Learning modal force: evidence from children's production and input*

Anouk Dieuleveut[1], Annemarie van Dooren[1],
Ailís Cournane[2] and Valentine Hacquard[1]
[1] University of Maryland, USA
[2] New York University, USA
adieulev@umd.edu, avdooren@umd.edu, cournane@nyu.edu,
hacquard@umd.edu

Abstract

This paper investigates when and how children figure out that possibility modals express possibility, and necessity modals, necessity. Given that *necessary p* entails *possible p*, what prevents children from hypothesizing possibility meanings for necessity modals? We argue that this *entailment problem* is not a psychological one. On the basis of a corpus study of the modal productions of 2-year-old English children and their mothers and two *Human Simulation Paradigm* experiments (Gillette et al. 1999), we show that children can use cues from the conversational context in which modals are used to learn force, and do not need to rely on negative environments, nor on a bias for necessity meanings.

1 Introduction

This paper investigates how children figure out the force of the modals in their language: that *can* or *might* in (1a) express possibility, whereas *must* or *have to* in (1b) express necessity. English modals lexically encode force: they express either possibility or necessity. Standard semantic accounts analyze them as quantifiers over possible worlds: possibility modals introduce existential quantification, whereas necessity modals introduce universal quantification (Lyons 1977; Kratzer 1977). Note that the same modal can be used to express different types, or *flavors* of modality: (1b) can be used to express an epistemic necessity (a likelihood), or various types of 'root' (*i.e.* non epistemic) necessities: deontic (obligation), bouletic (desire), or teleological (need).

(1) a. Jo *can/may/might...* draw. possibility
b. Jo *must/should/have to/...* draw. necessity

Experimental literature on modal comprehension suggests that children struggle with modal force until at least age 4: they tend to accept possibility modals in necessity situations, and necessity modals in possibility situations (Noveck 2001; Ozturk and Papafragou 2015). Typically, these studies attribute errors to reasoning difficulties: children over-accept possibility modals in necessity situations because of difficulties reasoning about when a stronger modal would be more appropriate, and necessity modals in possibility situations because of difficulties reasoning about open possibilities. However, they take for granted that children already know modals' underlying force. In this paper, we address when and how children figure out force by investigating modal talk to and by young children: quantitatively,

* We thank our research assistants J. Lee, D. Whitcomb and M. Pizzo, the members of UMD/NYU Modality group, S-Lab and Acquisition Lab, especially A. Williams, J. Lidz and Y. Yang. Our project is supported by NSF grant #BCS-1551628.

through corpus analyses, and qualitatively, using two Human Simulation Paradigm experiments testing how well adults can guess the force of child and adult modal uses from the conversational context.

How do children distinguish necessity from possibility modals? Syntactic information might help narrow candidate meanings to modal meanings (in the spirit of Landau and Gleitman's 1985 *syntactic bootstrapping hypothesis*), but it cannot help distinguish force. Cues from the physical context are also bound to be limited, since modals express non-actual concepts, with few physical correlates. It thus seems that to figure out modal force, children must rely on cues from the conversational context. But how informative is context about modal force? How easily identifiable are possibility meanings when possibility modals are used, and necessity meanings when necessity modals are used? A second issue might be that necessity entails possibility (the *entailment* or *set/subset problem*).[1] If a possibility meaning is always true when a necessity modal is used, what prevents children from hypothesizing a possibility meaning for a necessity modal?

One solution to the *entailment problem* would be to use evidence from negative (or more generally downward-entailing) environments (Gualmini and Schwarz 2009),[2] as they reverse the direction of logical entailments. A potential issue with this solution is that some necessity modals (*e.g. must*) scope over negation (Iatridou and Zeijlstra 2013). Such cases may be particularly misleading, and reinforce a possibility meaning, as *must not* (*necessary not*) is truth-conditionally equivalent to *cannot* (*not possible*). Another solution would be to equip learners with a bias towards strong meanings, in the spirit of Berwick's (1985) *strongest meaning hypothesis*. We argue that neither are necessary to solve the problem, as the conversational context in which modals are used provide enough evidence as to their force. Such evidence involves situational cues (*e.g.* who the interlocutors are), pragmatic cues (what the speaker is trying to achieve, in particular, giving orders or permissions), and cues from world knowledge. We show that modal force can be inferred on the basis of these cues alone.

In **section 2**, we provide a quantitative and qualitative assessment of the modals children hear in their *maternal input*. We use the Manchester Corpus of UK English (Theakston et al., 2001) on CHILDES (MacWhinney 2000), and a *Human Simulation Paradigm* (HSP) study (Gillette et al. 1999). This HSP allows us to give a general measure of the informativity of natural conversational contexts about force. In **section 3**, we provide an assessment of *children's* modals productions, using the same methods: first a corpus study, then an HSP study which assesses the extent to which children's modal use is adult-like, a novel way of approaching whether children have adult-like grammars. **Section 4** discusses further implications of our findings.

2 Children's modal input

2.1 Input corpus study

The Manchester Corpus (Theakston et al., 2001) consists of 12 child-mother pairs (6 females; age range: 1;09-3;00) recorded in unstructured play sessions. We chose this corpus for its relative density, uniformity of sampling, and early age range, and focused on the period between 2;00 and 3;00 y.o. All utterances containing modals[3] (26,598 of 564,625 total utterances; adult: 20,755; child: 5,842; excluding repetitions (6.6%): adult: 19,986; child: 4,844) were coded for force (possibility vs. necessity), presence of negative element, and flavor (epistemic vs. root).

[1] We borrow the term from related discussions involving nouns or quantifiers (see *e.g.* Xu and Tenenbaum 2007).

[2] Gualmini and Schwarz propose this not for modals, but for any entailment problem. See also Musolino et al. (2019).

[3] Modal: possibility: *can, could, might, may; able to;* necessity: *must, should, need, have to, got to, supposed to, need to*. We do not differentiate amongst subtypes of root flavors, and exclude future modals, whose force is a matter of debate.

Overall, possibility modals are more frequent than necessity modals in adult speech (72.5% of adults' modal utterances) (**Table 1**). Epistemic uses of modals are overall very rare (**Table 2**). Possibility modals co-occur with negation more frequently than necessity modals (possibility: 20.9% vs. necessity: 10.1%). Most cases of necessity modals occurring with negation involve a modal overscoping negation ({*must, should, ought to*}: 19.4% vs. {*have to, got to, need to, supposed to*}: 7.4%). Negation is significantly more frequent on root than on epistemic modals (epistemic: 4.6% negated, vs. root: 19.1%). Modals rarely occur in other negative environments. Whether children can make use of the relatively infrequent negated necessity cases depends on whatever expectations they have about how modals scope relative to negation, and how clearly meanings of non-necessity or impossibility get conveyed with a negated modal, which we test in the HSP.

	ADULT (n=19,986)		ADULT (n=18,853)			
	all		no negation		negation	
possibility	14,491	72.5%	10,672	79.1%	2,828	20.9%
necessity	5,495	27.5%	4,814	89.9%	539	10.1%

Table 1: Counts and percentages of modal uses by force with and without negation, for adults (repetitions excluded: 3.7% of the data)

	all		no negation		negation	
root	17,190	91.2%	13,896	80.9%	3,293	19.1%
possibility	12,175	64.6%	9,414	77.3%	2,761	22.6%
necessity	5,015	26.6%	4,482	89.4%	533	10.5%
epistemic	1,662	8.8%	1,590	95.4%	73	4.6%
possibility	1,324	7.0%	1,257	94.9%	67	5.0%
necessity	341	1.8%	332	97.3%	6	2.6%

Table 2: Counts and percentages of modal uses, by force, flavor and negation, for adults (n=18,853)

2.2 Human Simulation Paradigm Study

To assess the general informativity of natural conversational contexts about force, we use a variant of the *Human Simulation Paradigm* (Gillette et al. 1999), using dialogue contexts extracted from the Manchester corpus. We investigate whether participants can guess the force of a modal based on excerpts of conversations in which it appears, and whether the context is equally informative for necessity and possibility modals, for epistemic and root modals, and for negative vs. positive contexts.

Procedure – The experiment was run online on Alex Drummond's IBEX Farm. Participants recruited via Amazon Mechanical Turk were asked to guess a redacted modal in a dialogue between a child and her mother by choosing between two options, corresponding either to a possibility or a necessity modal, as illustrated in **Figure 1**. All dialogue contexts consisted of the modal sentence with a blank and the 7 preceding utterances. There was first a short training where participants had to choose between the definite vs. indefinite article (*the* vs. *a*) (3 examples with feedback), and then the test phase, without feedback. Overall, each participant had to judge 40 different dialogues (20 trials: 10 possibility, 10 necessity; 20 controls using tense: 10 past, 10 future). The 20 trials were selected randomly from a list of 40 contexts originally extracted from the corpus; the 20 controls were always the same.

Conditions – We tested force (possibility vs. necessity) within participants, and flavor (root vs. epistemic) and negation (present vs. absent) between participants. Negation was tested only for root flavor, because negated epistemics were too rare in the corpus. **Table 3** presents the experiment design.

Test condition (between participants)		Modal lemma (within participants) possibility	necessity
EPI-AFF (epistemic affirmative)		*might*	*must*
ROOT-AFF (root affirmative)[4]	**ROOT-AFF-1**	*can*	*must*
	ROOT-AFF-2	*can/able*	*have to*
ROOT-NEG (root negative)		*can't/not able*	*not have to*

Table 3: Summary of experimental conditions (Human Simulation Paradigm)

Material and participants – 160 contexts (2*20 per condition) were extracted from the corpus for the different modals (*can, able, might, must, have to)*. ***Exclusion criteria.*** We excluded contexts where the target modal was used in preceding utterances. Contexts were not excluded when the adult (or the child) used another non-target modal. ***Controls.*** Participants had to choose between future and past (*e.g.* [*saw*] vs. [*will see*]). Importantly, the correct answer was not always guessable based on the target sentence only: this required participants to read the entire dialogue. Extraction procedure was the same as for targets. 289 participants were recruited on Amazon Mechanical Turk (4 groups (between participants): ROOT-AFF-1: 73, ROOT-AFF-2: 72; ROOT-NEG: 73; EPI-AFF: 71; language: US English; 156 females, mean age = 40.6 y.o.). We removed from analysis 8 participants (2.8%) who were less than 75% accurate on controls. We thus present results for 281 participants (ROOT-AFF-1: 71, ROOT-AFF-2: 69; ROOT-NEG: 70; EPI-AFF: 71).

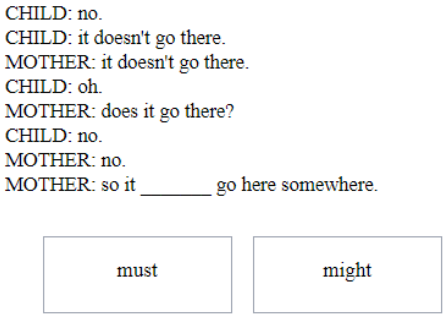

Figure 1 Human Simulation Paradigm stimuli: example trial (EPI-AFF, *must*)

Results – Data analyses were conducted using R (R Core Team, 2013), using the package lme4 (Bates *et al*. 2014a, 2014b). Overall, participants were highly accurate at guessing modal force (see **Table 4**: general mean accuracy 79.9%). We first ran binomial tests to see whether they differ from chance for each condition. Participants' accuracy significantly differs from chance in each condition. Their lowest performance is found for ROOT-NEG necessity modals (*e.g. not have to*) (61.3%).[5] **Force** (possibility vs. necessity). To test whether there was an effect of Force, we used binomial linear mixed effects models, built with a maximal random effect structure, testing Accuracy with Subject and Item as random factors (following Barr *et al.*, 2013),[6] first overall and then for each condition. We find a general effect of Force, in the direction of a higher accuracy for possibility contexts ($\chi^2(1)=20.49$, $p=5.9e\text{-}6$***). Restricting to each comparison group, we find a significant effect in ROOT-AFF-1 ($\chi^2(1)=61.1$, $p=5.5e\text{-}15$***) and ROOT-NEG ($\chi^2(1)=15.6$, $p=7.8e\text{-}05$***), again in the direction of a higher accuracy for possibility contexts, but not for ROOT-AFF-2 ($\chi^2(1)=6e\text{-}04$, $p=0.98$ *(NS)*) and EPI-AFF ($\chi^2(1)=3.73$, $p=.053$ *(NS)*). **Negation**: we compared ROOT-AFF-2 and ROOT-NEG (these conditions

[4] We implemented two versions of the **ROOT-AFF** condition: **ROOT-AFF-1** (*can* vs. *must*) allowed us to keep syntactic category of both options identical, **ROOT-AFF-2** (*can/able to* vs. *have to*) to avoid concerns related to the formality of *must* for US English speakers. In cases where *have to* was tensed, we used *able to* as the alternative to avoid losing tense information.

[5] Accuracy for controls was very high (94.6%). There was no difference in accuracy between groups.

[6] We sometimes had to step back to random-intercepts-only models when the model failed to converge with the full random-effects specification.

include the same lemmas). We find a significant effect of negation on necessity modals, which leads to lower accuracy (*have to* vs. *not-have to*: $\chi^2(1)=6.45$, $p=0.011$*). On possibility modals, negation leads to higher accuracy, but the effect is not significant (*can* vs. *can't*: $\chi^2(1)=2.29$, $p=0.13$ (NS)). We find a strong interaction effect between Force and Negation (**Interaction Force*Neg**: $\chi^2(1)=7.9$, $p=0.0047$**). **Flavor** (epistemic vs. root): no general effect of flavor ($\chi^2(1)=0.11$, $p=0.74$ *(NS)*).

	Mean accuracy [7] (se)		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	91.7% (0.027)	71.7% (0.054)	p <.001*** 95% CI [0.90, 0.94]	p <.001*** 95% CI [0.68, 0.75]
ROOT-AFF-2	81.5% (0.053)	82.0% (0.052)	p <.001*** 95% CI [0.79, 0.85]	p <.001*** 95% CI [0.79, 0.84]
ROOT-NEG	89.5% (0.031)	61.3% (0.065)	p <.001*** 95% CI [0.88, 0.92]	p = 8.95e-08 95% CI [0.56, 0.64]
EPI-AFF	87.2% (0.028)	74.3% (0.049)	p <.001*** 95% CI [0.84, 0.90]	p <.001*** 95% CI [0.71, 0.77]
all	87.5% (0.018)	72.3% (0.028)		
ALL	79.9% (0.018)			

Table 4: Accuracy by condition (adult, n = 281*10)

Analysis by contexts (post-hoc) – To get a sense of the kinds of contextual cues that were particularly helpful, we explored the contexts that led to lowest and highest accuracy, for possibility and necessity modals. This informal analysis pointed out two factors, depending on flavor. For root modals, cases where the proposition expressed by the prejacent was effortful or undesirable seem to lead to high accuracy for necessity modals. For epistemic modals, we found high accuracy for necessity modals in contexts that made salient strong evidence for the prejacent.

Discussion – Our results show that the conversational context is overall informative about force: adult participants can guess the force of the modal accurately, for both flavors (mean accuracy: 79.9%). This suggests that there are useful cues in the conversational context: if children are sensitive to these cues, they may not need to rely on negation, nor a bias towards necessity meanings. We find a general effect of **force**, with higher accuracy for possibility modals (general accuracy: possibility: 87.5%; necessity: 72.3%). This could be taken as showing that possibility contexts are more informative than necessity contexts, but the effect is found in only 2 sub-conditions, ROOT-AFF-1 and ROOT-NEG (it is near-significant in EPI-AFF: $\chi^2(1) = 3.73$, $p = .053$), but not significant in ROOT-AFF-2). This higher accuracy in possibility contexts might reflect a tendency to answer with possibility modals by default, maybe because of the relative frequency of possibility and necessity modals,[8] which itself might be due to alternative ways speakers can express necessity (e.g., imperatives for deontic necessity). As for **negation**, we find an opposite effect of negation on possibility and necessity modals: while negation leads to slightly higher accuracy for possibility modals (*can't*: 89.5% vs. *can*: 81.5%), it leads to significantly lower accuracy for necessity modals (*don't have to:* 61.3% *vs. have to:* 82.0%) (interaction effect Force*Negation). This suggests that negation may not be useful for necessity modals. Thus, contexts containing negated root necessity modals are not only infrequent, they are also not very informative. That said, negation may be useful for possibility modals: negation is frequent on root possibility modals (20.9%), and impossibility contexts are particularly informative. This could help at least for root possibility modals, if children assume that negation scopes over modals.

[7] Accuracy corresponds to the mean accuracy across the 20 contexts initially extracted for each condition. On average, each context was seen by 34.7 participants (ranging between 24 and 47).

[8] To control for the effect of frequency, we compared accuracy for *can* and *able-to* (used in ROOT-POS-2 and ROOT-NEG), which are both root possibility modals but strongly differ in frequency (3 *able* for 100 *can* in the Manchester corpus). The general accuracy on *able* was slightly but not significantly lower than on *can* (overall: *able*: 80.8% vs. *can*: 89.8%; vs. *have to*: 71.7%)

To summarize, possibility modals are more frequent than necessity modals in the input. Negation is rare on necessity modals, and most cases involve modals that scope over negation, which could be misleading if children assume that negation scopes over modals. Results from our HSP study however show that the conversational context in which modals are used is highly informative about both forces in affirmative contexts. Our posthoc examination of the contexts that led to highest and lowest accuracy in the HSP suggests that the cues adults use may vary with flavor: the desirability of the prejacent seems to matter for root modals, and the salience of evidential support for the prejacent for epistemics.

3 Children's productions study

3.1 Corpus study

Like adults, children produce more possibility modals than necessity modals, but the difference is even greater (79.3% of children's modal productions, vs. adults: 72.5%) (**Table 5**). Negated necessity modals are particularly rare in child productions (only 5.1% of necessity modals are negated), but negated possibility modals are extremely frequent: 51.0%. Epistemic modals are overall very rare in child productions (only 2.4% of their modal utterances).

	CHILD (n=4,844)		CHILD (n=4,800)			
	all		no negation		negation	
possibility	3841	79.3%	1861	49.0%	1937	51.0%
necessity	1003	20.7%	950	94.8%	52	5.2%

Table 5: Counts and percentages of modal uses by force, ordered by lemma frequency, with and without negation, for children (repetitions excluded: 17.0% of the data)

We find that children use (root) possibility modals frequently, both with and without negation, which we can take as initial evidence of productivity (Stromswold 1990). Children produce fewer necessity modals, and rarely with negation. This difference might be explained in part by their input, as children might grasp more frequent words first, and perhaps by differences in what adults and children like to talk about. Quantitative data about children's productions can only provide a partial picture of whether children use and understand modals correctly. To assess these productions in a more qualitative way, we ran an HSP study on children's modals.

3.2 Human Simulation Paradigm Study on child modal usage

The goal of this second HSP study was to investigate children's early modal productions, and see whether they use modals in an adult-like way by comparing their usage to adult usage. Can adults guess the force of a modal intended by children, given the conversational context in which they use it?[9]

The experiment was identical to the HSP on adult productions, except that we used children's utterances and made small changes in the instructions. An example of the display is given in **Figure 3**. We implemented the exact same conditions: ROOT-AFF-1; ROOT-AFF-2; ROOT-NEG; EPI-AFF. Controls were also based on tense. ***Extraction procedure*** **–** Given the low frequency of negated necessity modals and epistemic necessity modals in child productions, we could test only 10 different contexts for ROOT-NEG necessity and 12 contexts for EPI-AFF necessity. This did not make a difference for the participants, who always had 10 contexts to judge per condition. In all the other conditions, the 10 contexts were still selected randomly out of a list of 20 contexts, randomly extracted from the Manchester corpus. ***Exclusion criteria*** **–** We didn't remove cases where a modal already appeared in the preceding dialogue.

[9] We make the assumption that adults rely on their own competence to judge usage patterns.

Participants – 289 participants were recruited on Amazon Mechanical Turk (EPI-AFF: 74, ROOT-AFF-1: 72, ROOT-AFF-2: 73; ROOT-NEG: 72; language: US English; 173 females, mean age = 40.2 y.o.). We removed 18 participants (6.2%) who were less than 75% accurate on controls.[10] We thus present results for 273 participants (EPI-AFF: 68; ROOT-AFF-1: 70; ROOT-AFF-2: 70; ROOT-NEG: 65).

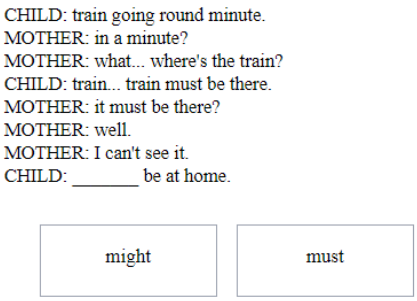

Figure 3: Human Simulation Paradigm stimuli: example trial (*must*), child productions

Analysis – Table 6 reports mean accuracy in each condition. We first ran the same binomial tests as for the adult version. Participants performed better than chance in all conditions involving possibility, but not necessity modals: for ROOT-AFF-2 (*have to*) (mean accuracy: 42.6%) and ROOT-NEG necessity (*not have to*) (mean accuracy: 32.3%), they performed lower than chance. We again used binomial linear mixed effects models (built with a maximal random effect structure testing Accuracy with Subject and Item as random factors), and find an effect of **Force** in all conditions, with higher accuracy for possibility modals (all: $\chi 2(1)=20.49$, $p=5.9e\text{-}6$***; ROOT-AFF-1 (ftc with full spec): $\chi^2(1)=60.4$ $p=7.7e\text{-}15$***; ROOT-AFF-2: $\chi^2(1)=7.37$ $p=0.0066$**; ROOT-NEG: $\chi^2(1)=38.1$, $p=6.6e\text{-}10$***; EPI-AFF: $\chi^2(1)=7.93$ $p=0.0048$**). We find an effect of Negation, significant for possibility ($\chi^2(1)=3.65$, $p=0.056$*) *and* necessity conditions ($\chi^2(1)=6.74$, $p=0.0093$**) (Interaction Force*Neg: $\chi^2(1)=9.2374$, $p=0.0024$**). No effect of **Flavor** ($\chi^2(1)=0.14$, $p=0.71$). **Age** (adult vs. child productions): we find a general effect of Age, with lower accuracy for child usage ($\chi^2(1)=260.52$, $p<.001$***). Among possibility conditions, only ROOT-AFF-1 is significant; among necessity conditions, all comparisons were significant, except EPI-AFF (**Figure 3**). We find a strong interaction Force*Age: the difference in accuracy between possibility and necessity modals for child productions is larger than for adult productions ($\chi^2(1)=32.1$, $p=1.45e\text{-}08$***).

	Mean accuracy (se)		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	85.1% (0.026)	42.6% (0.039)	p <.001*** 95% CI [0.82, 0.88]	p = 5.085e-05 95% CI [0.39, 0.46]
ROOT-AFF-2	79.6% (0.041)	60.2% (0.060)	p <.001*** 95% CI [0.77, 0.83]	p = 2.05e-07 95% CI [0.56, 0.63]
ROOT-NEG	88.2% (0.027)	32.3% (0.050)	p <.001*** 95% CI [0.86, 0.91]	p <.001*** 95% CI [0.29, 0.36]
EPI-AFF	75.6% (0.050)	56.8% (0.047)	p <.001*** 95% CI [0.73, 0.80]	p = 0.000194 95% CI [0.53, 0.61]
all	82.1% (0.019)	50.1% (0.028)		
ALL	67.4% (0.021)			

Table 6: Accuracy rates and significance tests by condition (child, n = 273*10 observations per cell)

[10] For the adult version, errors on controls was very low (5.4%). For the child version, the initial proportion of errors on controls was high (21.6%): post-hoc analysis revealed that this came from 5 control contexts for which accuracy was very low. We removed these 5 controls from our exclusion criteria, as they were particularly difficult, and probably did not indicate that subjects were not doing the task correctly. After restricting to the 15 remaining controls, mean accuracy on controls was 90.0%.

	possibility	necessity
ROOT-AFF-1	χ^2 (1) = 3.12, p = 0.078 (NS)	χ^2 (1) = 35.8, p = 2.1e-09 ***
ROOT-AFF-2	χ^2 (1) = 5.80, p = 0.016 *	χ^2 (1) = 51.8, p = 6.3e-13 ***
ROOT-NEG	χ^2 (1) = 2.78, p = 0.096 (NS)	χ^2 (1) = 21.1, p = 4.37e-06 ***
EPI-AFF	χ^2 (1) = 3.76, p = 0.053 (NS)	χ^2 (1) = 0.22, p = 0.64 (NS)
all	χ^2 (1) = 15.9, p = 6.7e-05 ***	χ^2 (1) = 175.7, p <.001***
ALL	χ^2 (1) = 231.4, p < 2.2e-16 ***	

Table 7: Results of the model testing effect of Age (adult usage vs. child usage)

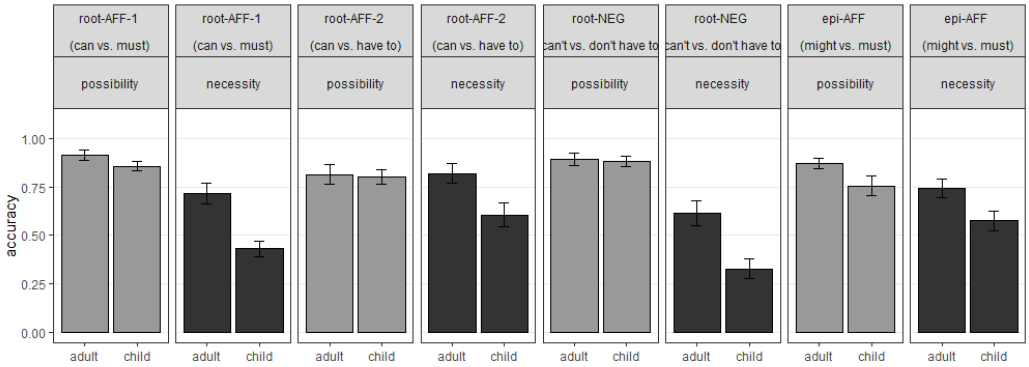

Figure 3: Accuracy by condition: comparison between adult usage and child usage

We find that participants are good at identifying *possibility* modals used by children for both flavors, even if they are less accurate than with adult modals. However, accuracy was much lower for children's *necessity* modals: while mean accuracy on child possibility modals is 82.1%, it is only 50.1% for necessity modals. The unclarity of children's necessity modals uses, illustrated in (1) and (2) below, suggests that children may lack adult competency.

(1) […] CHILD: what shall I put first?
CHILD: that.
CHILD: what's that?
MOTHER: pardon?
CHILD: I **have to** see a cat. (Becky, 2;08.16) (HSP mean accuracy: 2.9%)

(2) […] MOTHER: I thought we had all of these eggs.
CHILD: they not.
CHILD: they go in the bag.
CHILD: they going in there.
CHILD: they go in there.
MOTHER: oh you're putting them back in there now, are you?
CHILD: you **don't have to** eat them. (Carl, 2;08.07) (HSP mean accuracy: 20.0%)

4 General discussion

When do children figure out the force of modals? Our results suggest that children master possibility modals early: by age 2, they produce them frequently and productively, with and without negation, in an apparently adult-like way given the high accuracy found in the child HSP study. It is less clear, however, whether they master necessity modals at this early age: the few necessity modals they produce

do not seem adult-like: accuracy on necessity modals in our child HSP study was below chance, suggesting that children use them in situations where adults would rather use possibility modals. While production data may not fully reflect comprehension, this early asymmetry in children's mastery of modals might help explain earlier results from comprehension studies with older children, if it persists into the preschool years: if children are uncertain about the underlying force of necessity modals, they will accept them in possibility contexts, but they will also over-accept possibility modals in necessity contexts as they lack a reliable stronger alternative.

How do children eventually figure out modal force? Our results suggest that the conversational context in which modals are used might provide sufficient cues to help them figure out their force. Children don't need to—and probably can't—rely on negative contexts, nor do they need a bias toward necessity meanings. Looking at the actual input to children, our results show that negative contexts are not particularly helpful for necessity modals, and in fact it might even be responsible for some of the difficulties children have with these modals. First, negated necessity modals are rare in the input—perhaps for functional reasons, as speakers can express non necessity meanings from the mere use of a possibility modal, via *scalar implicatures* (Horn 1972). Second, negation doesn't behave uniformally with all modals: some necessity modals like *must* outscope negation (Iatridou and Zeijlstra 2013). If children were to rely on negation to figure out force, they might be misled into thinking that *must* expresses possibility, if they assume that negation scopes over modals. The problem is further complicated by the fact that epistemic and root modals interact differently with negation. Thus, whether children can use negation to figure the force a modal then depends on what they expect about its scope.[11] Finally, we find in our input HSP that intended force is the least clear for negated necessity modals: adults sometimes use them in situations corresponding to impossibility meanings (*e.g.* 'you *don't have to* break those things', used as a prohibition). Negation might be more helpful for possibility modals: they cooccur frequently in the input, and negated possibility contexts are particularly informative.

Our input HSP results show that a necessity bias is not necessary for children to solve the *entailment problem*, as the conversational context provides ample cues to figure out force. Our child HSP results further suggest that such a bias is just not at play: if it were, why would children master possibility modals earlier, and struggle with necessity modals? Should our results, in turn, be taken to show that children have a bias towards *possibility* meanings? We do not believe that this has to be the case. Children's difficulties with necessity modals can be explained by various aspects of the input: first, learners will have fewer opportunities to hear necessity than possibility modals, maybe for functional reasons: speakers have other ways to give orders (using imperatives), or to express certainty (using bare assertion of the prejacent). Second, possibility modals occur in a diverse set of environments (with and without negation, in declarative and interrogative sentences), while necessity modals occur mostly in declarative sentences. Given that the context is highly informative for both possibility and necessity modals in our adult HSP study, we believe that children's early mastery of possibility compared to necessity modals may be more of a matter of quantity rather than quality of the input.

We have shown that the conversational context is highly informative about force. But what exactly gives away force? One aspect of the context that might be particularly helpful for deontic modality is the kind of (indirect) speech acts that modals are used for: listeners might easily discern orders from permissions, by relying, in part, on the perceived undesirability or effortfulness of the prejacent (*e.g.* 'you *#can/have to* eat your broccoli' vs. 'you *can/#have to* take a cookie'). For epistemic modals, our posthoc analysis suggests that salient evidence in favor of the prejacent biases towards necessity. We plan to explore this further in future work.

[11] A number of studies assess children's comprehension of sentences containing modals and negation (*e.g.* Moscati and Crain 2014; Koring et al. 2018), but more research still needs to be done to really understand how and when children learn the right scope interpretations between modals and negation, across force and flavors.

References

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4*, 328.

Berwick, R. C. (1985). *The acquisition of syntactic knowledge* (Vol. 16). MIT press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176.

Gualmini, A., & Schwarz, B. (2009). Solving learnability problems in the acquisition of semantics. *Journal of Semantics*, *26*(2), 185-215.

Horn, L. R. (1972). On the semantic properties of logical operators. Doctoral dissertation, *UCLA*.

Iatridou, S., & Zeijlstra, H. (2013). Negation, polarity, and deontic modals. *Linguistic inquiry*, *44*(4), 529-568.

Koring, L., Meroni, L., & Moscati, V. (2018). Strong and Weak Readings in the Domain of Worlds: A Negative Polar Modal and Children's Scope Assignment. *Journal of psycholinguistic research*, *47*(6), 1193-1217.

Kratzer, A. (1977). What 'must'and 'can' must and can mean. *Linguistics and philosophy*, *1*(3), 337-355.

Landau, B., & Gleitman, L. R. (1985). Cognitive science series, 8. *Language and experience: Evidence from the blind child. Cambridge, MA, US: Harvard University Press*.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge University Press.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Moscati, V., & Crain, S. (2014). When negation and epistemic modality combine: The role of information strength in child language. *Language Learning and Development*, *10*(4), 345-380.

Musolino, J., Laity d'Agostino, K., & Piantadosi, S. (2019). Why We Should Abandon the Semantic Subset Principle. *Language Learning and Development*, *15*(1), 32-46.

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*(3), 191-214.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2013). Modeling the acquisition of quantifier semantics: a case study in function word learnability. *Under review*.

Portner, P. (2009). *Modality* (Vol. 1). Oxford University Press.

Stromswold, K. J. (1990). *Learnability and the acquisition of auxiliaries*, Doctoral dissertation, MIT).

Team, R. C. (2013). R: A language and environment for statistical computing.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of child language*, *28*(1), 127-152.

White, A. S., Hacquard, V., Resnik, P., & Lidz, J. (2017). The contextual modulation of semantic information.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological review*, *114*(2), 245.