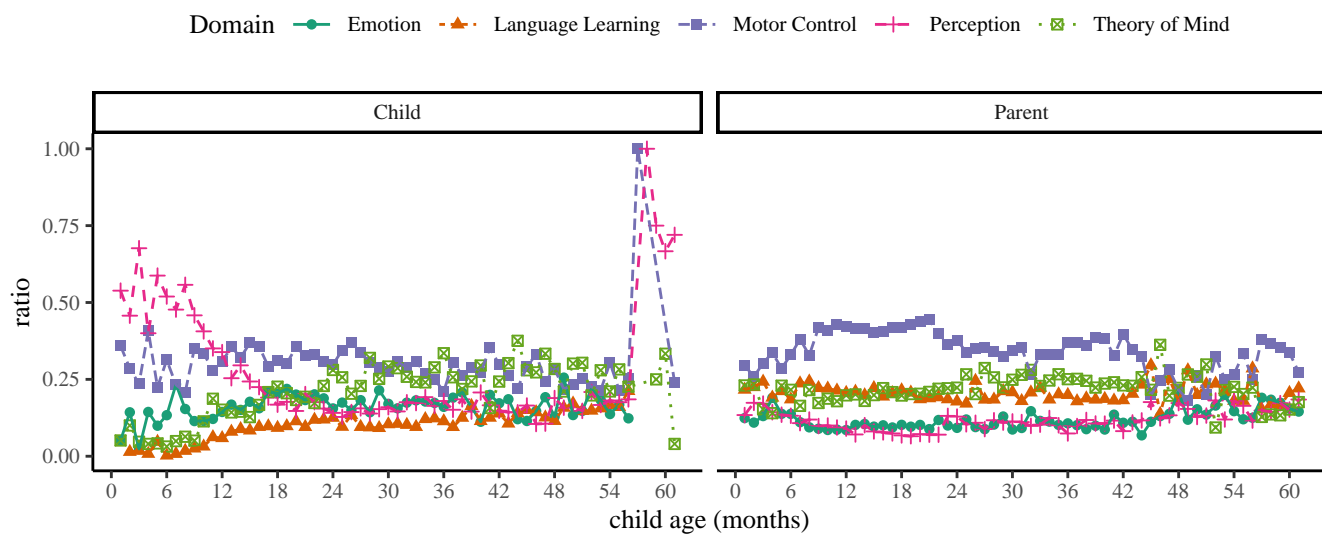


Figure 7: ALI domains.