

Devoir, ou pouvoir, that is the question

Anouk Dieuleveut & Ira Noveck*

Abstract. In languages like French and English, modals express either possibility (e.g., “you can”) or necessity (e.g., “you must”). Previous acquisition research has shown that English-speaking children have particular difficulty with necessity modals: comprehension experiments show that they tend to accept *must* or *have-to* in possibility scenarios (Noveck 2001, Özturk & Papafragou 2015, a.o.); production studies show that they use them less frequently than possibility modals, and when they do, their usage is not always adult-like (Dieuleveut et al. 2022). But the cause of this “Necessity Gap” remains debated. One challenge is that past studies have focused primarily on English, where necessity modals are much rarer than possibility modals in parental speech, which could suggest that the delay is simply due to less exposure. In this study, we demonstrate through a corpus analysis of French young children’s modal use and their linguistic input, as well as experiments based on this data (following the methods of Dieuleveut et al. 2022), that the delay cannot be attributed solely to limited exposure: despite more exposure, French-speaking children experience the same difficulties with necessity modals. Furthermore, we show that these difficulties persist until children are five years old.

Keywords. Modal acquisition; Corpus Study; Human Simulation Paradigm; French/English comparison; Necessity Gap.

1. Introduction. In languages like English and French, modals express either possibility or necessity. For instance, *peux* (‘can’), in (1a), means that it is possible for you to sleep—but you could just as well stay awake. *Dois* (‘must’), in (1b), means that it is necessary for you to sleep—with no other option. In this paper, we investigate when and how children figure out the “force” of their modals: that *pouvoir* (‘can’) expresses possibility, whereas *devoir* (‘must’) expresses necessity.

- | | | | |
|-----|---------------------------|-------------------------|-------------|
| (1) | a. Tu peux dormir. | ‘You can sleep’ | POSSIBILITY |
| | b. Tu dois dormir. | ‘You must sleep’ | NECESSITY |

The meaning difference between (1a) and (1b) appears obvious, but from the perspective of the child, figuring out modal force may not be so simple. Indeed, the entailment relation between necessity and possibility creates a logical *Subset problem* (Berwick 1985, Wexler & Manzini 1987, a.o.). Whenever a necessity modal statement like (1b) is true, the paired possibility statement (1a) is also true. If a learner mistakenly assumes that a modal means necessary, when it actually means possible, they will get evidence that their hypothesis is wrong, seeing it used in ‘possible but not necessary’ situations. However, if they assume a modal means possible when it actually means necessary, they won’t find direct evidence to falsify their hypothesis, as necessity always implies possibility. So what stops children from assuming that necessity modals like *must* mean possible?

* We would like to thank Morgan Moyer, Ailís Cournane, Valentine Hacquard, Annemarie van Dooren, Keny Chatain, Tom Meadows, David Mueller, Dominique Juffin, members of the GRISP at the LLF, members of the of the NYU Language Acquisition Lab, audiences at the Linguistic Evidence conference in Paris and audiences at BUCLD47. This project is supported by a Fyssen Foundation post-doctoral study grant and by the Chaire d’Excellence Université de Paris Cité. Authors: Anouk Dieuleveut, Université de Genève (anouk.dieuleveut@unige.ch) & Ira Noveck, Laboratoire de Linguistique Formelle, UMR 7110, Université de Paris, CNRS, France (ira.noveck@gmail.com).

This problem, which arises for any pair of words that entertain entailment relations (like *dog/animal*, Piantadosi 2011, Xu & Tenenbaum 2007, or *some/every* Rasin & Aravind 2021), is exacerbated by modal semantics. As modals are words used to talk about non-actual states of affairs, they lack clear reliable physical correlates—similarly to attitude verbs such as *think* or *want* (Gleitman et al. (2005)’s “hard words”). Moreover, in languages like French and English, modals can be used to express various types (or ‘flavors’) of modality: for instance, *doit* in (1b) can mean that you are *required* to sleep, but could also mean in a different context, that it’s *likely*. Children have to acquire these two dimensions, force and flavor, in tandem.

Previous research has shown that English-speaking children struggle with necessity modals, drawing insights from both comprehension studies—which show that 4-year-olds tend to both over-accept possibility modals in necessity situations and necessity modals in possibility situations (Noveck 2001, Öztürk & Papafragou 2015, a.o.), and studies of their productions—which show that children start producing necessity modals later than possibility modals, use them less frequently, and when they do, do so in a non-adult-like way (Dieuleveut et al. 2019, 2022). The origin of this “Necessity Gap” (Dieuleveut 2021) is a matter of debate. Children’s non-adult behavior has been attributed to a) conceptual difficulties reasoning with indeterminacy (cf the “Premature Closure” hypothesis from Acredolo & Horobin 1987; Öztürk & Papafragou 2015, Moscati 2017), b) semantic difficulties, with the meaning of necessity modals (children would not have figured out their underlying force) (Dieuleveut 2021, Cournane et al. submitted), or to c) pragmatic immaturity. But one issue limiting the conclusions we can draw is that these studies have focused primarily on English, where necessity modals are quite rare in the input as compared to possibility modals (Dieuleveut et al. 2022). Crucially, we do not know whether those difficulties are specific to English; the question of how children learn modal force has not been assessed in other languages.

The goal of this study is to fill in this gap by comparing French to English, using the exact same methods in both languages. What makes this comparison particularly interesting is that we find in French parental speech the opposite pattern from English: *necessity* modals are more frequent. This allows us to directly test the effect of quantity of exposure on children’s mastery. We will show that despite hearing more necessity modals, French children face the same difficulties as their English counterparts: they produce necessity modals later, less frequently and tend to “overuse” them, i.e. to use them in situations where adults find possibility modals more appropriate. We conclude that the delay cannot be due only to low quantity of exposure.

The rest of the paper is structured as follows. In **Section 2**, we give a brief overview of modals’ semantics and review existing results from the acquisition literature, showing that English-speaking children struggle with necessity modals. We then turn to our study. First, in **Section 3**, we report quantitative corpus data on French children’s modal productions and input, comparing our results to Dieuleveut et al. (2022)’s for English.¹ Then, in **Sections 4** and **5**, we report the two experiments run on the corpus data. The first one re-adapts Dieuleveut et al. (2022)’s paradigm to French, allowing us to replicate their results. We extend the experiment to older children (4 to 5 y-o), where difficulties persist. The second experiment is a follow-up study carried out to exclude the possibility that our results stem from participants’ expectations that children use more possibility modals. We show they do not. In **Section 6**, we reflect on the origin of children’s misuses, and identify questions that research should address next. Additionally, we discuss our experimental paradigm’s advantages and limitations, and how it could be used for other cases of word learning.

¹ We used the exact same methods as Dieuleveut et al. (2022) for their corpus study on English. We present them on a par to ease comparison. Preliminary results for French, with details about modals’ interaction with negation and flavor, can be found in Dieuleveut (2023).

2. Background. The meaning of modals is typically captured along two axes: force and flavor. The force corresponds to whether a modal expresses possibility (e.g., *pouvoir* ‘can’ in (1a)), or necessity (e.g. *devoir* ‘must’ in (1b)). In formal semantics, this difference is standardly captured by treating modals as quantifiers over (contextually determined) sets of possible worlds, paralleling quantifiers over sets of individuals (*some/all*). Possibility modals express existential quantification (in some possible worlds, *p* is true). Necessity modals express universal quantification (in all possible worlds, *p* is true). This analysis nicely captures the logical entailment between *must* and *can*: whenever (1b) is true, (1a) is also true.

The flavor corresponds to the type of modality the modal conveys: possibility/necessity given what is known, as in (2a) (*epistemic* modality), or based on some rules, as in (2b) (*deontic* modality), or based on some goals (*teleological* modality) (Kratzer 1981, 1991; see von Stechow 2006, Hacquard 2011, for overviews). Typologies of modal flavors vary, but most authors agree on a major distinction between epistemic and non-epistemic, based on both syntactic and semantic criteria. Here, we label non-epistemic as ‘root,’ following the terminology from Hoffman (1966).

- (2) Ira **peut/doit** dormir. ‘Ira **may/must** sleep.’
- a. Epistemic: according to what we know, it is possible/certain/likely that...
 - b. Deontic (root): according to the rules, it is allowed/required that...
 - c. Teleological (root): according to his goals, it is possible/necessary to...

In languages like French and English, modals (auxiliaries, semi-auxiliaries, as well as some adjectives and adverbs) lexically encode force, but vary in flavor.² As highlighted in the Introduction, modals thus raise a particularly complex learning problem. First, because the fact that necessity entails possibility creates a logical *Subset problem*: given that every time they will encounter a necessity scenario, the possibility scenario is also true, what prevents learners from thinking that necessity modals just mean possibility?³ With modals, this basic problem is complicated by the fact that modals lack clear reliable physical correlates, as discussed for attitude verbs (Gleitman et al. (2005)’s “hard words”). Plus, children must learn polysemous modal flavor—which might help or hinder force acquisition (see Dieuleveut 2021, for discussion).

Behavioral experiments show that preschool-aged children have difficulty with modal force. 4-year-olds tend both to (i) over-accept possibility modals in necessity situations—e.g., accepting “There might be a bear in the box” in a situation where it is certain that there is one; and to (ii) accept necessity modals in possibility situations, where they are false—e.g., accepting “There has to be a bear in the box” when it is merely possible that there is one (Noveck 2001, Öztürk & Papafragou 2015, Cournane et al. submitted, a.o.). The first result has received much attention in the context of children acquisition of scalar implicatures. Indeed, the use of a possibility statement (like “There might be a bear in the box”) typically triggers a scalar implicature that the stronger statement doesn’t hold (as the speaker would otherwise have used it; see Grice 1975). Various studies have argued that children “difficulty” with implicatures depends on their ability to access alternatives (Chierchia et al. 2001; Barner et al. 2011, Skordos & Papafragou 2016, a.o.). They generally take for granted that children already know the meaning of scalar terms. But the second result, children’s acceptance of necessity modals in possibility situations, suggests it is not an innocuous step.

² Some other languages have modals with “variable force”, i.e., that can be used in situations where English speakers would use either a possibility or necessity modal (see e.g., Deal 2011; see Yanovich 2016 for a summary).

³ Note that the statement ‘Necessity entails possibility’ only holds keeping flavor constant. For instance, it is not the case that ‘it is certain that *P*’ entails ‘it is allowed’, or that ‘It is required that *P*’ entails ‘it is likely’.

Studies of children’s modal productions often focus on the flavor dimension (see Papafragou 1998, Cournane 2020, for overviews). It has been shown in various languages that epistemic uses are “delayed”: Children start producing root modals by age 2 (e.g., abilities, obligations), but produce epistemics only from around age 3 (the so-called *Epistemic Gap*, Cournane 2015). Corpus analyses focusing on force are rarer. One study on English (Dieuleveut et al. 2019, 2022, based on the Manchester corpus, Theakston et al. 2001), using a combination of corpus analyses and experiments based on the corpus, reveals a “delay” in children’s mastery of necessity modals (*Necessity Gap*, Dieuleveut 2021). Their study shows that by the age of 2, children use possibility modals frequently, productively, and in an adult-like way, but they start using necessity modals later on, use them less often, and in a non-adult-like way: as when adults would expect possibility modals. To show this, they use a variant of the *Human Simulation Paradigm* (Gillette et al. 1999), where adult participants are asked to guess the force of modals uttered by children in dialogues extracted from the corpus. As they highlight in their discussion, their findings ground another hypothesis: that children are confused about the underlying force of their necessity modals—either they are merely uncertain or have encoded them with possibility force. As they point out, this explanation would also capture results from comprehension experiments with older children, which rarely ask whether children know the force of their modals.

One crucial question is thus to determine whether children’s difficulty with necessity modals is specific to learners of English, or if it is more general. Our study aims to address this question, comparing children’s developmental trends in French and in English using the exact same methods in both languages. One aspect that makes the comparison particularly relevant is the increased frequency of necessity modals in French children’s input, which allows us to directly assess the effect of quantity of exposure.

3. Corpus study

3.1. METHODS. For French (Fr), we use the Lyon Corpus (Demuth & Tremblay 2008) (5 child-mother pairs; age range: 1;00-3;00; 3 females, 2 males) and the Paris corpus (Morgenstern & Parris, 2007) (6 child-mother pairs; 3F, 3M; age range: 0;7-6;03) (CHILDES database, Mac Whinney 2000). For English (En), we use data from Dieuleveut et al. (2022), based on the Manchester corpus (Theakston et al. 2001) (12 child-mother pairs; 6F, 6M; age range: 2;00-3;00). Children were all recorded at home in unstructured play sessions with their parents.

3.2. CODING. To stay close to Dieuleveut et al. (2022) study, we focused on modal (semi)-auxiliaries and therefore excluded some other means to encode modality (adverbs like *peut-être* ‘maybe’, adjectives like *possible*, verbs like *penser* ‘think’ or *vouloir* ‘want’, modal verbal inflection (imperative, conditional, subjunctive). We excluded the semi-aux *aller* (‘go’), which (like *will*) conveys future. In French, *pouvoir* is the only modal expressing possibility; *devoir*, *falloir*, and *avoir à* all express necessity (Chu 2008).⁴ All utterances containing modal (semi)-auxiliaries were extracted and coded for force (3), flavor (epistemic vs root) (4), complement (5) (French: adult: 5,231 utterances; 2-3-year-olds: 1,514; 3-5-year-olds: 1,404; excluding repetitions⁵: adult (2.2%): 5,114; 2-3-y-o: (8.9%): 1,379; 3-5-y-o: 1,296 (7.7%); English: adult: 20,755 utterances; child: 5,842; excluding repetitions and tag-questions: adult (9.1%): 18,853; child (17.8%): 4,800).

⁴ French modals are typically considered as semi-auxiliaries (Hacquard 2010, Borgonovo & Cummins 2007). Contrary to English auxiliaries, they inflect for tense, mood, aspect, and agree with their subject. Note that *falloir* has a peculiar syntax: it only appears in impersonal constructions (with the expletive subject *il*) (*Tu faux venir), and can take both infinitival and CP complements (Il faut [venir]/Il faut [que tu viennes]).

⁵ Repetitions are cases where the speaker repeats a sentence uttered right before by herself or by another speaker.

- (3) Functional modal lemmas by force:
 - a. POSSIBILITY: Fr: *pouvoir*; En: *can, could, might, may, able to*
 - b. NECESSITY: Fr: *falloir, devoir, avoir à*;
En: *must, should, need (to), have to, got to, supposed to, ought to*
- (4) Flavor:
 - a. ROOT: MOTHER: ‘y a plein d’habits sales !’ (‘There are many dirty clothes!’)
MOTHER: ‘elle **doit** laver tout le linge.’ (‘She **must** do all the laundry’)
 - b. EPISTEMIC: CHILD: ‘je trouve pas la grosse.’ (‘I can’t find the big one’)
MOTHER: ‘elle **doit** être restée dans la voiture.’ (‘It **must** be in the car’)
- (5) Complement: + NP (excluded): Fr: ‘Il **faut** du pain’ En: ‘We **need** bread’
+ INF: Fr: ‘Il **faut** venir’ (‘it is necessary to come’)
+ CP: Fr: ‘Il **faut** que tu viennes’ (‘it is necessary that you come’)

3.3. QUANTITATIVE RESULTS. Overall, utterances containing modals represent 3.8% of all adult utterances in French (5.8% in English), and 1.9% of child utterances (2.4% in English). As reported in other languages, epistemic uses are rare, both in adult and child speech (adults: Fr: 5.9%; En: 8.8% of all modal uses; children: 2- to 3-y-o: Fr: 0.4%; En: 2.4%; 3- to-5 y-o: Fr: 1.8%; En: not assessed), except for *devoir*, which (like *must*) is more often used to convey epistemic than root modality (*must*: 64.2%; *devoir*: 60%) (see Dieuleveut 2023). **Table 1** summarizes counts of French adult and child modal productions by force, with English as a comparison. In adult speech, we see a strong difference between French and English in the relative frequency of possibility and necessity modals, with French parents using necessity modals more often (62% of all their modal utterances, vs 28% in English). *Falloir* is particularly frequent (53%). Despite this difference in input, children in both languages produce necessity modals less often (Fr: 38% of all modal utterances between 2 and 3; En: 21%), with only a slight increase for older children (42%).

| FRENCH (n = 7,379) | | | | ENGLISH (n = 23,653) | | |
|--------------------|---------------------|-------------------|-------------------|----------------------|-------------------|-------------------|
| | Adults ⁶ | 2-3-year-olds | 3-5-year-olds | | Adults | 2-3-year-olds |
| | count (% mod utt) | count (% mod utt) | count (% mod utt) | | count (% mod utt) | count (% mod utt) |
| POSS | 2008 (38%) | 850 (62%) | 516 (58%) | POSS | 13500 (72%) | 3798 (79%) |
| NECE | 3108 (62%) | 529 (38%) | 370 (42%) | NECE | 5353 (28%) | 1002 (21%) |
| <i>falloir</i> | 2659 (53%) | 492 (36%) | 298 (34%) | | | |
| <i>devoir</i> | 403 (8%) | 21 (2%) | 66 (7%) | | | |
| <i>avoir-à</i> | 46 (1%) | 16 (1%) | 6 (1%) | | | |
| ALL | 5114 (100%) | 1379 (100%) | 886 (100%) | ALL | 18853 (100%) | 4800 (100%) |

Table 1: Counts and percentages of modal uses by Force and Age group in French and English. Frequency by lemma for English is available in Dieuleveut et al. (2022).

3.4. AGE OF FIRST PRODUCTION. Both French and English children start producing possibility modals quite early (first *pouvoir/can*: around 1 year 11 months), on average 4 months before their first necessity modals (first *falloir/have to*: around 2;03; *devoir/must*: 2;11; *avoir-à*: 5;06).

⁶ Note that data for French *adult* talk corresponds to talk towards 2 to 3-yos, to make it comparable to the English study. However, there is not much variation in adult’s talk when they talk to older children (see Appendix on <https://osf.io/3cwqy/>). We also do not see much variation between mothers, and no correlation between child and mother frequency of use—thought this might simply be due to the low sample size (11 children). Note that since Dieuleveut et al. (2022) includes children up to 3;3, the actual boundary between age groups is 3;03 and not 3;00.

3.5. DISCUSSION. We find that both French and English children use necessity modals less frequently than possibility modals, and start using them later on, with a “gap” of almost 4 months. Given the difference in their input, this is striking. However, sparse production does not imply lack of understanding. Several factors might contribute to the prevalence of possibility modals in child early speech. First, differences in children and adults’ conversational status and goals: Children might be less in a position to give orders or express certainty than adults, therefore less likely to use necessity modals. It could also be that young children avoid using necessity modals because they are harder to produce, or are cognitively costlier (there are few studies on processing of possibility and necessity modals even in adults). Another factor that might play a role is the syntax of *falloir*, the most frequent French necessity modal, which could make it harder to acquire.

Beyond quantitative differences, Dieuleveut et al. (2022) have shown, using a corpus-based experiment, that English children tend to use their necessity modals in a non-adult-like way: in situations where adults expect possibility modals (e.g., when adults expect “I can see it,” children say “I have to see it”). We adapted their method to French to see if we can generalize this result.

4. Experiment 1. To get a finer-grained assessment of children’s uses, we used the method introduced by Dieuleveut et al. (2019) (itself a variant of the Human Simulation Paradigm, Gillette et al. 1999).⁷ The goal is to determine whether children use their possibility and necessity modals in an adult-like way, i.e., in the same contexts as adults would, by asking adult participants to guess the force of a redacted modal uttered by children in dialogues extracted from the corpus.

4.1. METHODS. In the Experiment, run online, adult participants read a series of mother-child dialogues randomly extracted from the corpus. Their task is to guess the force of a blanked-out modal, by picking between two options, either a possibility (*pouvoir*) or a necessity modal (*devoir*; *falloir*). **Figure 1a** illustrates a trial. We use participants’ accuracy in guessing the correct modal (uttered by children) as the measure of how ‘adult-like’ children’s uses were (indicating whether children used their modals in contexts where adults would). As a baseline, we use the same experiment on Mothers’ modal utterances (**Figure 1b**). **Procedure.** All experiments were coded using PennController for Ibex (Zehr & Schwarz, 2018) (<https://www.pcibex.net/>) and hosted on the LLF IbexFarm server (<https://ibex.llf-paris.fr/>). Overall, each participant had 40 dialogues to judge, presented in a randomized order: 20 controls using tense (past/future); 20 trials (10 possibility, 10 necessity, randomly selected out of a list of 20 dialogues randomly extracted from the corpus). A demo is available at https://ibex.llf-paris.fr/ibexexps/adiuleveut/ELM3_demo/experiment.html (2 to 3 y-o, Exp1D). **Conditions.** We ran two versions varying the necessity modal (**Exp1D**: *devoir* vs *pouvoir*; **Exp1F**: *falloir* vs *pouvoir*; *avoir à* was too rare to be tested). We had three groups based on the speaker’s age: 2-to 3-year-olds, 4-to 5-year-olds, Mothers (used as baseline). Force was tested within subjects, Age and Lemma between subjects. We tested only ‘root’ modals because epistemic uses are too rare in children’s production, and we excluded negated utterances because of issues with scopal irregularities with negated modals (Iatridou & Zeiljstra 2013).⁸ For controls, participants had to pick between future and past (e.g. [*a vu*] vs. [*va voir*]). For these, we

⁷ The original goal of the *Human Simulation Paradigm* is to compare different kinds of cues available to the child to figure out words’ meanings (Gillette et al., 1999). Here, we use with a different goal: as a way to evaluate children’s production. We use the HSP on adult’s production as a baseline, indicating that force can be guessed from conversational context. This method—using adults judgments to assess children usage—was developed in English (Dieuleveut 2021, Dieuleveut et al. 2022), and has also been used for definite description (*the* vs *a*) (Ying et al., 2024).

⁸ Dieuleveut et al. 2022 have two instantiations of Root-Positive (rootP1: *can* vs *must*; rootP2: *can* vs *have to*). They also test negated root modals (*can’t* vs *don’t have to*) and epistemic modals (*might* vs *must*).

ensured the correct answer was not always guessable based on the target sentence alone, to force participants to read the entire dialogues. **Material.** Contexts were randomly selected from the corpus database for each modal lemma (*pouvoir, devoir, falloir*) (excluding negated cases, epistemic uses and repetitions) and manually checked. In cases where another modal was used in the dialogue (either by the child or the adult), we left it as such, regardless of its force. In cases where the same modal was used several times in the dialogue, we put several blanks. To replace *falloir* with *pouvoir* (in **Exp1F**), we proceeded as follows: (i) for cases like ‘Il faut [V+INF]’ (e.g., ‘il faut manger’ ‘(you) have to eat’), we used French impersonal subject *on* (participants had to pick between ‘il faut’ and ‘on peut’); (ii) for cases like ‘il faut [_{CP} que tu manges]’, they had to pick between ‘tu peux manger’ and ‘il faut que tu manges’. When the expletive subject *il* was dropped, we put it back in the answers (e.g. ‘faut manger’ → ‘__ manger’: pick between ‘il faut’ and ‘on peut’). We used the same lists in **Exp1F** and **Exp1D** for possibility modals as a check, expecting no difference in accuracy. The final lists consisted of 20 dialogues for possibility and 20 dialogues for necessity for each three age groups (except for 2-3 y-o child *devoir*, where we could test only 17 contexts) (total: 237 test; 60 controls). They are available at <https://osf.io/3cwqy/>.

| | |
|--|---|
| <p>ENFANT : ... t'en laisses un petit coup. MAMAN : merci. ENFANT : voilà. MAMAN : merci. ENFANT : arrête d'aller là avec le ptit chevaux ENFANT : vous arrêtez d'aller là. ENFANT : parce que c'est après. ENFANT : qu'on _____ aller après.</p> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px;">doit</div> <div style="border: 1px solid black; padding: 2px 10px;">peut</div> </div> | <p>AUTRE ADULTE : oui. MAMAN : oui AUTRE ADULTE : une soucoupe. ENFANT : sont un peu vieilles. MAMAN : oui sont un peu abîmées tordues. MAMAN : ah celle-là elle marche bien. AUTRE ADULTE : merci beaucoup. MAMAN : tu _____ souffler dessus.</p> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px;">dois</div> <div style="border: 1px solid black; padding: 2px 10px;">peux</div> </div> |
| <p>CHILD: ... you leave a little. / MUM: thank you. / CHILD: there you go. / MUM: thank you. / CHILD: stop going there with the horsie. / CHILD: you stop going there. / CHILD: because it's after. / CHILD: that we _____ go after.</p> | <p>OTHER ADULT: yes. / MOTHER: Yes. OTHER ADULT: A saucer. / CHILD: They're a bit old. / MUM: Yes, they're a bit bent damaged. / MOTHER: Oh, this one works well. / OTHER ADULT: Thank you very much. / MOTHER: you _____ blow on it.</p> |

Figure 1: Example trials, Experiment 1 (1D: *pouvoir* vs *devoir*). On the left: Experiment on children’s production (2-3yo). On the right: Experiment on mothers’ production (MOT, baseline).

4.2. PARTICIPANTS. 358 French participants were recruited on Prolific (60 per condition, 2 failed to record data) (166 F, 186 M, 6 NB; mean age: 32.8yrs; age range: 18 to 74yrs). We removed 11 participants whose accuracy scores on controls were <75% (3.1%). We thus report results for 347 participants (**Exp1D**: MOT: 59; 2-3yo: 56; 4-5yo: 59; **Exp1F**: MOT: 59; 2-3yo: 54; 4-5yo: 58).

4.3. RESULTS. Data analyses were conducted using R (R Core Team, 2013), using the package lme4 (Bates et al. 2014a, 2014b). Accuracy for controls was high (91.2%) (**Exp1D**: MOT: 96.1%; 2-3yo: 88.1%; 4-5yo: 90.8%; **Exp1F**: MOT: 96.0%; 2-3yo: 87.6%; 4-5yo: 91.8%). **Figure 2** shows the mean accuracy by force and age group after participant exclusion. On average, each context was seen by 38.9 participants, ranging between 19 and 59 times. ANALYSIS. Overall, participants were accurate at recovering the original modal force (**Table 2**). Binomial tests (reported in **Appendix**) show they differ from chance in all conditions. To test for the effect of Force, we ran generalized linear mixed effects models, built with a maximal random effect structure allowed by our experimental design, testing Accuracy (dependent variable, binomial), with Force as fixed

effect and Subject and Item as random factors, and compare them with reduced models without Force as a fixed effect (following Barr *et al.*, 2013) (glmer syntax of the full model: Accuracy ~ Force + (Force|Subject) + (1|Item), reduced model: Accuracy ~ 1 + (Force|Subject) + (1|Item)).⁹ **Effect of Force.** For mother’s production, we find no difference in accuracy between possibility and necessity contexts (both overall and by experiment) (general mean accuracy: Poss: 78%; Nec: 77%). For child production, in both age groups, we find higher performance on possibility than necessity contexts (both overall and by experiment) (2-3yo: Poss: 75%; Nec: 60%; 4-5yo: Poss: 82%; Nec: 64%). Results of the models’ comparison are given in **Table 3. Effect of Age.** Comparing Child groups to Mother (glmer syntax of the full model: Accuracy ~ Age + (1|Subject) + (1|Item); reduced model: Accuracy ~ 1 + (1|Subject) + (1|Item)), we find no difference between children and adults for possibility modals (for both age groups), but the difference is significant for children necessity modals in Exp1F (almost significant in 1D), and significant overall for 4 to 5-year-olds. There is no significant difference when comparing 2- to 3-year-olds with 4- to 5-year-olds. **Table 4** summarizes the results. We checked that there was no effect of modal lemma (comparing Exp1D to Exp1F for all groups; glmer syntax of the full model: Accuracy ~ Experiment + (1|Subject) + (1|Item) (results are reported in Appendix). Finally, looking at the interactions Age*Force (full model: Accuracy ~ Force * Age + 1|Subject + 1|Item)), we find a significant difference comparing 4- to 5-year-olds to mothers, indicating that the difference in accuracy between possibility and necessity modals for child productions is larger than for mothers productions. The difference is almost significant comparing 2- to 3-year-olds with mothers (p=0.06). We find no effect between the two child age groups. Results are summarized in **Table 5.**

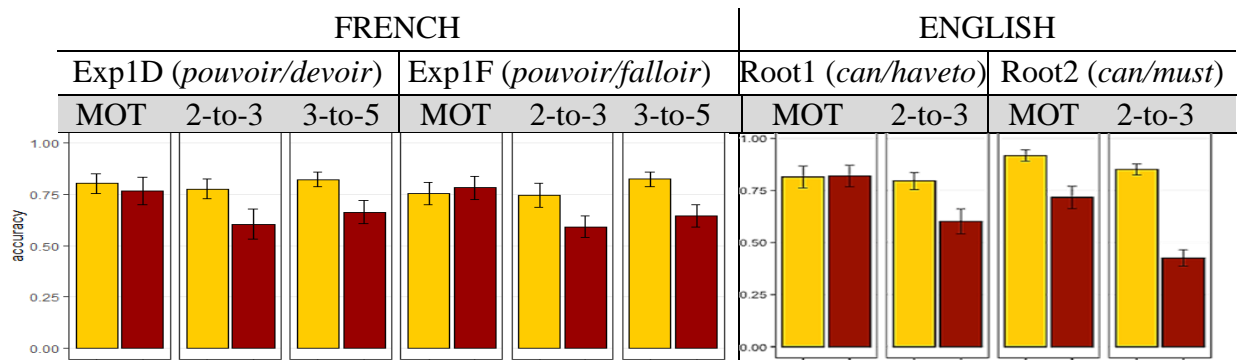


Figure 2: Mean accuracy by Force and Age Group on Exp1D and Exp1F, with English (Dieuleveut et al 2022, condition root_P1 and root_P2) as comparison.

| | MOTHER | | 2-3 y-o | | 4-5 y-o | |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|
| | POSS | NECE | POSS | NECE | POSS | NECE |
| Exp1D | 80.4% (0.048) | 76.6% (0.068) | 77.7% (0.049) | 60.5% (0.074) | 82.3% (0.036) | 66.4% (0.057) |
| Exp1F | 75.3% (0.055) | 78.2% (0.055) | 74.5% (0.058) | 59.3% (0.053) | 82.5% (0.036) | 64.6% (0.054) |
| ALL | 77,9% (0,036) | 77,4% (0,043) | 76,1% (0,038) | 59,8% (0,044) | 82,4% (0,025) | 65,5% (0,039) |

Table 2: Mean accuracy (se) by Age and Force, Experiment 1D and 1F. Accuracy corresponds to the mean accuracy (how good participants were to guess correctly the force of the modal given the context) across the 20 contexts initially extracted from the corpus for each condition of Force and Age. Each participant saw only 10 contexts (10 possibility, 10 necessity), randomly picked within PCIBex. On average, each context was seen 39 times, ranging between 19 and 59 times.

⁹ Answers were coded as 1 if the response was accurate, and 0 otherwise. The same procedure based on model comparisons was used for all subsequent experiments, so we don’t systematically report the reduced model.

| | MOTHER | 2-3 y-o | 4-5 yo |
|-------|-----------------------------------|----------------------------------|----------------------------------|
| Exp1D | $\chi^2(1) = 0.51, p = 0.47$ (NS) | $\chi^2(1) = 4.05, p = 0.04$ * | $\chi^2(1) = 5.28, p = 0.02$ * |
| Exp1F | $\chi^2(1) = 0.26, p = 0.61$ (NS) | $\chi^2(1) = 4.1, p = 0.04$ * | $\chi^2(1) = 6.52, p = 0.01$ * |
| ALL | $\chi^2(1) = 0.02, p = 0.88$ (NS) | $\chi^2(1) = 8.16, p = 0.004$ ** | $\chi^2(1) = 11.6, p < .001$ *** |

Table 3: Results of the model testing effect of Force (possibility vs necessity), by Age and Lemma, Experiment 1D and 1F.

| | MOT vs 2-3yo | | MOT vs 4-5 yo | | 2-3yo vs 4-5yo | |
|-------|------------------------------------|-----------------------------------|-----------------------------------|------------------------------------|-----------------------------------|-----------------------------------|
| | Possibility | Necessity | Possibility | Necessity | Possibility | Necessity |
| Exp1D | $\chi^2(1) = 0.21, p = 0.643$ (NS) | $\chi^2(1) = 2.21, p = 0.14$ (NS) | $\chi^2(1) = 0.01, p = 0.91$ (NS) | $\chi^2(1) = 1.67, p = 0.20$ (NS) | $\chi^2(1) = 0.17, p = 0.68$ (NS) | $\chi^2(1) = 0.2, p = 0.66$ (NS) |
| Exp1F | $\chi^2(1) = 0, p = 0$ (NS) | $\chi^2(1) = 6.88, p = 0.01$ ** | $\chi^2(1) = 0.67, p = 0.41$ (NS) | $\chi^2(1) = 3.71, p = 0.054$ (NS) | $\chi^2(1) = 0.63, p = 0.43$ (NS) | $\chi^2(1) = 0.61, p = 0.43$ (NS) |
| ALL | $\chi^2(1) = 0.10, p = 0.75$ (NS) | $\chi^2(1) = 8.21, p = 0.004$ ** | $\chi^2(1) = 0.24, p = 0.63$ (NS) | $\chi^2(1) = 5.12, p = 0.02$ * | $\chi^2(1) = 0.72, p = 0.40$ (NS) | $\chi^2(1) = 0.77, p = 0.38$ (NS) |

Table 4: Results of the model testing effect of Age (adult vs. child usage) Experiment 1D and 1F.

| | MOTHER vs 2-3yo | MOTHER vs 4-5 yo | 2-3 yo vs 4-5 yo |
|-------|------------------------------|---------------------------------|------------------------------|
| Exp1D | $\chi^2(1) = 0.67, p = 0.41$ | $\chi^2(1) = 0.76, p = 0.38$ | $\chi^2(1) = 0, p = 0.98$ |
| Exp1F | $\chi^2(1) = 3.43, p = 0.06$ | $\chi^2(1) = 3.84, p = 0.05$ * | $\chi^2(1) = 0.01, p = 0.92$ |
| ALL | $\chi^2(1) = 3.54, p = 0.06$ | $\chi^2(1) = 3.95, p = 0.047$ * | $\chi^2(1) = 0, p = 0.97$ |

Table 5: Results of the model testing Interactions Force * Age, Experiment 1D and 1F.

4.4. DISCUSSION. For the baseline on mothers’ productions, we find no effect of Force: participants are accurate at guessing force based on the dialogue, with no difference between possibility and necessity contexts. But in children’s productions, we find one, always in the direction of children’s necessity modals being harder to identify (**Table 3**). This indicates that children tend to “over-use” necessity modals: they use them when adults would rather use possibility modals. We therefore replicate results for English, and further, we show that they are still present among 4- and 5-year-olds (see **Figure 2**). Examples (6) and (7), which led to particularly low accuracy rates, illustrate some children’s non-adult-like uses of necessity modals.

- (6) [...] MAMAN : et alors tu y arrivais bien ? MOTHER: So you were good at it?
 ENFANT : oui ! CHILD: Yes!
 MAMAN : c'est ce que t'as fait aujourd'hui ? MUM: Is that what you did today?
 ENFANT : oui ! CHILD: Yes!
 MAMAN : ah ! MUM: Oh!
 ENFANT : et puis [**il faut/on peut**] attraper des papillons dans... dans un filet à papillons.
 CHILD: And then [you **have to/you can**] catch butterflies in a butterfly net].
 Paris Corpus, Madeleine, 30028; mean accuracy: 17.4% (nobs = 23)
- (7) [...] ENFANT : oh là c'est encore brûlant CHILD: Oh, it's still hot.
 MAMAN : oh ça va. MUM: Oh, that's all right.
 ENFANT : non là CHILD: No!
 ENFANT : arrête ! CHILD: Stop!
 MAMAN : voilà ! MUM: That's it!
 MAMAN : on s'essuie les mains. MUM: wipe your hands.
 ENFANT : et là aussi je [**peux/dois**] lécher. CHILD: And I [**can/have to**] lick there too.
 Lyon Corpus, Marie, 40005b; mean accuracy: 12.5% (nobs = 32)

- (9) [...] MAMAN : je veux manger dans ma chaise. ‘MOTHER: I want to eat in my chair’
 ENFANT : tu sais aller dans la chaise toute seule ? ‘CHILD: You know how to go in
 your chair by your own?’
 MAMAN : c'est toi tu _____ mettre dans ma chaise. ‘MOTHER: It's you you **have to** put
 me in my chair’ (Paris, Julie, 20122)

5.3. PARTICIPANTS. 120 French participants who had not taken part in Experiment 1 were recruited on Prolific (30 per group) (66 M, 49 F, 2 NB, 3 unknown; mean age: 32.5yrs). Accuracy on controls was high (93.2%, with no difference between Switch/NoSwitch control contexts: MOT: 95.2%; MOT-to-CHI: 94.3%; CHI: 91.7%; CHI-to-MOT: 91.8%). We excluded 2 participants due to low accuracy on controls (<75%).

5.4. RESULTS. Table 6 summarizes mean accuracy in each condition. The first two rows provide results from Experiment 1, first on all contexts, second after the context selection. The two last rows give results from Experiment 2, first without the switch, then with the switch. We tested the effect of the Switch using generalized linear mixed effects models, built with a maximal random effect structure allowed by our experimental design (details of the model are reported in Appendix). We find no effect except for mother’s possibility modals, which overall and in Exp2F lead to lower accuracy with the switch, indicating that participants tend to use more necessity modals when the modal is attributed to the child. This is the opposite of what we would obtain if participants were expecting children to use more possibility modals.

| | ExpD (<i>pouvoir</i> vs <i>devoir</i>) | | | | ExpF (<i>pouvoir</i> vs <i>falloir</i>) | | | |
|-------------------------|--|---------|---------|---------|---|---------|---------|---------|
| | MOTHER | | 2-3 yo | | MOTHER | | 2-3 yo | |
| | POSS | NECE | POSS | NECE | POSS | NECE | POSS | NECE |
| i Exp1 (all contexts) | 80% | 77% | 78% | 61% | 75% | 78% | 75% | 59% |
| ii Exp1 (kept for Exp2) | 80% | 72% | 80% | 66% | 73% | 79% | 78% | 57% |
| iii Exp2 (no switch) | 79,4% | 71,4% | 81,8% | 64,8% | 70,0% | 84,8% | 75,7% | 59,0% |
| | (0,054) | (0,099) | (0,061) | (0,105) | (0,095) | (0,039) | (0,088) | (0,095) |
| iv Exp2 (switch) | 73,7% | 64,4% | 83,7% | 65,2% | 62,4% | 83,0% | 79,2% | 52,8% |
| | (0,074) | (0,109) | (0,048) | (0,094) | (0,085) | (0,028) | (0,068) | (0,102) |

Table 6: Mean accuracy (se) by Age and Force, Experiment 2D and 2F (n=118), compared to Experiment 1D/1F. On average, each context was seen 41 times in Experiment 2 (20.9 times with original speaker, 20.2 times with switched speakers), ranging between 16 and 60 times.

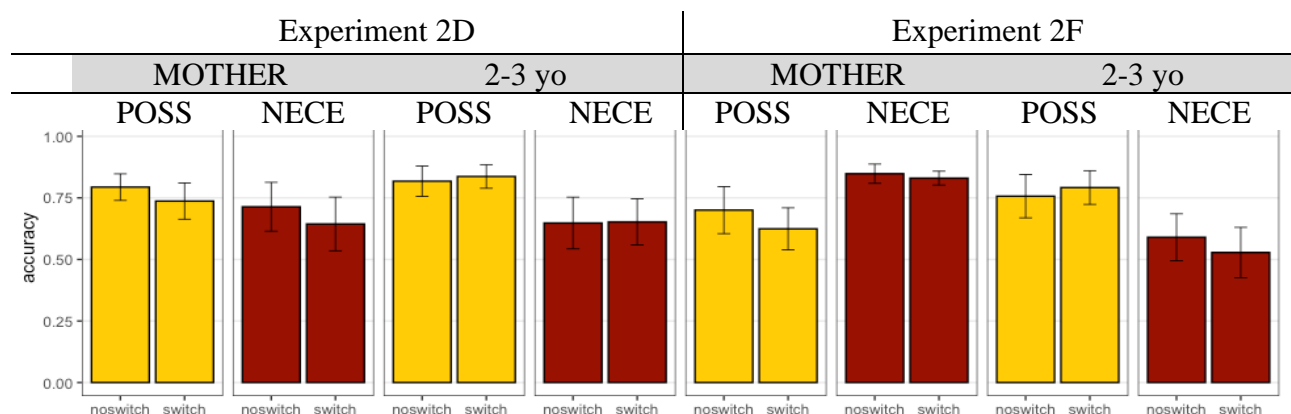


Figure 5: Effect of role switch, Experiment 2D and 2F.

| | MOT | | 2-3 yo | |
|-------|--------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | Possibility | Necessity | Possibility | Necessity |
| Exp2D | $\chi^2(1)=0.15, p=0.70$ (NS) | $\chi^2(1)=2.28, p=0.13$ (NS) | $\chi^2(1)=0.03, p=0.85$ (NS) | $\chi^2(1)=0.04, p=0.83$ (NS) |
| Exp2F | $\chi^2(1)=5.21, p=0.02$ * | $\chi^2(1)=0.32, p=0.57$ | $\chi^2(1)=0, p=1$ (NS) | $\chi^2(1)=1.34, p=0.25$ (NS) |
| ALL | $\chi^2(1)=4.15, p=0.04$ * | $\chi^2(1)=0.55, p=0.46$ (NS) | $\chi^2(1)=0.01, p=0.93$ (NS) | $\chi^2(1)=0.73, p=0.39$ (NS) |
| ALL | $\chi^2(1)=3.21, p=0.073$ (NS) | | | |

Table 7: Results of the model testing effect of Switch (no switch vs switch), by condition and age group, Experiment 2D and 2F.

6. Discussion. Our study makes three important points. First, we have shown that children’s difficulties with necessity modals are not limited to learners of English. We replicated Dieuleveut et al. (2022)’s findings for English in French: children master possibility modals early, but use necessity modals them later on, less frequently, and, crucially, use them in a non-adult-like way: they “overuse” them. Second, we have shown that these difficulties persist with older children (a point not assessed in Dieuleveut et al. 2022 on English): our experiment shows that 4- to 5-year-olds still use their necessity modals in a non-adult-like way. Third, our study shows that the “delay” for necessity modals cannot be due only to low quantity of exposure. French children actually hear more necessity than possibility modals in their input: Hearing more necessity modals doesn’t help.

Are children confused about the meaning of necessity modals? Or, is it simply that they do not know yet in which contexts they are appropriate? More research will be needed to answer these questions. Children might also fail to use necessity modals as adults do because of problems determining the modal’s domain of quantification (specifically, consider a smaller set of worlds than adults). Politeness could also be involved: adults tend to use possibility modals to express requests, in contexts in which children might more directly use necessity modals. One of our goal is to tease these hypotheses apart, acknowledging that they are not necessarily mutually exclusive.

Taken together, our results call for extension to other logical scales, like *some/all* or *sometimes/always*, where similarly *Subset problems* arise—do children have similar difficulty with *all* and *always*? The method we used, which uses adult’s judgments to evaluate children’s spontaneous productions, based on existing corpus data, has several advantages: it is quite easy to deploy, simple from the participant’s perspective, and therefore easily generalizable to other languages and to other cases of word learning.

References

- Acredolo, Curt & Karen Horobin. 1987. Development of relational reasoning and avoidance of premature closure. *Developmental Psychology* 23(1): 13–21. <https://doi.org/10.1037/0012-1649.23.1.13>.
- Barner, David, Neon Brooks & Alan Bale. 2011. Accessing the unsaid: The role of scalar alternatives in children’s pragmatic inference. *Cognition* 118(1): 84–93. <https://doi.org/10.1016/j.cognition.2010.10.010>.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. <https://doi.org/10.18637/jss.v067.i01>
- Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge: MIT Press.

- Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini & Luisa Meroni. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University Conference on Language Development*, eds. A.H.-J. Do, L. Domínguez, and A. Johansen, 157–168. Somerville: Cascadilla Press.
- Chu, Xiaoquan. 2008. *Les verbes modaux du français*. Editions OPHRYS.
- Cournane, Ailís & Sandrine Tailleur. 2020. La production épistémique chez l'enfant francophone: Complexité syntaxique et ordre d'acquisition. *Arborescences : revue d'études françaises*, 10, 47–72.
- Cournane, Ailís. 2015b. *Modal development: Input-divergent L1 acquisition in the direction of diachronic reanalysis*, PhD dissertation, University of Toronto.
- Cournane, Ailís. 2021. Revisiting the epistemic gap: Evidence for a grammatical source. *Language Acquisition* 28(3): 215-240. <https://doi.org/10.1080/10489223.2020.1860054>.
- Crain, Stephen & Rosalind Thornton. 1998. *Investigations in universal grammar*. Cambridge: MIT Press.
- Deal, Amy Rose. 2011. Modals without scales. *Language* 87(3): 559–585. <https://doi.org/10.1353/lan.2011.0060>
- Demuth, Katherine & Annie Tremblay. "Prosodically-conditioned variability in children's production of French determiners." *Journal of child language* 35.1 (2008): 99-127.
- Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane & Valentine Hacquard. 2019a. Learning modal force: Evidence from children's production and input. In *Proceedings of the 2019 Amsterdam Colloquium*, 111–122.
- Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane & Valentine Hacquard. 2019b. Acquiring the force of modals: Sig you guess what sig means? In *Proceedings of the 43rd Annual to the Boston University Conference on Language Development (BUCLD43)*, ed. Megan M. Brown and Brady Dailey, 189-202.
- Dieuleveut, Anouk. 2021. *Finding modal force*, PhD dissertation, University of Maryland.
- Dieuleveut, Anouk. 2023. Je peux, ou je dois? Faudrait savoir! Acquiring Modal Force: Evidence from French. In *Proceedings of the 47rd Annual to the Boston University Conference on Language Development (BUCLD47)*.
- Gillette, Jane, Henry Gleitman, Lila Gleitman & Anne Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73(2): 135–176. [https://doi.org/10.1016/s0010-0277\(99\)00036-0](https://doi.org/10.1016/s0010-0277(99)00036-0).
- Gleitman, Lila R., Kimberly Cassidy, Rebecca Nappa, Anna Papafragou & Trueswell, John C. 2005. Hard words. *Language learning and development*, 1(1), 23-64.
- Grice, Herbert Paul. 1975. Logic and conversation. In *Speech acts*, 41–58, ed. J.P. Kimball. Syntax and Semantics Vol. 3. Leiden: Brill. https://doi.org/10.1163/9789004368811_003
- Gualmini, Andrea & Bernhard Schwarz. 2009. Solving learnability problems in the acquisition of semantics. *Journal of Semantics* 26(2): 185–215. <https://doi.org/10.1093/jos/ffp002>.
- Hacquard, Valentine. 2011. Modality. In *Semantics: An international handbook of natural language meaning*, eds. C. Maienborn, K. von Stechow, and P. Portner, 1484–1515. Berlin: Mouton de Gruyter.
- Horn, Laurence R. 1972. *On the semantic properties of logical operators*, PhD dissertation, UCLA.
- Iatridou, Sabine & Hedde Zeijlstra. 2013. Negation, polarity, and deontic modals. *Linguistic Inquiry* 44(4): 529–568. http://dx.doi.org/10.1162/ling_a_00138.

- Jeretič, Paloma. 2018. Evidence for children's dispreference for weakness: A corpus study. Manuscript, New York University.
- Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10(2): 201–216. <https://doi.org/10.1007/BF00248849>.
- Kratzer, Angelika. 1991. Modality. *Semantics: An international handbook of contemporary research*, eds. C. Maienborn, K. von Stechow, and P. Portner, 639–650. Berlin: Mouton de Gruyter.
- MacWhinney, Brian. 2000. *The CHILDES project: The database*. Vol. 2. Hove, UK: Psychology Press.
- Manzini, M. Rita & Kenneth Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18(3): 413–444. <https://www.jstor.org/stable/4178549>.
- Morgenstern, Aliyah, and Christophe Parisse. "The Paris Corpus." *Journal of French language studies* 22.1 (2012): 7-12.
- Moscato, Vincenzo, Likan Zhan & Peng Zhou. 2017. Children's on-line processing of epistemic modals. *Journal of Child Language* 44(5): 1025–1040. <https://doi.org/10.1017/S0305000916000313>.
- Noveck, Ira A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2): 165–188. [https://doi.org/10.1016/s0010-0277\(00\)00114-1](https://doi.org/10.1016/s0010-0277(00)00114-1).
- Ozturk, Ozge & Anna Papafragou. 2015. The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development* 11(3): 191–214. <https://doi.org/10.1080/15475441.2014.905169>.
- Papafragou, Anna. 1998. The acquisition of modality: Implications for theories of semantic representation. *Mind & Language* 13(3): 370–399. <https://doi.org/10.1111/1468-0017.00082>.
- Piantadosi, Steven T. 2011. *Learning and the language of thought*, PhD dissertation, MIT.
- R Core Team. 2013. R: A language and environment for statistical computing. <https://www.r-project.org/>.
- Rasin, Ezer & Athulya Aravind. 2020. The nature of the semantic stimulus: The acquisition of every as a case study. *Natural Language Semantics* 29: 339–375. <https://doi.org/10.1007/s11050-020-09168-6>.
- Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153: 6–18. <https://doi.org/10.1016/j.cognition.2016.04.006>
- Theakston, Anna L., Elena V. Lieven, Julian M. Pine & Caroline F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28(1): 127–152.
- van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane & Valentine Hacquard. 2022. Figuring out root and epistemic uses of modals: The role of the input. *Journal of Semantics*.
- Xu, Fei & Joshua B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review* 114(2): 245–272. <https://doi.org/10.1037/0033-295X.114.2.245>.
- Yanovich, Igor. 2016. Old English **motan*, variable-force modality, and the presupposition of inevitable actualization. *Language* 92(3): 489–521.
- Zehr, Jeremy & Florian Schwarz. (2018). PennController for Internet Based Experiments (IBEX). <https://doi.org/10.17605/OSF.IO/MD832>