BUCLD 47 Proceedings
To be published in 2023 by Cascadilla Press
Rights forms signed by all authors

Assessing how the linguistic input affects children's mastery of modals

Anouk Dieuleveut, Ailís Cournane, and Valentine Hacquard

1. Introduction

How do children figure out that modals like *can* mean possible, and *have to* necessary? Acquisition corpus studies show an asymmetry in young English children's mastery of modals: while already by age 2 children use possibility modals like *can* frequently, productively, and in an adult-like way, they seem to have difficulties with necessity modals like *have to*, which they start producing later on, less frequently, and in environments where adults prefer possibility modals (Dieuleveut et al. 2019; 2022). The goal of this study is to investigate how children can figure out the force of modals (possibility vs. necessity), by identifying which features of children's linguistic input are most predictive of their mastery of modal force. Building on Dieuleveut et al.'s (2022) corpus study of 12 English children's modal production and input in the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000), we first examine how much variation we find between children for mastery of necessity modals, and then probe for factors in their input that correlate with this variation.

We focus on *can* and *have to*,[1] the earliest and most frequent possibility and necessity modals respectively produced by children. To assess children's mastery of these modals, we use a variant of the Human Simulation Paradigm (Gillette et al., 1999) where adults are asked to guess the force of modals used by children in the corpus. This allows us to get a qualitative assessment of their usage. We then test how various aspects of children's linguistic input might influence their mastery. We first examine general features of the input (here, maternal speech), known to influence children's language development in general: **mothers' complexity of speech**, as indexed by their *Mean Length of Utterance* (MLU, Brown, 1973), and **mothers' talkativeness**, as indexed by the mean number of utterances per one hour of recording time. We then turn to more specific aspects

* Dieuleveut, Laboratoire de Linguistique Formelle, Université Paris Cité, CNRS. Cournane, New York University. Hacquard, University of Maryland. Corresponding email: anouk.dieuleveut@gmail.com (Dieuleveut). We thank our research assistants Joon Lee, Jan Michalowski and David Whitcomb, and the UMD/NYU ModSquad & Language Acquisition Lab for relevant comments and discussion, especially Annemarie van Dooren, Jeff Lidz and Alexander Williams. This project is supported by NSF grant #BCS-1551628.

[1] We focus on 'root' uses of modals (i.e., non-epistemic: deontic, ability…), as children of this corpus produce too few epistemic modals to test ('epistemic gap', Cournane, 2015).

of children's *modal* input: **frequency of general modal talk** in maternal speech, **frequency of the modals *can* and *have to*** specifically, and **frequency of *can* and *have to* with negation** (*can't*, *don't have to*). Our results point out two factors that correlate with children's mastery of *have to*: exposure to negated occurrences, and exposure to general modal talk. We find no evidence that other factors matter.

In the next section, we describe the learning problem modals pose, and justify the various input measures we examine in our study. We present our study in section 3, and discuss its results and implications in section 4.

2. The learning problem

English modals express different 'forces': possibility (e.g. *can* (1a)) or necessity (e.g. *have to*, (1b)). Studies of young children's natural productions (Dieuleveut et al. 2019, 2022), and behavioral experiments testing their comprehension (e.g., Noveck, 2001; Öztürk & Papafragou 2015; Cournane et al. in prep.; Leahy & Zalnieriunas 2022) show that English-speaking children struggle with necessity modals, but seem to master possibility modals early. Specifically, they tend to produce—and accept—necessity modals in possibility situations; though, as these studies tend to report results in the aggregate, it is not clear how much variation there is between children in their mastery of modals.

(1)	a. You **can** leave.	Possibility
	b. You **have to** leave.	Necessity
	c. You **sig** leave.	Possibility? Necessity?

Necessity modals may be intrinsically harder to acquire than their possibility counterparts, as modals give rise to a "subset (or entailment) problem", originally discussed for the acquisition of syntax (Chomsky & Lasnik, 1977; Dell, 1981, Berwick, 1985; Wexler & Manzini, 1987, a.o.). What is necessary is also possible: thus, in necessity situations, both possibility (1a) and necessity modal statements (1b) are logically true. If learners don't yet know the force of a novel modal, *sig*, (1c), and hear it used in necessity situations, how can they determine whether it means necessity or just possibility? What is possible, however, is not necessary, thus possibility modals can be used in situations where other possibilities are open, which could allow children to realize that they do not express necessity.

Various hypotheses have been proposed in the literature for how children overcome the subset problem for modals (see Dieuleveut, 2021, for an overview). One possibility is that learners rely on negative (Downward-Entailing) environments (Gualmini & Schwarz, 2009), which reverse entailment patterns. Upon hearing a necessity modal under negation, for instance, "The dog *doesn't have to* go outside", used in a situation where it is clear that the dog can either go outside or stay inside (a situation of possibility, but not necessity), learners might be able to infer that *have to* cannot have a possibility meaning, reasoning *ad absurdum* that if *have to* meant 'possible', *not have-to* would mean 'not possible',

but then, could not be used in a possibility situation, and thus, has to mean 'necessary'. This solution faces several challenges (Dieuleveut et al., 2022): first, necessity modals do not always scope under negation; second, negated necessity modals are rare in the input, and when they do occur, their pragmatic uses can obscure their force: they are often used to soften prohibitions ("you don't have to shout" to convey that you should not shout, not merely that it is not necessary). Thus, truly informative uses of negated necessity modals may be very rare in the input. However, children might be able to learn from a few highly informative cases. We can thus ask whether children who hear '*don't have to'* relatively more often master necessity modals earlier.

But, perhaps the subset problem of modals is not really a learning problem (Piantadosi, 2011, Dieuleveut et al., 2022): children may not need to hear necessity modals under negation, they may just need to hear enough necessity modals. Children may need a certain amount of exposure to a necessity modal like *have to* to realize that they never occur in environments where different possibilities are open. And frequency may well play a role in children's later mastery of necessity modals, independently of the subset problem, at least in English. Indeed, Dieuleveut et al. (2022) found a strong asymmetry in children's input: necessity modals are much rarer than possibility modals (*have-to* represents 12% all adult's modal utterances in the Manchester corpus vs. *can* 57%).

Frequency of exposure may also matter in a more subtle way: it may not be enough to hear *have to* frequently to master its force, but it may be helpful to hear it amongst other modals. Hearing necessity claims contrasted with possibility claims (e.g., "you can, but you don't have to") could help children infer that the former expresses possibility, and the latter necessity. Even if necessity and possibility modals are not contrasted explicitly, hearing more modal talk in general could raise the salience of possibilities and necessities as notions that can be talked about, and thus prompt learners to expect words that express necessity.[2] Finally, it could be that necessity modals do not present a particular problem, and that their mastery chiefly relies on general factors, such as hearing more, or more varied maternal speech, which have been showed to affect vocabulary growth and language learning more generally (Hart & Risley, 1995; Hoff, 2003; Hoff & Naigles, 2002; Hsu et al., 2017; Rowe, 2008, 2012; see Anderson et al. 2021 for a meta-analysis of studies comparing the impact of quality vs. quantity for child language skills).[3] In this study, we thus investigate both quantitative and

[2] Other aspects of the input might provide useful cues. Dieuleveut et al. (2022) propose that (un)desirability of the modal's complement could help children figure out the meaning of root necessity modals like *have-to* (e.g. 'You sig do your homework' suggests a necessity meaning for *sig)*. We plan to investigate this with same methods in future work.

[3] The linguistic input is typically described in terms of 'quantity' and 'quality'. 'Quantity' is usually defined in terms of number of words/tokens (Rowe, 2012) or number of utterances (Laks et al., 1990) spoken to the child. 'Quality' involves features such as mothers' talkativeness, vocabulary diversity, syntactic complexity, and general interactive features like responsiveness or joint attention (see Rowe & Weisleder, 2020).

qualitative aspects of children' input, at a general and at a modal-specific level, to see which are predictive of children's mastery of necessity modals.

3. Study

Building on Dieuleveut et al. (2022)'s study of children's modal production and input in the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000), we assess children's modal use and modal input. This corpus consists of 12 child-mother pairs, recorded at home in unstructured play sessions. Child ages range from 1;8.22 to 2;0.25 at the start and 2;8.15 to 3;0.10 at the end of the corpus. Here, we only include data starting from 2;0, and focus on modals *can* and *have-to*, the first and most frequent modals produced by both children and mothers. The relative homogeneity of child-mother pairs allows us to test fine-grained properties of their input. The children are all monolingual, English-speaking, firstborn children; mothers are their primary caregivers. Section 3.1. reports quantitative measures of children's input and their modal productions. Section 3.2. reports our experiment, based on the Human Simulation Paradigm, aimed at providing a qualitative assessment of children's modal use.

3.1. Quantitative measures

Table 1 presents general metadata (unspecific to modal talk) about the twelve child/mother dyads. We use MLU as a measure of syntactic complexity in maternal speech, and mean number of utterances in one hour of recording time as a measure of "talkativeness".

Table 2 summarizes modal uses in maternal speech. It includes (iii) **frequency of use** of *can* and *have to*, distinguishing between **'overall' frequency** (i.e., proportion relative to total number of mother utterances), and **'relative' frequency** (proportion of *have to* relative to *can* utterances). *Can* is much more frequent than *have to*: on average, it occurs in 3.4% of all maternal utterances, vs. 0.75% for *have to.* The frequency of *have to* relative to total counts of *can* and *have to* is 18.1% (about 4 *can*s per 1 *have to*), ranging from 11.2% (9 *can* per 1 *have to*) (Becky's mother) to 27.4% (3 *can*s per 1 *have to)* (Liz's mother). **Table 2** further reports (iv) proportion of uses of *can/have to* in **negative environments**, and (v) frequency of **general modal talk** (modals included: possibility: *can, could, might, may; able to;* necessity: *must, should, need; have to, got to, be supposed to, need to*).

Table 1. 'General' input measures for each child: (i) mother's MLU; (ii) mothers talkativeness. In grey are children not tested in the experiment, because they produced too few *have to* (see child data in Table 3).[4]

child	gender	MLU 2;0.0	MLU 3;0.0	CDI score[5]	# recordings	(i) Mother's MLU	(ii) Mother's talkativeness (# mor/min)[6]	(# utt./min)
anne	F	2.15	3.16	180	30	4.47	149.5	33.5
aran	M	1.75	3.46	153	31	6.18	220.6	35.7
becky	F	1.54	3.04	138	34	4.26	108.1	25.3
carl	M	2.12	3.66	187	22	4.60	93.0	20.2
domin	M	1.61	2.62	153	31	4.16	143.1	34.4
gail	F	1.7	3.26	262	33	4.41	114.2	25.9
joel	M	1.83	3.17	122	31	4.33	115.9	26.8
john	M	2.05	2.81	191	30	4.65	92.9	19.9
liz	F	1.66	3.76	359	32	4.51	85.2	18.9
nic	F	1.14	3.09	102	34	4.67	133.6	28.6
ruth	F	1.46	2.99	44	31	4.36	158.4	36.3
warr	M	2.23	3.84	124	31	5.54	125.3	22.6
ALL	**6F/6M**	**1.77**	**3.24**	**168**	**370**	**4.68**	**128.9**	**27.5**

Table 2. Summary of specific input measures: (iii) overall and relative frequency of *can* and *have to*; (iv) frequency of *can* and *have to* in negative environments; (v) proportion of general modal talk. In grey are children not tested in the experiment (they produced too few *have to*).

child	Total # utt.	Total # modal utt.	(v)Prop. modal talk	(iii) Overall frequency				(iii) Relative frequency (*have*/ *can+have*)	(iv) Neg. envir.	
				can		*have to*			*can*	*have to*
anne	30110	1461	4.9%	766	2.5%	133	0.4%	14.8%	17.8%	1.5%
aran	33238	2245	6.8%	1222	3.7%	220	0.7%	15.3%	24.8%	4.1%
becky	25836	1488	5.8%	926	3.6%	117	0.5%	11.2%	19.4%	5.1%
carl	13322	660	5.0%	405	3.0%	83	0.6%	17.0%	20.7%	8.4%
domin	32016	1878	5.9%	1026	3.2%	278	0.9%	21.3%	26.0%	9.0%
gail	25601	1560	6.1%	803	3.1%	301	1.2%	27.3%	23.4%	2.7%
joel	24939	1236	5.0%	647	2.6%	218	0.9%	25.2%	42.3%	3.2%
john	17947	1233	6.9%	883	4.9%	131	0.7%	12.9%	7.6%	6.1%
liz	18101	1068	5.9%	516	2.0%	195	1.1%	27.4%	18.4%	2.1%
nic	29208	1882	6.4%	1269	4.3%	170	0.6%	11.8%	24.5%	3.5%
ruth	33777	1327	3.9%	686	2.0%	219	0.7%	24.2%	20.7%	3.7%
warr	21028	1811	8.6%	1053	5.0%	184	0.9%	14.9%	20.3%	5.4%
ALL	**305123**	**17849**	**5.8%**	**10202**	**3.4%**	**2249**	**0.8%**	**18.1%**	**22.2%**	**4.5%**

[4] Total numbers of morphemes and utterances and MLU were calculated using the MLU function of the CLAN program (MacWhinney, 2000). Only transcripts from 2-year-olds are included. Note that duration of recordings is not systematically reported: to compute mother's talkativeness, we assumed that all recordings were 30 minutes long.

[5] The language level of children was assessed through the MacArthur Communicative Development Inventory (CDI, Toddlers; Fenson et al., 1993) at the beginning of the study.

[6] We use morphemes rather than words. As a point of comparison, the average speaking rate in English ranges from 152 to 170 words per minute according to Google.

Table 3 summarizes children's data. Like their mothers, children use *can* significantly more than *have to: can* occurs in 1.8% of all children's utterances vs. *have to* in 0.2%. Relative frequency of *can* and *have to* for children is approximately 9 *can* per 1 *have to* (we report % *have to* out of total counts of these two modals). Looking at age of first production, we find that *have to* tends to be sampled considerably later than *can*, with a 4-month difference on average. Note that this likely under-estimates the gap between the two modals: most children actually start producing *can* in earlier transcripts than the ones included in this study, under 2-years-old.[7]

Table 3. Measures of children's modal mastery: 'overall' and 'relative' frequency of *can* and *have to*; age of first production; modal mastery as measured by the experiment.

child	Total # utt.	'Overall' frequency				'Relative' frequency (have to vs. can)	Age of 1st prod.		Experiment	
		can		*have to*			*can*	*have to*	*can*	*have to*
anne	16405	374	2.3%	22	0.1%	5.6%	2;00.15	2;00.29	80.2%	39.6%
aran	16144	560	3.5%	30	0.2%	5.1%	2;01.00	2;05.03	82.4%	57.0%
becky	23398	598	2.6%	84	0.4%	12.3%	2;00.07	2;05.08	78.3%	57.5%
carl	16998	438	2.6%	12	0.1%	2.7%	2;00.26	2;06.19	64.9%	57.7%
domin	19145	267	1.4%	11	0.1%	4.0%	2;00.28	2;04.11	70.3%	73.8%
gail	16396	317	1.9%	93	0.6%	22.7%	2;01.08	2;00.19	87.5%	61.9%
joel	16410	214	1.3%	17	0.1%	7.4%	2;01.23	2;0325	69.8%	43.9%
john	12464	46	0.4%	1	0.01%	2.1%	2;00.13	2;09.12	*Not tested.*	
liz	15501	253	1.6%	20	0.1%	7.3%	2;00.28	2;04.03	82.2%	44.1%
nic	16937	221	1.3%	45	0.3%	16.9%	2;01.01	2;08.20	69.3%	45.2%
ruth	19282	16	0.1%	3	0.02%	15.8%	2;04.01	2;07.10	*Not tested.*	
warr	14226	316	2.2%	14	0.1%	4.2%	2;01.14	2;03.08	67.0%	68.4%
ALL	**203306**	**3620**	**1.8%**	**352**	**0.2%**	**8.9%**	**2;00**	**2;04**	**75.2%**	**54.9%**

The first two measures – children's quantitative productions and age of first production – can be used as indicators of children's mastery, but in limited ways. Some children (e.g., John, Ruth) do not produce many *have to*, but that doesn't mean that they don't understand it; conversely, others (e.g. Gail, Becky) may use *have to* more often, but not understand it. Moreover, while direct repetitions were excluded from our analyses, we did not control for children repeating sentences used earlier in the conversation. To test if children use their modals in an adult-like way and get a more qualitative assessment of their uses, we run an experiment, using the paradigm introduced by Dieuleveut et al. (2022).

[7] We only looked at first *have to* sampled in the corpus. A more conservative measure would be *First of Repeated Uses* (FRU) (Snyder, 2007), aiming to count uses that show the onset of productivity. Modal verbs are relatively infrequent so FRU (developed for morphosyntactic productivity) is perhaps overly restrictive (see Cournane, 2021).

3.2. Experiment
3.2.1. Procedure

The experiment was identical to Dieuleveut et al.'s (2022) Experiment 5 (condition 2: *can* vs. *have to*).[8] In the experiment, adult participants are given dialogues extracted from the Manchester corpus, and asked to guess the force of a redacted modal that had been spoken by a child. They can choose between two options, corresponding to either a necessity (*have-to*) or a possibility (*can/able to*) modal. An example is given in **Figure 1**. As in Dieuleveut et al. (2022), each participant had 40 dialogues to judge: 20 test trials (10 possibility; 10 necessity) and 20 controls using tense (10 past; 10 future). The 20 test contexts were randomly selected for each participant out of a list of 200 (2*10*10 children). As a baseline, we use Dieuleveut et al.'s (2022) experiment on *adult* modal uses (Experiment 1, condition 2), which shows that the force of modals used by adults is guessable from context (we assume that dialogues preceding the modal sentence are equally informative for adult and child uses). Participants' accuracy at guessing children's modals (as compared to their accuracy at guessing adults') is taken as a measure of how 'adult-like' children's uses are.

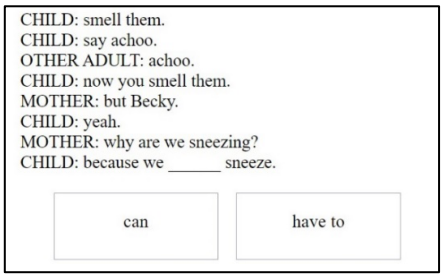

Figure 1. Experiment stimuli: example trial (*have to*)

Material. 20 contexts per child (10 *can*, 10 *have-to*) were initially randomly extracted from the corpus. Because of the low number of *have-to* utterances in John's and Ruth's productions (see **Table 3**), we could not include them in the experiment. We thus had a list of 200 contexts in total, and 20 controls. Controls were the same as in Dieuleveut et al. (2022), excluding their 5 problematic cases. ***Exclusion criteria.*** As in Dieuleveut et al.'s (2022), we excluded contexts where the modal already appeared in the preceding dialogue when it was a direct repetition (e.g., '[…] *CHILD: and you **have to** go and get a pan. OTHER ADULT: pardon? CHILD: you _____ get a pan*.'), but kept them otherwise.[9] We made

[8] We could not run condition 1 (*must* vs. *can*), because of the infrequency of *must* in children's productions, nor the other two conditions.

[9] Out of the final 200 contexts, 47 have the modal appear in the preceding dialogue (23.5%) (necessity contexts: 21; possibility contexts: 26) (child speech: 24, adult speech: 12; both: 11). Some contexts have the other (non-target) modal in the preceding dialogue (necessity

sure to include examples in the training (*the/a*) and control items where it was also the case that the right or wrong answer appeared in the preceding dialogue. We removed Briticisms (e.g. tag-questions), but did not correct children's ungrammatical utterances (e.g., *comed* for *came*), except in the case of *have to* when they omitted *to* or made an agreement mistake (e.g., *Mummy have to do it* was corrected to *Mummy has to do it*), so that participants would not reject the answer because of its ungrammaticality. In cases where *have to* was tensed (*had to*) or under another modal (e.g., *might have to*) (7 cases), we used *able to* as the alternative. However, in this experiment, we could not extract enough contexts for *be able to*, because of the overall low number of *able* in children's production (3 cases overall, and only 1 in affirmative sentence). This means that there was only one case where *able* was the right answer, vs. 7 for (tensed) *have to*. ***Expectations***. First, we expect participants to perform better when judging children's possibility modals (*can/able*) than for their necessity modals (*have to*) (as in Dieuleveut et al. 2022). Second, we expect *effects of input*, which can tell us about what may matter for learning possibility vs. necessity from the input. If children's mastery of necessity modals depends on general features of maternal speech, we expect that (i) children whose mother's MLU is higher and whose mothers are more talkative (ii) will show better mastery. If sheer quantity of exposure matters, we expect that children more exposed to necessity modals will show better mastery. If children need negative environments to resolve the subset problem, we expect that children exposed to more *negated have-to* will show better mastery. Finally, if children need greater modal input, we expect that children exposed to more general modal talk will show better mastery. These expectations concern performance on necessity modals only, as we expect ceiling performance for children's possibility modals. **Caveat**: It should be noted that this paradigm only tells us whether children use their modals in the same contexts as adults. It does not allow us to test whether children fully know the meaning of necessity modals, only whether they know how to use them in ways adult participants deem appropriate.

3.2.2. Participants

351 adult participants were recruited on Amazon Mechanical Turk (US English; 190 females, mean age=41.9 years). Proportion of errors on controls was low (8.9%). We removed 19 participants (5.4%) who were less than 75% accurate on controls. We thus present results for 332 participants.

3.2.3. Analysis

Data analyses were conducted using R (R Core Team, 2013), using the package lme4 (Bates et al. 2014a, 2014b). First, we replicate the main finding from

contexts: 14 with *can/can't*, 7 with *have to;* possibility contexts: 24 with *can/can't*, 1 with *have to*, 1 with both).

Dieuleveut et al. (2022): participants perform significantly better when guessing children's possibility modals than their necessity modals (mean accuracy for possibility modals (P): 75.2%; for necessity modals (N): 54.9%; vs. in Dieuleveut et al.'s (2022) condition ROOT-AFF-2: P: 79.6%; N: 60.2%; in the adult baseline, P: 81.5%; N: 82.0%). **Figure 2** shows results on the aggregate, with Dieuleveut et al.'s (2022: D) experiments 1 and 5 as comparison points.

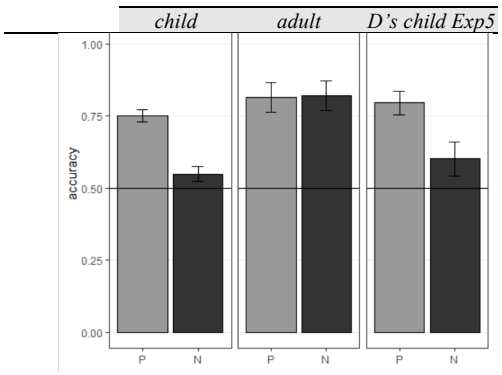

Figure 2. Mean accuracy by force condition: Comparison between experiments (n=332 participants; 200 contexts in total, 100/force; D's Exp. 5: n=70 participants; 40 contexts in total, 20/force; D's Input Exp. 1: n=69 participants; 40 contexts in total, 20/force). Grey bars ('P'): possibility contexts (*can/able*). Black ('N'): necessity contexts (*have to*).

Binomial tests revealed that participants' performance differs from chance for both possibility and necessity modals (**Table 4**). To test whether children's uses are adult-like, we compared our results for child productions to the experiment on adult productions. We used binomial linear mixed effects models built with a maximal random effect structure measuring Accuracy testing for the effect of Age group (child vs. adult usage), with Subject and Item as random factors. The result is only significant for necessity modals ($\chi^2(1)=47.1$, $p=6.8e-12$ ***): we find no difference for possibility modals ($\chi^2(1)=2.17$, $p=0.14$) (**Table 4**). The interaction Force*Age group is significant ($\chi^2(1)=20.9$, $p=4.9e-06$ ***).

Analysis by child. **Table 5** reports accuracy in each force condition (possibility vs. necessity), for the 10 children (see **Figure 3**). For each child, we ran both binomial tests to see whether participants differ from chance, and binomial linear mixed effects models to see whether child uses are adult-like, comparing them to the adult baseline. For possibility modals, we find performance above chance for all children, and for necessity modals, we find performance above chance for 5/10 children, at chance for 4, and below chance for 1 (39.6% accuracy on Anne's *have-to*). Binomial linear mixed effects models comparing their uses to the adult baseline show differences between child and adult usage for 4 children for possibility modals, and 7 for necessity modals.

Table 4. Results and statistical tests for our experiment, compared to Dieuleveut et al.'s (2022) Experiments 1 and 5

	Mean accuracy[10] (se)		Exact binomial tests (two-sided)		Model testing effect of Age (adult vs. child usage)	
	possibility	necessity	possibility	necessity	possibility	necessity
CHILD	75.2% (0.022)	54.9% (0.025)	p <.001*** 95% CI [0.74, 0.77]	p=1.3e-09 *** 95% CI [0.54, 0.56]	χ^2 (1)=2.17, p=.14 (NS)	χ^2 (1)=47.1, p=6.8e-12***
CHILD (D's EXP5)	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]	χ^2 (1)=5.80, p=0.016 *	χ^2 (1)=51.8, p=6.3e-13***
ADULT (D's EXP 1)	81.5% (0.053)	82.0% (0.052)	p <.001*** 95% CI [0.79, 0.85]	p <.001*** 95% CI [0.79, 0.84]	*Not relevant*	*Not relevant*

Table 5. Accuracy rates (se); results of binomial tests; and results of models testing effect of Age Group (adult vs. child usage), by child

child	Mean accuracy (se)		Exact binomial tests (two-sided)		Model testing effect of Age (adult vs. child usage)	
	possibility	necessity	possibility	necessity	possibility	necessity
anne	80.2% (0.074)	39.6% (0.052)	p < 2.2e-16*** 95% CI [0.77, 0.85]	p=1 (NS) 95% CI [0.34, 0.43]	χ^2(1)=2.74, p=0.098 (NS)	χ^2(1)=31, p =2e-8***
aran	82.4% (0.043)	57.0% (0.077)	p < 2.2e-16*** 95% CI [0.79, 0.87]	p=0.0049 * 95% CI [0.52, 0.60]	χ^2(1)=0.17, p=0.68 (NS)	χ^2(1)=36, p=2e-09 ***
becky	78.3% (0.088)	57.5% (0.112)	p < 2.2e-16*** 95% CI [0.72, 0.79]	p=0.0038 * 95% CI [0.53, 0.59]	χ^2(1)=0.01, p=0.92 (NS)	χ^2(1)=3.7, p=0.055 (NS)
carl	64.9% (0.071)	57.7% (0.067)	p=8.8e-09*** 95% CI [0.61, 0.69]	p=0.00023*** 95% CI [0.55, 0.63]	χ^2(1)=4.03, p=0.044*	χ^2(1)=11.7, p=0.0006***
domin	70.3% (0.069)	73.8% (0.067)	p < 2.2e-16*** 95% CI [0.67, 0.75]	p < 2.2e-16*** 95% CI [0.70, 0.77]	χ^2(1)=8.6, p=.0034**	χ^2(1)=6.4, p=0.012 *
gail	87.5% (0.056)	61.9% (0.047)	p < 2.2e-16*** 95% CI [0.84, 0.90]	p=1.6e-06*** 95% CI [0.58, 0.66]	χ^2(1)=4.5, p= 0.033 *	χ^2(1)=0.044, p=0.84 (NS)
joel	69.8% (0.071)	43.9% (0.076)	p < 2.2e-16*** 95% CI [0.67, 0.75]	p=0.98 (NS) 95% CI [0.40, 0.48]	χ^2(1)=24, p=9e-7***	χ^2(1)=22.5, p=2e-06***
liz	82.2% (0.056)	44.1% (0.071)	p < 2.2e-16*** 95% CI [0.77, 0.83]	p=0.99 (NS) 95% CI [0.39, 0.47]	χ^2(1)=0.0069, p=0.93 (NS)	χ^2(1)=14.9, p=0.00012***
nic	69.3% (0.060)	45.2% (0.080)	p=8.5e-13*** 95% CI [0.65, 0.74]	p=0.94 (NS) 95% CI [0.41, 0.49]	χ^2(1)=2.55, p=0.11 (NS)	χ^2(1)=13.3, p=0.00026***
warr	67.0% (0.088)	68.4% (0.081)	p=6.2e-11*** 95% CI [0.63, 0.71]	p=1.4e-11 *** 95% CI [0.64, 0.73]	χ^2(1)=2.66, p=0.10 (NS)	χ^2(1)=2.46, p= 0.12 (NS)
ALL	**75.2% (0.022)**	**54.9% (0.025)**	**p <.001*** 95% CI [0.74, 0.77]**	**p=1.3e-09 *** 95% CI [0.54, 0.56]**	**χ^2 (1)=2.17, p=.14**	**χ^2 (1)=47.1, p= 6.8e-12 *** **

Correlation analysis. Table 6 summarizes the results of correlation analyses testing the relation between the different input factors under consideration (see

[10] Mean accuracy: how good participants are at guessing the force of the modal based on context, e.g. answering *can* in a possibility context) across the 100 contexts extracted from the corpus for each force condition. On average, we have 31.6 observations per context (ranging between 27 and 49). The number of participants was determined so that we had about the same number of observations per context as in the previous experiments.

Tables 1 and **2**), and children's mastery as measured by the experiment.[11] We find no effect of (i) mothers' MLU, (ii) mother's talkativeness, nor (iii) quantity of exposure, for neither modal, regardless of whether we look at overall or relative frequency. However, we find an effect of (iv) negation for necessity modals (r=0.32, t(98)=3.35, *p=0.00114****; 95% CI: [0.13; 0.49]), and of (v) modal talk (r=0.22, t(98)=2.2, *p=0.027* **; 95% CI: [0.026 ; 0.40]**)**. For possibility modals, performances being at ceiling, we find no effect of either.

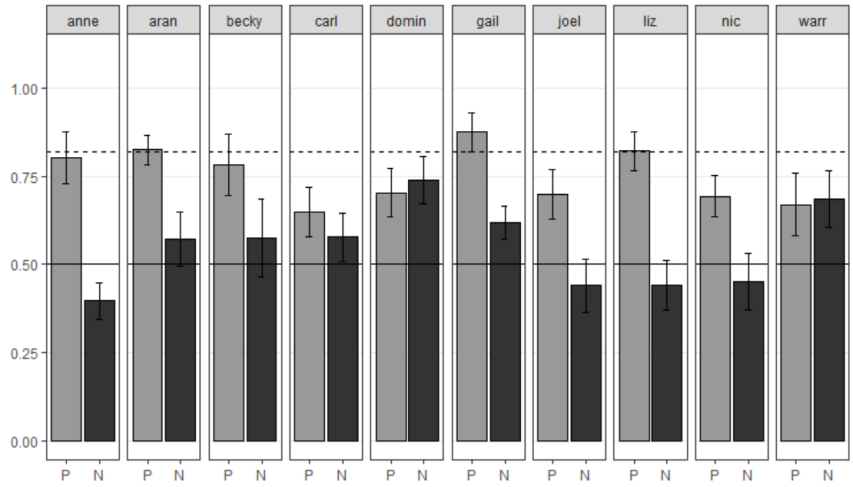

Figure 3. Mean accuracy by force condition by child The black dashed line corresponds to the adult baseline (accuracy on adult production contexts: 81.5% possibility and 82% necessity). The black line indicates chance level.

Effect of Age and children's MLU. We find no significant effect of children's age or MLU on their mastery (**Table 7**). The effect of age is almost significant for possibility modals (r=0.18, t(98)=1.85, p=0.067 (NS)), but not necessity modals. We looked at the relation between children's frequency of production (see **Table 3**), and performance in the experiment, and find no evidence that children who use *have to* in a more adult-like way are the ones who produce then more often.

[11] We also ran correlations between the input factors and *change* in children's mastery over the time period (post-hoc), using Pearson's r coefficients (i.e., the slope of the learning curves) between Age and Accuracy to measure children's improvement. However, these analyses are limited by the extremely reduced age range we sample from (2- to 3-year-old), which is even more restricted since necessity modals are nearly absent until 2:06 (see first production age in **Table 3**). In typical studies that use measures of change (e.g., Hoff, 2003; Rowe, 2008), the child/mother dyads are recorded at two different time points in a more controlled way. Results of the correlation tests are not significant, and thus not reported.

Table 6. Effect of input measures on accuracy: Results of correlation analysis (Pearson's r)

		Possibility modals *(can/able)*	Necessity modals *(have to)*
Input measures	Mothers' MLU	r=0.017, t (98)=0.17, p=0.87 (NS) 95% CI: [-0.18; 0.21]	r=0.071, t (98)=0.71, p=0.48 (NS) 95% CI: [-0.13; 0.26]
	Mothers' talkativeness	r=0.067, t (98)=0.67, p=0.51 (NS) 95% CI: [-0.13; 0.26]	r=0.031 , t (98) =0.30 , p=0.76 (NS) 95% CI: [-0.17; 0.23]
	'Overall' frequency	r=-0.11, t (98)=-1.13, p=0.26 (NS) 95% CI: [-0.30; 0.09])	r=0.12, t (98)=1.20, p=0.23 (NS) 95% CI: [-0.08; 0.31]
	'Relative' frequency (*can* vs. *have to*)	r=0.12, t (98)=1.16, p=0.25 (NS) 95% CI: [-0.08; 0.31]	r=-0.017, t(98) =-0.17, p=0.87 (NS) 95% CI: [-0.21; 0.18]
	Negation	r=-0.089, t (98) =-0.89, p=0.38 (NS) 95% CI: [-0.28; 0.109]	r=0.32, t (98)=3.35, p=0.0011*** 95% CI: [0.13; 0.49]
	Mothers' modal talk	r= -0.032, t (98) =-0.32, p= 0.75 (NS) 95% CI: [-0.23; 0.17]	r=0.22, t (98)=2.2, p=0.027 ** 95% CI: [0.026 ; 0.40]

Table 7. Effect of children's factors: age of production and MLU

	Possibility modals	Necessity modals
Age	r=0.18, t (98)=1.85, p=0.067 (NS) 95% CI: [-0.01; 0.37]	r=-0.09, t (98)=-0.90, p=0.37 (NS) 95% CI: [-0.28; 0.11]
MLU	r=0.016, t (98)=0.16, p=0.87 (NS) 95% CI: [-0.18; 0.21]	r=0.11, t (98)=1.10, , p=0.27 (NS) 95% CI: [-0.09; 0.30]

4. Discussion

The goal of this study was to identify which aspects of children's linguistic input were the most predictive of their mastery of modals, focusing on their root necessity modals (given that children master possibility modals earlier, and produce modals with epistemic interpretation later on). First, we replicate findings from Dieuleveut et al. (2022), with a larger sample: children seem to use *can* in an adult-like way at age 2, but have difficulties with *have to*. Looking at individual patterns, we find that all children seem to master *can* early on (participants' accuracy ranges between 64.9% and 87.5%). For *have to*, we find performances above chance for only five children, with four of them at chance, and one below (accuracy range between 39.6% and 73.8%; we find differences from adult usage for seven children, with the test approaching significance for Becky (*p=0.055*), vs. only four children for possibility modals). We thus find some variation between children: our results suggest that some might even master necessity modals before 3-years-old (Warren and Dominic).

What factors in the input influence modal mastery? We found no evidence that mere quantity of lexeme exposure matters, nor that generic aspects of maternal speech do, such as syntactic complexity or talkativeness. However, results from the correlation test suggest that children who hear *have to* with **negation** more frequently show better mastery. This suggests that children make use of negative environments to learn *have to*. However, it does not tell us about the nature of the mechanism at play: is this effect due to children 'logically' using negative environments to infer the force of *have to*, following Gualmini &

Schwarz (2009), or might negation be useful in other ways? Negation could for instance help learners by putting focus on necessity modals. To tease apart these possibilities, we would need to see whether the effect of negation generalizes to other necessity modals, especially those that outscope negation like *must.* As discussed by Dieuleveut et al. (2022), in those cases, using negation could be misleading if children reason logically, since *mustn't* is logically equivalent to *cannot*. Would a higher frequency of *mustn't* in the input lead to later mastery? We couldn't run the experiment on child *must* because of how infrequent it is in this corpus. The second effect that we found was that children more exposed to **general modal talk** seem to master *have to* earlier. The effect could come from modal talk making possibilities and necessities more salient, showing that they are notions that can be talked about. It could also be due to knowing a dual, which might facilitate modal force learning (see Dieuleveut, Cournane and Hacquard, 2020, for a novel word experiment on modals with adults that tests the effect on knowing a scale-mate), though see footnote 2.

Of course, the conclusions we can draw from this study are limited by the small number of children and age window. But this study lays the groundwork for future research, by allowing the identification of factors that matter in the learning process. Future research will probe further other factors. In particular, Dieuleveut et al. (2022) propose that cues from the perceived **(un)desirability** of the event described by the complement of the modal might be useful for root modals: hearing necessity modals with complements describing an undesirable event (e.g., 'You *sig* clean your room', vs. 'You *sig* go play outside') might help children figure out that *have to* expresses necessity. We plan to investigate whether children actually make use of such cues using the same methods in future work.

References

Anderson, Nina J., Susan A. Graham, Heather Prime, Jennifer M. Jenkins, and Sheri Madigan. "Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis." *Child Development* 92, no. 2 (2021): 484-501.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. "Fitting linear mixed-effects models using lme4." *arXiv preprint arXiv:1406.5823* (2014).

Berwick, Robert C. *The acquisition of syntactic knowledge*. Vol. 16. MIT press, 1985.

Brown, Roger. "A first language: The early stages." *Cambridge, MA: Harvard University* (1973).

Chomsky, Noam, and Howard Lasnik. "Filters and Control', Linguistic Inquiry8, 425-504." *Chomsky4258Linguistic Inquiry1977* (1977).

Cournane, Ailís. "Revisiting the epistemic gap: Evidence for a grammatical source." In *Proceedings of the 39th annual BUCLD*. 2015.

Dell, François. "On the learnability of optional phonological rules." *Linguistic Inquiry* 12, no. 1 (1981): 31-37.

Dieuleveut, Anouk, Ailís Cournane, and Valentine Hacquard. "Finding the force: a novel word learning experiment with modals." *ELM* 1 (2021): 136-146.

Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane, and Valentine Hacquard. "Learning modal force: evidence from children's production and input." In *Proceedings of the 2019 Amsterdam Colloquium*, pp. 111-122. 2019.

Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane, and Valentine Hacquard. "Finding the force: How children discern possibility and necessity modals." *Natural Language Semantics* (2022): 1-42.

Dieuleveut, Anouk. "Finding modal force." PhD diss., University of Maryland, College Park, 2021.

Fenson, Larry. *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company, 2007.

Gillette, Jane, Henry Gleitman, Lila Gleitman, and Anne Lederer. "Human simulations of vocabulary learning." *Cognition* 73, no. 2 (1999): 135-176.

Gualmini, Andrea, and Bernhard Schwarz. "Solving learnability problems in the acquisition of semantics." *Journal of Semantics* 26, no. 2 (2009): 185-215.

Hart, Betty, and Todd R. Risley. *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing, 1995.

Hoff, Erika, and Letitia Naigles. "How children use input to acquire a lexicon." *Child development* 73, no. 2 (2002): 418-433.

Hoff, Erika. "The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech." *Child development* 74, no. 5 (2003): 1368-1378.

Hsu, Ning, Pamela A. Hadley, and Matthew Rispoli. "Diversity matters: Parent input predicts toddler verb production." *Journal of child language* 44, no. 1 (2017): 63-86.

Laks, Dan R., Leila Beckwith, Sarale E. Cohen, and Sarale E. Cohen. "Mothers' use of personal pronouns when talking with toddlers." *The Journal of genetic psychology* 151, no. 1 (1990): 25-32.

Leahy, Brian, and Eimantas Zalnieriunas. "Might and might not: Children's conceptual development and the acquisition of modal verbs." In *Semantics and Linguistic Theory*, vol. 31, pp. 426-445. 2022.

MacWhinney, Brian. The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs. Psychology Press, 2014.

Noveck, Ira A. "When children are more logical than adults: Experimental investigations of scalar implicature." *Cognition*78.2 (2001): 165-188.

Ozturk, Ozge, and Anna Papafragou. "The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation." *Language Learning and Development* 11.3 (2015): 191-214.

Piantadosi, Steven Thomas. "Learning and the language of thought." PhD diss., Massachusetts Institute of Technology, 2011.

Team, R. Core. "R: A language and environment for statistical computing." (2013): 201.

Rowe, Meredith L. "A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development." *Child development* 83, no. 5 (2012): 1762-1774.

Rowe, Meredith L. "Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill." *Journal of child language* 35, no. 1 (2008): 185-205.

Rowe, Meredith L., and Adriana Weisleder. "Language development in context." *Annual Review of Developmental Psychology* 2 (2020): 201-223.

Snyder, William. *Child language: The parametric approach*. OUP Oxford, 2007.

Theakston, Anna L., Elena VM Lieven, Julian M. Pine, and Caroline F. Rowland. "The role of performance limitations in the acquisition of verb-argument structure: An alternative account." *Journal of child language* 28, no. 1 (2001): 127-152.

Wexler, Kenneth, and M. Rita Manzini. "Parameters and learnability in binding theory." In *Parameter setting*, pp. 41-76. Springer, Dordrecht, 1987.