ABSTRACT

Title of Dissertation:	FINDING MODAL FORCE
	Anouk Dieuleveut, Doctor of Philosophy, 2021
Dissertation directed by:	Professor Valentine Hacquard Department of Linguistics
	Associate Professor Alexander Williams Department of Linguistics

This dissertation investigates when and how children figure out the force of modals, that is, when and how they learn that *can*/*might* express possibility, whereas *must*/*have to* express necessity. Learning modal force raises a logical "Subset Problem": given that *necessity* entails *possibility*, what prevents learners from hypothesizing possibility meanings for necessity modals? Three main solutions to other Subset Problems have been proposed in the literature. The first is a bias towards strong (here, necessity) meanings, in the spirit of Berwick (1985). The second is a reliance on downward-entailing environments, which reverse patterns of entailment (Gualmini & Schwarz, 2009). The third is a reliance on pragmatic situational cues stemming from the conversational context in which modals are used (Dieuleveut et al., 2019). This dissertation assesses the viability of each, by examining the modals used in speech to

and by 2-year-old English children, through a combination of corpus studies and experiments testing the guessability of modal force based on their context of use.

I show that negative and other downward-entailing contexts are rare with necessity modals, making them impractical on their own. However, the conversational context in which modals are used in speech to children is highly informative about both forces. Thus, if learners are sensitive to these conversational cues, they, in principle, do not need to rely either on a necessity bias nor on negative environments to solve the Subset Problem.

Turning to children's own productions, I show that children master possibility modals very early: by age 2, they use them productively, and in an adult-like way. However, they struggle with necessity modals: they use them less frequently, and not in an adult-like way. Their modal uses show no evidence for a necessity bias.

To assess how children actually figure out modal force, and which of the available cues children use to figure out modal force, I then examine which aspects of children's input best predict their mastery of modals. Preliminary results suggest that negation is predictive of children's early success with necessity modals, and that frequency of modal talk, but not of particular lexemes, also contributes to their early success.

FINDING MODAL FORCE

by

Anouk Dieuleveut

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Valentine Hacquard, Chair/Advisor
Associate Professor Alexander Williams, Co-Chair/Advisor
Professor Jeffrey Lidz
Assistant Professor Ailís Cournane
Associate Professor Paolo Santorio, Dean's Representative

Dedication

To Timmy, who can breathe,

And to Floriane.

Acknowledgements

My first thanks to Valentine, my advisor. It's been so rich to work with you during these five years! Now that I can take a step back, I realize even more how lucky I am. Thanks for everything I've learnt from you, for being so supportive along these years, always being so patient, always answering my too long emails in less than 24 hours, for helping me get rid of my footnotes[1] (and parentheses), and making me understand there's a good reason why Proust is not so famous for his scientific articles. For being great on both the human aspects and the professional ones, for finding the right tone, for helping my anxiety go away (almost), and for trusting me, while remaining exacting. Et comme en plus, je peux te le dire en français : merci ! J'ai une chance incroyable, d'avoir pu bosser avec toi.

To my second advisor, Alexander. Thanks for being such a unique mind, and always bringing novel, unexpected, perspectives on things. For your help and advice on how to present and write. Not that I became a good writer, but no doubt, I made progress. And, for knowing everything about French andouillette.

To Ailís, who may not officially be my advisor, but who surely {could/should} count as such. Thanks for your energy! For your dynamism, your positivity, your approachability. You've been there since the first year, and the least I can say is that I learnt a lot from you as well. I hope we {can/will} keep on working together!

To the other two members of my committee. To Jeff, for your Socratic and wise way of asking questions rather than providing answers. For being careful, and taking the time to think. And to Paolo, my Dean's representative. For getting interested, and for the extremely relevant points you raised.

Then, I'd like to thank Benjamin and Emmanuel. I can't be grateful enough, to both of you. Thanks for planting the seed, for showing me what research in linguistics is, for giving me the right advice at the right moments, and making me consider doing my PhD in the US, which I would never have done by myself. I know that working with you shaped a lot of what I became, so thank you for that.

[1]And keeping on reading them, still.

And looking ahead, I also want to thank Ira Noveck, for answering so positively when I reached out earlier this year. I'm looking forward to working with you!

We learn from mentors, but we learn from friends as well, so I have two very special thanks to give: one to my dissertation older sister, the other one to my dissertation little sister.

To Annemarie. Well, you know why! It's a rare chance, to work with such a friend. The studies presented in this dissertation would not exist if we hadn't worked together. I'll never say enough how much I owe to you! Thanks for being enthusiastic about modals, for teaching me so many things, for Harpers Ferry, for the Amsterdam colloquiums, for being so quick when learning R, for being a European lost in the US like me, for being stressed and upset—*bref,* for being someone I could talk to in French (quoi, tu comprends pas?)

To Yu'an. You too, you know why! I would not have finished this without you. Thanks for being there last summer when I broke my leg, thanks for helping me with the crutches. And thanks for being there soooo many other times. Thanks for your positivity, your incredible energy, your craziness, your support, for all the presentations we rehearsed together, for trusting me and daring telling me when I'm wrong. And, my apologies to your shoulder. It didn't deserve to see me cry so much.

Five other special thanks, to the greatest cohort, SMAART. To Sig, the spring, for letting me borrow your name so many times. To Mina, the Fairy. I know why the garden just loves you! To Aaron, the bigger A, who cooks the best lasagnas—thanks for reminding me two weeks ago why once upon a time, I wanted to be a syntactician. To Rodrigo, the mango thief, for your support, especially during the last three months (was it a year?), but already when we went to Guatemala; for our discussions—it's so easy to talk with you. To Tyler, the master juggler, for your great sense of humor, and for systematically answering my questions about research, papers, experiments, and life in less than 20 minutes. I will miss you all back in France. A lot.

To Lambda-Tea-ers. To Yu'an, once again (you deserve to go twice, anyway!) To Jessica, for your elegance when you talk about semantics, and for adopting my bike and standing desk now that I leave. I'm happy to know you'll take care of them. And

to Jad, for all our discussions about actuality entailments. I hope one day we can work out something together!

To the members of the ModSquad, both from UMD and NYU. Annemarie, Ailís, Dan, Tyler, Zoe, Maxime, Chiara, Adam, Caleb, Jessica, Jad, Yu'an, Quinn—I probably cannot list you all. To my research assistants, with a special mention to Joon and to Monica who worked with us for over two years. To my LING419T seminar students—I wish we hadn't had to finish the semester on Zoom, but it was great working with you. To other people in the department, students and faculties, older and younger. Paulina, for your calm, for the yoga, and for leaving me your room in the Hill House. Rachel, for all the good vibes, and for showing the way. Thanks for all your help! To so many people, Chia-Hsuan, Laurel, Alice, Anissa, Nick, Gesoel, Anton, Adam, Thomas. A special thanks to Kim. And to all the faculties who're taking off of their time to help us. Masha and Omer, for encouraging me to go to Guatemala in my first year, and meeting with me almost every week. Andrea and Colin, for the parkruns. Special thanks for your pictures, Andrea! You're a talented photographer. Peggy, for your relevant advice about teaching. Naomi—sorry I've kept disappearing all the time! I wish we had another year to finish what we started.

I also want to thank the people I met in McGill, with a special mention to Jessica Coon who was such a great syntax teacher (why move?). And other people from Montreal, linguists, and non-linguists: Martha—I hope we can go biking again one day! My roommates, Nawal and Kemi. And my Quebec family: Marie-France, Annie, Antoine and Elaine, Laurence, and all the others.

Thanks to friends from the Cogmaster, Hyppolite, Arnaud, Sébastien, Nicolas, Mora, Milica. Paloma, for hosting me at NYU, and being interested in the same topics as me! People from Numi: Cici, Kelsey, Rebekah, Sara, Evan, Eden, Lorraine; especially Cici, thank you for the sun! and the other yoga teachers at the studio, who helped me find some space. Urs and Anahi, my French-speaking refuge in the US. Some (but clearly, not all) of my roommates: Hao, for pre-preparing coffee every single morning for three years, and cutting the pineapple in such a perfect way. Joey, for your help with the bike, and for your eye-saver-screen. Nika, even though you had to leave earlier than expected, for the hikes, for the bread, and because you were there during

the year that's probably been the most difficult. Jacques, Evan. And Otis, of course, the best of us.

To Kerpagnonnes, Juliette, fidèle cothésarde surlepont, Clémentine, maître yogi, Louise, bucheronne, and Laura, for being one of the only brave enough to come visit me in the US (sorry I was working all the time). For still being there after 6, 8, 12 years, and hopefully, still at least as much. And to other friends in France, Jad, Vincent, Iryna, Caroline.

Last, my family. Maman, merci de toujours avoir été là, et encore plus ces derniers mois, d'avoir accompagné chaque pas ; merci pour tous les -s rajoutés à la troisième personne, et puis, merci de croire en nous, tellement. Papa, merci de m'avoir appris bien avant mon studio de yoga que le plus important dans la vie, c'est pas le boulot, et qu'on arrivera tous ensemble à Noël. Daphné, merci pour le soutien, le calme, et d'être la meilleure des grandes sœurs. Aymeric, merci d'être mon frérot. Merci pour toutes les mains dans le dos pour aider les montées, merci d'être super comme tu l'es, d'être un exemple sur pleins d'aspects (hé, hé). Rémi, merci pour toutes les fois où tu m'as débloquée avec R, IBEX, python, excel, html, css, sans jamais laisser l'once d'une moindre chance d'espoir à ma panique. Merci pour tes jeux de mots d'eau, et de connaître ce que ça peut être la vie aux US, et de savoir me dire simplement que si, en fait, je compte. Jordane, et Nicolas, merci aussi, même si c'était avec plus de distance, vous étiez là. Et puis enfin, Floriane, merci, vieux, gros, large merci, large, gros et vieux ; pour toutes les fois où t'as été là, pour le soutien, la patience, pour toutes les crises d'angoisse pas forcément jugulées mais accompagnées, tous les matins où tu t'es levée pour relire mes handouts de sémantiques, toutes les présentations répétées à 6h de décalage horaire, tous les tirets entre adult et like qui se sentiraient roulés sans toi, pour always pointing out les trucs pas pertinents, même quand c'est relou. Pour être brillante comme tu l'es, pour être ma petite sœur, et la meilleure des dessinatrices.

Si j'ai appris quelque chose en 5 ans, c'est que quand vous êtes pas là, c'est (trop) vide.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Imagine you are on your way to visit an old friend, who lives in a nearby city. You arrive at a crossing, with two roads: one goes through a forest, the other through the mountains. But you don't know whether both would allow you to get to your friend's place. It's your lucky day, another person joins you, so you ask her for directions. She tells you:

(1) You **have to** take the road through the forest.

Hearing (1), you will understand that the road through the forest is the only way for you to arrive: the road through the mountains is not a possible option. However, if instead, she says (2):

(2) You **can** take the road through the forest.

You will not conclude that the other road, through the mountains, is not a possible option. Maybe the speaker doesn't know the status of that road, or maybe she does, and means to suggest that it is also possible, but the forest road is preferable.

For adult speakers familiar with English, the difference in meaning between *must* and *can* may seem obvious. These words differ in 'force', i.e., in their logical strength: **possibility** modals such as *can* or *might* have a logically weaker meaning; **necessity** modals such as *must* or *have to* have a logically stronger meaning. *Must(p)* entails *can(p)*: If you *have to* go through the forest, you also *can*; conversely, if you

can't, you also *don't have to*. But from the perspective of the child learning language, figuring out this difference might not be so easy.

To see this, imagine now that you are travelling to a faraway island where the language is similar to English, except for a few words, "modals": words people use to talk about possibilities and necessities. You are again on your way to visit an old friend, who lives on the other side of the island; you arrive at a crossing, and there are two roads to follow, but you don't know whether both would allow you to get to your friend's place. Another lucky day, another person joins you, so you ask her for directions. She answers:

(3) You **sig** take the road through the forest.

What does *sig* mean? Do you think it means that it is necessary for you to go through the forest, i.e., that it is the only option for you to reach your friend's place? Or does it mean that it is just a possibility, and that there might be other ways to go? How would you be able to tell?

This is the main question this dissertation investigates. How and when do children figure out the force of modals? How do they learn the mapping between *must/should/have to* and necessity meanings, and *can/could/might* and possibility meanings? What kind of evidence can they rely on?

Figuring out modals' meaning presents particular challenges. As with any other words, children have to learn meanings just based on the way they are used by speakers around them. But modals are words used to talk about what is possible or necessary, which goes beyond what is actual. What is possible need not be actual. And what is

necessary is, so to speak, more than actual: it is something that could not be otherwise. Talk of possibility or necessity therefore involves reference to a range of alternatives to the actual. And for this reason, these properties have few physical correlates, that can be seen or sensed in any way: when a speaker describes a situation as possible or necessary, there is no signpost indicating the range of options they are considering. Cues from the physical context of speech as to meanings of modal words are thus bound to be limited. This makes them what Gleitman et al.'s (2005) call "hard words".

It has been argued that to learn such words that express abstract meanings, cues from the syntax could play a crucial role. This is Landau & Gleitman's (1985) *syntactic bootstrapping hypothesis*: the hypothesis according to which learners can home in on a word's meaning, by exploiting principled links between this meaning, which is closed to observation, and the word's syntactic distribution, which is easier to observe. Indeed, the modal's syntactic position, before a verbal complement, might help narrow candidate meanings as expressing some kind of modal meaning—something that expresses a relation to (or property of) a situation or proposition. But the question this dissertation focuses on is not how children figure out that *must* and *can* have a modal meaning in general. It is how they figure out their difference in *force*, possibility and necessity. Here, syntax cannot help, since possibility and necessity modals belong to the same syntactic categories, and can appear in all the same syntactic environments.

A main problem with figuring out modal force is that necessity entails possibility, provided we are considering the same range of options. What prevents children from hypothesizing a weaker possibility meaning for necessity modals? This kind of problem is called a "Subset Problem". If you *must* go through the forest,

considering the range of passable routes, then you *can* go through the forest, given that same set of options. Likewise, if you *must* eat with your right hand, given the rules of etiquette, then those same rules imply you *can* eat with your right hand. So, if you think that *sig* means 'possible' but in fact it means 'necessary', it is unclear how you can discover that in fact, *sig* has a stronger meaning: in situations where a necessity modal is used, a possibility statement is also systematically true. But then what prevents learners from postulating possibility meanings for necessity modals like *must* or *have to*? How do they know that they express necessity and not possibility? Are there situational cues that give away the contrast between possibility and necessity?

To answer this question, I will explore in depth the speech young English children are exposed to, on the basis of a detailed corpus study of their linguistic input between 2 and 3-year-old, complemented by experiments based on the corpus. This study will allow us to identify what kinds of cues are in principle available in the input as to modal force. Identifying the kind of information that is available—and the kind of information that is not—will put us in a better position to assess the kinds of capacities and biases children need to make use of the available information successfully.

The second question I address in this dissertation is the question of *when* children figure out force. How early do they use possibility and necessity modals? And do they use them appropriately? To answer it, I will explore 2- to 3-year-old English children's own spontaneous productions of modals, again, by bringing together corpus and experimental methods.

In the remainder of this chapter, I break down the learning problem for force in more detail, to show what elements make it challenging. In **section 1.1**, I focus on the logical Subset Problem, explain its logic and situate it with respect to previous discussions of similar Subset Problems in word learning. In **section 1.2**, I focus on the semantics and pragmatics of modals, to get a better sense of the target grammar children need to acquire, and circumscribe the range of possible modal systems that they could in principle acquire, given what we currently know about the cross-linguistic diversity of modal systems. In **section 1.3**, I come back to the Subset Problem, and consider three possible solutions to it based on previous discussions of other Subset Problems. The goal of the input study (presented in Chapter 2) will be to evaluate the viability of each of these three solutions. Finally, in **section 1.4**, I describe the methods used in the dissertation, and present its outline.

1.1 The Subset Problem[2]

Figuring out modal force raises a logical Subset Problem for the learner. The problem comes from the fact that necessity entails possibility—if it is necessary to go through

[2] Originally, the *Subset Problem* (also known as the "entailment problem") was discussed for the acquisition of syntax (see Chomsky & Lasnik, 1977; Dell, 1981, Berwick, 1985, Wexler & Manzini, 1987, a.o.). In the case of syntax, the problem unfolds in a different way: The question is how learners choose between two grammars G1 and G2, when both G1 and G2 can generate their input, in the absence of negative evidence directly informing them that G2 overgenerates. A classic example is the so-called *that trace* effect (see Chomsky & Lasnik, 1977). For syntax, the set/subset are the sets of sentences generated by the grammar. The child prefers the grammar G1 that does not generate some ungrammatical sentences, over the more permissive grammar G2 that does generate it. The logic of the Subset Problem has then been applied to semantic issues, both at the level of word learning (e.g. *fruit/apple*), and at the sentence level (e.g., the acquisition of scopally ambiguous sentences) (e.g., Wexler & Manzini, 1987; Crain & Thornton, 1998; Musolino, 2006, a.o.). Applied to semantic issues, the set/subset correspond to meanings relation, meaning being defined as extensions (sets of entities a word can refer to). In this

the forest, then it is also possible to go through it, in relation to the same set of options. This means that if learners think that necessity modals like *must* or *have to* mean 'possible'—or if on the island, *sig* actually means 'necessary' but you think it means 'possible'—they have no direct evidence that it is wrong, since in a necessity situation, both *can* and *must* are *logically* true. So what keeps them from thinking that necessity modals mean 'possible'? How can they determine that necessity modals have stronger meanings?

This kind of logical Subset Problem has previously been discussed for other cases of word learning, in particular for content words, like *apple/fruit* (see Quine's 1960 *gavagai* problem; Xu & Tenenbaum, 2007),[3] and function words: quantifiers like *some*/*every*, and numerals (see Piantadosi, 2011; Piantadosi et al., 2013; Rasin & Aravind, 2021). I will take first content words and quantifiers *some/every* as an example, to see the logic of the problem; then, we will see how modals raise additional challenges.

Imagine you are still on the island, but now you are unsure about what *mala*, a noun, means—you just know people use it to refer to objects. You hear it once, used to refer to what looks like an apple. If you think *mala* means 'apple' but it actually means 'fruit', there is no problem: you might be able to revise your incorrect hypothesis, by

---

section, I leave historical considerations aside, and focus on explaining the logic of the problem, in the case of word learning. I come back to these debates in **section 1.3**, when describing the proposed solutions to such Subset Problems.

[3] Quine's *gavagai* problem includes our Subset Problem, but is more general. He proposes the following thought experiment: Imagine hearing *gavagai* used when a rabbit crossed a road. How can you tell whether it means "rabbit", "white", or "animal" (or something else), since your observation is in principle compatible with an infinite number of hypotheses? (see Quine, 1960, Ch. 2, pp. 26).

hearing it in some other situations, used to refer to objects that are not apples, but still fruits. But on the other hand, if you think *mala* can refer to any type of fruit but it actually means 'apple', a problem arises: you might have no reason to change your hypothesis, as you will only see it used to refer to objects that are fruits. The problem arises because all apples are fruits: the set of apples is a subset of the set of fruits.[4]

The same goes for quantifiers like *some* and *all*. *All* asymmetrically entails *some*: if the sentence "**all** roads are closed" is true, then it is also true that *some* of them are. Imagine you are unsure about what *gleeb*, a determiner, means, and you hear the sentence "**gleeb** roads are closed." If you think that *gleeb* means 'all' (i.e., that it has the logically stronger meaning) but it actually means 'some', you might be able to revise your hypothesis, by hearing it used when *only some* roads are closed. But on the other hand, if you think that *gleeb* means 'some' (the logically weaker meaning) but it actually means 'all', you might have no reason to change your hypothesis, since you will only hear it used in situations *logically* compatible with 'some'.

The same logic applies to modals like *can* and *must*. Necessity entails possibility: if you are unsure about the meaning of a novel modal, *sig*, and if you think it means necessity (the logically stronger meaning) but it actually means possibility, you might be able to revise your hypothesis, by hearing it used in situations that express possibility but not necessity: for instance, if you can see that it is possible to go both ways. But on the other hand, if you think *sig* means 'possible' but it actually means

[4] Word (and sentence) meanings are here defined as their extension. For words like *apple*, this is the set of things to which the word applies truly, namely the apples. For words like *must*, this might be, on one common understanding, the set of 'situations' or 'worlds' to which it applies truly, namely those that are necessary.

'necessary', you might have no reason to change your mind, since you will only hear it used in situations logically compatible with a possibility interpretation.

Such Subset Problems thus arise whenever the meaning of one word entails the meaning of another. Various solutions have been proposed in the literature for how children solve such problems, which I present in **section 1.3**. Before describing them, and seeing how they apply to the case of modals, I provide some background about modals' semantics and pragmatics. This will allow us to see what elements make it challenging to solve the Subset Problem in the case of modals. More generally, it will allow us to get a better sense of the target grammar that children need to acquire, as the question I address in not only how children solve the Subset Problem: it is how children figure out force in general.

1.2 Modal force in English and beyond

1.2.1 Standard analysis of modal force and flavor

We typically distinguish two main forces for English modals: possibility and necessity. Standardly, this is captured by treating modals as either existential or universal quantifiers over possible worlds, following the modal logic tradition (Carnap, 1947; von Wright, 1951; Prior, 1957, Kripke, 1963). Possibility modals are treated as existential quantifiers over possible worlds: *can(p)* means that in **some** worlds accessible from the actual world, the modal's prejacent *p* is true. Necessity modals are treated as universal quantifiers over possible worlds: *must(p)* means that in **all**

accessible worlds, *p* is true. Possible worlds correspond to possible 'ways things could have been' (Lewis, 1973).

This quantificational analysis captures the logical entailment relationships we described between possibility and necessity modals: they show the same patterns of entailments and logical equivalences as the quantifiers *some* and *every* in the nominal domain. This is illustrated in (4) and (4) (from von Fintel & Heim, 2011; see also Hacquard, 2020). Just like (4)a) entails (4)b), (5)a) entails (5)b); assuming that the domain is not empty, (4)a/c) and (4)b/d) are logically equivalent, as well as (5)a/c) and (5)b/d).

(4) a. **Everyone** left. c. It's **not** the case that **someone** stayed.

b. **Someone** left. d. It's **not** the case that **everyone** stayed.

(5) a. You **must** leave. c. It's **not** the case that you **may** stay.

b. You **may** leave. d. It's **not** the case that you **must** stay.

Further force distinctions can be made for modals. In particular, necessity modals are often split into strong (e.g. *must*) vs. weak (e.g. *should*) necessity modals, a difference illustrated in (6) (see von Fintel & Iatridou, 2008). While functional modals seem restricted to possibility and (weak and strong) necessity, other grammatical categories such as nouns (e.g., *possibility*) and adjectives (e.g., *likely*) can encode even finer-grained strength distinctions.

(6) Employees **must** wash their hands. Everyone else **should**.

In this dissertation, I leave aside the learnability issues raised by these finer-grained distinctions, to focus on the main contrast between possibility and necessity modals.

Up to this point, I described modals' meaning focusing on the contrast in force, between possibility and necessity (i.e., in logical strength). But modals vary along a second dimension, 'flavor' (for a cross-linguistic perspective, see van der Auwera et al., 2005; for overviews, see Portner, 2009; Hacquard, 2011). Functional modals can express various types of possibilities and necessities: For instance, *might* and *must* in (7)a/b) express possibilities and necessities given some evidence (called epistemic modality), and *can* and *must* in (8)a/b) express possibilities and necessities given some rules (called deontic modality). In English, a given modal always expresses the same force (possibility or necessity), but it can be used for different flavors. Depending on the context, *must* in (9) can be used to mean that it is likely given what is known (or what can be inferred from the situation) that Anne sleeps a lot (epistemic modality, (9)a)), or that given the rules, it is necessary for her to sleep (deontic modality, (9)b)), or that to reach her goals, it is necessary for her to sleep (teleological modality, (9)c)). It is common in the semantic literature to distinguish between epistemic uses and non-epistemic ones, the latter covering many different types: rule-based, goal-based, or based on physical capacities. Following Hoffmann (1966), I will use the term 'root' modality to subsume all non-epistemic flavors.

(7) Given what we know…,

a. The dog **might** be outside.	possibility (◊)
b. The dog **must** be outside.	necessity (□)

(8) Given the rules…,

a. The dog **can** go outside. possibility (◊)
b. The dog **must** go outside. necessity (□)

(9) Anne **must** sleep a lot.

a. Epistemic: according to what is known, …
b. Deontic (root): according to the rules, …
c. Teleological (root): according to the goals, …
d. Ability (root): according to physical capacities, …
e. Etc.

In the now standard Kratzerian framework (Kratzer, 1981, 1991), these two dimensions of modals' meaning are captured by having modals be lexically specified for force, but not for flavor. Flavor gets determined contextually by what Kratzer calls conversational backgrounds. These specify the set of worlds that the modal quantifies over, as the lexical entries in (10), slightly modified from Kratzer (1991), illustrate. The domain of quantification of the modal (i.e., the set of worlds it quantifies over) is determined by context.

(10) For any world w, conversational background f:[5]

a. $[[\text{can}]]^{w,f} = \lambda q_{<st>} . \exists w' \in \cap f(w): q(w') = 1$
b. $[[\text{must}]]^{w,f} = \lambda q_{<st>} . \forall w' \in \cap f(w): q(w') = 1$

This flavor variability has important consequences when we focus on the question of how children acquire a modal's force. A first important clarification point is that the Subset Problem described in **section 1.1** holds within a given flavor. Logical entailment relations hold within flavor: for instance, being *able* to does not mean being

[5] I ignore the ordering source here, which can derive further gradability and flavor differences amongst root modals.

allowed, and conversely, being *allowed* does not mean you are *able* to; epistemic necessity (e.g. 'given what we know, he **must** be outside.') does not entail deontic possibility (e.g. 'given the rules, he **can** be outside.'), and conversely.

From the learner's perspective, the fact that modals can express different flavors could make it easier to figure out force. If children expect that a modal like *must* always expresses the same force, having figured out that *must* expresses *deontic* necessity might allow them to conclude, by extension, that it also expresses necessity when used for another flavor. This might be especially helpful if the contrast in force is easier to grasp for some flavors than for others. But this flavor variability could also make the task harder, since the flavor intended by the adult can't always be interpreted straightforwardly. For instance, imagine a child who knows only the deontic use of *must* and not its epistemic use. She hears an epistemic use of *must*: "Al **must** be enjoying himself." By assumption she will understand this, wrongly, as deontic. But what the speaker says is necessary epistemically (Al enjoying herself) need not be necessary deontically: Al is not required to enjoy herself. And in that case the child might be misled and conclude that *must* expresses possibility. Therefore, due to the logical independence of epistemic and deontic modality, the fact that a single word can be used for both, might sometimes cause trouble for the learner.

1.2.2 Modals and negation

Possibility and necessity modals are not uniform in their interaction with negation, neither force-wise nor flavor-wise. Cross-linguistically, epistemic possibility modals tend to be interpreted above negation, and roots below it (Coates, 1983; Cinque, 1999;

Drubig, 2001; Hacquard, 2011; for a typological overview, see de Haan, 1997; van der Auwera, 2001). This is illustrated for English in (11)a), (11)b) and (11)c): (root) *can* is interpreted below negation, (epistemic) *might* is interpreted above negation; *may* is interpreted under negation with a root interpretation, and over negation with an epistemic interpretation.

(11)	a. Jo **can't$_{root}$** sleep.	not > possible; *possible > not
	b. Jo **might$_{epistemic}$ not** sleep.	*not > possible; possible > not
	c. Jo **may$_{root/epistemic}$ not** sleep.	
	root: 'it is not possible that Jo does'	not > possible; *possible > not
	epi: 'it is possible that Jo does not'	*not > possible; possible > not

Necessity modals, on the other hand, seem to always keep the same scopal behavior with respect to negation, regardless of flavor: they either systematically scope over negation, like *must*/*should* in (12)a) (Dutch *moeten*, German *müssen*) (a behavior Iatridou & Zeijlstra (2013) attributes to their being Positive Polarity Items), or systematically scope under negation, like *need* in (12)b) and *have to* in (12)c). English *need*, as well as Dutch *hoeven* and German *brauchen*, are commonly analyzed as Negative Polarity Items (NPIs).[6] I come back to the consequences of these scopal interactions for the learner in **section 1.3.2**.

(12)	a. Jo **must not/should not** sleep.	necessary > not; *not > necessary
	epistemic/root: 'it is necessary that Jo does not sleep'	

[6] NPIs are words that typically occur in the scope of negation, and more generally, are licensed by DE environments (they are ungrammatical in non-DE environments) (see Fauconnier, 1975; Ladusaw, 1979; Van der Wouden, 1994; a.o.). Other examples of NPIs are English *any* or *ever* (e.g. 'Jo didn't eat anything', vs. * 'Jo ate anything'). PPIs are in a sense the mirror image of NPIs: they tend to escape the scope of negation, and seem 'banned' from negative contexts (Szabolcsi, 2004). Other examples of PPIs are *someone* or *something*: in "Jo didn't eat something", *some* can only take scope over negation ('There is something that Ann didn't eat' vs. *'There is nothing Ann ate').

b. Jo **needn't** sleep. *necessary > not; not > necessary
epistemic/root: 'it is not necessary that Jo sleeps.'

c. Jo **doesn't have to** sleep. *necessary > not; not > necessary
epistemic/root: 'it is not necessary that Jo sleeps.'

1.2.3 Modals and scalar implicatures

Possibility modal sentences are logically true in necessity situations: a speaker would not be considered a liar uttering (2), "You **can** go through the forest", even if she actually thinks that going through the forest is the only option for you. However, uses of this sentence would generally be interpreted as conveying a stronger meaning: that the speaker does not believe that you *have to* take this road (i.e., she believes that there might be other options for you, or she does not know about them). This is an example of a *scalar implicature*, and common example to illustrate how sentence and speaker's meaning differ (for various theoretical perspectives on scalar implicatures, see Horn, 1972; Grice, 1975; Gazdar, 1979; Hirschberg, 1985; Sperber & Wilson, 1986; Spector, 2006; Chierchia, Fox and Spector, 2012; Levinson et al., 2000; Sauerland, 2012; see Chemla & Singh, 2014, Noveck, 2018, for recent overviews of experimental approaches to the phenomenon). As we will see in Chapter 3, the study of scalar implicatures has generated many acquisition experiments focusing on children's understanding of modal force (e.g. Noveck, 2001): modals were used as a case study to see whether children were able to understand scalar implicatures. I will now briefly describe the phenomena and explain how standard accounts capture these inferences. Readers familiar with this topic can skip to **section 1.2.4**.

A possibility modal statement such as (2), "You **can** go through the forest" can convey a stronger meaning: that you *don't have to go* through the forest ('it is not the

case that you have to'). The same kind of inferences arises with quantifiers like *some*: use of a sentence such as "**Some** of the roads are blocked" is generally interpreted as conveying that it is not the case that *all* of them are. But this is not part of these sentences' literal meaning: these inferences can be cancelled (e.g., one can say "**some** of the roads are blocked, in fact, all of them are"; "you **can** go this way; in fact, you have to') or reinforced ("**some** of the roads are blocked, but not all of them"; "You **can** go this way, but you don't have to'). The same applies with epistemic modals: *might* in (7)a) ("the dog **might** be outside") can implicate that the dog *doesn't have to* be outside (i.e. that the speaker is not certain that he is).

According to the standard Gricean approach to scalar implicatures (Grice, 1975), these inferences arise because the sentence containing the logically weaker term, (*can*/*might*/*some*) competes with another more informative sentence the speaker could have chosen instead, with the logically stronger term (*have to/must/all*). According to Horn (1972), this is because these terms form scales (e.g., <$can_{deontic}$, have $to_{deontic}$>, <$might_{epi}$, $must_{epi}$>, <some, all>, etc.),[7] which he defines as conventionalized associations of lexical items that can be ordered in term of informational (logical)

[7] Modal scales are defined within a given flavor. Note that ability modals (e.g. "Anne **can** speak Dutch") raise an interesting puzzle here. As discussed by Horn (1972, Chapter 2), there is no clear scale for ability modals (see also Hackl (1998), who claims that "ability modals have no dual"). As noted by Hackl, necessity modals seem never used for ability ascriptions (Ability modal generalization (Hackl, 1998): "If a language uses a modal auxiliary to express ability, then it is always an existential modal and never a universal modal". This gap might not be real: it has also been argued that as other modal flavors, ability modals have duals, compulsion modals (for instance, "I **have to** sneeze right now; see Mandelkern et al., 2015, 2017). That duals of ability modals are hard to identify would then come from the fact that people don't often talk about compulsions. One consequence is that scalar implicatures almost never occur with ability modals. Note that, partially because of their lack of a clear necessity counterpart, the force of ability modals has been a matter of debate in the semantic literature: It is even sometimes argued that they have universal force (e.g. Kenny, 1976; Brown, 1988; Giannakidou, 2001), which is doubtful given their behavior when expressing different flavor (see Hacquard, 2010).

strength. Other standard examples involve logical connectives (<or, and>), or numerals (<1,2,3, …>). The implicature arises from the assumption that participants in a conversational exchange are cooperative agents who try to make their contribution as informative as required for the current purposes of the exchange, and do not say things for which they lack evidence (following Grice's *maxims of quantity* and *quality*). Cooperative speakers should always prefer to use a logically stronger sentence (i.e., maximally informative) when it is relevant, if they believe it to be true. Listeners can thus infer, from the fact that the speaker did not choose the stronger (more informative) sentence, that it is not the case that the speaker believes it: otherwise, why not use it?

Some scalar implicatures are triggered routinely, perhaps even conventionally. But even these are not utterly automatic or necessary. Their likelihood is still contingent on how relevant the stronger alternative is in the conversation. For instance, hearing (2), "You **can** take the road through the forest", you might, but you won't necessarily infer that you *don't have to* go the other way (i.e., that it is also an option for you to go through the mountains), as what is crucial in this situation is for you to reach your goal, rather than how many different options would allow you to reach it.

There are debates about the nature of scalar implicatures, which I won't review here (see Sauerland, 2012, for a summary). Note that all approaches rely on the notion of scales, and take them for granted: The derivation of scalar implicatures depends on speaker and addressee shared knowledge of the scales.

How do children learn those scales, given that these terms systematically lead to the kind of Subset Problems described in **section 1.1**?

1.2.4 Modal scales and crosslinguistic variation in the expression of force

In Indo-European languages like English, possibility and necessity 'duals' are common. However, outside Indo-European languages, there are languages where we find no such scales. Instead, the same modals can be used both in situations where English speakers would use a possibility modal, and in situations where they would use a necessity modal.

Such "variable force" modals have been described in a number of languages, and do not behave uniformly across different languages (see Yanovich, 2016, for a summary). In Nez Perce (Niimiipuutimt), the modal *o'qa* has been analyzed as a possibility (existential) modal, whose apparent variable force is due to the lack of a lexicalized stronger necessity dual in the language: *o'qa* does not belong to a Horn-scale, therefore its use is never associated with a scalar implicature (Deal, 2011), and is thus deemed appropriate in situations of necessity. Gitksan (Tsimshianic) *=ima* is similarly analyzed as a possibility modal (Matthewson, 2013; Peterson, 2010). On the other hand, in St'˝at'imcets (Lillooet Salish) and Washo (Hokan/isolate), "variable force" modals have been analyzed as being underlyingly necessity (universal) modals, which can be weakened by contextually restricting their domain of quantification to derive possibility readings (Rullmann et al., 2008; Bochnak, 2015). Other analyses take these modals to neither be underlying possibility, nor underlying necessity. In particular, Kratzer (2012) proposes that they can be analyzed as upper-end degree modals, roughly equivalent in meaning to 'it is somewhat probable (/desirable) that p' (Kratzer, 2012); see also Stalnaker's (1991) proposal for *would*).

The range of cross-linguistic variation we find suggests that there are few constraints on the space of hypotheses learners entertain for modals. They can neither expect that their language must have a possibility modal, nor a necessity modal. Moreover, they can't expect modals to come in duals. And even in a language with duals like English, knowing the force of *one* modal doesn't guarantee that the next modal will express a different force, given that several lexemes can express the same force (e.g., *can, might* and *may*): children will thus have to figure out force for each modal anew.

1.2.5 'Polite' uses of modals

As we saw in English, a language with scales, the use of a possibility modal can convey a scalar implicature. For example, "you can go" can be used to convey that 'you don't have to'. While logically true, possibility modal sentences are inappropriate in necessity situations, because under informative. This aspect of the pragmatics of modals could in principle help (English) learners distinguish possibility from necessity modals. If speakers systematically refrain from using possibility modals in necessity situations (because necessity modals would be more informative), situations in which possibility modals are used might never overlap with those in which necessity modals are used.

However, speakers do not always aim for maximal informativity: other conversational principles can interfere. And in particular, issues for learners might arise from 'polite' uses of modals: Possibility modal statements can also be used to soften statements in a polite way, or to perform speech acts whose force is stronger. For

instance, *could* in (13) can be used to perform an order to be quiet, and *might* in (14) can be used to convey that it *is* too late (see Searle, 1975; Grice, 1975; Austin, 1975, Brown & Levinson, 1987, a.o.).[8]

(13) "You **could** be a little more quiet" → 'Be quiet!' (order)

(14) "It **might** be too late" → 'It is too late.' (assertion of certainty)

Here, there is a mismatch between the force of the speech act performed (for instance in (13) an order, closer to necessity) and the actual force of the modal (its literal meaning: possibility). On the island, the speaker might just be polite, and perfectly well know that the forest is the only way to go. This aspect could, in principle, raise a problem in acquiring force. If children are often exposed to sentences like "the toy might be in the box" when parents clearly know it is in the box, what will they infer about the meaning of *might*? If these uses were systematic, children could be led to lexicalize necessity meanings for possibility modals.

How do children interpret these uses? Are they able to 'objectively' evaluate a given situation (e.g., by discovering afterwards that the other option was not available), or are they more sensitive to the speech act performed? Do they have expectation for adults to be informative, or polite, or is this something they learn?[9] There is no definite

[8] Note that these examples don't necessarily involve politeness. In other contexts, (14) could be used ironically, or as a guessing game, etc. The general point is that the 'force' of the speech act and the force of the modal don't align.

[9] According to Searle (1975), the understanding of a request e.g., in "*Can* you pass the salt?" arises by reasoning that the ability in question is so trivial that it cannot be intended as a real question about whether the addressee has this ability. But here, we put the problem backward and ask how learners can figure out the meaning of *can* based on these uses.

answer to those questions in the literature, and I will only start addressing them in the dissertation, but this is an aspect of the problem we cannot ignore. And these challenges, coming from the distinction between sentence and speaker meanings, are particularly acute in the case of modals.[10]

The existing acquisition literature suggests that children are sensitive to the kind of speech acts speaker performs early on, and are able to track the goals and intentions of their interlocutors (see Baldwin, 1991; Bloom, 2002; Clark & Amaral, 2010, a.o.). But while the literature shows that children are sensitive to the difference between main types of speech acts (assertions/questions/requests), it has not been shown that they can tell the difference in their 'force'—between, for instance, orders and permissions (for deontic modality), or between different levels of certainty and degree of commitment of the speaker (for epistemic modality).

1.3 Solving the Subset Problem with modals

How can children solve the Subset Problem with modals, given both what they mean and how they are used? That is, how do English learners figure out that necessity modals, like *have to* and *must*, have logically stronger meanings than their possibility counterparts, *can* and *might*?[11] In this section, I will introduce three possible

[10] For a discussion of the role of pragmatics in the acquisition of attitude verbs such as *think/want*, see Hacquard and Lidz, 2018.

[11] Recall that according to the logic of the Subset Problem, it is words that have stronger meanings that are potentially problematic (e.g.: *all* does not mean 'some', *apple* does not mean 'fruit', *must* does not mean 'possible'). In this section, as before, I focus the discussion on how children can solve the Subset Problem, i.e., how they learn necessity modals, to relate it better to the existing acquisition literature and

"solutions" to the Subset Problem, based on what has previously been proposed in the literature for other instances of Subset Problems, and discuss their applicability given what we have just seen about the semantics and pragmatics of modals. The first solution we will consider is that learners would have a bias towards strong (here, necessity) meanings, in the spirit of Berwick (1985) (see also Manzini & Wexler, 1987). The second one is that they would rely on downward-entailing environments, which reverse patterns of entailment (Gualmini & Schwarz, 2009). I then put forward a third solution, where learners use pragmatic cues from the conversational context in which modals are used to draw inferences about modals' force (Dieuleveut et al., 2019; see Rasin & Aravind, 2021, for a similar proposal for the acquisition of *every*).[12]

1.3.1 A bias for strong (necessity) meanings

A first way to solve Subset Problems is for learners to have an (innate) bias towards positing logically stronger meanings. This kind of idea, originally proposed for syntax Subset Problems, has many variants (see in particular Crain et al., 1994; Crain, 2012;

---

debates. But our learning problem could, and maybe should, be framed in a more general way: how children can figure out the force of the modals of their language.

[12] Here, I won't discuss two other possibilities found in the literature. First, that children would be directly corrected by adults when they "incorrectly" produce necessity modals. This type of solution has been criticized for syntactic Subset Problems (see Pinker, 1979), but could prove more relevant for semantic Subset Problems (Brown & Hanlon (1970) argue that while parents do not tend to correct their children for grammaticality, they do object to false statements made by children). In the case of modals, this kind of corrections would correspond to a child saying "It **must/has to** be in the box", and being corrected by the adult: "no, you should have said 'can', it is not *necessary* that it is: it can also be in the drawer." In their study on *every*, Rasin & Aravind (2021) report, out of 72 *every* uttered by children, 1 case of such explicit negative feedback. I leave open the question of whether children (systematically) encounter such correction with modals. Second, I will not fully address the question of potential effects of contrast, even though knowing a dual might affect learning (following Clark's *principle of contrast*, Clark & MacWhinney, 1987). Even thought, as mentioned, learners cannot expect two distinct lexemes to necessarily express different forces, since several lexemes can express the same force (e.g., *must/have to/got to/need to* all express necessity), such contrasts (seeing necessity modals explicitly contrasted with possibility modals, e.g., "you can, but you don't have to") could be helpful for learners.

Crain & Thornton, 1998 *Semantic Subset Principle*).[13] In the case of modals, this would mean that by default, children would assume necessity meanings for new modals—at the island crossing, you would by default assume that *sig* means that the speaker believes that it is the only option for you to reach your goal to go through the forest. Children would revise their hypothesis only for possibility modals, when hearing them used in situations of possibility but non-necessity—where a necessity modal cannot be used.

Take a child who would have to learn two modals, *sig* and *gorp*. *Sig* means 'possible', while *gorp* means 'necessary', but she does not know this. If she assumes by default that both express 'necessity', she never needs to change her hypothesis in the case of *gorp* (~ *must*), as it is the correct one. In the case of *sig (~ can)*, she might be able to revise her hypothesis, by hearing it used to describe a situation that is plainly not necessary: For instance, hearing the sentence "You *sig* go through the forest" in a situation where she can see that it is possible to use either road.

Do children need such a bias to solve the Subset Problem with modals? As we will see in the next two subsections, this type of proposal has been strongly criticized for other instances of the Subset Problem, both on conceptual (e.g., Gualmini & Schwarz, 2009) and empirical grounds (Xu & Tenenbaum, 2007; Musolino, 2006;

[13] The *Subset Principle* was originally proposed for the acquisition of syntactic phenomena (Baker, 1979; Pinker, 1979; Dell, 1981; Berwick, 1985; Manzini & Wexler, 1987, a.o.). Later on, the *Semantic Subset Principle* (SSP) was introduced by Crain and Thornton (1998) to account for semantic Subset Problems, at the sentential level (the acquisition of ambiguous sentences whose readings are related by entailment) (see also Crain et al. 1994; Crain 2012; Crain et al. 1994). Crain et al. (1994) discuss the case of the acquisition of *only*, and claim that the absence of truth-conditional evidence makes it impossible for a child to learn a strong reading after first learning a weak reading. They posit the *Semantic Subset Principle* as a constraint that preempts the Subset Problem.

Piantadosi, 2011, Piantadosi et al., 2013; Rasin & Aravind, 2021; for a summary, see Musolino et al., 2019). Such a bias has been shown to be unnecessary in those cases, but the question is still open in the case of modals, where the learning problem might be more acute.

1.3.2 Using Downward-Entailing environments

Gualmini and Schwarz (2009) propose a general solution to semantic Subset Problems, which only requires relying on truth-conditional evidence. Children just need to observe words that express strong meanings (for us, necessity modals) in downward-entailing environments (that is, environments that reverse patterns of entailment), for instance, under negation. They use this to argue that there is no Subset Problem, once we take downward-entailing (DE) environments into consideration.

Again, the logic might be easier to grasp using first the example of content words. Recall our thought experiment with nouns, where the problem arises if you are unsure about the meaning of *mala (~ apple)*, but think it means 'fruit'. In fact, there exists a logical way to solve your Subset Problem, even just using truth-conditional evidence: hearing the sentence "This is **not** a mala" used when pointing at a banana, you should be able to infer that *mala* cannot mean 'fruit'.

So again, take a child who would have to learn two new modals, *sig* and *gorp*; *sig* means 'possible', *gorp* means 'necessary', but she does not know this. As before, there is no problem for *sig* (~ *can*): if she wrongly assumes that it means 'necessary', she should be able to revise her hypothesis by hearing it used to describe a situation that is plainly not necessary. But if she thinks that *gorp (~ must)* means 'possible', she

in fact has a ‘logical’ way to solve the Subset Problem: by hearing 'you **don’t gorp to** (go down the forest)’, used to describe a situation where it plainly *is* possible to go down the forest. She should be able to infer that the meaning of *sig* cannot be ‘possible’: if it meant ‘possible’, then, under negation, it would mean ‘not possible’, and it could not be used in this situation. If we further assume that learners only consider a restricted space of hypotheses about possible meanings for modals ({possible, necessary}), they may further infer that, since *gorp* cannot mean ‘possible’, it has to mean ‘necessary’.

If children can observe necessity modals in such negative environments, they thus have a systematic way to solve Subset Problems. In this case, as argued by Gualmini and Schwarz (2009), they don’t need a bias towards strong meanings. This is an elegant solution, since it only requires from learners to rely on truth-conditional evidence, and can in principle apply to any instances of semantic Subset Problems.

However, problems arise from the irregularities of scope between negation and modals. As we saw in **section 1.2.3**, the scope interactions of modals and negation are not well-behaved. In particular, some necessity modals, for instance English *must* or *should*, systematically take scope over negation.

We can thus separate two types of cases, depending on the scope interpretation that necessity modals receive when they occur with negation. First, for necessity modals that outscope negation, like *must* or *should* (“she mustn’t go”: necessary > not), the reasoning proposed by Gualmini and Schwarz (2009) cannot be applied. Indeed, these necessity modals do not occur in negative environments: Learners will never hear “she mustn’t go” in possibility situations, since this is not a possible meaning in the adult grammar. If children already know that *must* expresses necessity, they might infer

its scope relation (and maybe, its polarity restrictions) from negative environments; but then, they need to have figured out force first. But from the learner's perspective, these cases could even be misleading for force: if children use negation to infer force and expect the same scope behavior for *must* and *can*, they should conclude that they express the same force, since from a pure truth conditional point of view, "you mustn't go" and "you can't go" are equivalent (given the logical equivalence between *necessary > not* and *not > possible)*. For these necessity modals, using a learning strategy based on negation might thus even add to the learning problem. And the same issue will arise for learners of other languages as well, since the tendency of necessity modals to outscope negation and yield 'strong' interpretations is found across several languages (e.g., Dutch *moeten*, German *müssen*).[14]

In the second type of case—namely, when necessity modals do scope below negation, such as *don't have to* and *don't need to/needn't*, which mean 'not necessary'—the rationale proposed by Gualmini and Schwarz at least can apply: negation could in principle be useful. The question is then how frequently these necessity modals occur in Downward-Entailing environments in the actual input, and whether those are clearly used to convey non-necessity rather than impossibility.

It seems that for negation to be helpful in figuring out force, learners would need to have already figured out how the modal scopes relative to negation, and expect negation to scope differently depending on force, flavor and modal idiosyncrasies. But

[14] This is also reported in languages where "variable force" modals are analyzed as underlying necessity modals, such as St'át'imcets and Washo (Rullmann et al., 2008; Bochnak, 2015): in those languages, the fact that modals cannot occur under negation is discussed as one of the things that makes the force diagnostic more challenging.

how would they figure out the right scope relations between modals and negation without knowing the force of the modals? Children might need to figure out force *first*—and then use it to learn scope and potential polarity restrictions.[15]

The consequence for us is that using Downward-Entailing environments cannot be the solution to the Subset Problem with modals: at least, not a general one.

1.3.3 Using cues from the conversational context

Given the irregularity of scope of necessity modals and negation, using Downward-Entailing environments cannot be a general solution to the Subset Problem. The rationale proposed by Gualmini & Schwarz can in principle apply for cases like *have to*, but cannot for cases like *must*. Then, how do children "solve" the Subset Problem?

Let's consider another kind of solution: Learners may be able to learn modal force using contextual information, rather than pure logic (Dieuleveut et al., 2019; for a similar proposal for the acquisition of *every*, see Rasin & Aravind, 2021). Children would use pragmatic cues, information from the situational context in which modals are used. This, in a way, dissolves the Subset Problem: if these kinds of cues are available, and if children are able to use them, they do not need to rely on negative environments, and they do not need a necessity bias.

---

[15] According to Iatridou and Zeijlstra (2013), only necessity modals, but not possibility modals, are subject to polarity restrictions. This type of polarity-type analysis has interesting consequences for the learner: it suggests that children may need figure out force before polarity restrictions. How are these two properties, force and polarity restrictions, learnt in tandem? This is an open question (for work and discussion of the role of the input on the acquisition of the NPI *any*, see Tieu, 2010, 2015; Tieu & Lidz, 2016; for work on the acquisition of the Dutch modal NPI *hoeven* 'need', see Lin, Weerman & Zeijlstra, 2018).

As we saw in **section 1.2.3**, we can expect some aspects of the pragmatics of modals to help English learners. If, following Gricean maxims, speakers systematically refrain from using possibility modals in necessity situations (because necessity modals would be more informative), situations in which they use possibility modals might never overlap with those in which they use necessity modals. This could help learners distinguish possibility from necessity modals. However, as we then saw in **section 1.2.5**, the extent to which adults always choose to use necessity modals over possibility modals in necessity situations is not entirely clear: the same sentence, "you can go", can depending on the context, be used (or interpreted) in opposite ways: either as 'you don't have to' (scalar implicature), or as 'you really should' ('polite' use). These politeness considerations (which are peculiar to modal scales, and do not arise, for instance, with quantifiers like *some/all*) could be misleading for learners: if frequent, they might blur the distinction between possibility and necessity modals.[16]

Do speakers use possibility and necessity modals in clearly distinct situations, and in ways that can be informative to the child? Conversational contexts include diverse components children might be sensitive to: the social status of participants (the

[16]Again, one crucial question is what aspects of a situation children are sensitive to. Xu & Tenenbaum (2007) have convincingly shown that a Bayesian learning model better captures the learning of content words (*Labrador/dog*), than traditional approaches to modeling word learning, based on deductive hypothesis elimination and associative learning. They show that learners can generalize from just a few positive examples of a novel word's referents, by making rational inductive inferences (see Tenenbaum's 1999 *size principle*; see also Piantadosi, 2011; Piantadosi et al., 2013 for a Bayesian model of the acquisition of quantifiers like *every*). However, one of the prerequisites for these Bayesian models to work is that speakers systematically use logically stronger sentences when they are true (and that learners expect them to). This assumption might not be problematic in the case of quantifiers, but with modals, the situation might be more complex, given potential 'polite' uses of possibility modals described in **section 1.2.5**. Learners might thus have to deal with a potentially more noisy input than when learning nouns or quantifiers.

child herself, her parents, maybe other people around them), the activity they are involved in (for instance, playing with toys), the goals of the speaker when she utters a specific sentence (the speech act performed: asking a question, giving a suggestion, giving an order or a prohibition), etc. But what aspects correlate with the distinction in force? Are children sensitive to them, and able to use them when learning meanings?

One example of such a cue, which I will explore in the dissertation, is the desirability of the possible events described. Let's picture our crossing differently: now the road through the mountains looks threatening, but the road through the forest is enticing. Compare (15)a) and (15)b). What would you infer about the meaning of *sig* and *gleeb*?

(15) a. You **sig** take this lovely road through the enchanted forest.

b. You **gleeb** take this treacherous road through the mountains.

If children expect permissions/possibilities to be more often associated to desirable predicates, and obligations/necessities to be more often associated to undesirable predicates, children could use cues based on the desirability of the event described: whether they are positive or negative.[17]

Take a child who would have to learn two new modals, *sig* and *gorp*; *sig* means 'possible', *gorp* means 'necessary', but she does not know this. If she hears *sig* often used with desirable prejacents (e.g., "You sig have a candy", "You sig play in the garden", "You sig not eat the cookie"), she might infer that it means 'possible', because

[17] This example applies to deontic modality. For other flavors, cues might be different.

permissions and possibilities are usually associated to desirable events. Conversely, if she hears *gorp* often used with undesirable prejacents (e.g., “You gorp brush your teeth”, "You gorp do your homework”, “You gorp not help with the dishes”), she might infer that it means ‘necessary’, because obligations and necessities are usually associated with undesirable events.[18]

Can children learn the contrast in force just on the basis of these kinds of conversational cues? Is the conversational context rich enough for learners to infer their force, without having to rely on either negation or a necessity bias? To answer these questions, and see what strategies are available for children to learn force, we need to look at their input.

1.3.4 Summary

We saw that what makes the mapping of modal form to force particularly challenging is that necessity entails possibility. This creates a Subset Problem for necessity modals, which have a logically stronger meaning than possibility modals. How can children solve this problem with modals? We consider three solutions. The first is for them to have a bias towards strong (necessity) meanings, in the spirit of Berwick (1985). The second is for them to rely on downward-entailing environments, which reverse patterns of entailment (Gualmini & Schwarz, 2009). The third is that they exploit cues from the

[18] The reason for this association, which seems intuitive for adults, deserves further exploration.

conversational context, if speakers use possibility and necessity modals in clearly distinct situations in ways that are informative as to their force for children.

This dissertation assesses each of these solutions, to better understand how and when children eventually figure out the force of modals. To do so, I conduct a series of corpus and experiments to examine how modals are used in speech to and by children. By looking at children's modal input, we can assess the viability of each of these solutions: Do modals regularly occur in DE-environments, like in the scope of negation? Are cues from the conversational context like desirability available and exploitable? If neither is available or reliable in the input, a necessity bias might be necessary for learners to solve the Subset Problem. By looking at children's early modals, we can assess when they appear to master the force of each, and whether we find evidence for a necessity bias in their modal uses.

In a nutshell, results from these studies will show that the conversational context in which modals are used is highly informative about both forces. This means that if children are sensitive to the same conversational cues as adults, they can in principle use them to figure out modal force, and therefore don't need to rely on a necessity bias or on downward-entailing contexts. The nature of these conversational cues could be quite diverse, and might vary with modal flavor. In particular, I show that children might be able to use desirability to figure out the force of deontic modals.

Looking at young children's own spontaneous productions of modals, we find an asymmetry between their early mastery of possibility and necessity modals, which I'll call a 'Necessity Gap'. Children seem to master possibility modals early: at age 2, they use them frequently and productively, and in an adult-like way. However, they

seem to struggle with necessity modals: they produce these much less frequently, hardly ever with negation, and often in a non adult-like way. These results cast a new light on prior results from comprehension experiments: as we will see, if this difficulty with necessity modals persists into the preschool years, it could explain children's tendency to both accept possibility modals in necessity contexts (if they lack a relevant stronger alternative), and necessity modals in possibility contexts (if are unsure that these modals express necessity). It remains unclear, however, whether children's difficulties with necessity modals stem from not knowing their underlying force, or whether children have successfully learned their force, but have either conceptual or pragmatic difficulty deploying them in ways that adults would.

1.4 Outline of the dissertation

To answer our question of how children figure out modal force, and to test the viability of each of these three solutions, we need to study children's input: how modals are used by their parents. How often do children hear possibility and necessity modals? Are they often used with negation? And how informative is the conversational context about force?

Chapter 2 presents a corpus study of the input to 2- to 3-year-old English children, based on the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney 2000). Its main goal is to evaluate the viability of each of the three solutions we have for the Subset Problem: to determine how, in principle, children can figure out modal force, by closely examining how modals are used in the speech of

their parents. It gives both quantitative data about parents' modal productions, and insights from three experiments that assess the general informativity of context about force, based on the *Human Simulation Paradigm* (Gillette et al. 1999). They show that the conversational context is highly informative, about both forces, which suggests that learners, at least in principle, needn't rely either on a necessity bias or on downward-entailing environments.

Chapter 3 focuses on the when question, and assesses children's own productions of modals, using the same corpus. It first reviews what we currently know about children's understanding of modal force, which is mostly based on comprehension experiments, and typically targets older children (from age 4). It then assesses more directly the question of when children figure out force, by providing a detailed corpus study of 2- to 3-year-old children's spontaneous productions of possibility and necessity modals, as compared to their parents'. The study is novel in three ways. First, it targets very young children, between 2 and 3: much younger than children tested in experiments. Moreover, it focuses on children's spontaneous productions: it is thus complementary to comprehension experiments. Last, besides quantitative measures about children's productions (i.e., how frequently they use possibility and necessity modals), it uses a novel method to test whether children use their modals in an adult-like way: the experiment is based on the HSP paradigm used for adults' productions, but evaluates children's productions. Results from the corpus study suggest that children master possibility modals early, but struggle with necessity modals: we find a 'Necessity Gap'.

Chapter 2 shows that conversational context informative, and thus that in principle, children do not need to rely on negation or a necessity bias. But do children actually make use of this information? To start addressing this question, and how children eventually master necessity modals, **Chapter 4** presents a study that relates the input study (Chapter 2) and the output study (Chapter 3). It aims at identifying what factors in the speech of parents influence children's mastery of force in practice, by focusing on variation between children in their mastery. Its results point out two factors that seem to play a crucial role in the learning process: first, children who hear *have to* with negation more frequently in their input seem to master it earlier. Second, while we don't find that the mere frequency of exposure to *have to* has an effect, we find that children more exposed to modal talk in general seem to master *have to* earlier.

Chapter 5 summarizes our main findings and discusses their implications for the acquisition of force.

Chapter 2. Children's input: How can children figure out force?

This chapter focuses on children's input. Its aim is to determine how, in principle, children can figure out modal force, by closely examining how modals are used in the speech of their parents. I do so via a corpus study, based on the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000), which provides a quantitative assessment of the modals children hear (which modals parents use, how frequently, and how frequently they use them with negation), and three experiments, that assess the informativity of natural conversational contexts about modal force, using a variant of the *Human Simulation Paradigm* (HSP) (Gillette et al., 1999).

Remember our thought experiment from Chapter 1: You arrive at a crossing, ask for directions, and your informant tells you: "You **sig** go through the forest." How do you determine whether this new modal, *sig*, expresses necessity (there is no other way you can go), or possibility (you can go this way, but there may also be other ways you can take)? Cues from the physical context are bound to be limited, since modals express non-actual states of affairs, with no physical correlates (Landau & Gleitman, 1985). Syntactic cues will be limited as well, since possibility and necessity modals belong to the same syntactic category: the syntactic position of *sig* might help you narrow candidate meanings as expressing some kind of modal meaning (in the spirit of Landau and Gleitman's 1985 syntactic bootstrapping hypothesis), but it does not help distinguish force.

The goal of this chapter is to clarify what kind of information children can, in principle, rely on to learn modal force, by looking at how modals are used in parents'

speech to children. If physical and syntactic cues are limited, what kind of cues can they use? What strategies are available to them, given their input? How can children map {*can/could/might*} to possibility meanings, and {*must/should/have to*} to necessity meanings?

As we saw in Chapter 1, what might make this learning problem particularly challenging for children is that necessity entails possibility, which creates a Subset Problem. What is necessary is also possible: this means that whenever adults use necessity modals in declarative affirmative sentences (e.g., "The dog **must** go outside"), a possibility modal statement (e.g., "The dog **can** go outside") is also logically true—even if potentially inappropriate. Then, what prevents children from thinking that necessity modals like *must* mean possible? There is no such issue for possibility modals, since children can in principle observe them used in situations of mere possibility, where a necessity modal is logically false, and therefore, cannot be used (e.g., "The dog **can/#must** go outside, but she is also allowed to stay inside").

How can children learn that necessity modals have a stronger meaning than possibility modals? In Chapter 1, we considered three possible "solutions" to this Subset Problem, which I now briefly reintroduce.

According to the first solution, children have a bias towards 'strong' (here, necessity) meanings (in the spirit of Berwick, 1985).[19] They would by default assume necessity meanings, and then revise their hypothesis only for possibility modals, when

[19] As discussed in Chapter 1, many variants of this idea, which was originally proposed for the acquisition of syntactic phenomena, can be found in the literature. See Musolino et al. (2019), for a summary.

hearing them used in situations of non-necessity (for instance above, by being told explicitly that the dog is also allowed to stay inside).

According to the second solution, such a bias is not necessary, because children have a logical way to solve the problem: they can figure out the force of necessity modals just by observing them in downward-entailing (DE) environments, for instance under negation (Gualmini & Schwarz, 2009).[20] Children should be able to infer that necessity modals cannot express possibility, by hearing them used with negation in situations of possibility: for instance, by hearing "The dog **doesn't have to** go outside" used in a situation where it is clear that the dog can either stay inside or go outside. If *have to* meant possible, its negation would mean 'not possible', and it could not be used in such a situation (it *is* possible for the dog to go outside). Therefore, *have to* has to mean necessary.

According to the third solution, children do not need to have a bias, nor to rely on negation: they can figure out force relying on pragmatic situational cues, from the context in which modals are used. Such information could involve the goals of the speaker in the context of her utterance (e.g., if they are performing orders, prohibitions, or granting permission), or properties associated with the types of events described in the prejacent (e.g., desirable events with permissions; undesirable ones with obligations).

[20] As made clear in Chapter 1, Gualmini and Schwarz do not propose this solution for modals specifically, but as a general solution for any Subset Problem. Their main goal is to show that from a logical stance, a strong bias is not necessary, since children can in principle use truth-conditional evidence from Downward-Entailing environments.

Do children need to have a bias towards necessity meanings? Can they rely on downward-entailing environments? Or is the conversational context informative enough about force for them to learn without such a bias or without relying on downward-entailing environments? To assess the viability of each of these three possible learning strategies, we need to look at children's input, to see if children are exposed to right kind of information (negation or pragmatic cues), or if a necessity bias is necessary. In this chapter, we will answer the following questions: How are possibility and necessity modals used by adults in conversations with children? How frequently are they used with negation? Then, how informative is the conversational context about force? Do adults use possibility and necessity modals in clearly distinct situations? And last, can we identify aspects of their conversations that are crucial to distinguish force?

We will start with a study of parent's productions, which is based on the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000). In **section 2.1**, I provide a descriptive quantitative assessment of the modals children hear: which modals parents use, how often, and in particular, how frequently they occur in downward-entailing environments. In **sections 2.2, 2.3** and **2.4**, I then present three experiments that allow us to assess the informativity of natural conversational contexts about modal force. The goal of the first two experiments (**section 2.2**: *Input Experiment 1* and **section 2.3**: *Input Experiment 2*) is to assess the general informativity of conversational contexts, by asking adult participants to guess a modal blanked out from an adult's sentence in dialogue extracted from the corpus, following the

Human Simulation Paradigm (HSP) (Gillette et al., 1999).[21] In the first experiment, the blanked modal statement is presented in context (i.e., with the preceding dialogue); in the second experiment, it is presented without context. Our results show that the conversational context in which modals are used is highly informative, about both forces. The last experiment (**section 2.4**: *Input Experiment 3*) explores further one specific situational cue for root modals, the desirability of the event described by the prejacent. The results show that this cue is available in the input: root necessity modals are more often used with undesirable prejacents than their possibility counterparts. Last, in **section 2.5**, I discuss implications of our findings for how children might acquire modal force and solve the Subset Problem.

To preview, I will show that:

(i) Our experimental results show that the conversational context is highly informative, about both forces. This means that learners don't need to rely on either a necessity bias or on negative environments to solve the Subset Problem: they can in principle use conversational cues to figure out force.

(ii) The nature of these conversational cues might vary with modal flavor. In particular, children might rely on cues from desirability to figure out the force of root deontic modals. We show that this cue is available in the input:

[21] In the original HSP paradigm (Gillette & al., 1999), adult participants were shown silenced videos of mother-child interactions, with a tone occurring when a particular word was used. They are asked to guess the word that was uttered. The paradigm has also been used to compare directly the informativity of different types of contexts, by manipulating the kind of information participants have access to (see White et al., 2017, for a summary)**.** The accuracy with which participants can recover the actual word given the context is taken as a general measure of the informativity of that context.

adults use (root) necessity modals more frequently with undesirable events, and (root) possibility modals more often with desirable events.

(iii) Using evidence from negative environments might not be sufficient to solve the Subset Problem for necessity modals. Depending on the modal, cases for necessity modals are either informative but extremely rare (e.g., *don't have to*), or potentially misleading because of the scope irregularities between necessity modals and negation (e.g. *mustn't*), as discussed in Chapter 1. However, using evidence from negative environments might be more helpful for children to figure out the force of possibility modals: our results show that negated possibility modals (e.g., *can't*) are frequent in the input, and that context is highly informative about their force.

2.1 Corpus study[22]

2.1.1 Methods

We used the Manchester Corpus (Theakston et al., 2001) of UK English (CHILDES database, MacWhinney, 2000), which consists of 12 child-mother pairs (6 females; age range: 1;09-3;00) recorded in unstructured play sessions. This corpus was chosen for its relative density and uniformity of sampling, as well as its early age range. We focused on the period between ages 2;00 and 3;00. All utterances containing modal

[22] This study is based on joint work with Annemarie van Dooren, Ailís Cournane and Valentine Hacquard.

auxiliaries and semi-auxiliaries (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) (16), presence of negation (17), flavor (epistemic vs. root) (18), subject (first/second/third person) (19), clause type (declarative/ interrogative/tag question) (20), and whether they occur in the antecedent of *if*-clauses (21). We did not include *will, would, shall* and *going to,* as they primarily express future, for which force is a matter of debate (Stalnaker, 1968; Cariani & Santorio, 2018, a.o.).

(16) Modal lemmas by force:

Possibility: *can, could, might, may; able to*
Necessity: *must, should, need; have to, got to, be supposed to, need to*

(17) Negation:

No negation: 'I **can** go to the pub now.'
Negation:
on main verb: 'I ca**n't** get it'/ 'I must **not** forget Whispy.'
on higher auxiliary: 'we **don't** have to play with your toys.'
on embedding verb: 'I **don't think** you have to look for it.'
other negative quantifier: '**nobody** can reach it.'

(18) Flavor:

Root:
MOTHER: we won't do that.
CHILD: I want her.
CHILD: I want her.
MOTHER: well you **must** treat her nicely then. (Aran, 2;07.14)

Epistemic:
MOTHER: oh.
MOTHER: somebody's done a neat pattern, haven't they?
MOTHER: goodness me.
MOTHER: that **must** have taken a long time. (Anne, 2;02.10)

(19) Subject:

1st person: 'I **can** see a bucket.'/ 'We **can** fit a cow through there.'
2nd person: 'You **can** do it.'
3rd person: 'He **can** go in the cart.' / 'The cat **can** go in the house.'

(20) Clause type:

Declarative clause: 'I **can** see a bucket.'
Interrogative clause: 'What **can** you see?'; '**Can** you see any chickens?'
Tag question (excluded): 'You can wash it later, **can't** you?'

(21) Conditionals:

No *if*-conditional: 'He **can** go in that one.'; 'see if you **can** balance it on your head.'
Modal in antecedent: 'they drink milk **if they can** get milk.'; '**if you can** open that you'll find a dog.'
Modal in consequent: 'you **can** make it big **if** you want to.'; '**if** I really want to get it I **can**.'

2.1.2 Results

Modal utterances (i.e., utterances containing a modal auxiliary or semi-auxiliary) represent 5.9% of all mothers' utterances (possibility modals: 4.2%, necessity modals: 1.7%). Overall, parents use possibility modals more frequently than necessity modals: possibility modals represent 72.5% of all their modal utterances (**Table 2.1**). *Can* is their most common modal (57.3% of all modal utterances), and *have to* their most frequent necessity modal (12.0%).

Downward-entailing environments. We find that negated possibility modals are quite frequent in adult speech, but negated necessity modals are rarer (possibility: 20.9% negated; necessity: 10.1% negated). Moreover, most of the cases where we find necessity modals with negation correspond to modals that outscope negation (*must, should, ought to*: 19.4% vs. *have to, got to, need to, supposed to*: 7.4%). Cases of

negated *have to/got to* (i.e., where using negation could in principle be useful to infer force) are thus rare (4.5% of adults' *have to* utterances are negated; *got to*: 1.1%). Modals occur rarely with other negative quantifiers (e.g., *nobody*/*nothing/never*), with no difference between possibility and necessity (necessity: 0.1% of adults' modal utterances; possibility: 0.2%), neither do they occur under negated embedding verbs (*e.g. don't think*), again with no difference between possibility and necessity (necessity: 2.1%; possibility: 1.5%).

Both possibility and necessity modals occur rarely in other downward-entailing environments such as the antecedents of *if*-conditionals (overall: 0.6%) (**Table 2.2**). Necessity modals hardly ever occur in such environments: we find only 15 occurrences in the whole corpus (with 7 of them corresponding to 'if you must') (vs. 106 possibility modals). (As a point of comparison, we find 135 necessity modals occurring in the *if*-conditionals consequent, vs. 432 possibility modals).

Table 2.1 Counts and percentages of modal uses by force for adults, ordered by lemma frequency, with and without negation (repetitions excluded: 3.7% of the data).[23] * indicates necessity modals that outscope negation.

	ADULT (n=19,986)		ADULT (n=18,853) [24]			
	all		no negation		negation	
POSSIBILITY	**14,491**	**72.5%**	**10,672**	**79.1%**	**2,828**	**20.9%**
can	11,472	57.4%	8,383	77.7%	2,396	22.2%
could	1,449	7.3%	1,116	96.6%	39	3.3%
might	1,216	6.1%	1,005	82.8%	208	17.1%
able	315	1.6%	134	42.5%	181	57.4%
may	39	0.2%	34	89.5%	4	10.5%
NECESSITY	**5,495**	**27.5%**	**4,814**	**89.9%**	**539**	**10.1%**
have to	2,398	12.0%	2,290	95.5%	108	4.5%
got to	940	4.7%	926	98.8%	11	1.1%
*should**	793	4.0%	537	77.1%	159	22.8%
need (to)[25]	493	2.5%	409	82.9%	84	17.0%
*must**	452	2.3%	346	84.1%	65	15.8%
supposed to	335	1.7%	230	68.6%	105	31.3%
*ought to**	84	0.4%	76	91.5%	7	8.4%

Table 2.2 Counts and percentages of modal uses by force in if-conditionals, for adults (excluding tags and repetitions).

	ADULT (n=18,853)					
	no if-clause		modal in antecedent		modal in consequent	
POSSIBILITY	12,962	96%	106	0.8%	432	3.2%
NECESSITY	5,201	97.2%	15	0.3%	135	2.5%

Flavor. We find that epistemic uses of modals are rare in parents' speech: they represent only 8.8% of all their modal utterances (**Table 2.3**) (a breakdown by modal is provided in Appendix A; see van Dooren et al., 2017, for details and discussion). Note that negation is considerably less frequent with epistemic than with root modals: epistemics are rarely negated (epistemic: 4.6% negated, vs. root: 19.1%).

[23] Are considered as repetitions cases where the speaker repeats a sentence uttered right before by herself or by another speaker with no significant change.

[24] Excluding tags and repetitions. Tag questions (e.g. "you can wash it later, *can't you*?") are very frequent in this corpus (4.7% of all modal utterances). We exclude modals in the tags, as they do not matter for our purposes.

[25] There are only 5 occurrences of the NPI *need* (e.g. "you **needn't** whisper").

Table 2.3 Counts and percentages of modal uses, by force, flavor and negation (adults, excluding tags and repetitions)**.**

ADULT (n=18,853)						
	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%

Sentence type. To get a sense of the kind of speech acts modals tended to be used for, we looked at clause type and subject person. They are included here, but I leave further analysis for future research. **Figures 2.1a** and **2.1b** summarize the distribution of possibility and necessity modals with and without negation per sentence type and subject, for root (2.1a) and epistemic (2.1b) modals. Necessity modals are rare in questions, especially epistemic ones, but parents use many possibility modals in interrogative sentences (e.g., "can you see it?").

Figure 2.1a Distribution of possibility and necessity modals with and without negation, by sentence type and subject for adults' root modals (n=17187):

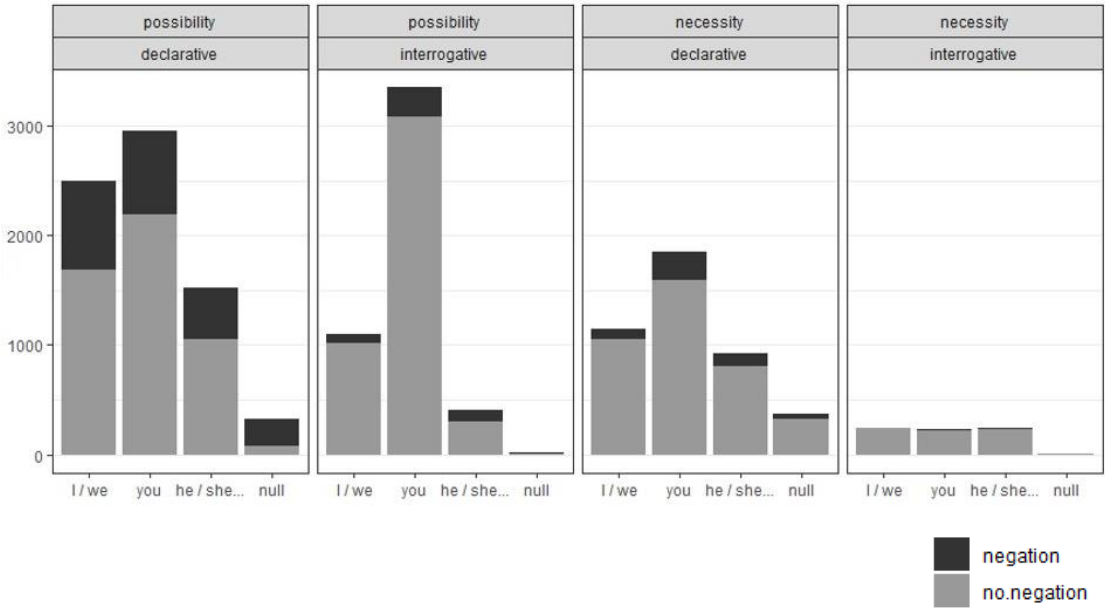
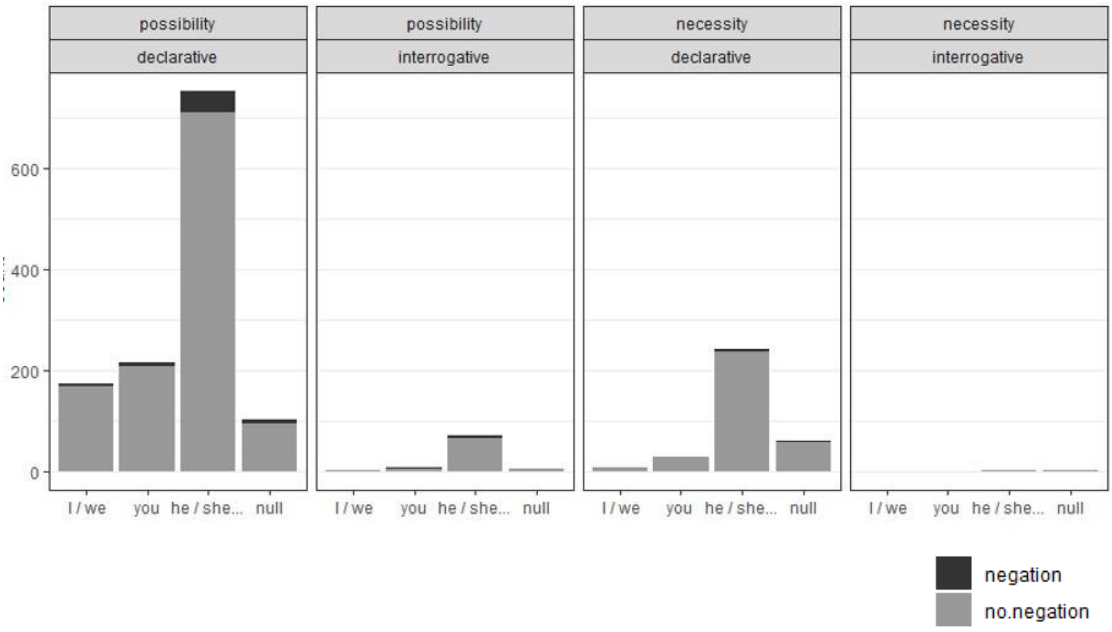


Figure 2.1b Distribution of possibility and necessity modals with and without negation, by sentence type and subject for adults' epistemic modals (n=1666):



2.1.3 Interim summary

Overall, we find that possibility modals are significantly more frequent than necessity modals in children's input. Note that this difference in frequency between possibility and necessity modals could be specific to English: in other languages, we might encounter different proportions. However, it might also be a more general phenomenon, due to alternative ways speakers can express necessity instead of using necessity modals: using imperatives for deontic necessity (e.g. "Be careful!" for "You must be careful"), or directly asserting the prejacent for epistemic necessity (e.g., "You have left a piece" for "You must have left a piece.").

Necessity modals are not frequent with negation, let alone in other downward-entailing environments: on the aggregate, only 10.1% of all necessity modals cooccur

with negation (vs. 20.9% of possibility modals). Moreover, most of these cases correspond to necessity modals that outscope negation (*must, should, ought to*: 19.4%, vs. *have to, got to, need to, be supposed to*: 7.4%). Cases of negated *have to/got to*, where using negation could in principle be useful to infer force, following Gualmini & Schwarz's rationale, are quite rare (*have to*: 4.5%; *got to*: 1.1%).

2.2 Input Experiment 1: adults' modal productions

In order to assess the general informativity of natural conversational contexts about force, we implemented a variant of the *Human Simulation Paradigm* (Gillette et al. 1998). One of the goals of the original *Human Simulation Paradigm* (Gillette et al., 1999; see also (Snedeker et al., 1999; Snedeker, 2000; White et al., 2017) was to compare the effect of different kinds of contextual information on the ability to recover a word's meaning: extralinguistic scenes, associated words and morphemes, or syntactic-frame information. The accuracy with which participants can recover the actual word given the context is taken as a general measure of the informativity of that context. We used a variant of the original paradigm, with two main modifications: first, following Orita et al. (2013), all participants were given written conversation transcripts from the corpus, with a blanked-out word (they had no visual or acoustic information). Second, they were asked to make a forced choice, between a possibility and a necessity modal. This allowed us to have a general measure of the informativity

of conversational context (as in the original paradigm), and to test the force contrast in a more controlled way.[26]

How easy is it for adults to guess the force of blanked-out modals based solely on excerpts of conversations in which they appear? Are there differences between root and epistemic modals? Is the conversational context equally informative for necessity and possibility modals? Last, how informative are negative uses of modals?

2.2.1 Methods

Procedure. The experiment was run online on Alex Drummond's IBEX Farm.[27] Participants recruited via Amazon MechanicalTurk were asked to guess a redacted modal in a dialogue between a child and mother by choosing between two options, corresponding either to a possibility (e.g. *might*) or a necessity modal (e.g. *must*), as illustrated in **Figure 2.2a**. All dialogue contexts consisted of the modal sentence with a blank and the 7 preceding utterances, with the two options displayed at the bottom of the screen. There first was a short training where participants had to choose between the definite vs. indefinite article (*the* vs. *a*) (3 examples with feedback), followed by 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

[26] A previous version of the experiment, where instead of making a choice between forces, participants had to 'fill in the blank', is reported in (Dieuleveut et al., 2019). In this experiment, we also tested the effect of sentence type (declarative/negative/interrogative sentence/tag), Results show that adults were overall quite good at guessing modal force from natural contexts, with no significant difference between necessity and possibility (overall accuracy for necessity modal contexts: 68.6% vs. for possibility modal contexts: 63.6%). Negation and interrogative sentences were both found to decrease accuracy.

[27] An example of the experiment can be accessed below (EPI-AFF condition): http://spellout.net/ibexexps/modsquad/HSP_FC_epiP/experiment.html

future), presented in random order. The 20 trials were randomly selected for each participant from a list of 40 contexts originally extracted from the corpus; the 20 controls were the same for all participants. Further details about the instructions and material are provided in Appendix B.

Figure 2.2a Input Experiment 1 stimuli: example trial (EPI-AFF, *must*):

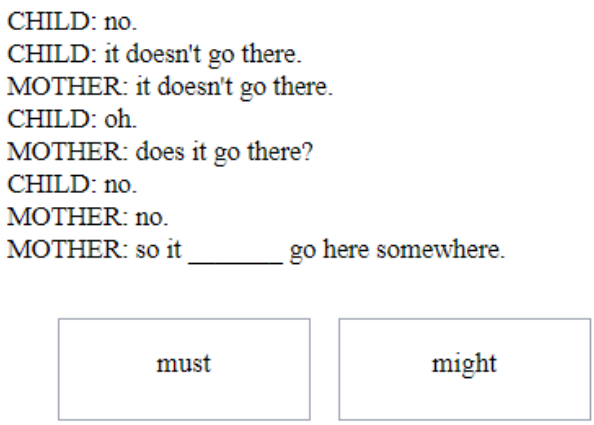
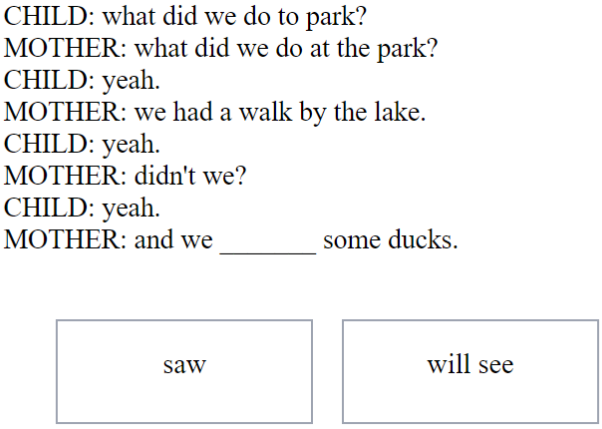


Figure 2.2b Input Experiment 1 stimuli: control trial (*saw*):



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 to sample (**Table 2.2**). **Table 2.4** summarizes the experimental design.

Table 2.4 Summary of experimental conditions (Input Experiment 1):

Test condition (between participants)		Modal lemmas: possibility	Modal lemmas: necessity
EPI-AFF (epistemic affirmative)		*might*	*must*
ROOT-AFF (root affirmative)[28]	**ROOT-AFF-1**	*can*	*must*
	ROOT-AFF-2	*can/able*	*have to*
ROOT-NEG (root negative)		*can't/not able*	*not have to*

Material. ***Extraction procedure*** **–** 160 contexts (2*20 per condition) were randomly extracted from the corpus for the different modals (*can, able, might, must, have to)*. ***Exclusion criteria*** **–** We excluded contexts where the adult or the child used the target modal in preceding utterances. Contexts were not excluded when the adult or the child used another non-target modal. Briticisms, such as *willn't*, were removed from the dialogue and replaced with American English equivalent (e.g. *won't*). We didn't exclude contexts where there were tag questions (e.g., *'..., mustn't she?*'), but removed the tags when they occurred in the target sentence. ***Controls*** **–** Participants had to choose between past and future (*e.g.* [*saw*] vs. [*will see*], see **Figure 2.2b**). Importantly, the correct answer was <u>not</u> always guessable based on the target sentence alone: it

[28] 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, while **ROOT-AFF-2** (*can/able to* vs. *have to*) allowed us to avoid concerns related to the formality of *must* for US English speakers. As in Experiment 1, in cases where *have to* was tensed, we used *able to* as the alternative to avoid losing tense information: for example, participants had to choose between [*will have to*] and [*will be able to*]. We extracted the same number of contexts from *able to*, to avoid having the *able to* option always be the wrong answer. Same principles applied for **ROOT-NEG** condition: participants had to choose between [*didn't have to*] and [*wasn't able to*] when *have to* was tensed.

required participants to read the entire dialogue. Extraction procedure and data cleaning were the same as for targets.

2.2.2 Results

Participants. 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-years-old). 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).

Analysis. Overall, participants were highly accurate at guessing modal force (general mean accuracy: 79.9%). We first ran binomial tests to see whether they differ from chance for each condition (**Table 2.5**). 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%). **Figure 2.3** summarizes the mean accuracy for each condition.[29] **Force.** 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),[30] 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

[29] Accuracy for controls was very high (94.6%). There was no difference between groups in accuracy. (Controls were the same across all groups).

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

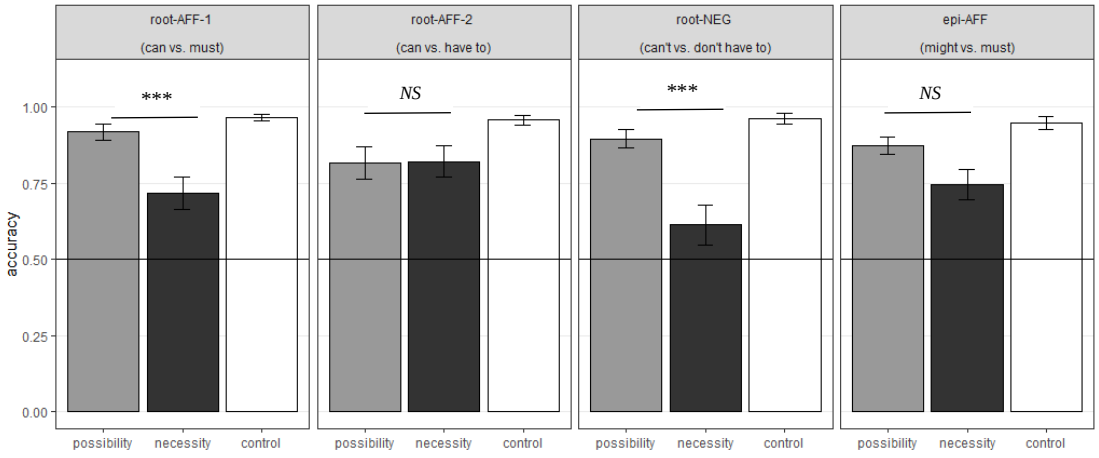
to each comparison group, we find a significant effect in ROOT-AFF-1 ($\chi^2(1)=61.1$, $p=5.5e\text{-}15$***) and ROOT-NEG ($\chi^2(1)=15.6$, $p=7.8e\text{-}05$***), again in the direction of a higher accuracy for possibility contexts, but not for ROOT-AFF-2 ($\chi^2(1)=6e\text{-}04$, $p=0.98$ *(NS)*) and EPI-AFF ($\chi^2(1)=3.73$, $p=.053$ *(NS)*). **Negation.** We compared ROOT-AFF-2 and ROOT-NEG, as these conditions included 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.** There was no general effect of flavor ($\chi^2(1)=0.11$, $p=0.74$ *(NS)*).

Table 2.5 Accuracy rates and significance tests by condition (Input Experiment 1: adults' productions) (n=281, 10 observations per cell):

	Mean accuracy (se)[31]		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=8.9e-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]
Total	87.5% (0.018)	72.3% (0.028)		
ALL	79.9% (0.018)			

[31] Accuracy corresponds to the mean accuracy (how good participants were to guess correctly the force of the modal, e.g. to answer can in a possibility context) across 20 contexts initially extracted from the corpus for each condition of force and flavor. Each participant saw only 10 out of the 20 contexts (10 for possibility, 10 for necessity). On average, each context was thus seen by 34.7 participants (ranging between 24 and 47).

Figure 2.3 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 looked at the contexts that led to lowest and highest accuracy, both for root and epistemic flavors. We focused on necessity modals as there was more variability in accuracy for them, as shown in **Figure 2.4** (distribution of accuracy for possibility and necessity modals in each condition). This informal analysis revealed two factors, depending on flavor. For root modals, cases where the proposition expressed by the prejacent seemed clearly undesirable (e.g., going to the hospital) or effortful (e.g., lifting a heavy object) seemed to lead to high accuracy for necessity modals (see (22)). For epistemic modals, we found high accuracy for necessity modals in contexts that made salient strong evidence for the prejacent (see (23)).

Our post-hoc analysis also pointed out a particularly high accuracy for possibility root modals interrogative sentences (e.g. ___ *you see*?) (mean accuracy for root possibility modals in interrogative: 98.0%). Note that in this case, accuracy may not reflect pure informativity, as participants may rely on idiomatic turns of phrases.

However, further analyses show that they are still accurate restricting to contexts that do not involve interrogatives: the mean accuracy for root possibility, restricted to declarative sentences, is 76.3%.[32]

(22) CHILD: Mummy.
CHILD: Mummy.
MOTHER: Mummy?
CHILD: that Mummy.
MOTHER: what... what happened to Mummy?
CHILD: poorly.
MOTHER: she's poorly, is she?
MOTHER: she... she _____ go to the hospital.

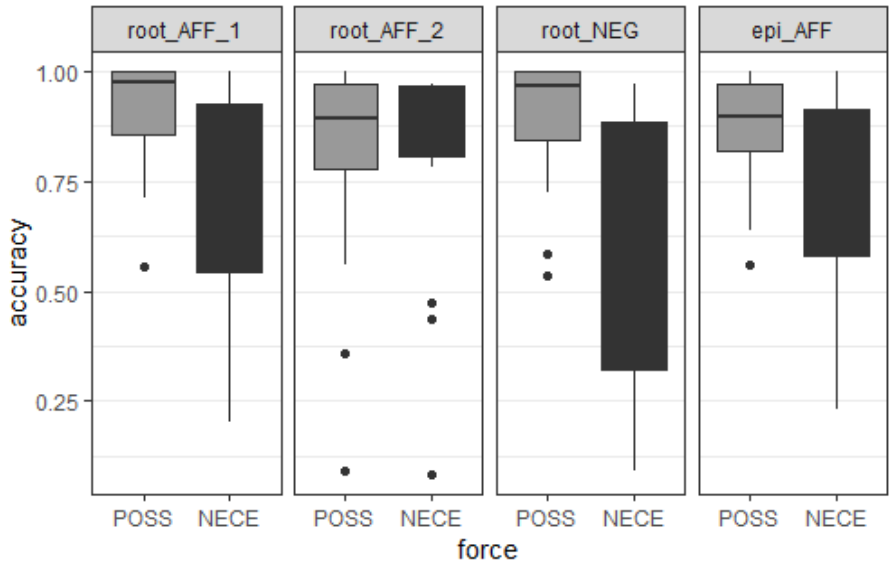
(*has to,* 'undesirable'; HSP mean accuracy: 96.6%)

(23) MOTHER: ...
CHILD: yeah.
MOTHER: but Bertie was very close behind, wasn't he?
MOTHER: it was a near thing I think.
CHILD: he's lost his hat.
MOTHER: he has.
MOTHER: yeah.
MOTHER: it _____ have been windy eh?

(*must,* 'strong justification'; HSP mean accuracy: 92.1%)

[32] Note that contexts involving interrogative sentences appeared almost exclusively in ROOT-AFF-1 and ROOT-AFF-2, as epistemic and negated modals are rare in interrogatives (see Figure 1a and 1b). Out of 80 contexts for root-AFF, there were 21 interrogative sentences (19 involving possibility modals; 2 necessity modals).

Figure 2.4 Accuracy for possibility and necessity contexts for each condition:



2.2.3 Discussion

Results from this first experiment show that the conversational context is informative about force: participants were able to guess the force of the modal accurately with a single exposure, just from short conversation transcripts, for both forces (general mean accuracy: 79.9%; possibility modals: 87.5%; necessity modals: 72.3%). This means that it is possible, at least in principle, for learners to figure out the force of modals based on conversational context alone. If children are sensitive to the same cues as adults, they at minimum don't need to rely on a bias towards necessity meanings, nor on negation, to figure out the force of modals.

Of course, some of the cues available to adults in this experiment might not be usable by children: for instance, children might lack some world knowledge. This limitation is intrinsic to any paradigm where adults are used to simulate word learning

(adults are asked to guess a word they already know, whereas children have to guess the meaning of a *new* word from the context in which it is used) (see White et al., 2017; Orita et al., 2013, for discussion). That said, children also have access to a substantially richer context than participants in our experiment, who had no visual nor prosodic information, and no common ground with the child and the mother.

We find a general effect of force, with participants being more accurate on possibility modals. This could be interpreted as possibility contexts being more informative than necessity contexts. However, this effect should be taken with caution, as it is carried by only 2 sub-conditions (ROOT-AFF-1 and ROOT-NEG; it is not significant in EPI-AFF ($\chi^2(1)=3.73, p=.053$), and in ROOT-AFF-2), and it is not significant once we take into account the effect of interrogative sentences, which lead to a very high accuracy for root possibility modals: if we restrict to declarative contexts only, participants don't perform significantly better on possibility contexts. This higher accuracy in possibility contexts might in principle also reflect a general tendency to answer with possibility modals by default, maybe because of the higher frequency of possibility modals as compared to necessity modals. To see whether this could explain our results, we compared accuracy for *can* and *able* (used in ROOT-POS-2 and ROOT-NEG), which are both root possibility modals but strongly differ in frequency (3 *able* per 100 *can* in the Manchester corpus). We find no significant difference in accuracy between *able* and *can* (overall: *able*: 80.8% vs. *can*: 89.8%; vs. *have to*: 71.7%); moreover, participants are still more accurate for *able* than for *have to*. This suggests that frequency does not fully predict participants' performances.

What are the cues making the conversational context useful for guessing the right modal force? Multiple factors may play a role: situational cues (e.g., who the interlocutors are), cues from world knowledge (e.g., what is culturally allowed or prohibited; or physical laws), or pragmatic cues (conversational goals, what the speaker is trying to achieve; for instance, performing orders, permissions or prohibitions). Our post-hoc exploration suggests that the nature of these cues may vary depending on modal flavor. It appears that the (un)desirability and effortfulness of the prejacent could be particularly useful with roots, and some explicit supporting evidence for epistemics. Note that this finding echoes Mandelkern (2019) who independently argues that epistemic necessity 'must' claims require *Support*: some argument in support of the claim that needs to be accessible to all interlocutors. He takes this constraint to explain the generalization that epistemic necessity claims are felicitous only if the speaker's evidence for them is in some sense indirect. We probe the effect of desirability more directly in the last experiment presented in this chapter (*Input Experiment 3*), and leave the case of epistemics for future research.

Finally, we find opposite effects of negation on (root) possibility and necessity modals: while negation leads to a slightly higher accuracy for possibility modals (*can't*: 89.5% vs. *can*: 81.5% (NS)), it leads to lower accuracy for necessity modals (*don't have to*: 61.3% vs. *have to*: 82.0%, *p=0.011**) (significant interaction effect Force*Negation: *p=0.0047***). Further exploration of the contexts that led to highest and lowest accuracy shows that in our corpus, mothers often use *don't have to* in 'polite' ways to perform prohibitions (orders to not do something), as in (23) or (24), with intended meanings that seem closer to impossibility. However, even that does not

fully explain their lower performance in the experiment: participants were actually extremely good at guessing these polite uses (100% accuracy for the dialogue in (25)). There might remain a few cases where *don't have to* is used in contexts that make it clear that the impossibility interpretation does not hold, which would be 'logically' informative for a learner following Gualmini & Schwarz's rationale, but among the 20 contexts that were tested in the experiment, it is hard to find any example of context with high accuracy not involving polite uses. 2- to 3-year-old children would need to already know about these conventional polite uses to use these contexts to learn force.

(24)
CHILD: break.
MOTHER: you want me to break it?
CHILD: yeah.
MOTHER: no.
MOTHER: we **don't have to** break these things.
MOTHER: oh.
MOTHER: you've broken it.
CHILD: yeah. (Aran, 2;0.28)

(25)
CHILD: knock off again.
MOTHER: that's gonna fall.
CHILD: no.
MOTHER: yes.
CHILD: no.
MOTHER: oh.
MOTHER: we are noisy, aren't we?
MOTHER: _____shout. (*don't have to*)

(HSP accuracy: 100%)

Our findings however suggest that negative environments could be more helpful to figure out the force of possibility modals: negated possibility modals are frequent in the input (22.6% of root possibility modals are negated), and the experiment

shows that impossibility contexts are particularly informative (mean accuracy for *can't*: 89.5%). Children may make use of these occurrences to infer the force of possibility modals, if they expect negation to scope over modals. I will come back to this point in more detail in the general discussion (**section 2.5**).

2.3 Input Experiment 2: Isolating the role of context

Our first input experiment shows high accuracy for both possibility and necessity. We take these results to mean that the context is informative as to force. But could it be that participants succeed at the task not by relying on the context, but through biases, which could also be at play in children's modal learning? In particular, could their high accuracy be due to a necessity bias that allows them to correctly guess necessity meanings?[33] To isolate the contribution of the dialogue context, we ran a second experiment, presenting only the target sentence without its discourse context. We expect that participants' performance should decrease in this new experiment, if their successes in our first experiment are due to a reliance on context, rather than a bias.

2.3.1 Methods

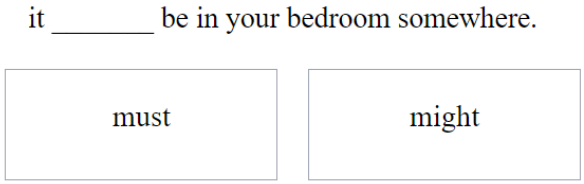
Procedure. *Input Experiment 2* was identical to *Input Experiment 1*, except that participants only saw the target sentence, and not the preceding dialogue (see **Figure 2.5**).[34] As the task was shorter, they judged all 40 contexts (60 trials: 20 possibility; 20

[33] For possibility modals, participants would have enough cues indicating that the necessity interpretation does not hold, but their accuracy with necessity modals would not come from the context being informative.

[34] An example of the experiment can be accessed at (EPI-AFF condition):

necessity; 20 controls using tense). We removed from target sentences any repetitions (*e.g.* '*dolly... dolly* _____ *use her pottie'* was corrected to '*dolly* _____ *use her pottie'*), as well as phatic words (*e.g. oh, yeah, well*). We did not remove logical words (*e.g. so, but, then, now, because, if*-clauses). In order to make sure that participants kept paying attention, we also had 8 attention checks (simple additions and subtractions, *e.g.* 1+3=__). Conditions were the same as in *Input Experiment 1*. Instructions are provided in Appendix B.

Figure 2.5 Input Experiment 2 stimuli: example trial (EPI-AFF, *must*):



2.3.2 Results

Participants. 123 participants were recruited on Amazon Mechanical Turk (ROOT-AFF-1: 31, ROOT-AFF-2: 33; ROOT-NEG: 30; EPI-AFF: 29; language: US English; 66 females, mean age: 44.0 years-old). We removed from the analysis 1 participant who was less than 75% accurate on attention checks and 6 participants who were less than 75%

https://spellout.net/ibexexps/modforce/modforce_FC0cxt_epiP/experiment.html

accurate on tense controls (5.7%).[35] We thus present results for 116 participants (ROOT-AFF-1: 30, ROOT-AFF-2: 28; ROOT-NEG: 30; EPI-AFF: 28).

Analysis. Overall, participants were still good at guessing force (**Table 2.6**), but their overall accuracy is lower without dialogue than when they saw the entire dialogue (binomial linear mixed effects models comparing general accuracy in Experiment 1 vs. Experiment 3*: $\chi^2(1)=48.2$, $p=3.9e\text{-}12$ ****). Looking at the 8 subcomparison groups, we see decreased performance for necessity contexts in ROOT-AFF-1, ROOT-AFF-2 and EPI-AFF, and for possibility contexts in ROOT-AFF-2 and ROOT-NEG. We find no difference for possibility ROOT-AFF-1 and EPI-AFF and necessity ROOT-NEG. Results are summarized in **Table 2.7**. We ran interaction tests to see whether the effect of the dialogue differed for possibility and necessity modals. The general interaction Experiment*Force is not significant ($\chi^2(1)=1.1$, $p=0.29$), but when restricted to affirmative conditions (i.e., excluding ROOT-NEG) (post-hoc), we find a significant effect, dialogues being more helpful for necessity modals ($\chi^2(1)=4.0.4$, $p=0.044$*). Looking at the 4 groups, the interaction effect is significant for EPI-AFF ($\chi^2(1)=5.08$, $p=0.024$*), but not ROOT-AFF-2 ($\chi^2(1)=0.015$, $p=0.90$). Problems with the model do not allow us to conclude for ROOT-AFF-1 and ROOT-NEG.[36]

[35] Accuracy on attention checks and tense controls was very high (attention checks: 99.4%; tense controls: 95.8%), with no difference between groups. To compute accuracy on tense controls, we only included sentences that could not lead to an ambiguity (e.g. because of containing a temporal adverb) (10 out of 20 cases).

[36] The problem (*singular fit*) appears to be due to variances of one linear combination of effects being close to zero. This is a relatively common problem with complex mixed effect modals used here, but it could not be solved by simplifying the model (see footnote 17).

Table 2.6 Accuracy rates and significance tests by condition (Input Experiment 2) (n=116, 20 observations per cell):

	Mean accuracy (se)		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	90.2% (0.030)	62.0% (0.062)	p <.001*** 95% CI [0.88, 0.92]	p <.001*** 95% CI [0.59, 0.65]
ROOT-AFF-2	71.8% (0.052)	73.0% (0.054)	p <.001*** 95% CI [0.68, 0.74]	p <.001*** 95% CI [0.70, 0.76]
ROOT-NEG	84.8% (0.036)	57.3% (0.061)	p <.001*** 95% CI [0.82, 0.87]	p=.00019 95% CI [0.54, 0.61]
EPI-AFF	88.6% (0.021)	64.6% (0.054)	p <.001*** 95% CI [0.86, 0.90]	p <.001*** 95% CI [0.61, 0.68]

Table 2.7 Results of the model testing effect of the Dialogue (Experiment 1 vs. Experiment 2):

	possibility	**necessity**
ROOT-AFF-1	$\chi^2(1)=0.903$, $p=0.34$ (NS)	$\chi^2(1)=14.9$, $p=0.00012$ ***
ROOT-AFF-2	$\chi^2(1)=15.5$, $p=8.0e-05$ ***	$\chi^2(1)=11.7$, $p=0.00064$ ***
ROOT-NEG	$\chi^2(1)=6.4$, $p=0.011$ *	$\chi^2(1)=1.81$, $p=0.18$ (NS)
EPI-AFF	$\chi^2(1)=0.31$, $p=0.57$ (NS)	$\chi^2(1)=9.25$, $p=0.0023$**
all	$\chi^2(1)=11.5$, $p=0.0007$***	$\chi^2(1)=32.6$, $p= 1.1e-08$ ***
Overall	$\chi^2(1)=48.2$, $p=3.9e-12$ ***	

2.3.3 Discussion

This control experiment allows us to isolate the contribution of the preceding dialogue, and shows that context is informative beyond potential biases. Note that we did not expect participants to be at chance for this version of the experiment, as the information conveyed by the prejacent contributes to the context. Furthermore, it is sometimes possible to recover the modal from the clause type (*e.g.* interrogative sentences with *can:* mean accuracy in Experiment 1: 97.8%; in Experiment 2: 96.4%). Despite that,

we find that participants are overall better at identifying force when presented with the dialogue, for both forces (overall accuracy in Experiment 1: 79.9%; in Experiment 2: 74.0%; effect of the dialogue: overall +5.9%; necessity: +8.1%; possibility +3.7%).

Interestingly, we find that having the dialogue context is more helpful for necessity modals than for possibility modals. In all affirmative conditions, the effect of having the dialogue is significant for necessity modals, but only for one of the possibility conditions. The overall interaction Force*Experiment is not significant, but the interaction Force*Experiment is significant when we restrict the analysis to the affirmative conditions. In the negated condition (ROOT-NEG), the effect of the dialogue seems to go in the opposite direction: having the dialogue is slightly more helpful for *can't* than for *don't have to* (NS). But if participants' high accuracy in the first experiment was due to a necessity bias, we would expect their performance to remain the same in this follow-up (participants should guess necessity meanings, unless presented with direct evidence against it). Altogether, participants' high accuracy on possibility modals, even with context reduced only to what is in the prejacent, suggests that if they bring a force bias to the task, it is more likely to be a possibility bias, rather than a necessity one.

The results from these two experiments show that the conversational context in which modals are used is informative about their force, and might be even more informative for necessity modals. But what is it about the context that is particularly informative? As discussed in **section 2.2.3**, several factors could be at play. Our post-hoc analysis of the contexts that lead to higher and lower accuracy suggested that the cues may vary with flavor: for root modals, necessity modals seem associated with

undesirable and effortful events; for epistemics, necessity modals seem to occur in contexts that highlight strong evidence that supports the proposition expressed by the prejacent. I now turn to an experiment that directly tests the hypothesis that (un)desirability matters for root modals, as an initial proof of concept, and leave a more systematic probing of additional features of the context for future research.

2.4 Input Experiment 3: Desirability

Desirability is a feature likely to be conceptually accessible to young children: the cognitive developmental literature suggests that children can reason about desires quite early on, and understand that people can have incompatible desires (Wellman & Woolley, 1990; Repacholi & Gopnik, 1997; Rakoczy et al., 2007; Ruffman et al., 2018, a.o.). Moreover, the link between desirability and force seems quite intuitive, at least for adults, though it is an open question whether children also have such associations: desirable activities or events are usually associated to permissions, whereas undesirable activities are associated with orders and prohibitions. The goal of this last experiment is first to assess the availability of this cue in the input: do adults actually use necessity modals more frequently with undesirable events (e.g., 'You *must/#can* clean your room'), and possibility modals with desirable events (e.g., 'You *can/#must* go play in the garden')? Second, does this contribute to participants' performance in *Input Experiment 1*, i.e., do participants actually rely on this cue to guess force?

2.4.1 Methods

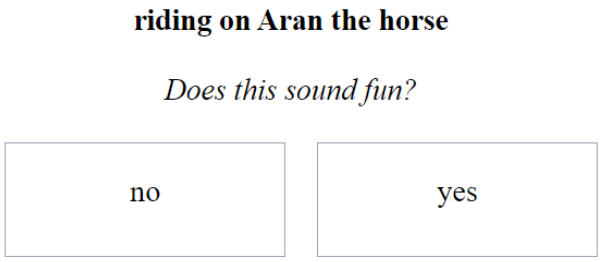
Procedure. Participants were asked to indicate whether various activities (*e.g.* 'doing a puzzle') sounded fun or not (see **Figure 2.6**). They were told that the activities involved two-year-old children and their mothers. The different activities corresponded to the prejacents[37] of the modals tested in *Input Experiment 1* and *2*.[38] We used the prejacents, rather than the full modal sentences to avoid biases towards positive responses for possibility modals, and negative responses for necessity modals. For example, for 'Can the dolly ride on Aran the horse?', participants were asked whether '*riding on Aran the horse*' sounded fun ('yes') or not ('no'). Referential pronouns (e.g. *it*) were replaced with the full nominal whenever they could be recovered from the context (*e.g.* '*Finding the green marker'* for 'Can you find it?'). In each group, participants judged all 40 prejacents (42 trials: 20 possibility, 20 necessity; 2 initial practice items, which were removed from the analysis). To make sure participants kept paying attention, we had 10 attention checks (*e.g.* 1+3=__). Instructions are given in Appendix B. As our hypothesis concerns root modals, we ran the experiment only on ROOT-AFF-1 (*can* vs. *must*) and ROOT-AFF-2 (*can/able* vs. *have to*). ***Rationale***. This experiment allows us first to assess the desirability of the different events in an objective way, to see if there is a relation between desirability (measured by the proportion of yes answers to 'being fun', a child-friendly way of assessing what is desirable) and force usage in the corpus. We can then probe whether adults used this

[37] This is not true *stricto sensu*, as participants also lose the information about the subject (e.g. *I/you/Caroline*).

[38] An example of the experiment (Root-AFF-1 condition) can be accessed at: https://spellout.net/ibexexps/modforce/modforce_hspdesF_rootP1/experiment.html

cue to infer force in Experiment 1 by looking at the correlation between the desirability score in Experiment 3 and accuracy in Experiment 1. We expect a negative correlation for necessity modals (fewer 'yes' responses for accurate guesses of necessity uses) and a positive correlation for possibility modals (more 'yes' responses for accurate guesses of possibility uses).

Figure 2.6 Input Experiment 3 stimuli: example trial (ROOT-AFF-1, *can*):



2.4.2 Results

Participants. We recruited 70 participants on Amazon Mechanical Turk (ROOT-AFF-1: 35, ROOT-AFF-2: 35; language: US English; 35 females, mean age: 40.4-years-old). Accuracy on attention checks was very high (99.6%), and we did not have to remove any participant from the analysis based on attention checks.

Analysis. We find a general effect of force: participants judged prejacents extracted from possibility statements overall more 'desirable' than those extracted from necessity statements (overall mean of 'yes' answers: 40.7%; possibility: 52.9%; necessity: 28.6%) (**Table 2.8**). **Figure 2.7** shows the distribution of ratings for possibility and necessity for the two groups. We first checked that there was no significant difference

between groups (comparing ROOT-AFF-1 (*must* vs. *can*) and ROOT-AFF-2 (*have to* vs *can/able*): overall: $\chi^2(1)=0.22$, $p=0.64$; possibility: $\chi^2(1)=0.126$, $p=0.72$; necessity: $\chi^2(1)=0.16$, $p=0.69$). We find a general effect of Force, with predicates extracted from necessity statements rated as less desirable than their possibility counterparts ($\chi^2(1)=15.5$, $p=8.2e\text{-}05$ ***). The effect is significant for both groups (ROOT-AFF-1: $\chi^2(1)=8.2$, $p=0.0041$** ; ROOT-AFF-2: $\chi^2(1)=6.2$, $p=0.012$*). Last, we computed correlations between the desirability score (*Input Experiment 3*) and accuracy in *Input Experiment 1* (see **Figure 2.8**). For possibility, we find a weak positive correlation (Pearson's $r=0.12$) ($t(1398)=4.42$, $p < 0.001$; 95%-CI: [0.065; 0.168]); for necessity, a weak negative correlation (Pearson's $r=-0.073$) (t(1398)=-2.74, $p= 0.0063$; 95%-CI: [-0.125; -0.021]).

Figure 2.7 Distribution of 'desirable' answers for possibility and necessity contexts for each group (ROOT-AFF-1 (*must* vs. *can*) and ROOT-AFF-2 (*have to* vs. *can/able*):

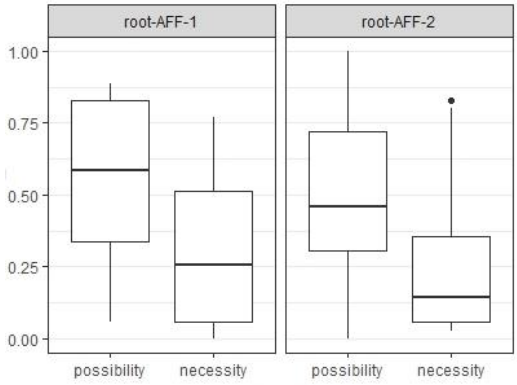
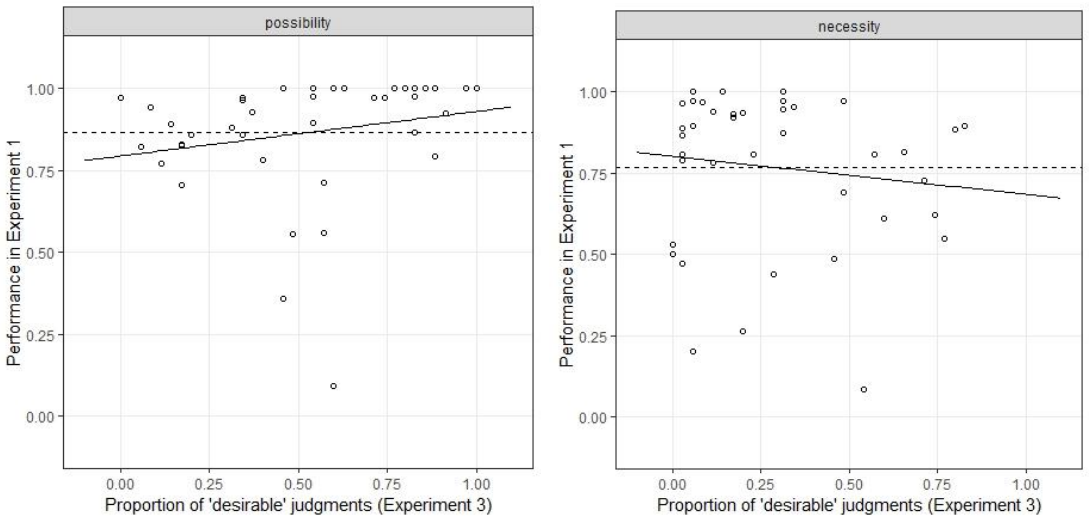


Table 2.8 Desirability scores and significance tests (binomial linear mixed effects models comparing possibility/necessity) for possibility and necessity modals**:**

	Mean of desirable ('yes') answers (se) possibility	necessity	Effect of Force
ROOT-AFF-1	56.0% (0.063)	31.4% (0.060)	$\chi^2(1)=8.22$, $p=0.0041$ **
ROOT-AFF-2	49.7% (0.067)	25.7% (0.057)	$\chi^2(1)=6.2562$, $p=0.012$ *
ALL	52.9% (0.045)	28.6% (0.041)	$\chi^2(1)=15.5$, $p=8.2e\text{-}05$ ***
ALL	40.7%		

Figure 2.8 Relation between accuracy in Input Experiment 1 (y-axis) and desirability score in Input Experiment 3 (x-axis) by force. Black lines correspond to Pearson's r. Dashed lines correspond to the mean accuracy in Experiment 1, for possibility and necessity contexts.



2.4.3 Discussion

Our results confirm our initial observations for *Input Experiment 1*, and show that there is a relation between the desirability of the prejacent (evaluated by participants that were blind to the force of the modal originally used) and force in adults' speech. Adults use root possibility modals more frequently with desirable events, and root necessity modals with undesirable events (mean desirability score for possibility modals (*can/able*): 52.9%; for necessity modals (*must/have to*): 28.6%). Furthermore, the lower accuracy in Experiment 1 for possibility modals with undesirable prejacents and for necessity modals with desirable prejacents suggests that adult participants made use of desirability in their force judgments. Together, this suggests that children can conceivably use this cue: it is available in the input, the cognitive developmental literature suggests they are sensitive to it, and adults participants make use of it. Note

however that, to be able to use this relation, children would further need to expect a link between, on the one hand, orders and undesirable events, and on the other hand, permissions and desirable events: the relation seems quite intuitive for adults, but this might be something children have to learn.

2.5 Summary and general discussion: children's modal input

Let's come back to our problem: how can children figure out force, and in particular, how can they solve the Subset Problem for necessity modals? In this last section, I'll summarize the results from our study of the input, and discuss the conclusions we can draw for each of the solutions we considered at the beginning.

The first solution we considered is that children have a bias towards strong necessity meanings, in the spirit of Berwick (1985): they would, by default, assume necessity meanings for modals. What this study of the input shows is children might not need such a bias. Indeed, children have other strategies available given the way modals are used in their input: we find that speakers use possibility and necessity modals in clearly distinct situations, that are reflective of force, for both root and epistemic modality. If the conversational context is highly informative about both forces, and if children are able to use those situational cues when learning, they don't need to have a necessity bias.

Of course, from this does not follow that they lack such a bias. However, as we saw in Chapter 1, the justification for this kind of proposals is often that it is the only way to explain how children learn. Once we show that given their input, children don't

need such a bias, the justification for it becomes less compelling. We will come back to this question in the next chapter, where focusing on children's own modal productions, we will see that we don't find evidence for such a bias in their early uses of modals. On the contrary, children already seem to have mastered possibility modals by age two, but they do not use necessity modals appropriately.

The second solution we wanted to assess is that children can use evidence from downward-entailing environments, following Gualmini & Schwarz (2009). From hearing necessity modals under negation, which is by far the most prevalent DE context in this corpus, for instance "She doesn't have to go" used in clear situations of non-necessity ('she can go, and she can also stay'), children might be able to infer that *have to* cannot mean possible, since if it did, its negation would mean impossible, and therefore it could not be used in a situation where it *is* possible for her to go. Assuming that children are able to apply such a reasoning, the question was whether they get exposed to such informative cases.

Even before looking at input, we saw that relying on negation to learn necessity modals could be confusing given that some, but not all, necessity modals scope above negation. As we saw in **section 1.2.3** of Chapter 1, while English *have to* scopes under negation, *must* or *should* do not (Iatridou & Zeijlstra, 2013).

We can thus separate two types of cases, depending on the scope interpretation that necessity modals receive with negation. First, for necessity modals that outscope negation, like *must* ("she **mustn't** go": *necessary* > *not*), the reasoning proposed by Gualmini & Schwarz (2009) cannot be applied by children. Learners will never hear "she mustn't go" in possibility situations, since this is not a possible meaning for the

adult. And as explained in Chapter 1, these cases could even be misleading, since "she **mustn't** go" and "she **can't** go" are truth-conditionally equivalent: if children expect the same scope behavior for *must* and *can*, uses of *mustn't* for impossibility should lead them to infer that *must* expresses the same force as *can*. If children already know that *must* expresses necessity, they might infer its scope relation (and maybe, its polarity restrictions) from negative environments; but then, they need to have figured out force first: We get into a vicious circle here, where learners would need force to figure out the scope, but need the scope to figure out the force.

Turning to the second type of case, necessity modals that do scope below negation like *have to* ("she **doesn't have to** go": *not* > *necessary*), negation could in principle be 'logically' useful: the rationale proposed by Gualmini and Schwarz at least can apply. Looking at the actual input, we find that they don't often occur with negation or in other Downward-Entailing environments: on the aggregate, they are found in those environments 7.4% of the time (vs. 19.4% for potentially 'problematic' modals scoping over negation like *must*). They could still be useful, if children are able to use them. But more problematically, our experimental results show that the rare cases they hear might not always be so informative: even when adults use *don't have to*, they do not systematically convey non-necessity. We find that the context is the least clear about force for negated necessity modals (the accuracy for *don't have to* in our Experiment 1 is 61.3%). And most instances involve 'polite' uses, where *don't have to* is used to perform a prohibition (e.g., "you don't have to break those things"), the meaning of which may seem closer to impossibility. Do children have the pragmatic sophistication to understand that the necessity modal is used for prohibition? Can they

expect speakers to be 'polite'? These are open questions, but these uses question the applicability of Gualmini and Schwarz's solution, even in the cases where it is in principle usable, since overall, it seems that children are hardly ever exposed to cases where *don't have to* is clearly used to express non-necessity, and at least, not in a systematic way.

Why is negation rare with those necessity modals? There could be a principled reason for this. It may be due, in part, to a competition with the use of a bare possibility modal, which can convey non-necessity via a scalar implicature ("She can go, but she doesn't have to") (see Horn, 1972): speakers have another potentially simpler way to convey 'non-necessity' meanings.[39] From alogical point of view, there is only one case where *can* and *don't have to* take opposite truth-values (i.e., where speakers could use *don't have to*, but not *can*): cases of impossibility (*not* > *possible* entails *not* > *necessary*). But then, speakers should prefer to use *can't*, as it is logically stronger (more informative). *Don't have to* thus pragmatically competes with either *can* (simpler) or *can't* (more informative). Coming back to our learner, this is important to keep in mind: indeed, if this is the reason negated necessity modals are rare, we can expect the same problem to arise for learners of other languages, who won't get exposed to many occurrences of negated necessity modals either.

[39] Horn focuses on a different but related problem, namely the fact that cross-linguistically, the 'O' corner of the Aristotelian square of opposition (corresponding to negated universals, here, non-necessity meanings) seems to never be lexicalized, whereas the other three corners (corresponding to possibility, necessity and impossibility) can be. Horn argues that this follows from the fact that there is no functional pressure to lexicalize non-necessity meanings: speakers already have a way to express non necessity, using scalar implicatures.

Note that the few other corpus studies that report data on the distribution of possibility and necessity modals with respect to negation also suggest that negated necessity modals are infrequent. However, these studies are quite rare, and very few focus on child-directed speech. De Haan (2011) reports that negation is rare with *must*: 2.5% in the Brown corpus (written English), and 1.4% in the Switchboard corpus (spoken English). Thornton & Tesan (2013) report the frequencies of some negative auxiliary verbs in the input to children in the Providence corpus, but don't specify their frequency relative to the positive forms. Last, Jeretič (2018) also reports that negation on necessity modals is not frequent in the input to French and Spanish children (necessity modals in French: 15.5% with negation; in Spanish: 6.2%).

Those results thus suggest that negation may not be sufficient to solve the Subset Problem. First, for a number of necessity modals like *must/should*, children cannot observe them in negative environments, as they systematically scope above negation; in those cases, negation might even be potentially misleading as to their force (*mustn't* is truth-conditionally equivalent to *can't*, which might drive children to infer that they express possibility, if children assume that negation scopes over (root) modals by default). In principle, children could still use negation to figure out force for these cases, but they would have to have figured out their polarity restrictions first—which they might not be able to do if they haven't figured out force yet. Second, for the necessity modals that can scope under negation (e.g. *have to, got to*), where negation could in principle be helpful, we find that they rarely occur with negation, and that it is unclear how explicit context is as to their force: their uses seem potentially misleading as well, notably because of 'polite' uses. This does not mean that children cannot use

negation to learn force: in principle, they could still use these rare cases to figure out the force of necessity modals like *have to*.

Negation could be quite useful for honing in on the force of possibility modals, at least root ones. Here again, we need to separate two types of cases, depending on the scope interpretation that they receive with negation. First, for possibility modals that scope under negation, like *can* ("she **can't** go": not > possible), negation could be quite useful. Our corpus results show that negated possibility modals like *can't* are frequent in the input (22.6% of root possibility modals are negated), and our experimental results show that they are used in contexts particularly informative with respect to force (mean accuracy for *can't*: 89.5%). In the second type of case, possibility modals that outscope negation like *might* (and *may* in its epistemic uses) ("she **might not** go": *possible* > *not*), again, negation could in principle be misleading as to force if children expect the same scope behavior for *might* and *can*: occurrences of *mightn't* to convey non-necessity should lead them to infer that *might* and *can* express different forces (children would then be misled into thinking that *might* means *necessary*). However, looking at the input, we find that these potentially misleading cases are extremely rare, maybe too rare to be a serious issue: out of the 67 epistemic possibility modals with negation we found, only 42 correspond to potentially misleading cases such as *might/may not* (e.g., "**might not** be in there"), and 25 correspond to *can't/couldn't*, which are not misleading since the possibility modal scopes under negation (e.g., "**can't** have been Anne": *not* > *possible*). This means that only 3.4% of uses of *might* would be problematic, if context

is informative about force—which we didn't test experimentally, precisely because of how infrequent they are.[40]

Our third solution to the Subset Problem was that children use cues from the conversational context. Our study supports this last possibility: we show that the conversational context in which modals are used is informative about force. If children are sensitive to these conversational cues, and able to use them when learning, they don't need to have a necessity bias, nor to rely solely on negative environments.

What exactly, in the conversational context, signals modal force? We saw that these conversational cues may vary with flavor: for epistemic modality, our post-hoc analysis suggests that contexts that explicitly highlight salient evidence in favor of the prejacent may bias interpretations towards necessity; for root modals, that the perceived (un)desirability of the prejacent could be particularly helpful. *Input Experiment 3* confirms the potential usefulness of such a cue. It shows that it is available in the input: necessity modals tend to occur with undesirable prejacents (e.g., 'you **must/have to** clean your room'), and possibility modals with desirable prejacents (e.g., 'you **can** go play in the garden'). Moreover, participants in our *Input Experiment 1* seem to make use of this cue to determine force: they were better at guessing necessity modals when they occur with undesirable prejacents, and possibility modals when they occur with desirable prejacents. Other aspects of the context could also prove useful, including situational cues (e.g., who the interlocutors are), cues from world knowledge (e.g., what

[40] Moreover, if learners expect negation to scope over root modals but under epistemic modals given some more general assumptions about flavor and scope (and if they have already figured out flavor for *might*, which might not be straightforward), *might* not is no longer problematic.

is allowed or prohibited), pragmatic cues (what the speaker is trying to achieve, in particular performing orders, permissions or prohibitions), and prosody. This is an avenue for future research.

Before we conclude, let's briefly discuss potential implications of these findings for how children acquire modal force in languages beyond English, and in particular, in languages with 'variable force' modals. As discussed in **section 1.2.3** of Chapter 1, in a language like English where modals come in both forces, we can expect speakers to use possibility and necessity modals in fairly distinct situations, and notably, to avoid using possibility modals in necessity situations (modulo politeness considerations). And indeed, our input results show that speakers use possibility and necessity modals in distinct situations that are highly reflective of force. But in a language that lacks modal duals, speakers are more likely to use particular modals in both possibility and necessity situations. For variable force modals that are underlying possibility modals, like Nez Perce *o'qa*, it seems that negation would thus be crucial for learners to hone in on its underlying force—just as it was for Deal (2011) to argue for a possibility analysis. For variable force modals that are underlying necessity modals as in St'át'imcets or Washo, the challenge may be much greater. Not only might speakers use the same modals in possibility and necessity situations, but learners may not be able to rely on negation, given that—similarly to what happens with *must*—those modals can't scope over negation in these languages. Yet, speakers seem to have converged on necessity meanings for those modals, as evidenced by their preferred translations using English necessity modals (Rullmann & Matthewson, 2008; Bochnak, 2015). Here fieldworkers can and do rely on such translations as evidence for the

modals' underlying force, but this strategy is obviously inaccessible to the child. How do learners figure out their underlying force? At first blush, this situation might argue for a necessity bias. However, it could also be that while these modals can in principle be used in possibility situations, in practice, variable force modals are mostly used in contexts where English speakers use necessity modals, in which case, their acquisition could involve the same reliance on contextual cues that we've proposed for the acquisition of English modals.

To conclude, we saw that in principle, a strong necessity bias may not be necessary, that negation may not be that helpful, but that aspects of the context, like properties of the prejacent such as desirability, could provide useful cues to the force of a modal. But even if some cues are extremely helpful, children could be oblivious to them—unable to perceive them, or unable to see their utility. In Chapter 3, we will focus on young children's own modal productions, to assess their early mastery of possibility and necessity modals. Then, in Chapter 4, we will probe directly how helpful different aspects of children's input are in practice, by relating this input study to young children's modal mastery, to see whether variation in mother's speech can predict variation between children.

Chapter 3. Children's modal productions: A necessity gap?

This chapter focuses on children's modals. What modals do 2- to 3-year-old children spontaneously produce? How early and how frequently do they use possibility and necessity modals with and without negation? And do they use them in an adult-like way?

Traditionally, studies of young children's modal productions focus on their acquisition of flavor (e.g., how early they start producing ability, deontic, or epistemic modal flavors). What we know about the acquisition of modal force mostly draws on behavioral experiments, which typically target children from at least age 4. Existing comprehension studies show that children struggle with both forces: they tend to both accept possibility modals (e.g. *can/might*) in necessity situations, and necessity modals (e.g. *have to/must*) in possibility situations, for both epistemic and root modality (e.g., Noveck, 2001; Öztürk & Papafragou, 2015; Cournane et al. in prep.). Typically, these errors have been attributed to reasoning difficulties: children over-accept possibility modals in necessity situations because of difficulties reasoning about when a stronger modal is more appropriate (i.e., they have trouble with scalar implicatures, the inference from '"it might" that 'it doesn't have to' see **section 1.2.3**); they over-accept necessity modals in possibility situations because of difficulties reasoning about open possibilities (so-called *Premature Closure*, Acredolo & Horobin, 1987; Öztürk & Papafragou, 2015).

In the present study, I focus on younger (2 to 3-year old) children's naturalistic spontaneous productions of possibility and necessity modals, in a way that

complements existing studies and results from the literature. When do children start using possibility and necessity modals? And do they use them appropriately? In addition to standard quantitative corpus measures about children's frequency of productions (how frequently they use root and epistemic possibility and necessity modals, and in which syntactic environments), I assess experimentally whether children use their modals in an adult-like way, using a novel method which I call the *Guess the Force* (GF) paradigm.[41] The experiment borrows the HSP paradigm used to assess the informativity of context for adults' productions in Chapter 2 (*Input Experiment 1*, **section 2.2**) in order to assess children's own productions: adult participants have to guess the force of modals uttered by children, blanked out from corpus dialogues. The adult HSP experiment is used as a baseline, showing that force is guessable from context. This allows us to measure children's modal mastery: given that adults can guess the force of other adults' modals based on context, if children also know the force and use them in the same situations as adults, adults should be able to guess force from their utterances. If they cannot, that suggests that children are non adult-like in some way. I use it to test how adult-like young children's modal uses are for root and epistemic flavors, with and without negation.

Results from the study will show an asymmetry in English children's early mastery of possibility and necessity modals. We will see that children seem to master possibility modals early: at age 2, they use them frequently and productively, both with

[41] In this chapter, I use a different name for the paradigm than for the adult HSP experiment presented in **section 2.1** of Chapter 2 essentially for ease of presentation. The paradigm is the same: the only difference is that *children*'s productions are tested, which is a novel method of assessing young children's mastery.

and without negation, and in an adult-like way (they do not use them in situations adults treat as necessity). However, they seem to have more difficulties with necessity modals: they tend to produce them later on, less frequently, rarely with negation, and often in a non adult-like way: they use them in situations where adults would prefer using possibility modals, and with negation, in situations where adults would prefer negated possibility modals.

What is the nature of children's struggle with necessity modals? One possibility that is tacitly assumed in the existing literature is that they know the force, but have trouble dealing with open possibilities for conceptual reasons. Another possibility is that they know the force, but have trouble figuring out when to use necessity modals. Finally, another possibility is that they have lexicalized the wrong force, or they are uncertain about their force. If this were the case, and if the difficulty persists in the preschool years, it could explain why children tend to both over accept necessity modals in possibility situations (they wouldn't know their force), and possibility modals in necessity situations (they wouldn't have a reliable stronger alternative to make scalar implicature).

The rest of this chapter is structured as follows. In **section 3.1**, I review the existing literature on modal acquisition, both in terms of flavor and force. Then, in **sections 3.2** and **3.3**, I turn to the study of children's productions, using the Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000). As for the input study presented in Chapter 2, I break it down into two parts: first in **section 3.2**, I provide quantitative results about how frequently children use possibility and necessity modals, with and without negation, for different flavors; then, in **section 3.3**,

I present the GF (*Guess the Force*) experiment on their productions, which assesses how adult-like their uses of possibility and necessity modals are. Finally, in **section 3.4**, I discuss possible explanations for the 'Necessity Gap' we seem to find—whether it might reflect conceptual, semantic, or pragmatic issues, or a combination of these, and relate these new results to the previous literature.

3.1 Background: What we know about children's understanding of modals[42]

In this section, I first review what we know about children's understanding of flavor, which mostly draws on corpus studies of their spontaneous productions. Then, I review what we know about their mastery of force, with comprehension experiments initially motivated by work on scalar implicatures, as well as work on their interpretation of negated modal sentences.

3.1.1 Modal flavor in child productions

The literature on modal acquisition initially explored flavor, to see if there could be conceptual asymmetries reflected in children's modal productions. Initial corpus studies reveal a strong asymmetry in children's modal productions: children tend to produce roots (e.g. ability, deontic, teleological) about a year before epistemics (e.g. Kuczaj & Maratsos, 1975; Wells, 1979; Stephany, 1983; Cournane, 2015a,b; Van Dooren et al., 2017; for overviews, see Papafragou, 1998; Hickmann & Bassano, 2016; Cournane, 2021). This 'delay' with epistemics has been reported in several languages,

[42] See Cournane, 2020, for a more comprehensive review.

and was called the "Epistemic Gap" (Cournane, 2015) (for Dutch: van Dooren et al., 2019; for Bosnian/Croatian/Serbian: Veselinović & Cournane, 2020; for French: Cournane & Tailleur, 2020).

The Epistemic Gap was first blamed on children's conceptual development (Sweetser, 1990; Perkins, 1983; Shatz & Wilcox, 1991; Astington, 1993; Papafragou, 1998). Epistemic reasoning (i.e., having to do with inferences and knowledge) would be delayed, as compared to reasoning involved for root modality (i.e., having to do with goals, rules, desires, or physical abilities) (Shatz et al., 1983; Bartsch & Wellman, 1995; Asplin, 2002; de Villiers, 2005; Papafragou et al., 2007, a.o.).[43] However, research has shown since that this cannot be the entire explanation. First, already by age 2, children produce lexical epistemics like *maybe* in an adult-like way: this epistemic 'delay' is specific to functional modals (auxiliaries and semi auxiliaries that can express variable flavors; see Cournane, 2021). Moreover, studies using denser corpus show that children already produce some epistemics before age 3, even though they are rare (see van Dooren et al., 2017). This suggests that the problem is not conceptual: an alternative explanation is that the Epistemic Gap relates to children's grammatical development (Heizmann, 2006; Cournane, 2015a,b; Veselinović & Cournane, 2020), since epistemic

[43] Similar explanations were proposed to capture the asymmetry in acquisition between attitude verbs *think* and *want* (for overviews, see Harrigan et al., 2018a; Hacquard & Lidz, 2019). Likewise, there is an asymmetry in how early children seem to master verbs like *think*, that express notions related to belief (~ epistemic), and verbs like *want*, that express notions related to goals or desires (~ root): *think* seems to be acquired later than *want*. Children's failures with *think* were first attributed to problems with Theory of Mind, in particular to ascribing (false) beliefs to others (see e.g. Wellman et al., 2001). However, research has since shown that children might not have such difficulty with Theory of Mind (see e.g. Onishi & Baillargeon, 2005; Southgate et al., 2007), and suggested that this asymmetry is better explained by other factors (for a summary and alternative explanations for the asymmetry, see Hacquard & Lidz, 2019).

modals have more complex grammatical representations than root modals (Cournane, 2015a,b; see also Heizmann, 2006; de Villiers, 2007). Children's ability to produce epistemic modals would depend on their mastery of propositional embedding: then, they are not expected to start producing them before age 3, since the first markers of propositional embedding are found a little before age 3 (see Cournane (2015a,b); de Villiers & Roeper, 2016). This would also explain that the duration of the Epistemic Gap is found to differ from one language to the other: while first epistemic uses are found around age 3 or earlier in English, they remain absent until age 4 in Bosnian/Croatian/Serbian, where modal verbs require CP embedding (see Veselinović & Cournane, 2020; Cournane, 2020, for discussion).

Early studies focused on children's productions. They didn't look at their input: what modals children hear in their parents' speech. Importantly, later work focusing on the input shows that modals are used much more frequently to express root modality than epistemic in parents' productions as well: in English for instance, 90.7% of parents' modal uses involve root meanings, vs. only 9.3% epistemic (see van Dooren et al., 2017). Similar input asymmetries are reported in Dutch and Bosnian/Croatian/Serbian. The Epistemic Gap might thus also just reflect the asymmetry found in parents' speech: the 'delay' might be a matter of exposure and clarity of the input, rather than a real "gap" related to conceptual development.

As discussed in Chapter 1 (**section 1.2.2**), the fact that modals can express different flavors could in principle make it easier to figure out force: if children expect a modal like *must* always to express the same force, having figured out that it expresses *deontic* necessity might allow them to conclude by extension that it also expresses

necessity when used for another flavor. What the acquisition literature on flavor shows is that there is an asymmetry in children's production—and maybe comprehension—of root and epistemic modals: they use root modals productively by age 2, but epistemics are much rarer. The fact that children don't produce epistemic modals doesn't show that they don't understand them,[44] but it suggests that young children may not have a robust mastery of epistemic modals, and may not systematically realize when a modal like *must* is used to express epistemic flavor. This may have some repercussions on what we can conclude from comprehension experiments on children's understanding of force.

3.1.2 The acquisition of force

Most of what we know about children's mastery of force comes from behavioral comprehension experiments, which typically target older children, from age 4. These studies are generally embedded in other research questions, in particular, work about children's acquisition of scalar implicatures (the inference from "it might p" that 'it doesn't have to p') or focusing on their understanding of the difference between the bare and the modal statement ("it might/must be" vs "it is"). A number of other experiments, which I will also review but more briefly, focus on children's acquisition of scope interpretations of sentences that contain modals and other logical operators, in particular negation or disjunction.

[44] here I don't review the behavioral experiments that focus on older children's comprehension of flavor, but see (Coates, 1988; Heizmann, 2006; Cournane & Pérez-Leroux, 2020).

Initial studies (e.g., Noveck, 2001) stemmed from work on scalar implicatures: modals were used to see whether children could compute scalar implicatures (see Noveck, 2018; Horowitz et al., 2018 for recent overviews). Typically, these experiments focus on epistemic modals, and involve felicity judgment tasks, where children have to judge whether a possibility or a necessity modal statement is appropriate or not (e.g. "The cow **might/must** be in the blue box"), in scenarios where the speaker is more or less certain about the location of a toy (see e.g. Byrnes & Duff, 1989; Noveck, 2001; Bascelli & Barbieri, 2002; Öztürk & Papafragou, 2015; Moscati et al., 2017, a.o.).[45] Results of these experiments show that children struggle with both forces: they don't behave in an adult-like way, neither for possibility nor for necessity modal sentences.

First, children tend to over-accept possibility modals when necessity modals are more appropriate. For instance, they over-accept sentences such as "The cow **might** be in the blue box" when the speaker can be certain that it is (the blue box is the only option) (e.g. Noveck, 2001; Öztürk & Papafragou, 2015). This first result has been discussed in the context of children's (more general) issues with scalar implicatures, also found with other scalar terms—for instance, children also over-accept under-informative sentences such as "**Some** of the horses jumped over the fence" when the stronger alternative sentence ("**All** of the horses jumped over the fence") is true (e.g., Papafragou & Musolino, 2003; Noveck, 2001). It has been shown that this non adult-

[45] The reason why experiments tend to focus on epistemic modality may be that a large part of this literature originates in work on logical inferences. Others have tested force in deontic (rules) root modality (e.g., Hirst & Weil, 1982; Bascelli & Barbieri, 2002), using contrasting statements (*must* vs. *may*), but not questioning directly whether children know modals' forces.

like behavior is, for the most part, due to issues in accessing the relevant scalar alternatives (necessary to make the implicature), and holding them in memory (Chierchia et al., 2001; Barner & Bachrach, 2010; Barner et al., 2011; Skordos & Papafragou, 2014, a.o.).[46] Importantly, children's performance improves (i.e. they accept under-informative statements less often when the contrast between alternatives is made salient by the experimental design, and when they are explicitly given the alternatives and have to choose between two statements (for modals, see in particular Hirst & Weil, 1982; Noveck et al., 1996).

The second result is more puzzling, and specific to modal scales. Children tend to accept necessity modals in possibility situations: for instance, they accept "The cow **has to** be in the blue box" when the speaker cannot be certain about the location of the cow (e.g. for English, see Öztürk & Papafragou 2015; Noveck, 2001; Leahy, 2021; for Dutch, Koring et al., 2018). But, while possibility modals are logically true in necessity situations, and rejected by adults for pragmatic reasons, necessity modals are not true in possibility situations, and always rejected by adults. And this is not reported with other scalar terms: children don't accept "All dogs have spots", even if it is true that some of them do (see e.g. Noveck, 2001).

[46] Note that as discussed by Öztürk and Papafragou (2015), children's behavior with possibility modal sentences could come from the polite uses of modals we discussed in Chapter 1 (**section 1.2.5**) ("A possible explanation for this pattern comes from the fact that weak epistemic modals are sometimes used when the speaker believes a stronger statement to be true but wants to hedge or be polite (Brown & Levinson, 1987, among others). For instance, even if a speaker has definitive evidence that it is raining outside, he/she might say to a stranger, "It may be raining so you may need to bring an umbrella." The role of politeness might make it more difficult for children (and, occasionally, adults) to compute scalar implicatures for modals compared with other expressions"; Öztürk and Papafragou, 2015).

Why do children over-accept necessity modals in possibility situations? A first explanation, proposed by Öztürk and Papafragou (2015) to explain their finding (47% acceptance of *have to* statements in possibility situations) is that this reflects conceptual difficulties, namely, trouble dealing with open possibilities (see also Moscati et al., 2017). The developmental cognitive literature shows that young children have issues with situations that involve alternative possible outcomes, and tend to commit to a conclusion before evidence is available and arbitrarily select one possibility over the other. This tendency is known as *Premature Closure* (see Acredolo & Horobin, 1987; Bindra et al., 1980; Piéraut-Le Bonniec, 1980, a.o.). When asked to judge necessity modal statements in possibility situations, children would perfectly understand the modal statement, but they would have trouble considering simultaneously various possibilities: they would toss one out at random, and as a consequence judge the statement as true 50% of the time, when tossing out the unmentioned location.

Leahy & Carey (2020) propose that infants and toddlers have difficulties *representing* possibilities; they start with only 'minimal representations' of possibility and therefore struggle when having to consider multiple possibilities simultaneously (for critics, see e.g. Cesana-Arlotti et al., 2018). Leahy & Carey (2020) focus on younger children, and want to identify developmental milestones in reasoning ability. They propose that under 4, children cheat on modal reasoning: it is 'minimal', i.e., similar to probabilistic guessing among options, but not 'modal' (actually holding two

incompatible possibilities in mind simultaneously); however, after 4, children would be able to do the modal reasoning.[47]

However, another possible explanation for children's non adult-like behavior with necessity modals is that they would have issues with their meaning, in particular, they would fail to interpret correctly the flavor intended by the adult. Initial studies tend to focus on epistemic modals. In these experiments testing epistemic modality, children may over-accept necessity modals because they assume a deontic interpretation: they would interpret "The cow **has to** be in the blue box" as 'the rules are such that the cow has to be in the blue box', instead of 'it is certain that the cow is in the blue box' intended by the experimenter. Then, they might accept the sentence, if they think the deontic necessity statement is true, or if they don't know the rules and are assuming that the speaker is speaking truthfully (Crain & Thornton, 1998). And as we saw in the preceding section, young children may have a less robust mastery of epistemic modals than roots, the latter representing 90% of their modal input between 2- and 3-year-old.

Even if conceptual difficulties or difficulties with flavor may partly explain children's behavior, recent work suggests that this still cannot be the whole story. Children's tendency to accept necessity modals like *must* or *have to* in possibility situations was replicated with teleological root modals (goal-oriented) (Cournane, Repetti-Ludlow, Dieuleveut, & Hacquard, in prep.): 3- to 4-year-olds also tend to

[47] *Premature closure* concerns older children, above age 4, and works on the assumption that these children are able to entertain more that one possibility at once (outcome for future, or current epistemic possibility), but not very well (e.g., not as many as adults, not as easily and/or not for that long): difficulties are in *maintaining* the different options. The question addressed by Leahy and Carey is whether (younger) children can, or not, deal with possibilities.

accept “Cat **has to** go down the yellow road” in intended possibility situations, where two roads are open. This result may still in principle be explained by a problem with flavor: children would interpret *have to* as deontic (rule-based) instead of teleological (goal-based). But the justifications provided by some children (when asked why they accepted *have to* in possibility scenarios) seem to suggest that they understand that there are two open possibilities, and might suggest that they treat *have to* as expressing possibility, rather than interpret it as deontic (e.g., one child said “because they're both the same and they're both not blocked so we don't know which one”). Last, the acceptance rate reported by Cournane et al. for root *have to* in possibility situations is significantly higher (87%) than the one reported for (5-year-old) children tested in epistemic modality, who are at chance (47% of acceptance, in both experiments by Öztürk and Papafragou 2015 and Noveck 2001): Cournane et al. find no evidence that children treat *can* and *have to* differently. This difference may come from the difference in that age group (they test younger children: mean age 4;1, vs. 5;1 for both Öztürk & Papafragou 2015 and Noveck 2001), but does not seem to come from the difference in the flavor they test.

Leahy (2021) shows that children aged 4-year-old also perform below chance when answering to *have to* epistemic statements. Leahy uses an experiment involving both a behavioral and a linguistic measure (in the behavioral task, children are asked to place a wagon to catch a marble, in different situations of necessity and possibility; in the linguistic task, they are asked a modal question of the form, “If I drop a marble in here, {can/will/does it have to} it come out here?” Regardless of the age group, children give adult-like responses for can, but not for have to: they tend to answer ‘yes’ to *have*

to questions when tested in situations of possibility (4-year-old: app. 65%; 5-year-old: app. 47%).

Taken together, those results suggest that there must be some other sources than *Premature closure* or difficulties with flavor for children's difficulties. This raises the question of whether young children even know the underlying force of necessity modals. Children's behavior could also stem from not knowing that necessity modals like *must* or *have to* encode necessity: such difficulties with force would capture both their over-acceptance of necessity modals in possibility situations, and their tendency to accept their possibility counterparts in necessity situations: if they don't know that there is a stronger modal expressing necessity, they cannot access the stronger alternative necessary to make the scalar implicature (i.e., they don't know the scale). This may also explain the differences between age groups: by 5, children may have learned that *have to* is a necessity modal, and be more likely to correctly reject it.

Experiments testing children's productions are rare. One exception is an elicited production study by Hirzel et al. (in prep.), which shows, using a sentence repair task, that 3- to 4-year-old children have a general preference for using possibility modals such as *can*, even in (intended) necessity scenarios. They also report that the few children who use necessity modals tend to use them both in necessity and possibility scenarios, instead of using them only in necessity scenarios. These results are compatible with a lexicalization problem, but this opens a last possibility to explain children's behavior: that children fail to use and comprehend necessity modals correctly because of difficulties quantifying over the right domain, as has been argued in the case of definite descriptions (see Abbott, 2008; Brockmann et al., 2018), and

deploying them in the right situation.[48] In comprehension experiments, children's over-acceptance of necessity modals in possibility contexts could be due to their assuming a different (smaller) domain of quantification: in the box scenarios, they might assume that the speaker has reasons to rule out certain possibilities and thus accept the necessity statement.[49]

Table 3.1 summarizes these different possible explanations for children's difficulties with force, which I will discuss further in **section 3.4**. Note that these explanations are not mutually exclusive: they most likely combine to explain the full acquisition picture.

Table 3.1 Summary of possible explanations for children's difficulties with force reported in the existing literature. These explanations are not mutually exclusive.

Nature of the difficulties	**Explanation**
Conceptual difficulties with ***Premature Closure*** (4-year-old) (Öztürk and Papafragou, 2015)	When having to deal with several open possibilities, children arbitrarily toss one out, in order to reduce cognitive load. Specific to epistemic modality?
Conceptual difficulties with **modal reasoning** (infants) (Leahy and Carey, 2020)	Infants have difficulties representing possibilities: they start with only 'minimal representations' of possibility, and therefore struggle when having to consider multiple possibilities simultaneously. Difficulties are not specific to necessity modals.

[48] In the case of the definite article, for instance for sentences such as "Put **the** doll in the suitcase", a unique referent is required in the context (there should only be one doll; if there were two dolls, the indefinite article should be used). Children seem to be over-permissive of using and accepting definites in contexts where their uniqueness presupposition is not satisfied (Karmiloff-Smith, 1979; Schaeffer & Matthewson, 2005; Van Hout et al., 2010, a.o.)

[49] Note that outside experimental settings, adults would probably accommodate those kind of necessity statements: for instance, even when being at first certain that the speaker has no reliable source of information whatsoever, they might accept "It must be raining" and understand that the speaker probably in fact has a reliable source of information, since the necessity modal statement *conveys* that the speaker is certain.

Semantic difficulties with modals' **flavor**	Children fail to interpret the flavor intended by the adult: for instance, they would interpret "The cow **has to** be in the blue box" as 'the rules require that the cow be in the blue box', instead of 'it is certain that the cow is in the blue box'.
Semantic difficulties with the **force** of necessity modals	Children don't know the underlying force of necessity modals. Weak version: Children are unsure about the force. Strong version: Children think that necessity modals like *have to* mean 'possible'.
Pragmatic difficulties (determining the domain of quantification)	Children have issues deploying modals in the right situations (in the same situations as adults do); children and adults differ in how they interpret possibility/necessity situations.
Epistemic difficulties (knowing what is necessary or possible)	Children don't know what is possible and necessary (for instance, they might not know the rules)

3.1.3 The acquisition of modals and negation

A number of experimental studies focus on children's interpretation of sentences containing negated modals. Here again, these studies usually take for granted that children already know the underlying force of their modals, and that they have no deep conceptual or pragmatic issues, to focus on what they know of their scope relative to negation (i.e., whether children choose strong inverse scope interpretations rather than weak surface scope interpretations). I will focus on the literature on children's acquisition of negated modal sentences, and leave aside the literature on Free Choice (how modals interact with disjunction, e.g., "You may have sushi or pasta"), as well as the literature on Polar Sensitive Items (which asks, for example, how children learn the licensing restrictions on NPIs).[50]

[50] For work on Free Choice: see e.g., Tieu et al., 2016; Jasbi, 2018; for work on the acquisition of modal Polar Sensitive Items: see e.g. van der Wal, 1996; Lin et al., 2015, 2018). Note that the English NPI *needn't* is extremely rare in English (5 occurrences in the whole Manchester corpus, 0 by children), which is one of the reasons why I don't dive further into this issue.

As we saw in **section 1.2.3** of Chapter 1, possibility and necessity modals vary with respect to whether they scope over or under negation. For instance, necessity modals like *must* and *should* take wide scope with respect to negation; necessity modals like *have to* and *need to* take narrow scope. This means that depending on the force of the modal (possibility vs. necessity) and its scope with respect to negation (over vs. below), we have four possible interpretations. But given the logical equivalence between *not* > *possible* and *necessary* > *not*, we need to distinguish only two: a necessity modal scoping over negation (e.g., "you **mustn't** p") is logically equivalent to a possibility modal scoping under negation (e.g., "you **can't** p"), and a possibility modal scoping over negation (e.g., "you **might not** p") is logically equivalent to a necessity modal scoping under negation (e.g., "you **don't have** to p/you **needn't** p"). We have therefore two logical interpretations, a 'strong' one (*necessary* > *not* or *not* > *possible*), and a 'weak' one (*not* > *necessary* or *possible* > *not*).

Results of experiments testing children's understanding of negated modal sentences show that children systematically tend to prefer strong interpretations (*not* > *possible/necessary* > *not*), even when adults prefer weak ones (*possible not*/*not necessary*). For instance, children tend to reject "There **might not** be a horse in the box" (*possible* > *not*) when it is possible that there is no horse in the box (see e.g., Moscati & Crain, 2014, for Italian *potere* 'can'), or, when tested with teleological modality, they reject "Cat **doesn't have to go** down the yellow road" (*not* > *necessary*) when both ways are open (Cournane et al, in prep., Experiment 2) (see also Noveck, 2001 for *not have to*; Moscati & Gualmini, 2007 for epistemic *cannot* and Italian *non dovere* 'mustn't', Gualmini & Moscati, 2009 for Italian *non dovere*; Koring et al., 2018

for Dutch *niet hoeven* 'needn't'). The main proposal for this preference for strong interpretations is the Semantic Subset Principle (SSP), a variant of the Subset Principle (see Chapter 1, **section 1.3.1**). According to this principle, children acquire scope ambiguities by first assuming stronger meanings, regardless of their availability in the adult grammar (Crain et al., 1994; Crain & Thornton, 1998; Crain, 2012; Moscati et al., 2016; for critiques, see Musolino, 2006; Musolino et al., 2019).

However, it is not straightforward what to conclude as to the nature of children's difficulties just based on these experiments' results. First, negated necessity epistemic claims (e.g. "It might not be in drawer") are extremely unnatural, even for adults; as we saw in Chapter 2, when looking at children's input, we find that only 4.6% of epistemic modal statements are negated, and only 2.6% of *necessity* epistemic modal statements (i.e., 6 examples in the whole corpus). Root negated necessity modal statements are also quite rare (e.g., "You don't have to do it"), and often used in 'polite ways', rather than to clearly convey 'non necessity.' Second, here too, these findings could be explained by children being uncertain about the underlying force of modals. For instance, if children think *have to* expresses possibility, they might accept it in the same situations as *can't*, as they will understand *not have to* as meaning 'not possible'. Or if they think *might* expresses necessity, they might accept it in the same situations as *mustn't*, as they will understand *might not* as meaning 'necessary not'.

Corpus studies assess children's production of modals with negation are rare.[51] I'll come back to these debates, and the consequences they may have for learners figuring out force, in the discussion.

3.1.4 Summary and motivation for the study

Our current understanding of children's early mastery of modal force is limited. Comprehension studies on force tend to focus on older children. Corpus studies tend to focus on flavor, and while they note when particular lexemes first appear in children's spontaneous speech, to date, there hasn't so far been any systematic corpus investigation of force in English.

What can we learn from looking at very young children's spontaneous productions about when and how children master the force of modals? The study I turn to provides the first large-scale investigation of the development of modal force, by examining the modal production of twelve children between the ages of 2 and 3.

3.2 Corpus study: Modal force in child productions

The purpose of this corpus study is to provide a thorough description of the modals children produce between 2- and 3-year-old, focusing on the force dimension. What do young children's spontaneous productions of possibility and necessity modals look

[51] See Jeretič, 2018, for work on the acquisition of (root) modal force in French and Spanish and their interaction with negation.

like, as compared to their parents'? How frequently do they use possibility and necessity modals? And do they use them in an adult-like way?

To study children's modal productions, we used the same methods as in the input study (Chapter 2). Similarly, I first present the results from our corpus analysis, now comparing children's early productions to those of adults': how frequently children use possibility and necessity modals, in which environments, and how they interact with flavor and negation. Then, I present results from an experiment based on the same paradigm as the *Input Experiment 1* (see **section 2.3**, Chapter 2), that tests how adult-like children's uses of modals are, by asking (adult) participants to guess the force of modals uttered by children in dialogues extracted from the corpus, given the conversational context in which they use it.

3.2.1. Methods

As in our study of the input in Chapter 2, we use the Manchester Corpus (Theakston et al., 2001) of UK English (CHILDES database, MacWhinney, 2000). It consists of 12 child-mother pairs (6 females; age range: 1;09-3;00), recorded in unstructured play sessions. Our analysis focuses on the period between ages 2;00 and 3;00. Children's utterances containing modal auxiliaries and semi-auxiliaries (5,842; excluding repetitions (17%): 4,844) were coded for force (possibility vs. necessity) (26), negation (present vs. absent) (27), flavor (epistemic vs. root) (28), subject (first/second/third person) (29) and clause type (declarative/interrogative/tag question) (30). We applied the exact same coding scheme for children and adults; here, I present it again for ease of reference. Note that we do not include *will, would, shall* and *going to* as they

primarily express futurity, which force is debatable (Stalnaker, 1968; Cariani & Santorio, 2018). We do not differentiate amongst various subtypes of root flavors (e.g. ability, teleological, deontic).

(26) Modal lemmas by force:

Possibility: *can, could, might, may; able to*
Necessity: *must, should, need; have to, got to, be supposed to, need to*

(27) Negation:

No negation: 'I **can** play with you.'
Negation: 'You **can't** eat it.'

(28) Flavor:

Root:

CHILD:	I got crane out my box.	
MOTHER:	oh you've got your box as well yeah.	
CHILD:	I **must** get crane.	(Aran, 2;02)

Epistemic:

MOTHER:	oh we've got a bit of hair stuck, haven't we?	
CHILD:	look.	
CHILD:	it **must** be some of dolly's hair.	(Aran, 2;09)

(29) Subject:

1st person: 'I **can** see a bucket.'/ 'We **can** fit a cow through there.'
2nd person: 'You **can** do it.
3rd person: 'He **can** go in the cart.' / 'The cat **can** go in the house.'

(30) Clause type:

Declarative clause: 'I **can** see a bucket.'
Interrogative clause: 'What **can** you see?'; '**Can** you see any chickens?'
Tag question (excluded): 'You can wash it later, **can't** you?'

3.2.2 Results

Overall, modal utterances represent 1.6% of all children's utterances (vs. 5.8% for adults). Like adults, children produce more possibility modals than necessity modals, and the asymmetry is even stronger: possibility modals represent 79.3% of their modal productions (vs. 72.5% of adults' modal productions) (**Table 3.2**). As for adults, *can* is by far their most common modal (75.6% of their modal productions, vs. adults: 57.3%), and *have to* their most frequent necessity modal (7.3% of their modal productions, vs. adults: 12.0%).

Negation. Negated possibility modals (e.g. *can't*) are extremely frequent in children's productions: more than half of their possibility modal utterances are negated (vs. adults' negated possibility modals: 20.9%). Conversely, necessity modals are particularly rare with negation in children's productions: only 5.1% (vs. adults: 10.1%). Looking at the evolution of their productions between 2- and 3-year-old, we see that children tend to use necessity modals more frequently over time (relatively to possibility modals): as summarized in **Figure 3.1a**, while necessity modals represent only 12% of children's modal productions between 2;0 and 2;3-year-old, they represent 24.5% between 2;9 and 3-year-old, almost the same as in adults productions. **Figure 3.1b** confirms that for adults, the relative proportion of possibility and necessity modals does not significantly change over time: we only find a slight increase of necessity modals.

Table 3.2 Counts and percentages of modal uses by force, ordered by lemma frequency, with and without negation, for children (repetitions excluded:[52] 17.0%) (*X2 (1, N=24830)=92.6, p < 2.2e-16*).[53]

	CHILD (n=4844)		CHILD (n=4800)[54]			
	all		no negation		negation	
POSSIBILITY	**3841**	**79.3%**	**1861**	**49.0%**	**1937**	**51.0%**
can	3663	75.6%	1739	48.0%	1881	51.9%
might	86	1.8%	78	97.5%	2	2.5%
could	80	1.6%	34	39.5%	52	60.4%
able	3	0.1%	1	33.3%	2	66.6%
may	9	0.2%	9	100%	0	0%
NECESSITY	**1003**	**20.7%**	**950**	**94.8%**	**52**	**5.2%**
have to	352	7.3%	345	98.0%	7	1.9%
got to	288	5.9%	283	98.3%	5	1.7%
should	22	0.5%	17	80.9%	4	19.0%
need to	217	4.5%	204	94.0%	13	5.9%
must	114	2.4%	94	82.5%	20	17.5%
supposed	9	0.2%	6	66.7%	3	33.3%
ought to	1	0.0%	1	100%	0	0%

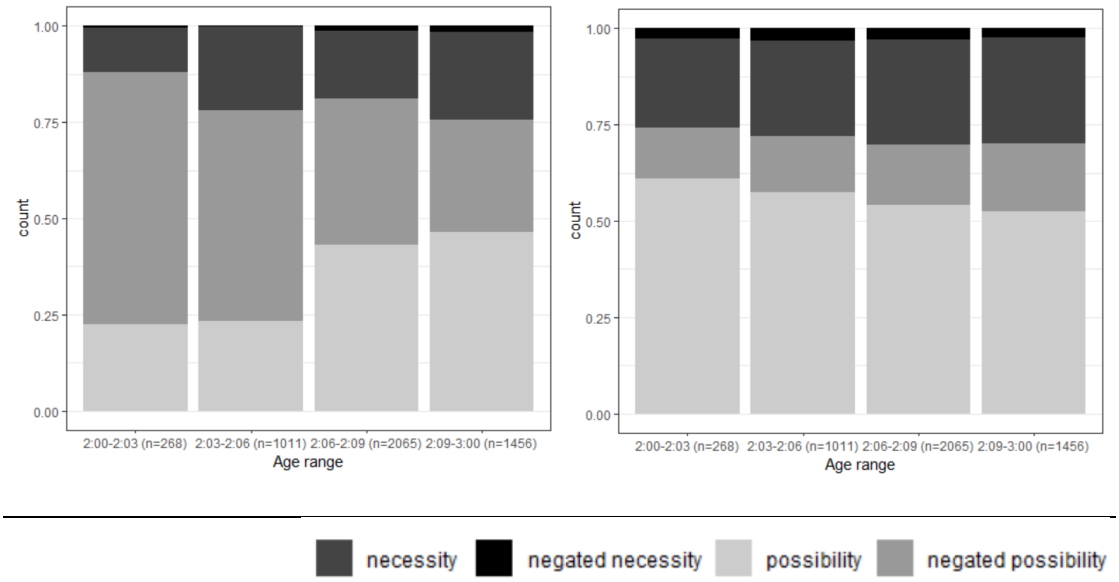
[52] Were considered as repetitions cases where the speaker repeated a sentence uttered right before by herself or by another speaker with no significant change.

[53] Note that the chi-square assumption of independence of observations is violated by corpus samples, as the same speaker supplies multiple uses per cell. However, this test metric is commonly used in corpus linguistics for simple distributional comparisons, and is not straightforwardly a violation as we are comparing spontaneous utterances, not individuals (each spontaneous production is taken as a proxy for independence).

[54] Excluding tag questions and repetitions. Tag questions (e.g. 'you can wash it later, can't you?') are more frequent in adults' speech (0.9% of all children's modal productions; 5.7% of adults'). We exclude modals in the tags from the analyses, as they do not directly matter for our purposes.

Figure 3.1 Evolution of children's modal productions from 2- to 3-year-old by force and negation, binned in 3-months periods:

a. Child productions (n=4,800) **b.** Adult productions (n=18,853)



Flavor. As reported earlier in the literature (see in particular, van Dooren et al., 2017 for results on the Manchester corpus), epistemic uses are overall very rare in child productions: they represent only 2.4% of all their modal utterances (114 cases, possibility: 93, necessity: 21) (vs. adults: 8.8%).

Sentence type. Most of children modals occur in declarative sentences, with first person subjects; they rarely use modals in interrogative sentences. **Figures 3.2a** and **3.2b** summarize the overall distribution of possibility and necessity modals in children's speech, with and without negation, per sentence type and subject, for root (2a) and epistemic (2b) modals.

Figure 3.2a Distribution of possibility and necessity modals with and without negation, by sentence type and subject for children's root modals (n=4686):

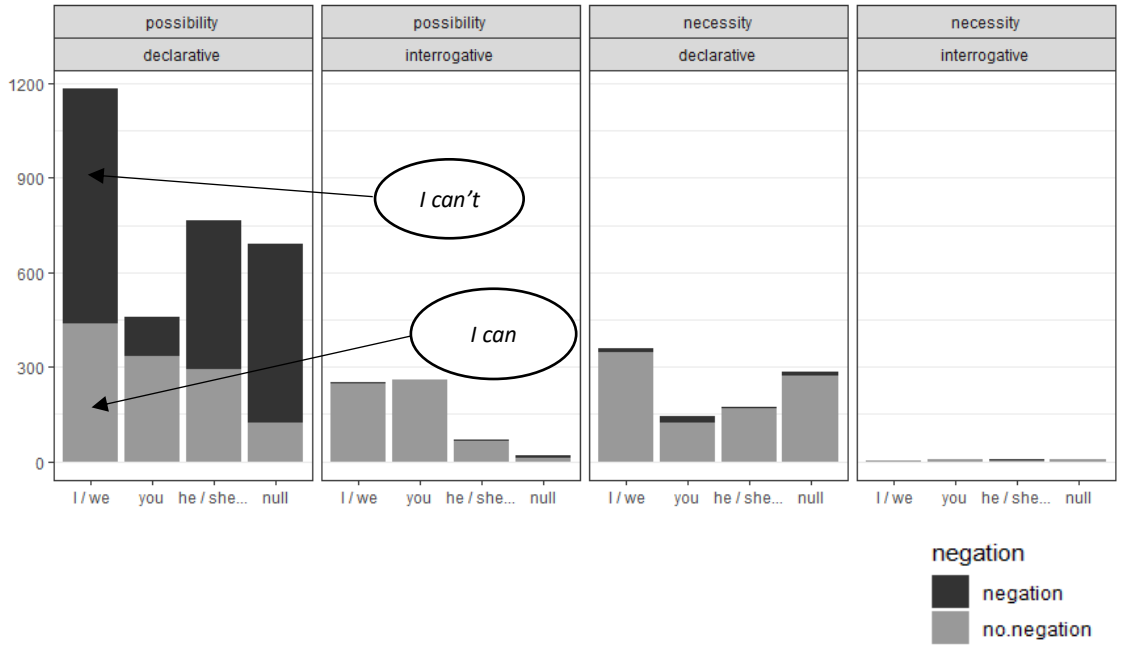
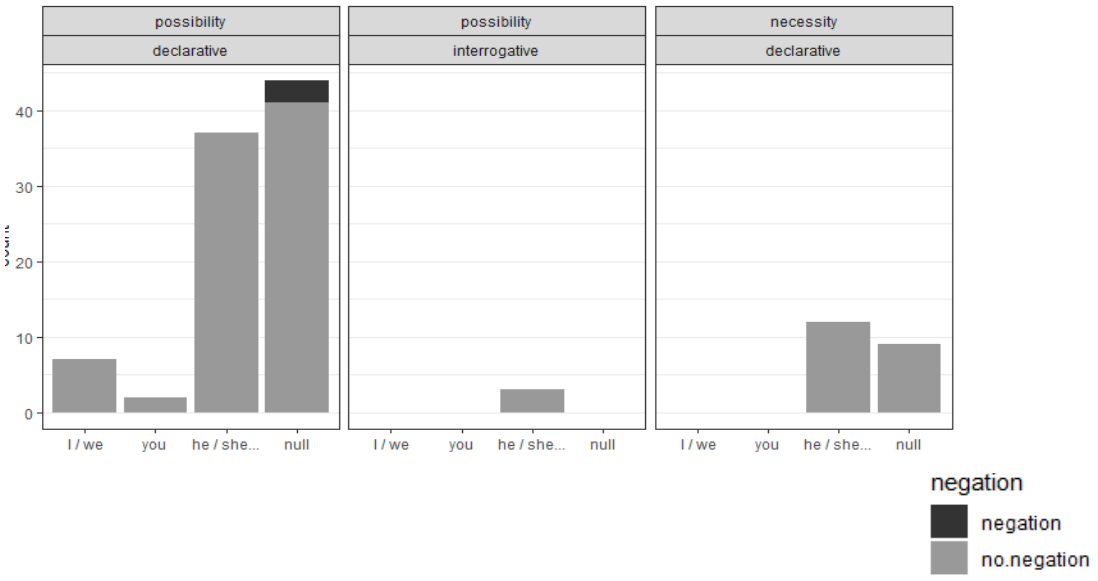


Figure 3.2b Distribution of possibility and necessity modals with and without negation, by sentence type and subject for children's epistemic modals (n=114):[55]



[55] There are no productions of interrogative necessity epistemic modals.

3.2.3 Discussion

Our results show that children use (root) possibility modals frequently, both with and without negation, which we can take as initial evidence of productivity (Stromswold, 1990). They use them both in declarative and in interrogative sentences. However, they use fewer necessity modals, rarely with negation, and almost never in interrogative sentences. Necessity modals also tend to be produced later on: while all 12 children in the corpus already produce possibility modals by age 2 (*can/can't*), 6 of them don't produce any necessity modal before 2:03 year of age. We will come back to this point in Chapter 4, when looking at individual variation between children.

The lower frequency of necessity modals might come from a combination of factors. First, recall that it is also found in parent's speech. Even though the asymmetry is more pronounced when looking at children's productions, adults as well use necessity modals relatively less frequently than possibility modals. If children tend to acquire more frequent words first, this might explain why they use them earlier on, and more frequently, at a younger age. Second, the difference between children and adults might come from social differences in status and topics of conversations: children may be less prone to giving orders than adults (therefore, less prone to using deontic necessity modals), or less in a position to express certainty (therefore, less prone to using epistemic necessity modals) (Hickmann & Bassano, 2016).

As discussed earlier, the link between production and comprehension is not straightforward: children may prefer alternative strategies, such as using imperatives to express orders, or bare sentences to express epistemic certainty. These production data only provide a partial picture of whether these very young children understand and

produce modals correctly. To assess whether they use them in an adult-like way, we ran an experiment on children's modals, using the *Guess the Force* (GF) paradigm.

3.3 *Guess The Force* Experiment 1: Children's modal productions[56]

The goal of this experiment is to investigate children's early modal productions to see whether they use modals in an adult-like way, for different flavors, with and without negation. Can adults guess the force of a modal used by a child, given the conversational context in which they use it, the way they're able to when the modals are used by adults?

The paradigm is identical to the one presented in **section 2.2** of Chapter 2 (*HSP Input Experiment 1*), used with parents' productions. Here, we use the experiment on adult productions as a control, which shows that force is in principle guessable from conversational context, for both possibility and necessity modals.

3.3.1 Methods

Procedure. *GF Experiment 1* is identical to *HSP Input Experiment 1*, except that we tested children's utterances instead of adults', and made small changes in the instructions (see Appendix B).[57] The experiment was run online on Alex Drummond's IBEX Farm. Adult participants recruited via Amazon MechanicalTurk were asked to

[56] In Chapter 4, I'll present results from a second GF experiment (*GF Experiment 2*), which uses the same paradigm as here (ROOT-P2 condition: *can* vs. *have to*), but focuses on differences between children of the Manchester corpus.

[57] An example of the experiment can be accessed below (ROOT-P2 condition): https://spellout.net/ibexexps/modsquad/HSP_FC_dilch_rootP2/experiment.html

guess a redacted modal in a dialogue between a child and mother by choosing between two options, corresponding either to a possibility (e.g. *can*) or a necessity modal (e.g. *have to*). An example of the display is given in **Figure 3.3**. We implemented the same conditions: ROOT-AFF-1; ROOT-AFF-2; ROOT-NEG; EPI-AFF, as summarized in **Table 3.3** 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. Controls were based on tense (past vs. future).

Table 3.3 Summary of experimental conditions (GF Experiment 1):

Test condition (between participants)		Modal lemmas	
		possibility	necessity
EPI-AFF (epistemic affirmative)		*might*	*must*
ROOT-AFF (root affirmative)[58]	**ROOT-AFF-1**	*can*	*must*
	ROOT-AFF-2	*can/able*	*have to*
ROOT-NEG (root negative)		*can't/not able*	*not have to*

Material. ***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 conditions (see **Table 3.2**).[59] This did not make a difference for the participants, who always had 10

[58] 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, while **ROOT-AFF-2** (*can/able to* vs. *have to*) allowed us to avoid concerns related to the formality of *must* for US English speakers. As in Experiment 1, in cases where *have to* was tensed, we used *able to* as the alternative to avoid losing tense information: for example, participants had to choose between [*will have to*] and [*will be able to*]. We extracted the same number of contexts from *able to*, to avoid having the *able to* option always be the wrong answer. Same principles applied for **ROOT-NEG** condition: participants had to choose between [*didn't have to*] and [*wasn't able to*] when *have to* was tensed.

[59] Because some of the negated *have to* in child productions were particularly unclear (e.g. 'I *can't have to* read it.'), we also used *not gotta* and *not need.*

contexts to judge per condition (40 dialogues in the whole experiment: 20 trials: 10 possibility, 10 necessity; 20 controls: 10 past, 10 future). In all other conditions, the 10 contexts were selected randomly out of a list of 20 contexts initially extracted from the corpus, in the same way as for the adult experiment. ***Exclusion criteria***. Given the low frequency of negated necessity and epistemic modals, we didn't remove cases where the modal already appeared in the preceding dialogue.[60] We made sure to include examples in the training (*the/a*) and control items (past/future) where it was also the case that the right (or wrong) answer appeared in the preceding dialogue. Briticisms, such as *willn't*, were removed from the dialogue and replaced with American English equivalent (e.g. *won't*), but we did not correct children's ungrammatical utterances (e.g. *comed* for *came*), except in the case of *have to* when children omitted *to* (so participants would not reject the answer because of its ungrammaticality). We didn't exclude contexts where there were tag questions (e.g., *'..., mustn't she?'*), but removed the tags when they occurred in the target sentence. ***Controls.*** Participants had to choose between future and past (*e.g.* [*saw*] vs. [*will see*], see **Figure 3.4**). ***Rationale***. We make the assumption that adults rely on their own competence to judge usage, and that the dialogues preceding the modal sentence are equally informative for adults' and children's utterances. If children use their modals in an adult-like way, we expect no difference in accuracy between the experiment on children's productions and the experiment on adults' productions. If they do not (e.g. they use *can* in a necessity

[60] Out of 148 contexts, 36 of them had the modal appear in the preceding dialogue (24.3%) (uttered by the child: 11, by the mother or another adult: 20; by both: 5).

situation, when adults would use *must*, or they use *must* in a possibility situation, when adults would use *can*), we expect a lower accuracy for children's utterances. Note that any observed differences could be explained either by issues with force or by issues with the pragmatics of using modals with those forces.

Figure 3.3 GF Experiment 1 stimuli (child productions): example trials for ROOT-P2 (*have to* vs. *can*):

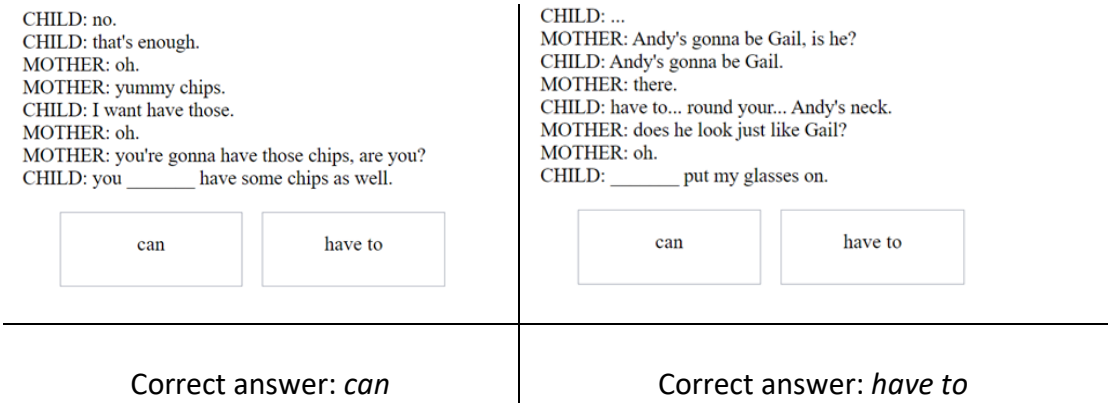
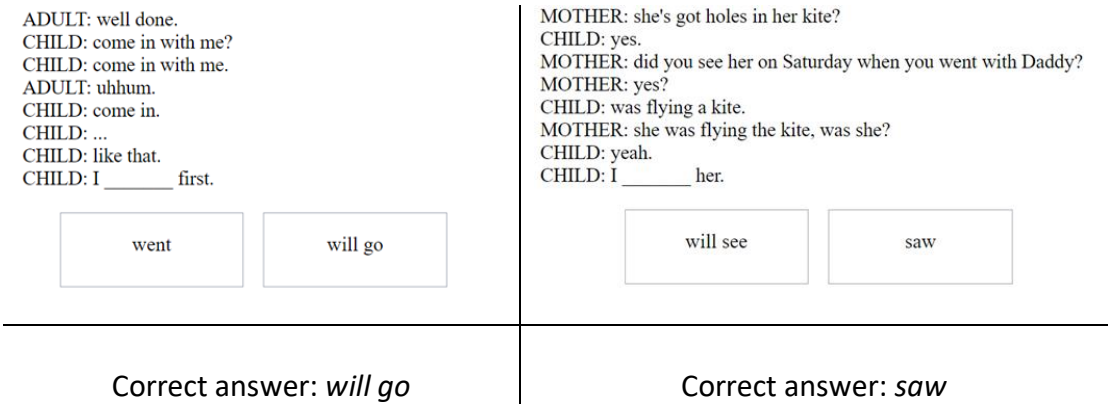


Figure 3.4 GF Experiment 1 stimuli (child productions): example control items (*past* vs. *future*):



3.3.2 Results

Participants. 291 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-year-old). We removed 18 participants (6.2%) who were less than 75% accurate on controls.[61] We thus present results for 273 participants (EPI-AFF: 68; ROOT-AFF-1: 70; ROOT-AFF-2: 70; ROOT-NEG: 65).

Analysis. Data analyses were conducted using R (R Core Team, 2013), using the package lme4 (Bates et al. 2014a, 2014b). We first run the same tests as for the adult experiment, then we compare between adult and child usage. **Table 3.4** reports mean accuracy in each condition (summarized in **Figure 3.5**). The overall accuracy on possibility modals is 82.1%; on necessity modals, 50.1%. We first run binomial tests to see whether performance differs from chance for each condition. Participants performed better than chance in all conditions involving possibility modals, but not for necessity modals. For ROOT-AFF-1 (*must*) (mean accuracy: 42.6%) and ROOT-NEG necessity (*not have to*) (mean accuracy: 32.3%), they performed lower than chance (**Table 3.4**).

Force. 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

[61] For the adult version, the proportion of errors on controls was very low (5.4%), with no difference between groups. For the child version however, the initial proportion of errors on controls was quite high (21.6%): post-hoc analysis revealed that this came from 5 control contexts for which the accuracy was particularly low, thus not reliable controls. We decided to remove these 5 controls from our exclusion criteria, as they were particularly difficult, and probably do not indicate that subjects were not doing the task correctly. After restricting to the 15 remaining controls, mean accuracy on controls was 90.0%, showing that participants were not answering randomly.

Subject and Item as random factors). We find an effect in all conditions, always with higher accuracy for possibility modals (all: $\chi^2(1)=20.4,\ p=5e\text{-}6$***; ROOT-AFF-1: $\chi^2(1)=60.4\ \ p=7.7e\text{-}15$***; 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: $\chi^2(1)=7.93\ p=0.0048$**).

Negation. We compare ROOT-AFF-2 and ROOT-NEG, as these conditions included the same lemmas. We find an effect for both possibility and necessity conditions: higher accuracy with negation for possibility modals, and lower accuracy with negation for necessity modals (*can* vs. *can't*: $\chi^2(1)=3.7,\ p=0.056$ *; *have to* vs. *not-have to*: $\chi^2(1)=6.7,\ p=0.0093$**; Interaction Force*Neg: $\chi^2(1)=9.2,\ p=0.0024$**).

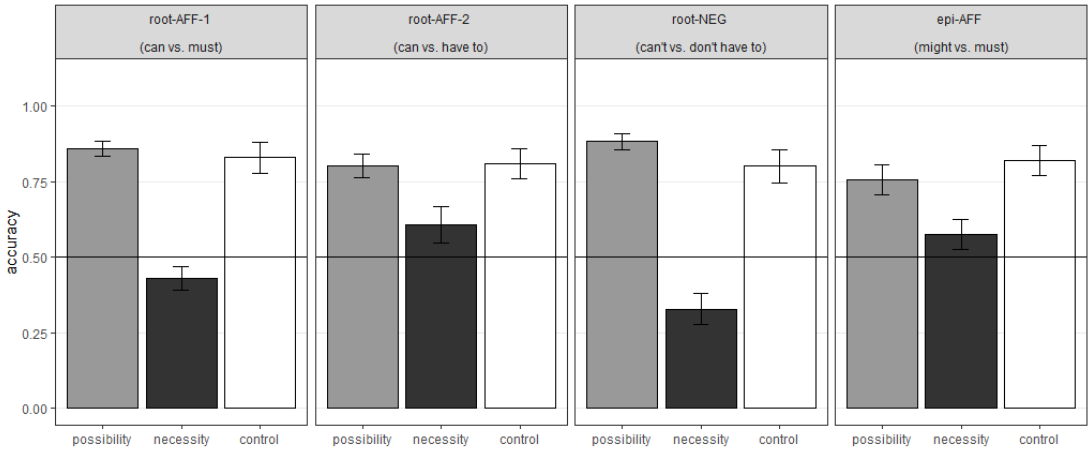
Flavor. There was no effect of flavor ($\chi^2(1)=0.14,\ p=0.71$).

Table 3.4 Accuracy rates and significance tests by condition (GF Experiment 1: children's productions) (n = 273, 10 observations per cell):

	Mean accuracy[62] (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=5.1e-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=2.0e-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=0.00019 *** 95% CI [0.53, 0.61]
Total	82.1% (0.019)	50.1% (0.028)		
ALL	67.4% (0.021)			

[62] 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).

Figure 3.5 Accuracy by condition, GF Experiment 1, children's productions (n=273):

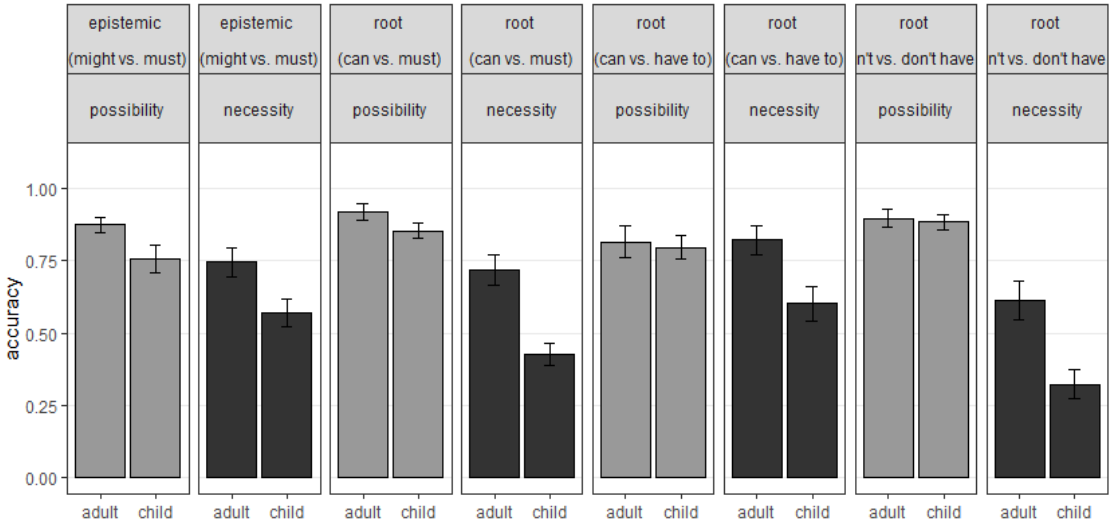


Age (adult vs. child productions). We then compared results for children's production to results on adults' usage (as reported in Chapter 2), which showed that force is guessable from context (we assume that conversational context should be equally informative in both cases). **Figure 3.4** summarizes the comparison. We find a general effect of Age, with lower accuracy for child usage ($\chi^2(1)=260.5$, $p <.001$***) (general mean accuracy for children's productions: 67.4%; vs. for adults' productions: 79.9%). Among possibility conditions, we find a significant difference only for ROOT-AFF-2; among necessity conditions, all comparisons are significant, except EPI-AFF (**Table 3.5**). 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 ($\chi^2(1)=32.1$, $p=1.45e-08$***).

Table 3.5 Results of the model testing the effect of Age (adult usage vs. child usage):

	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=4.4e-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 ***	

Figure 3.6 Accuracy by condition, comparison between GF Experiment 1 (children's productions) and Input Experiment 1 (adults' productions):



3.3.3 Discussion

Even if participants are overall slightly less accurate than when judging adults' modals, they are good at identifying possibility modals used by children, for both flavors (mean accuracy on all possibility modal contexts: 82.1%, vs. when guessing adults' modals: 87.5%; for root: affirmative: 82.4% (vs. 86.9%); negative: 88.2% (vs 89.5%); for epistemic: 75.6 (vs. 87.2%)). In only one of the conditions involving possibility modals

is the difference between child and adult usage significant. However, for necessity modal contexts, participants' performance is much lower: only 50.1% (vs. when guessing adults' modals: 72.3%), especially for negated ones (32.3%, vs. 61.3%), with a significant difference between child and adult usage for the three root conditions. The necessity contexts given in (31) to (34) illustrate some of children's non adult-like uses. (They led to the lowest accuracy for each condition).

(31) ROOT-AFF-1:

CHILD: Daddy repaired it.
CHILD: it off.
MOTHER: Daddy repaired it but he'll have to do it again I think, won't he?
MOTHER: come on.
MOTHER: you come and show Mummy.
MOTHER: show Mummy the truck.
MOTHER: oh.
CHILD: I {can/**must**} get a tractor.

(Aran, 2;03.02; HSP accuracy: 15.8%)

(32) ROOT-AFF-2:

CHILD: oh.
CHILD: just here.
CHILD: ...
CHILD: what shall I put first?
CHILD: that.
CHILD: what's that?
MOTHER: pardon?
CHILD: I {can/**have to**} see a cat.

(Becky, 2;08.16; HSP accuracy: 2.9%)

(33) ROOT-NEG:

CHILD: I am stuck now.
CHILD: no... no.
CHILD: no... no.

CHILD: get the petrol in.
CHILD: get the petrol.
CHILD: petrol.
CHILD: get some in.
CHILD: the hippos {can't/**don't have to**} go in.

(Carl, 2;08.01; HSP accuracy: 9.2%)

(34) EPI-AFF:

FATHER: that's a digger, is it?
CHILD: yes.
FATHER: is there another digger?
CHILD: yes.
FATHER: where's the other digger?
CHILD: don't know.
FATHER: you don't know?
CHILD: {might/**must**} be upstairs.

(Domin, 2;04.25; HSP accuracy: 23.5%)

3.4 General discussion

In this chapter, we focused on the question of *when* children figure out the force of their modals, by studying their early spontaneous productions. I'll first summarize our results, to then see how they shed new light on the studies of children's mastery of force reviewed at the beginning, and discuss how we can explain children's difficulties.

The previous literature shows potential struggles with both forces. Results from this corpus study show that children master possibility modals like *can* or *might* very early. Already by age 2, they use them productively, with and without negation, and our GF experiment suggests that they use them in an adult-like way: participants were able to guess the right force, both for root and epistemic uses, and both for non negated and negated uses. Children, however, seem to struggle with necessity modals like *have*

to and *must*. They use them later on, less often, hardly ever with negation, and crucially, they don't use them in an adult-like way: our GF experiment shows that they produce them in situations where adults prefer possibility modals. Their difficulties with necessity modals are even more salient when they use them with negation: participants' performance on negated modals like *have to* is particularly low, which suggests that children use them to convey impossibility.

This casts a new light on the results from both comprehension and production experiments reviewed in **section 3.1.2**. If the difficulty with necessity modals we observe for 2- to 3-year-olds persists into the preschool years, this could also explain results from comprehension experiments: why, both when tested with root and epistemic modals, 4-year-old children tend to over-accept possibility modals in contexts where adults prefer necessity modals (since they cannot make scalar implicatures if they lack a stronger alternative), and necessity modals in possibility contexts (since they treat them as possibility modals). And recall that the few production experiments that use more controlled experimental settings to test (older) children's productions draw similar conclusions: 3- to 4-year-old seem to have a general preference for using possibility modals (even in necessity situations), and the ones who use necessity modals tend to use them both in necessity and possibility situations (Hirzel et al., in prep.).[63] This aligns with our findings on younger children's

[63] See Lin et al. (2018) for a study on how Dutch children acquire the polarity restrictions on the modal verb NPI *hoeven* 'need', using an elicited imitation task.

spontaneous productions: overall, 2- to 3-year-olds use possibility modals more frequently, and they tend to use necessity modals in possibility situations.

There seems to be a period during which children do not use (and maybe comprehend) necessity modals in the same way as adults, which I'll call the Necessity Gap.[64] However, this does not tell us the nature of children's difficulties. The reason for children's non adult-like behavior with necessity modals, found both in comprehension and production studies, is compatible with various possibilities.

Let's come back to the different explanations we considered at the beginning (summarized in **Table 3.1**). A first possibility is that children's difficulties are conceptual in nature, and have to do either with *Premature Closure* (young children would have difficulty reasoning in situations that involve alternative possible outcomes, and tend to commit to a conclusion before evidence is available; see Acredolo & Horobin, 1987; Ozturk & Papafragou, 2015), or representing possibilities (children would start with only 'minimal representations' of possibility, and therefore struggle when having to consider multiple possibilities simultaneously; Leahy & Carey, 2020).

A second possibility is that children's difficulties are semantic in nature, and arise from issues with the underlying force of necessity modals like *must* and *have to*. There are two versions of this hypothesis: the weak version is that they are *unsure* about

[64] The expression should be taken with caution, as it is not clear yet whether results about children's production generalizes to other languages than English. Moreover, as in the case of the 'Epistemic Gap', it might not be a real "gap", but just a natural consequence of a difference in frequency of exposure, given that at least in English, necessity modals are significantly less frequent than their possibility counterparts.

the underlying force of necessity modals; the strong version, that they are *mistaken* about their force, and maybe think that they mean possible.[65] Either version would capture results from both productions and comprehension experiments: children's tendency to produce necessity modals in situations where adults use possibility modals, and as already explained, their tendency to accept possibility modals in necessity situations, as they cannot reason that there is a more appropriate sentence to use with a necessity modal.

Why would children struggle figuring out the meaning of necessity modals (and not of possibility modals)? Given the conclusion from the previous chapter that the information they need is available in the input, what would explain this late acquisition? Is it that they don't perceive the information in the input? Do they perceive it, but don't see its relevance? Or are they just not enough exposed to these informative cases?

A number of factors might combine to make the meaning of necessity modals more challenging to figure out for learners—and conversely, could help explain the early advantage for possibility modals. First, the Subset Problem—if it is a real one for the learner—holds for necessity modals, but not for possibility modals. Children might have many opportunities to realize that possibility modals like *can* cannot mean necessary—much more than to see that modals like *must* cannot mean possible. But even though children have strategies to obviate the Subset Problem (i.e., if they are able to use pragmatic situational cues), difficulties figuring out the meaning of necessity

[65] Another possibility is that children have a weaker meaning than strong necessity, that could still differ from possibility (e.g. weak necessity or a some kind of definite modal, as has been proposed for the future). In future research, we intend to run a similar experiment on children productions including other choices other than *can* and *have to*, in particular *should*.

modals could also come from other aspects of their input. First, it might be a matter of frequency of exposure: maybe even though the situational context is informative, children are not exposed to a sufficient number of the relevant cases. Our input study shows that necessity modals are quite rare in parents' productions: They represent 27.5% of adults' modal utterances, which means that children hear *can* (the most frequent possibility modal used by their parents) four times more frequently than *have to* (the most frequent necessity modal). This could be a reason for the asymmetry: If children grasp the meaning of most frequent words first, they will just need more time to figure out their meaning, and use them appropriately.

In some cases, children might also face conflicting cues. In particular, if as considered in Chapter 2, they use negative environments to figure out force, this might contribute to their difficulties—at least for necessity modals which outscope negation like *must*. As discussed, if children assume the same scope relations for *can* and *must*, uses of *mustn't* could suggest to them that it expresses possibility. And interestingly, when we compare ROOT-AFF-1 (*can* vs. *must*) and ROOT-AFF-2 conditions (*can/able* vs. *have to*) in our GF experiment, we find a lower performance for children's *must* than for *have to* (*must*: 42.6% vs. *have to*: 60.2%). The difference could come from differences in their input frequency (*have to* represents 12% of all parents' modal utterances; vs. only 2.3% for *must*), or from the different way they interact with negation.[66] The interaction with negation could also explain why conversely,

[66] Recall that following the idea from Gualmini and Schwarz (2009) (see **section 1.3.2** of Chapter 1), negation may be particularly helpful to figure out the force of modals like *have to* (but it is extremely rare in the input: 4.5% of *have to* utterances are negated) (see Chapter 2). However, as discussed in

possibility modals are mastered so early by children: our study of the input showed that negated possibility modals are quite frequent, and used in particularly informative contexts (e.g. to talk about prohibitions or physical impossibilities).

A last possibility is that children know the meaning of necessity modals, and have the right underlying concepts, but fail to use them in appropriate contexts because of difficulties grasping the intended domain of quantification (as has been argued in the case of definite descriptions, see Brockmann et al., 2018). The difficulty would be about not knowing what information is shared, and how it is shared, amongst conversation participants. Here again, this would capture results from both production and comprehension experiments. In comprehension, children would assume that the speaker must have additional information that is not part of the common ground: for instance, hearing "the toy must/has to be in the blue box", they might understand that the speaker has a different source of information. In production, they would make claims on the basis of private information, or information that adults wouldn't deem under consideration: for instance, they might say "the toy must/has to be in the blue box" when two boxes are closed and therefore should be equally possible choices, because they have eliminated the one whose color they don't like.

To summarize, the nature of children's difficulties with necessity modals could be conceptual (if children know the force, but have trouble dealing with open possibilities), semantic (if children don't know the underlying force of modals),

---

section 1.3.3 and **2.5**, negation could turn out to be misleading in the case of *must* (and is not as rare: 15.8% of *must* are negated).

pragmatic (if children know the force, but have trouble determining the domain of quantification in the same ways as adults), and maybe even epistemic (if children don't know what counts as possible or necessary, or have a different assessment of what counts as necessary than adults).

Coming back finally to the Subset Problem, it is important to note that we find no evidence in favor of a necessity bias. Children's highly adult-like uses of possibility modals might even suggest a bias towards possibility. Of course, this lack of evidence doesn't necessarily entail that children don't rely on a necessity bias when acquiring modals: It is conceivable that children use the bias to acquire necessity modals, but fail to use them in an adult-like way for independent reasons, as discussed above. However, the lack of evidence for a necessity bias in our results, together with its superfluity given our input results, suggests that a bias towards strong meanings is dispensable, even for modals. Our results are thus in line with recent discussions of other Subset Problems, which argue that strong meaning biases may have no place in acquisition (Musolino et al. 2019).

Chapter 4. Input/output study: How do children figure out the force of necessity modals?

In Chapter 2, we focused on children's modal input, to probe the viability of various strategies that could help children figure out modal force, and in particular solve the Subset Problem. In this chapter, we want to get a better sense of how children *actually* figure out modal force. Which of the cues that are in principle available in the input do children actually rely on?

This study addresses this question by looking more closely at individual variation between children, to test what aspects of the input are most predictive of children's mastery of necessity modals, as measured by the *Guess the Force* (GF) paradigm introduced in Chapter 3, our study of children's early modal productions. In Chapter 3, we saw that there is an asymmetry in children's early mastery of possibility and necessity modals. Children seem to master possibility modals such as *can* and *might* very early on: they produce them early (already by age 2), frequently and productively (with and without negation), and use them in an adult-like way. However, they seem to have more difficulties with necessity modals such as *have to* and *must*: they produce them later on, less often, hardly ever with negation, and they do not use them in an adult-like way: results from our *Guess The Force* (GF) experiment show that they tend to use them where adults would prefer possibility modals. And they seem to struggle using necessity modals for quite a long time: results from comprehension experiments with older children show that four year-old children tend to both over-accept possibility modals in necessity situations, and necessity modals in possibility

situations (e.g., Noveck, 2001; Öztürk & Papafragou 2015; Cournane et al. in prep.; Leahy, 2020).

As discussed in Chapter 3, there are several possible explanations for children's struggles with necessity modals: conceptual, semantic, pragmatic, or epistemic. And this 'Necessity Gap' might be expected, given the lower frequency of necessity modals in their input described in Chapter 2. The goal of this study is not to tease apart these different possible explanations, but to address more directly the question of how children actually figure out modal force in practice, in a way that might also help us better understand the nature of their difficulties. Can we identify the aspects of their linguistic input that affect children's early mastery of modals, and in particular of necessity modals?

The present study addresses this question by focusing on the root modals *can* and *have to*, as they are the first modals children produce, and the most frequent in this corpus. We will thus leave aside the question of how children figure out the force of epistemic modals, for which as discussed in Chapter 2, the relevant cues to force might be different, but for which children might figure out force via their root meanings. As we saw in Chapter 3, children between 2- and 3-year-old produce extremely few epistemic modals (the so-called Epistemic Gap, Cournane, 2015), too few for them to be tested using this corpus.

We will compare different aspects of children's input, to assess whether and how they affect children's mastery of modal force. Building on the acquisition literature and on our conclusions from Chapter 2, we might expect different factors to play a role

in the learning process.[67] First, we will look at how some general aspects of mothers' speech, known to potentially affect children's language skills, affect children's mastery of modals: mothers' talkativeness and complexity of speech, as indexed by their *Mean Length of Utterance* (MLU, Brown, 1973). Second, we will look at quantitative aspects of their input: first, the mere frequency of exposure of children to *can* and *have to* (i.e., how frequently their mothers use these modals), since we might expect that more exposure might lead to earlier mastery; second, the proportion of modal talk in general in mothers' speech (i.e., how frequently mothers use modals in general, including all their modal utterances), since we might expect children exposed to more modal talk to grasp modal force earlier. Last, we will look at two specific linguistic environments, the candidates we singled out as potentially helpful to learn root necessity modals such as *have to* in Chapter 2: how frequently they are used with negation (e.g., "you **can't/don't have to** go outside"); and how frequently they are used with undesirable prejacents (e.g., "you **?can/have to** brush your teeth" vs. "you **can/?have to** eat dessert").[68]

We use three indicators of children's modal mastery: their frequency of use of *can* and *have to*, the age at which they start using them, and, whether they use them in an appropriate (adult-like) way, which we measure using the GF (*Guess the Force*)

---

[67] Given our methods, we don't directly test whether children know the meaning of necessity modals, but whether they know how to use them. As made clear in Chapter 3, children might know the meaning of modals and fail to use them in the same ways as adults because of pragmatic issues determining the domain of quantification. But if they don't know the meaning, they will fail to use them in the right situations.

[68] We focus on variation in their mastery of necessity modals (*have to*), even though the question of how children figure out force is in principle relevant for possibility modals as well. However, given results from Chapter 3, we do not expect to find much variation in possibility modals' mastery (children seem to master them early), which makes it difficult to assess the effect of our input factors on their mastery.

paradigm from Chapter 3. This study thus also provides us with the opportunity to replicate the results from this first GF experiment.

Our results will point out two factors that influence children's mastery of *have to* more particularly and seem to play a role in the learning process. First, we find that children who hear *have to* with negation more frequently seem to master it earlier. This is an important result, given our discussion of how children might solve the Subset Problem using downward-entailing environments, which I will revisit in the first section of this chapter. However, this result does not tell us why negation has an effect, i.e., how children actually use it. Second, while we don't find that hearing more *have to* has an effect on its mastery, we find that general exposure to modals matters: children exposed to more modal talk in general seem to master *have to* earlier. This may relate to how knowing a dual might facilitate learning. We find no evidence that other general language factors, such as mother talkativeness and MLU, matter. Our preliminary findings also fail to find an effect of '(un)desirability', though further testing will be required to confirm this finding.

The rest of this chapter is structured as follows. In **section 4.1**, I review the motivation behind the choice of the input factors: what aspects of the input we expect to matter, and why. I then turn to the study itself. In **section 4.2**, I provide some general metadata, focusing on individual variation between children in the Manchester corpus, and describe all our input measures. In **sections 4.3** and **4.4**, I present results about how these factors affect children's modal mastery, starting with children's frequency of use of modals and age of first utterances (**section 4.3**), and then results from the GF experiment which assesses the adult-likeness of their uses (**section 4.4**). Finally, in

section 4.5, I discuss conclusions we can draw from these results, and open questions to address next.

4.1 Background and goals

How do children figure out that *can* expresses possibility, and *have to* expresses necessity? What aspects of their linguistic input do we expect to matter, and why, given what we know about the literature on word learning and modal acquisition, and our conclusions from Chapter 2?

Linguistic input is typically described in terms of both its quantitative and its qualitative aspects.[69] Studies on other cases of word learning tend to show that qualitative aspects of the input matter more than quantitative aspects (see in particular, Hart & Risley, 1995; Hoff, 2003; Hoff & Naigles, 2002; Hsu et al., 2017; Rowe, 2008, 2012; for a recent meta-analysis of studies of how quality vs. quantity of input impact child language skills, see Anderson et al., 2021). In particular, it appears that mothers' talkativeness, or the syntactic complexity of their speech, matter in developing language skills. Do these general factors affect children's early mastery of modals as well?[70]

---

[69] Quantity of input is generally defined in terms of the number of words (/tokens) (Rowe, 2012) or number of utterances (Laks et al., 1990) spoken to the child. I will use the number of utterances here. In contrast, quality of input has been used to refer to a number of aspects of language (see Rowe & Weisleder, 2020, for a review).

[70] In this study, I restrict our focus to general indicators of syntactic complexity (as evaluated by mother's MLU) and their 'talkativeness' (in number of utterances per minute), leaving aside for instance, vocabulary diversity and general interactive features that may prove relevant (e.g. responsiveness or joint attention), as they were harder to assess.

One of the initial motivations for this study was to see whether difference in frequency of exposure might be in part responsible for the asymmetry between necessity and possibility modals, since we saw in Chapter 2 that necessity modals are overall quite rare in children's input, as compared to possibility modals. All else being equal, the more exposure, the better children should perform. Moreover, quantitative aspects of the input have also been shown to matter in other cases of language learning (see for instance Hart & Risley, 1995). Do children who hear more *have to/can* overall master them earlier, and do they use them in a more appropriate way?

Another aspect of their input that might matter is how much exposure children have to modal talk in general: how frequently their mothers use modals in the input. Indeed, we might expect children more exposed to modals in general to master modals' meaning earlier, as it might raise the salience of possibilities and necessities as notions that can be talked about; the contrast with other modals that express different forces, might help them as well. Do children who hear more modal talk in their input master necessity modals earlier?

Finally, more subtle properties of the input have also been shown to affect children's word learning. For instance, in the case of the acquisition of the difference in meaning between attitude verbs like *think* and *know,* it has been shown that hearing *know* with first person subjects leads to worse performance on behavioral tasks, while hearing it used in questions predicts higher accuracy (e.g., see Howard et al., 2008; Dudley, 2017). In the case of modals as well, we might expect those 'qualitative' aspects to matter. Even though hearing many uses of *have to*, children might not learn

its meaning if these cases are not informative; conversely, children who hear fewer uses of *have to* might still learn its meaning, if they are exposed to highly informative cases.

In Chapter 2, we identified two factors that could in principle help children figure out force, for root necessity modals like *have to*, given their input: how frequently *have to* is used with negation, and how frequently it is used with undesirable predicates. I will briefly review our three "solutions" to the Subset Problem (i.e., how children figure out the force of necessity modals, given that necessity entails possibility), and what we concluded for each of them, to remind us of why we expect these two factors to matter more particularly.

The first "solution" we considered—which I will put aside in this chapter—was that child learners have a bias towards 'strong' necessity meanings (in the spirit of Berwick, 1985): by default, they would assume necessity meanings for modals, and revise their hypothesis only for possibility modals, when encountering them used in situations of non-necessity. Our study of the input suggests that such a bias is not necessary, since children may have other strategies available; moreover, we found no evidence for such a bias in Chapter 3, when looking at children's early productions.

The second "solution" was that children use evidence from Downward-Entailing environments, following Gualmini & Schwarz (2009). By hearing negated necessity modals used in situations of possibility (for instance, "The dog *doesn't have to* go outside" when it is clear that the dog can either go outside or stay inside), children might be able to infer that *have to* cannot have a possibility meaning: if *have to* meant *possible*, *not have to* would mean *not possible*; but then, it could not be used in a possibility situation). However, our study of the input shows that potentially

informative cases for *have to* are extremely rare in children's actual input: for instance, only 4.1% of mother's utterances involving *have to* are negated (and 1.1% for *got to*, the second most frequent necessity modal).[71] Moreover, *Input Experiment 1* shows that these rare cases of negated *have to* might not always be informative about force: in particular, they are often used in 'polite' ways, to soften prohibitions (e.g., "you don't have to shout"), the meaning of which seems closer to impossibility. But even though extremely rare, the remaining negated uses of necessity modals could be useful, if children can make use of them. Moreover, negation might be hard to exploit with a single exposure, but it could be more useful over time—an aspect we didn't test in *Input Experiment 1*.

The third "solution" to the Subset Problem was that children rely on pragmatic cues, from the conversational context in which modals are used. In particular, we explored the idea that for root modals, cues from the perceived '(un)desirability' of the event described might be especially useful: hearing necessity modals used with prejacents describing an undesirable event (e.g., 'You *have to/?can* clean your room', vs. 'You *can/?have to* go eat dessert') might help children figure out that *have to* expresses necessity. In *Input Experiment 3*, we showed that this cue is available in the input: adults do use necessity modals more frequently with undesirable events, and possibility modals with desirable events. Moreover, *adults* seem to make use of it to figure out modal force: it contributes to participants' performance in our *Input*

[71] In Chapter 1 and 2, we also discussed problems that arise from scope irregularities for necessity modals such as *must*: *must* systematically outscopes negation. In these cases, the strategy just cannot be applied. In this chapter, I leave these issues aside, as we focus on *have to* (*must* could not be tested because this corpus doesn't contain enough data per child).

Experiment 1. But while these cues are in principle available, children might not make use of them in practice.

In this study, we thus focus on these two aspects of the input—how frequently *have to* is used with negation, and the strength of the 'desirability' signal (how frequently *have to* is used with undesirable prejacents by mothers). Of course, other aspects of the input could also provide useful cues; I leave their investigation to future research. Are these aspects predictive of children's mastery of modal force for root modals? Do children who are exposed to more negated occurrences of *have to*, or who hear it more often with undesirable prejacents, master its force earlier?

To see how these various aspects of children's input affect their mastery, we will use three complementary indicators: children's own frequency of production of modals (how frequent *can* and *have to* are in their own speech); the age at which we find their first productions in the corpus; and whether they use them in an appropriate way, which I evaluate experimentally using the GF (*Guess the Force*) paradigm introduced in Chapter 3. The main difference with the first GF experiment is that we now directly test for differences between children, to see whether we can then identify the factors in their input that predict their mastery.

Investigating potential variation between children, at an individual level, may also help better understand the results from our first experiment, and more generally children's difficulties with necessity modals. How general are the difficulties we described? Do we find variation between children in their mastery of modals? Do some children master necessity modals earlier than others? Do we find that children's uses become more adult-like over time? In our first GF experiment on children's

productions, we did not have enough data to compare children directly. This new study also allows us to control for how age and grammatical development might affect children's performance (using the MLU of children's utterances (Brown, 1973) as a measure of general grammatical development).

The existing acquisition literature does not tell us much about variation that can be found at an individual level, even with older children. In general, comprehension studies (e.g., Noveck, 2001; Öztürk & Papafragou, 2015) do not report results besides aggregates. When reported acceptance rates correspond to chance level (e.g., 47% of acceptance of *have to* in possibility scenarios in Öztürk & Papafragou 2015, Experiment 1), we don't know whether they reflect distinct groups of children (some children being adult-like, and some others treating necessity modals as expressing possibility), or all children answering randomly.[72] Cournane et al. (in prep.) do report data about individual variation in their study on teleological modals *can/have to* (mean age: 4;01): out of the 12 children they test in Experiment 1's necessity condition, 9 systematically accept *have to* in possibility situations, and 3 children are more adult-like (mean: 2.3/5 'no' answers); none of them are perfectly adult-like.[73] However, it is hard to generalize given the small size of the sample (a problem that will also hold for us).

In short, the goal of this study is to identify what aspects of the input children use, in practice, to learn force. We focus on (root) necessity modals, since children

[72] Given that Öztürk & Papafragou (2015) defend a *Premature Closure* explanation, they probably did not have two distinct populations, but since they don't report individual data, we can't know for sure.
[73] There were 5 test cases per scenario per child. Force of the modal (*can* vs. *have to*) was tested between participants. Each modal sentence was tested in three scenarios: possibility, necessity, impossibility.

show early mastery of possibility modals, and produce modals with epistemic interpretation later on. By looking at individual patterns and variation between children, we can assess what factors may explain this variation, by seeing if there is a relation between input factors and children's mastery. Of course, the conclusions we can draw from this study are limited by the small number of children included in the corpus and the small age window we investigate (from age 2 to 3), but this study lays the ground for future research.

In the next section (**section 4.2**), I present general metadata and describe all of our input measures. In **section 4.3**, I turn to the assessment of children's productions, looking at their own frequency of use of the modals *can* and *have to*, and their age of first utterance. In **section 4.4**, I present results from the GF experiment which assesses the adult-likeness of children's modal uses. Finally, in **section 4.5**, I discuss conclusions we can draw from these results.

4.2 Study

The goal of this study is to see how different characteristics of the linguistic input impact children's early mastery of modals. I focus on children's *can* and *have to* (in their root uses), which are the first modals produced by children, and the most frequent modals between age 2- and 3 in this corpus, both in children's speech and in their input.[74] We use three indicators to evaluate children's mastery: (i) children's own

[74] As mentioned, we could not run the experiment on children's (root) *must* (condition ROOT-AFF-1 in our first GF experiment on children's productions), because we are limited by the low numbers of *must*

frequency of production (i.e., how frequent *can* and *have to* are in their own speech); (ii) the age at which we find their first productions of *can* and *have to;* and (iii) whether they use them in an appropriate way (adult-like), which I assess experimentally using the GF (*Guess the Force*) paradigm introduced in Chapter 3. We will focus on the third one, as it allows a more fine-grained assessment of children's understanding. Is mastery just a matter of quantity of exposure (i.e., do children who hear *have to* more often seem to master it earlier)? Is it helpful to hear (necessity) modals used with negation more often? To hear (necessity) modals used with undesirable prejacents? To be exposed to more modal talk overall?

Participants. The Manchester corpus (Theakston et al., 2001; CHILDES database, MacWhinney, 2000) consists of 12 children, all monolingual, English-speaking, firstborn children, whose mothers are the primary caregivers (see Chapter 2). No information about socioeconomic status was collected. Children were recorded with their mothers in unstructured playing sessions at home. 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 study: in this study, as in the preceding chapters, I only include data starting from 2-year-old. Children's MLUs range between 1.14 at the start of the sample (Nicole) to 3.84 at the end of the sample (Warren). One of the interests of using this corpus is thus the relative homogeneity in children's backgrounds, which allows us to test fine-grained properties of their linguistic input.

---

utterances in children's own productions: only 4 of them produce more than 10 *must* utterances. For the other two conditions (ROOT-NEG; EPI-AFF), there were even fewer utterances per child in the corpus.

General input measures: mothers' talkativeness and MLU. **Table 4.1** presents general metadata (unspecific to modal talk) about the twelve child/mother dyads in the corpus. The "talkativeness" of the mother corresponds to the mean number of utterances in one hour of recording time. The MLU is here used as a measure of mother's syntactic complexity.

Table 4.1 General data about the 12 children in the Manchester corpus.[75] Children are ordered by alphabetical order.

child	gender	MLU 2;0.0	MLU 3;0.0	CDI score[76]	total # of morphemes (mother)	total # of utterances (mother)	# recordings	Mother's talkativeness[77] (# morph/min)	(# utt./min)	Mother's MLU
anne	F	2.15	3.16	180	134515	30110	30	149.5	33.5	4.47
aran	M	1.75	3.46	153	205161	33238	31	220.6	35.7	6.18
becky	F	1.54	3.04	138	110236	25836	34	108.1	25.3	4.26
carl	M	2.12	3.66	187	61357	13322	22	93.0	20.2	4.60
domin	M	1.61	2.62	153	133082	32016	31	143.1	34.4	4.16
gail	F	1.7	3.26	262	113075	25601	33	114.2	25.9	4.41
joel	M	1.83	3.17	122	107760	24939	31	115.9	26.8	4.33
john	M	2.05	2.81	191	83620	17947	30	92.9	19.9	4.65
liz	F	1.66	3.76	359	81774	18101	32	85.2	18.9	4.51
nic	F	1.14	3.09	102	136288	29208	34	133.6	28.6	4.67
ruth	F	1.46	2.99	44	147277	33777	31	158.4	36.3	4.36
warr	M	2.23	3.84	124	116512	21028	31	125.3	22.6	5.54
ALL	**6F/6M**	**1.77**	**3.24**	**168**	**1430657**	**305123**	**370**	**128.9**	**27.5**	**4.68**

[75] 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-old are included. Note that duration of recordings is not systematically reported (e.g., they are for Anne, but not for Warr): to compute mother's talkativeness, I assume that all recordings were 30 minutes long.

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

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

'Quantitative' input measures (raw frequencies of *can*/*have to* in mother's speech, and frequency of modal talk overall in mother's speech). **Table 4.2** summarizes data about modal uses in mother's speech, for the twelve mothers. It first gives data about the frequency of exposure to *can* and *have to*, with (i) the bare frequency of *can* and *have to* (i.e., the proportion relative to the total number of utterances uttered by the mother, which I call **'**overall' frequency for ease of exposure) and (ii) the proportion of *have to* utterances relative to *can* utterances (which I call 'relative' frequency).[78] As already described in Chapter 3, mothers' most frequent possibility modal, *can*, is much more frequent than their most frequent necessity modal, *have to*: on the aggregate, it occurs in 3.4% of all mother's utterances (vs. only 0.75% for *have to*). Looking at variation between mothers, we see that the overall frequency of *have to* ranges from 0.4% (Anne's mother) to 1.2% (Gail's mother). On the aggregate, the frequency of *have to* relative to *can* is 18.1% (approximately 4 *can*s per 1 *have to*), ranging from 11.2% (app. 9 *can* per 1 *have to*) (Becky's mother) to 27.4% (app. 3 *can*s per 1 *have to)* (Liz's mother). **Table 4.2** also reports data on the frequency of modal talk in general in mothers' speech (i.e., the frequency of all their modal utterances, not only *can/have to* utterances) (modals included: possibility: *can, could, might, may; able to;* necessity: *must, should, need; have to, got to, be supposed to, need to*).

[78] In Chapters 2 and 3, we've been focusing on the proportion of each modal relative to other *modal* utterances; in this chapter, we focus on frequency with respect to all utterances, not only modal ones.

Specific linguistic environments. Table 4.2 also reports data on proportion of uses of *can/have to* in negative sentences, and the strength of the (un)desirability signal in mother's speech (how frequently they are used with undesirable prejacents). I explain below how we computed the desirability scores.

Table 4.2 Summary of input 'quantitative' and 'qualitative' measures (mother's modal talk): 'overall' and 'relative' frequency of *can* and *have to* in mothers' speech; proportion of modal talk (i.e., including other modals: *can, could, might, may; able to; must, should, need; have to, got to, be supposed to, need to*); proportion of modal necessity talk (i.e., including only other necessity modals: *must, should, need; have to, got to, be supposed to, need to);* frequency of uses of *can* and *have to* in negative sentences; desirability scores for *have to*. For each input measure (in columns), children who have the "highest" exposure (i.e., from whom we might expect earlier mastery) are highlighted in dark orange (e.g., for 'overall' frequency of *have to*: Gail: 1.2%; Liz: 1.1%)*;* children who have the "lowest" exposure (i.e., from whom we might expect later mastery) are highlighted in light orange (e.g., for 'overall' frequency of *have to*: Anne: 0.4%; Becky: 0.5%). In grey are children that are not tested in the experiment, because they produce too few *have to* (see children data reported in **Table 4.4**).

MOTHERS (n=12)

	Total # utt.	Total # modal utt.	Prop. modal talk	Prop. Nece	'Overall' frequency		'Relative' frequency (have to vs. can)	Negation		Desirability score	
					can	*have to*		*can*	*have to*	*(i) fun*	*(ii) notfun*
anne	30110	1461	4.9%	1.4%	766 2.5%	133 0.4%	14.8%	17.8%	1.5%	24%	21%
aran	33238	2245	6.8%	2.0%	1222 3.7%	220 0.7%	15.3%	24.8%	4.1%	23%	41%
becky	25836	1488	5.8%	1.4%	926 3.6%	117 0.5%	11.2%	19.4%	5.1%	17%	39%
carl	13322	660	5.0%	1.5%	405 3.0%	83 0.6%	17.0%	20.7%	8.4%	22%	36%
domin	32016	1878	5.9%	2.0%	1026 3.2%	278 0.9%	21.3%	26.0%	9.0%	19%	38%
gail	25601	1560	6.1%	1.9%	803 3.1%	301 1.2%	27.3%	23.4%	2.7%	36%	24%
joel	24939	1236	5.0%	1.8%	647 2.6%	218 0.9%	25.2%	42.3%	3.2%	17%	43%
john	17947	1233	6.9%	1.1%	883 4.9%	131 0.7%	12.9%	7.6%	6.1%	Nc.	Nc.
liz	18101	1068	5.9%	1.8%	516 2.0%	195 1.1%	27.4%	18.4%	2.1%	28%	37%
nic	29208	1882	6.4%	1.6%	1269 4.3%	170 0.6%	11.8%	24.5%	3.5%	32%	38%
ruth	33777	1327	3.9%	1.3%	686 2.0%	219 0.7%	24.2%	20.7%	3.7%	Nc.	Nc.
warr	21028	1811	8.6%	2.0%	1053 5.0%	184 0.9%	14.9%	20.3%	5.4%	26%	42%
ALL	**305123**	**17849**	**5.8%**	**1.7%**	**10202 3.4%**	**2249 0.8%**	**18.1%**	**22.2%**	**4.5%**	**24%**	**36%**

Desirability measures. ***Methods*.** To evaluate the strength of the (un)desirability signal in mother's speech, we coded for a sample of 500 *have to* utterances (50 per mother).[79] All utterances were coded independently by two coders (native speakers of English). Similarly to the way we assessed desirability in *Input Experiment 3* (**section 2.4**), coders were blind to the modal (it was replaced by a blank before coding). This is to avoid their being biased ('I **can** X' might suggest that *X* is desirable, and 'I **have to** X' might suggest that *X* is undesirable). Coders were asked to code sentences in two ways. First, they were asked to indicate whether the activity sounded 'fun' or not (yes/no). This allows us to keep a direct comparison point with our first 'desirability' experiment (which would not be the case if we were asking for a scaled judgment). Second, coders were asked to indicate whether the activity sounded 'not fun' or not (yes/no). The reason for this second coding is to avoid merging 'neutral' cases (i.e., activities such as "grabbing a pen", that are not necessarily considered as 'fun' but not clearly *un*desirable), with clear 'negative' cases (e.g., "going to the dentist"), given that our hypothesis for necessity modals focuses on *un*desirability.[80] Examples are given in (35) and (36). Coders were told that the activities come up in conversations between two-year-old children and their mothers. They were told that 'not fun' could correspond to anything annoying, effortful, or painful: anything they would not like to do, if they were the child. ***Cases of mismatch***. Coders were told to consider cases of potential

[79] Because of time constraints, we didn't run an experiment similar to the one presented in Chapter 2, **section 2.4** (*Input Experiment 3*), but intend to do so in the future.

[80] This double coding is unnecessary when running an experiment, since averaging across more participants allows to differentiate negative cases (~all participants answering "no") from neutral ones (~half of the participants answering "no") with the question "Does this sound fun?" alone. We didn't test the 'not fun' task in *Input Experiment 3*.

mismatch between the child and her mother from the perspective of the child: for example, "jumping in a puddle" was coded as *fun*, but "being quiet" was coded as not fun, even though it might be the opposite from the mother's perspective.

(35) Does it sound fun to ____?

- go to the dentist/wait for a long time → 0: '*No, not fun'*
- grab a pen/say hello → 0: '*No, not particularly fun'*
- play with a toy/get the sticker → 1: '*Yes, sure, it sounds fun*!'

(36) Does it sound not fun to ____?

- go to the dentist/wait for a long time → 1: '*Ugh, I would hate that*'
- grab a pen/say hello → 0: '*No, that's fine*'
- play with a toy/get the sticker → 0: '*No, happy to do that*'

Table 4.3 reports the details of the desirability scores obtained for each mother (also reported together with our other input measures in **Table 4.2**). For each mother, we computed two scores: (i) a desirability score, based on coders' answers to 'Does it sound fun to [X]', and (ii), an undesirability score, based on their answers to 'Does it sound not fun to [X]'. ***Comparison between coders***. The rate of agreement between coders was 66.2% (cases of direct agreement on 'fun': 352/500 (70.4%); cases of direct agreement on 'not fun': 310/500 (62.0%)). For each activity, the score corresponds to the mean of the ratings by the coders (a desirability score of 1 thus corresponds to cases when they agree on considering it as 'fun', respectively, 'not fun', for the undesirability score; 0, to when they agree on considering it as not 'fun'; 0.5, to when they disagree). The final score for each mother corresponds to the mean across 50 contexts, excluding cases that were not possible to code (e.g., when the prejacent is elided) (9 cases in total).

As expected, activities judged as 'fun' were almost always judged as not 'not fun', and conversely, activities judged as 'not fun' judged as not 'fun'.[81]

Overall, the mean desirability score is 24.1%, ranging from 17% (Becky's mother) to 32% (Nicole's mother); the mean undesirability score was 36%, ranging from 21% (Anne's mother) to 43% (Joel's mother). As a point of comparison, the mean rate of 'fun' judgments found in *Input Experiment 3* (**section 2.4**) was 25.7% on *have to* utterances (vs. 49.7% on *can* utterances). Note that since we focus on how children figure out the force of root necessity modals, we did not code for *can* utterances.

Table 4.3 Desirability scores ((i): fun; (ii): not fun) for *have to* utterances for all 10 mothers:

		(i) FUN score				**(ii) NOTFUN score**			
mothers	*excluded*	0	0.5	1	score	0	0.5	1	score
anne	1	28	18	3	24%	32	13	4	21%
aran		31	15	4	23%	20	19	11	41%
becky	1	35	11	3	17%	20	20	9	39%
carl	2	34	7	7	22%	21	19	8	36%
domin	2	33	12	3	19%	19	22	7	38%
gail	2	22	17	9	36%	29	15	4	24%
joel	1	35	11	3	17%	20	16	13	43%
liz		27	18	5	28%	20	23	7	37%
nic		24	20	6	32%	21	20	9	38%
warr		28	18	4	26%	18	22	10	42%
ALL	8	297	147	47	24.1%	220	189	82	36%

[81] There was only one case that was judged both as 'fun' and 'not fun' by both coders: 'pay the mother some money to go through'.

4.3 Children's modal productions

How do these different factors affect children's mastery of modals? I will start by looking at our first two measures of modal mastery: children's own frequency of production of modals and their age of first production.

4.3.1 Children's quantitative productions

Table 4.4 summarizes children's data. It reports the accuracy found in the GF experiment, for ease of future comparison. First, looking at children's bare frequency of production, we find that *can* is significantly more frequent than *have to*, as already described in Chapter 3*:* it occurs in 1.8% of all children's utterances (vs. 3.4% of all mother's utterances), vs. *have to*, only 0.2% (vs. 0.8% for mothers). Relative frequency of *can* and *have to* for children is approximately 9 *can* for 1 *have to* (vs. 4 *can* for 1 *have to* for mother's).

Table 4.4 Measures of children's modal mastery: 'overall' and 'relative' frequency; age of first productions; mastery as measured by the GF paradigm (see Table 4.6)). In grey are children not tested in the experiment, because they produce too few *have to*.

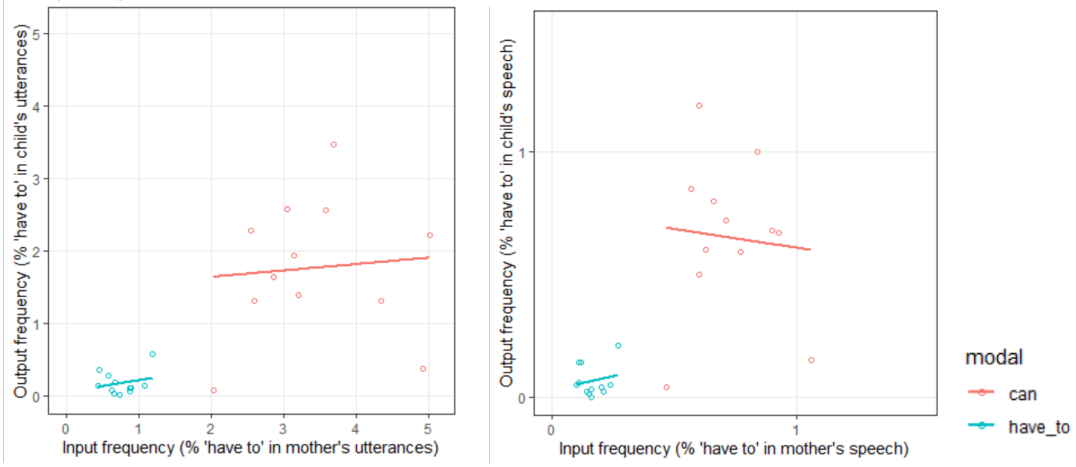
	Total # utt.	Total # of modal utt.	Prop. modal talk	'Overall' frequency		'Relative' frequency (have to vs. can)	Age of prod.		GF experiment	
				can	*have to*		*can*	*have to*	*can*	*have to*
anne	16405	501	3.1%	374 2.3%	22 0.1%	5.6%	2;00.15	2;00.29	80.2%	39.6%
aran	16144	729	4.5%	560 3.5%	30 0.2%	5.1%	2;01.00	2;05.03	82.4%	57.0%
becky	23398	788	3.4%	598 2.6%	84 0.4%	12.3%	2;00.07	2;05.08	78.3%	57.5%
carl	16998	583	3.4%	438 2.6%	12 0.1%	2.7%	2;00.26	2;06.19	64.9%	57.7%
domin	19145	322	1.7%	267 1.4%	11 0.1%	4.0%	2;00.28	2;04.11	70.3%	73.8%
gail	16396	449	2.7%	317 1.9%	93 0.6%	22.7%	2;01.08	2;00.19	87.5%	61.9%
joel	16410	313	1.9%	214 1.3%	17 0.1%	7.4%	2;01.23	2;0325	69.8%	43.9%
john	12464	59	0.5%	46 0.4%	1 0.01%	2.1%	2;00.13	2;09.12	*Not tested.*	
liz	15501	317	2.0%	253 1.6%	20 0.1%	7.3%	2;00.28	2;04.03	82.2%	44.1%
nic	16937	280	1.7%	221 1.3%	45 0.3%	16.9%	2;01.01	2;08.20	69.3%	45.2%
ruth	19282	39	0.2%	16 0.1%	3 0.02%	15.8%	2;04.01	2;07.10	*Not tested.*	
warr	14226	420	3.0%	316 2.2%	14 0.1%	4.2%	2;01.14	2;03.08	67.0%	68.4%
ALL	**203306**	**4800**	**3.1%**	**3620 1.8%**	**352 0.2%**	**8.9%**	**2;00**	**2;04**	**75.2%**	**54.9%**

Is there a relation between children's own frequency of production and frequency in the input?[82] **Figures 4.1a** and **4.1b** show the relation between the frequency of use of *can* and *have to* in children's input (mothers' speech) and in their own speech, computed either out of the total number of utterances (**4.1a**) or out of the total number of morphemes (**4.1b**) (both are commonly used in the literature). Correlations are not significant, whether we look at overall or relative frequency (overall frequency: *can*: Pearson's r=0.086 (t(10)=0.27, *p=0.79 (NS)*; 95% CI: [-0.51;0.63]); *have to*: Pearson's r=0.232 (t(10)=0.75, *p=0.47 (NS); 95%* CI: [-0.39;0.71]; 'relative' frequency: Pearson's r=0.29 *(t(10)=0.99, p=0.35 (NS);* 95% CI: [-0.33; 0.74]). We also find no effect when looking at frequency computed out of the number of morphemes: *can*: Pearson's r=-0.08 (t(10)=-0.26, *p=0.80 (NS)*; 95% CI: [-0.63;0.52]); *have to*: Pearson's r=0.19 (t(10)=0.61, *p=0.55 (NS)*; 95% CI: [-0.43;0.69]). However, this absence of result should be taken with caution: it is expected given the low sample size (only 12 child/mother dyads).

[82] We don't test the effect of qualitative input measures on the quantity of children's production.

Figure 4.1 Relation between overall frequency of use in the input (mother's speech) (x-axis) and frequency of use in children's productions (y-axis), out of total number of utterances (1a) and out of total number of morphemes (1b). Each dot corresponds to one child/mother dyad (n=12). Red dots correspond to frequency of *can*; blue dots correspond to frequency of *have to*.

Figure 4.1a (out of total # of utterances) **Figure 4.1b (out of total # of words)**



4.3.2 Age of first productions

Looking at our second indicator of children's mastery, age of first production, we find that *have to* tends to be sampled later than *can*, with on average a 4-months difference (**Table 4.4**).[83] Note that our measure is likely to under-estimate the gap between first productions of the two modals, notably because most children start producing *can* before 2-years-old: except for Ruth, the first occurrences of *can* are found in earlier transcripts than the ones included in this study. Focusing on their first production of

[83] These results are for now just indicative, since we only looked at the first *have to* sampled in the corpus. A more conservative measure to use is the *First of Repeated Uses* (FRU) (Snyder, 2007), that allows to avoid the possibility that this first use is just repeating something the child recently heard (a use counted as repeated if it was within two months or two transcripts as the previous use).

have to, we find some variation in how early children start using *have to*, but it does not seem to be predicted by input frequency.

As argued in Chapter 3, the fact that children don't produce *have to* doesn't mean that they don't understand it. Conversely, the fact that some children produce it doesn't mean that they understand it or use it in a proper way: this assessment does not control for children repeating sentences used earlier in the conversation, even though direct repetitions were excluded from our analyses. Last, we might easily miss some children that use *have to* earlier, but not during the time of the recording.

To summarize this section, we don't find that quantitative aspects of the input affect children's quantity of productions. However, the low sample size might also be responsible for this absence of effect.

4.4 Guess The Force Experiment 2

We now look at how both the quantitative and qualitative aspects of their input we described in **section 4.2** affect the quality of children's modals, using the *Guess The Force* (GF) paradigm introduced in Chapter 3 (condition ROOT-AFF-2, testing root *can* vs. *have to*). If we find variation between children in how adult-like their modal uses are, what are the aspects of their input that best predict this variation? In particular, do children who hear *can/have to* more frequently use them in an adult-like way earlier? Does it help to hear modals with negation? Does it help to hear them used more frequently with (un)desirable events? Does it help to hear more modal talk in general? And do other factors like mothers' MLU and talkativeness matter?

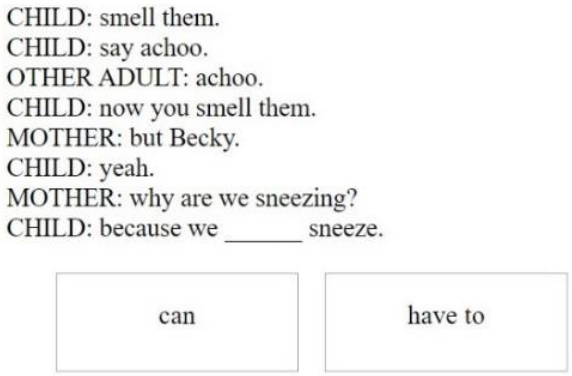
A secondary goal is to control for an effect of age and syntactic development: do children tend to become more adult-like with age, as expected?

4.4.1 Methods

In the experiment, adult participants are given dialogues extracted from the Manchester corpus, and asked to guess the force of a redacted modal that was uttered by a child, as illustrated in **Figure 4.2**.[84] They are asked to choose between two options, corresponding either to the necessity (*have to*) or the possibility (*can/able to*) modal. As in *GF Experiment 1*, each participant had to judge 40 dialogues overall: 20 trials (10 possibility; 10 necessity); 20 controls using tense (10 past; 10 future). The 20 trial contexts were randomly selected for each participant, out of a list of 200 contexts previously extracted from the Manchester corpus. Controls were the same as in Experiment 1, except that we changed the 5 problematic cases. ***Rationale.*** This allows us to evaluate how "adult-like" children's uses are. We use the experiment on adults' uses (*Input Experiment 1*, **section 2.3**, Chapter 2) as a baseline which shows that force is guessable from context, and assume that dialogues preceding the modal sentence are equally informative for adults' and children's uses. We thus take participants' accuracy at guessing children's modals (as compared to their accuracy at guessing adults') as a measure of how 'adult-like' children's uses are.

[84] An example of the experiment can be accessed below:
https://#spellout.net/ibexexps/modforce/modforce_FC_IK_rootP2/experiment.html

Figure 4.2 Experiment stimuli: example trial (*have to*):



Material. 20 contexts per child (10 *can*, 10 *have to*) were initially randomly extracted from the corpus. Note that because of the low number of *have to* utterances in John's and Ruth's productions (John: 1; Ruth: 3; see **Table 4.2**), we could not include them in the experiment. We thus had a list of 200 contexts in total (10 possibility and 10 necessity for each of the 10 children), and 20 controls. The full list is available at [LINK]. ***Exclusion criteria.*** As in the first experiment, we didn't remove contexts where the modal already appeared in the preceding dialogue, unless 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*.').[85] Again, we made 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. Again, we removed Briticisms, but did not correct children's ungrammatical utterances (e.g., *comed* for *came*), except in

[85] Out of the final list of 200 contexts, 47 of them have the modal appear in the preceding dialogue (23.5%) (necessity contexts: 21; possibility contexts: 26 cases) (uttered by the child: 24, by the mother or another adult: 12; by both: 11). A number of contexts have the other (non-target) modal appear in the preceding dialogue (necessity contexts: 14 where there is *can/can't* before, 7 where there is *have to;* possibility contexts: 24 where there is *can/can't* before, 1 where there is have to, 1 where it is both).

the case of *have to* when children 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 again 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* utterances in children's production (3 cases overall, and only 1 in a declarative sentence). This means that there was only one case where *able* was the right answer, vs. 7 for (tensed) *have to*.

Input measures. Our input measures are based on the data reported in **Table 4.1** and **4.2.** For 'quantitative' measures, as before, we use two measures (overall frequency, i.e., the proportion of *can*/*have to* with respect to the total number of utterances uttered by the mother, and relative frequency, i.e., the proportion of *have to* relative to *can*). For 'qualitative' measures, we use mothers' MLU (as an index of syntactic complexity); mothers' talkativeness; proportion of modal talk (i.e., including other modals); proportion of modal necessity talk (i.e., including only other necessity modals*)*; frequency of uses of *can* and *have to* in negative sentences; and desirability scores (for *have to*). **Child measures.** Finally, we control for the effect of age and MLU. ***Age*** corresponds to the age of production of the modal by the child, in months. Age of production in the extracted contexts was on average 31.8 months for possibility modals contexts, and 32.3 months for necessity modals contexts. ***MLU*** corresponds to the MLU of the child at the time of production of the modal. (37) summarizes the conditions. All conditions are tested within participants.

(37) Force: {possibility, necessity}
Child: {anne; aran; becky; carl; domin; gail; joel; john; liz; nic; ruth; war}
Age: from 26-month-old to 36-month-old
MLU: between 1.2 and 4.1

Expectations. We expect:

- Participants to perform better for children's uses of possibility modals (*can/able*) than for their necessity modals (*have to*) (i.e., we expect to replicate the results from our first GF experiment on children's productions).
- **Input frequency**: all else being equal, children who are exposed to more necessity modals in their input should use them in a more adult-like way (controlling for age and MLU).
- **Modal talk:** All else equal, we expect children who are more exposed to modal talk in general (i.e., whose mothers use modals frequently), to use them in a more adult-like way.
- **Negation**: We expect negation to be helpful for both possibility and necessity modals, as the limits of a strategy based on negation for *have to* might stem from their extreme low frequency. Children more exposed to *don't have to* should master *have to* earlier, if children are able to use negative environments. We expect negation to be useful for *can* as well, but as mentioned already, we might not be able to capture any effect if children's performances are at ceiling.
- **Desirability**: We expect children who have a stronger '(un)desirability' signal (i.e., whose mothers use necessity modals with undesirable prejacents more

frequently), to use them in a more adult-like way, if they are sensitive to this cue the way adults are.

- **Children's Age and MLU (control)**: Performance should improve with children's age and MLU (children's usage should become more adult-like with time).

4.4.2 Results

Participants. 351 adult participants were recruited on Amazon Mechanical Turk (language: US English; 190 females, mean age=41.9-year-old). The 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.

Analysis. All 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 our *GF Experiment 1*: participants perform significantly better for children's possibility modals than for their necessity modals (mean accuracy for possibility modals: 75.2%; for necessity modals: 54.9%; vs. in *GF Experiment 1*, condition ROOT-AFF-2: respectively, 79.6% and 60.2%; and in the adult baseline, condition ROOT-AFF-2: respectively, 81.5% and 82.0%). **Figure 4.3** shows results on the aggregate, with the experiment on adults' productions and our first experiment on children's productions as comparison points. As for the two previous experiments, we first run binomial tests to see if, overall, participants' performance differs from chance, for each condition of force. We find that participants differ from chance for both possibility and necessity modals, as reported in **Table 4.5**. Then, to test whether children's uses are adult-like,

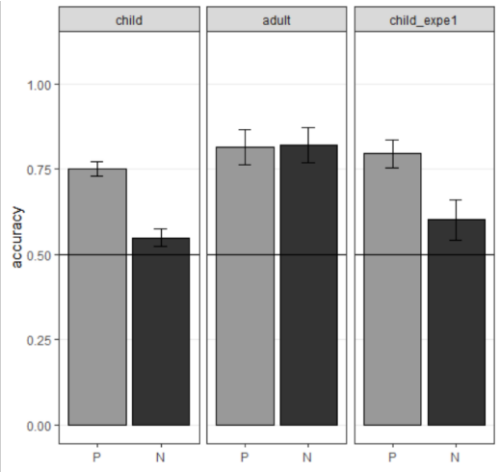
we compare them to the experiment on adults' productions (we assume that the conversational context should be equally informative in both cases). We use binomial linear mixed effects models built with a maximal random effect structure testing Accuracy with Subject and Item as random factors testing for the effect of Age group (child vs. adult usage). The result is only significant for necessity modals ($\chi^2(1)=47.1$, $p=6.8e\text{-}12$ ***): we find no difference for possibility modals ($\chi^2(1)=2.17$, $p=0.14$). The interaction Force*Age group is significant ($\chi^2(1)=20.9$, $p=4.9e\text{-}06$ ***).

Table 4.5 Results and statistical tests for GF Experiment 2, as compared to GF Experiment 1 and Input Experiment 1 (condition ROOT-AFF-2, respectively presented in Chapter 3, **section 3.3** and in Chapter 2, **section 2.2**).

	Mean accuracy[86] (se)		Exact binomial tests (two-sided)		Results of the model testing effect of Age (adult vs. child usage)	
	possibility	necessity	possibility	necessity	possibility	necessity
GF EXPE 2 (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 ****
GF EXPE 1 (CHILD)	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 ****
INPUT EXPE-1 (ADULT)	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*

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

Figure 4.3 Mean accuracy by condition of force: Comparison between experiments (GF Experiment 2: n=332 participants; 200 contexts in total, 100/force; GF Experiment 1: n=70 participants; 40 contexts in total, 20/force; Input Experiment 1: n=69 participants; 40 contexts in total, 20/force).. 'P': possibility contexts (*can/able*). 'N'=necessity contexts (*have to*). Accuracy corresponds to how good participants are at correctly guessing the force of the modal based on the context (e.g. to answer can in a possibility context).

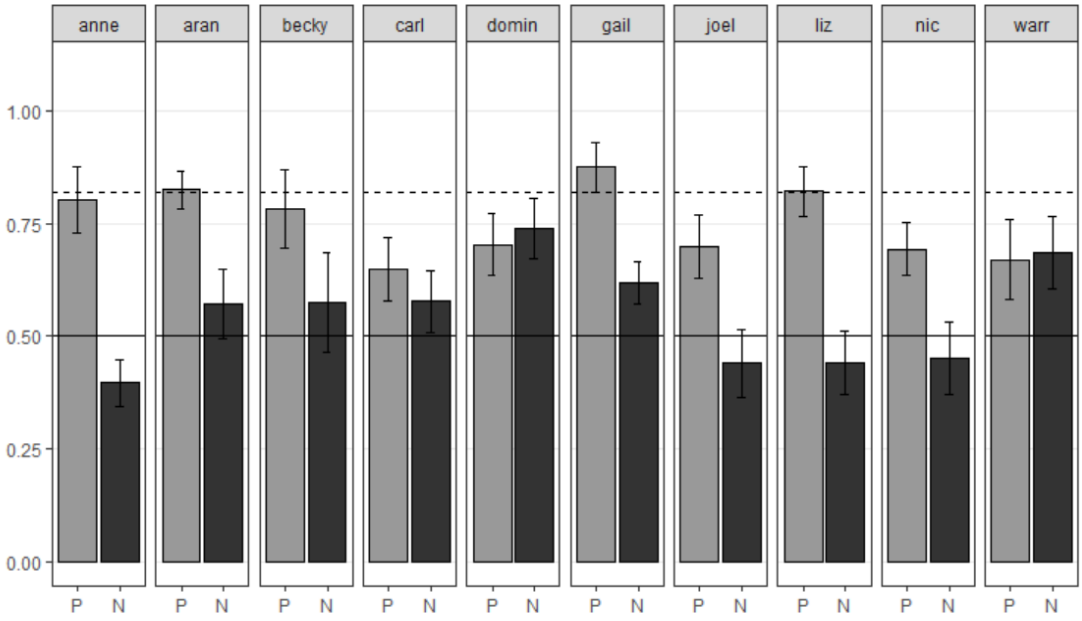


Analysis by individual child. **Table 4.6** reports participants' accuracy in each condition of force (possibility vs. necessity), for the 10 children (ordered alphabetically) (see **Figure 4.4)**. For each child, we ran both binomial tests to see whether participants differ from chance, and binomial linear mixed effects models to see whether children's uses are 'adult-like', comparing them to the adult baseline. For possibility modals, we find that all children perform significantly better than chance. For necessity modals, we find that 5/10 children perform better than chance, 4/10 don't differ from chance, and 1 (Anne) performs lower than chance (39.6%). Comparing their uses to the adult baseline, we find that for possibility modals, 4 out of 10 children are not adult-like, and for necessity modals, 7 children are not adult-like. The test approaches significance for Becky: *p=0.055*.

Table 4.6 Accuracy rates (se) and results of binomial tests; models testing effect of Age group (adult usage vs. child usage), by child condition (*GF Experiment 2*: n=332).

child	Mean accuracy (se)		Exact binomial tests (two-sided)		Results of the 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]	$\chi^2(1)=2.74$, $p=0.098$ (NS)	$\chi^2(1)=31$, $p=2e\text{-}8$***
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]	$\chi^2(1)=0.17$, $p=0.68$ (NS)	$\chi^2(1)=36$, $p=2e\text{-}09$ ***
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]	$\chi^2(1)=0.01$, $p=0.92$ (NS)	$\chi^2(1)=3.7$, $p=0.055$ (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]	$\chi^2(1)=4.03$, $p=0.044$*	$\chi^2(1)=11.7$, $p=0.0006$***
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]	$\chi^2(1)=8.6$, $p=.0034$**	$\chi^2(1)=6.4$, $p=0.012$ *
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]	$\chi^2(1)=4.5$, $p=0.033$ *	$\chi^2(1)=0.044$, $p=0.84$ (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]	$\chi^2(1)=24$, $p=9e\text{-}7$***	$\chi^2(1)=22.5$, $p=2e\text{-}06$***
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]	$\chi^2(1)=0.0069$, $p=0.93$ (NS)	$\chi^2(1)=14.9$, $p=0.00012$***
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]	$\chi^2(1)=2.55$, $p=0.11$ (NS)	$\chi^2(1)=13.3$, $p=0.00026$***
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]	$\chi^2(1)=2.66$, $p=0.10$ (NS)	$\chi^2(1)=2.46$, $p=0.12$ (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 *****

Figure 4.4 Mean accuracy by condition of force by child. The black dashed line corresponds to the adult baseline (participants' accuracy on adults' modal productions, which is approximately the same for possibility and necessity contexts: respectively, 81.5% and 82%). The black line indicates chance level.



Contexts in (38) to (41) illustrate some examples of worst and best performance for *have to* and *can*.

(38) Two necessity contexts with the lowest accuracy:

Becky, *have to* (0% accuracy)	**Carl, *have to* (7.7% accuracy)**
MOTHER: yeah.	MOTHER: it's not.
MOTHER: I'm not sure there is another man actually.	MOTHER: what color is it?
MOTHER: I think it's you that's got the other man.	CHILD: its pink.
MOTHER: there is more of the train here though, look.	MOTHER: that's right.
CHILD: no.	MOTHER: its pink.
CHILD: not.	CHILD: you can hold it.
CHILD: this man can't go on the grass.	MOTHER: thank you.
CHILD: I _____ go on the grass.	CHILD: you _____ make them.

(39) Two necessity contexts with the highest accuracy:

Becky, *have to* (100% accuracy)	**Domin, *have to* (100% accuracy)**
MOTHER: Michael was eating your lunch?	CHILD: choooo.
CHILD: yeah.	CHILD: that one does.
MOTHER: oh dear.	MOTHER: it goes with that one over there.
MOTHER: naughty Michael.	CHILD: mhm.
CHILD: oh.	MOTHER: near your red and pink car.
CHILD: he's sick now.	MOTHER: can you see it?
MOTHER: oh dear.	CHILD: yeah.
CHILD: he _____ go to hospital.	CHILD: _____ be careful.

(40) Two possibility contexts with lowest accuracy:

Warr, *can* (14.7% accuracy)	**Warr, *can* (27.3% accuracy)**
MOTHER: it's a good little flower, isn't it?	MOTHER: oh no.
MOTHER: can Mummy have a go at making one?	MOTHER: more work for Mummy.
CHILD: I have got some more to get in.	MOTHER: help.
MOTHER: no.	CHILD: I'm going to get you.
MOTHER: I'll just use this piece.	MOTHER: I can't get you off me.
MOTHER: I'll just use this piece thank you.	MOTHER: help.
CHILD: no.	MOTHER: oh those slavering jaws.
CHILD: _____ find some more.	CHILD: _____ get up now.

(41) Two possibility contexts with the highest accuracy:

Liz, *can* (100% accuracy)	Gail, *can* (100% accuracy)
MOTHER: yeah.	CHILD: ...
MOTHER: he got his hair cut, didnt he?	MOTHER: we might go to Run Riot with baby James.
MOTHER: and there's his mouth.	MOTHER: what do you think?
MOTHER: and he's got a beard, hasn't he?	MOTHER: will that be fun?
MOTHER: like that.	MOTHER: hm?
CHILD: and I do some eyes, Mummy.	CHILD: bye bye train.
MOTHER: okay.	CHILD: ...
CHILD: so he _____ see.	CHILD: _____ move out the way please?

4.4.3 Children's mastery: Relation with the linguistic input

Correlation analysis. Table 4.6 summarizes the results of correlation analyses testing the relation between the different input factors under consideration (reported in **Table 4.2** and **4.3**), and children's performance, as measured in the experiment (**Table 4.4**). We first run correlations between the input factors and children's mastery of necessity and possibility modals, and second, between the input factors and *change* in children's mastery over the time period (even though these post-hoc analyses are limited by the extremely reduced age range we sample from in this study).

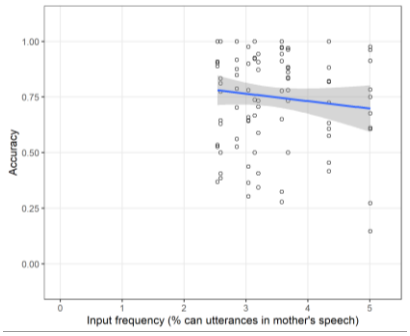
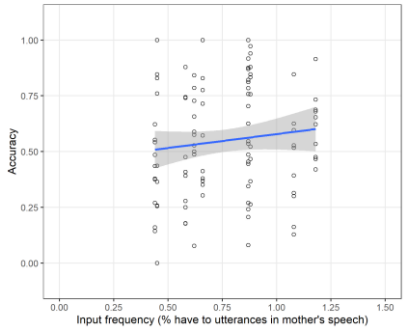
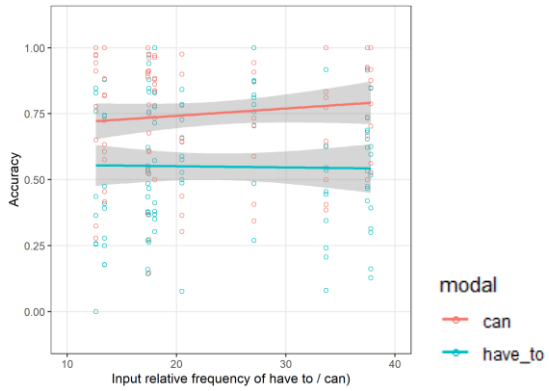
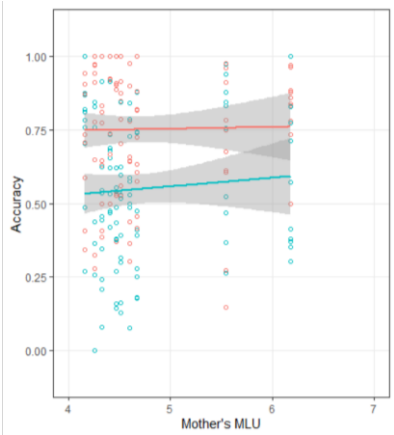
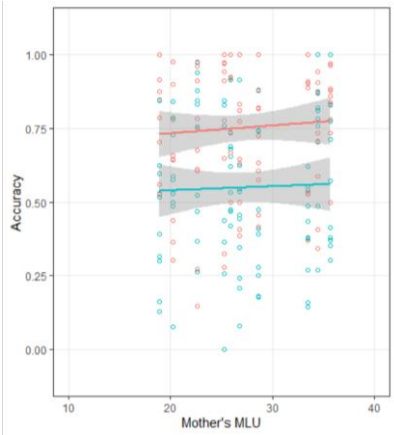
Quantitative measures. **Effect of frequency**. We do not find an effect of frequency, neither for possibility nor for necessity modals, regardless of whether we look at overall frequency or relative frequency (see **Figures 4.5a, 4.5b** and **4.6**). **General qualitative measures.** We find no effect of mothers' MLU and mother's talkativeness (see **Figures 4.7a** and **4.7b**). **Targeted qualitative measures.** For necessity modals, we find an effect of modal talk (r=0.22, t(98)=2.2, *p=0.027* **; 95% CI: [0.026 ; 0.40], see **Figure 4.9a)**, an effect of modal *necessity* talk (r=0.24, t(98)=2.4, *p=0.018***; 95% CI: [0.043; 0.41], see **Figure 4.9b**), and a strong effect of negation (r=0.32, t(98)=3.35,

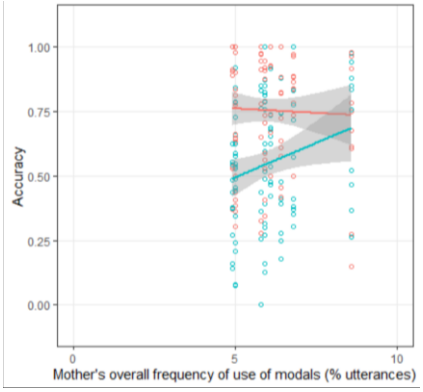
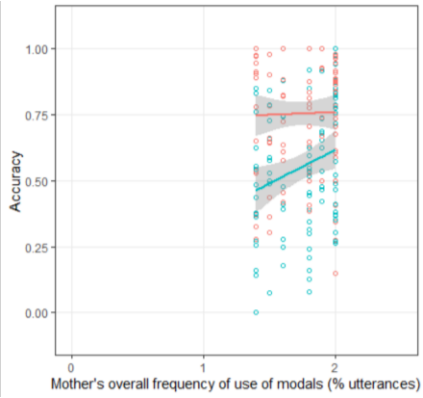
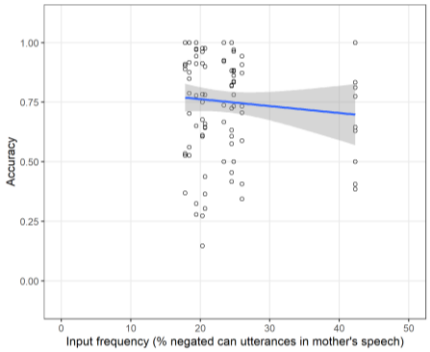
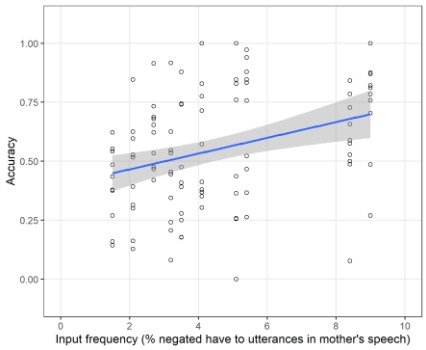
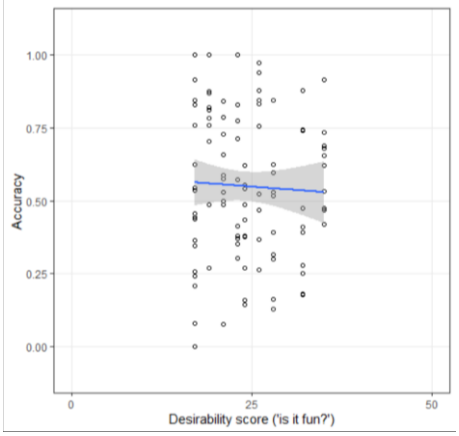
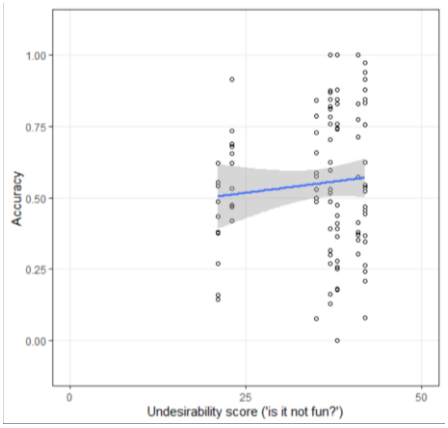
*p=0.00114****; 95% CI: [0.13; 0.49]; see **Figure 4.9b**). We find no effect of the strength of the (un)desirability signal (see **Figures 4.10a** (desirability score) and **4.10b** (undesirability score)). For possibility modals, we find no effect of either of these factors (see **Figures 4.8a, 4.8b, 4.9a, 4.9b** and **4.10a)**, which could be explained by performances being at ceiling.

Table 4.7 Effect of input measures on general accuracy: Results of correlation analysis (Pearson's r):

			Possibility modals *(can/able)*	**Necessity modals *(have to)***
Input measures	**Quantitative measures**	'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]
	Qualitative generic 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]
	Qualitative specific measures	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]
		Mothers' necessity modal talk	r=0.021, t (98)=0.2, p=0.83 (NS) 95% CI: [-0.18; 0.22]	r=0.24, t (98)=2.4, p=0.018** 95% CI: [0.043; 0.41]
		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]
		Desirability score	*Not coded.*	r=-0.43, t (98)=-0.43, p=0.67 (NS) 95% CI: [-0.24; 0.15]
		Undesirability score	*Not coded.*	r=0.089, t (98)=0.88, p=0.38 (NS) 95% CI: [-0.11; 0.28]

Figure 4.5 to 4.10 Relation between input factors (x-axis) and accuracy in the GF Experiment 2 (y-axis). When possible, *can* (in red) and *have to* (in blue) appear on a single plot; however, in cases where range of frequencies differ, I use separate plots. Red dots correspond to frequency of *can*; blue dots correspond to frequency of *have to*.

5	**5a** Overall frequency for possibility modals (NS):	**5b** Overall frequency for necessity modals (NS):
		
6	Relative input frequency for possibility and necessity modals (NS):	
		
7	**7a** Mothers' MLU (NS):	**7b** Mothers' talkativeness (NS):
		

8	**8a** Mothers' modal talk: (*have to*: r=0.22) 	**8b** Mothers' necessity modal talk: (*have to*: r=0.24) 
9	**9a** Negation for possibility modals (NS): 	**9b** Negation for necessity modals (r=0.32) 
10	**10a** Desirability score (NS): *Coded for necessity modals only.* 	**10b** Undesirability score (NS): *Coded for necessity modals only.* 

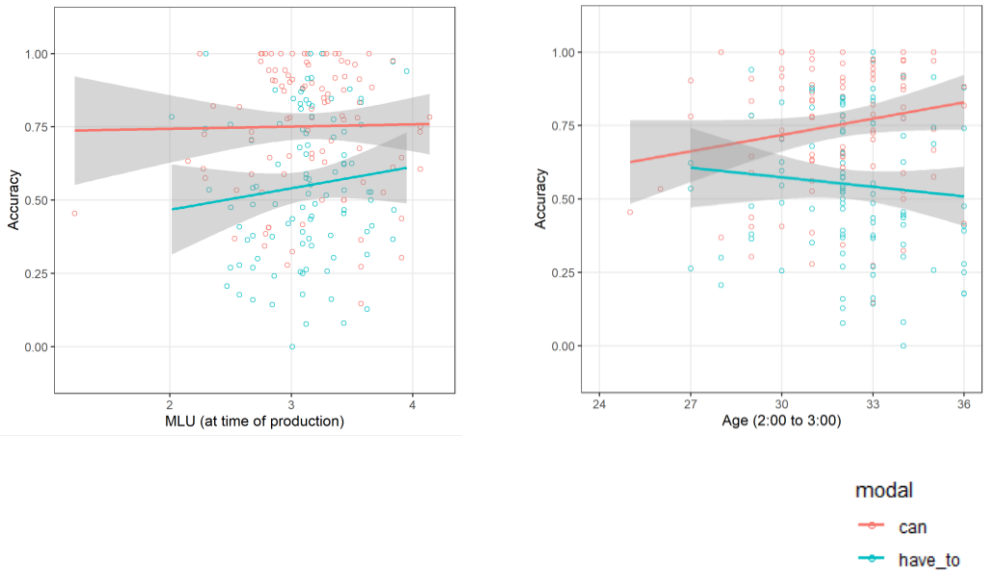
Effect of Age and children's MLU. We find no significant effect of Age and MLU (see **Table 4.8** and the corresponding **Figures 11a** and **11b**). The effect of age is almost significant for possibility modals (r=0.18, t (98)=1.85, p=0.067 (NS)), but for necessity modals, if anything, uses tend to become less adult-like. Note that we do find the expected relation between children's Age and MLU: r=0.16, t (198)=2.29, p=0.023**; 95% CI: [0.02; 0.29]).

Table 4.8 Effect of children's factors: age of production and MLU at the time of production.

	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]

Figure 4.11 Relation between output factors (x-axis) and accuracy in the GF Experiment 2 (y-axis):

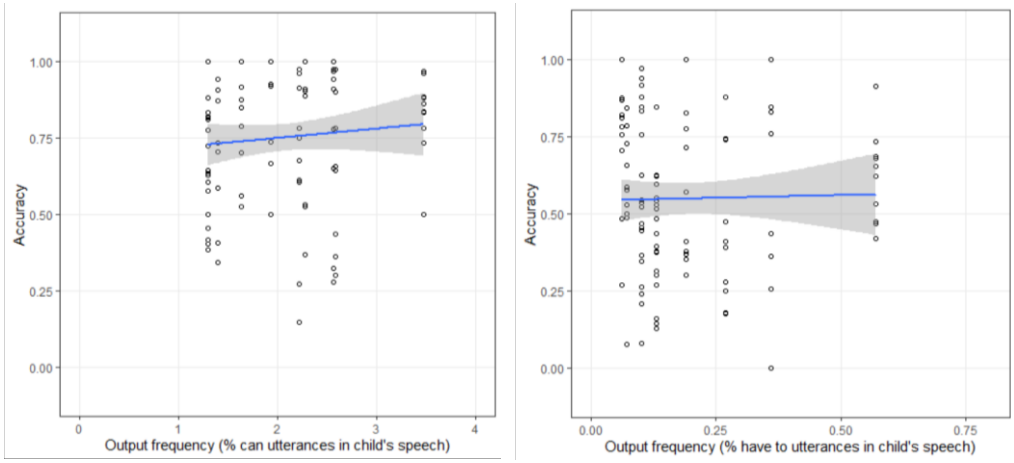
11	**11a** Relation between MLU and accuracy for possibility and necessity modals:	**11b** Relation between Age and accuracy for possibility and necessity modals:



We also looked at whether we find a relation between our different measures of children's mastery: their own frequency of production, and performance measured in the experiment. We find no relation between frequency in children's speech and their accuracy in the experiment: children who produce more *can/have to* are not necessarily the ones that use them in a more adult-like way.

Figure 4.12 Relation between output frequency (x-axis) and accuracy in GF Experiment 2 (y-axis):

12	**12a** Relation between frequency in child's speech and accuracy for possibility modals:	**12b** Relation between frequency in child's speech and accuracy for necessity modals:



Measure of change (post-hoc). We then looked at whether the different input factors might affect *change* in children's mastery of necessity and possibility modals (i.e., whether child usage improves over the time period). However, note that these analyses are limited in important ways, given the extremely reduced age range we are sampling from (2- to 3-year-old), which is even more restricted since as reported in Chapter 3,

necessity modals almost do not occur between 2 and 2:06 (see also ages of first productions reported in **Table 4.6**), and the fact that we sampled randomly over the corpus data, which makes it less easy to compare between children.[87]

Figure 4.13 shows learning curves for the 10 children based on age, for possibility and necessity modals. To measure children's improvement, we used Pearson's r coefficients (reported in **Table 4.9**), which correspond to the slope of the curves for each child in **Figure 4.13**). Results of the correlation tests between the different input measures (reported in **Table 4.2** and **4.3**) and the slope are reported in **Table 4.9**.[88] There is no significant effect.

Table 4.9 Slope of the learning curves (Pearson's r) for each child, based on the correlation coeffeicients of the relation between Age and Accuracy (mean of participants accuracy on the 10 contexts tested for each child). r > 0 indicates an improvement with age; r < 0 indicates a decrease in performance. Note that given the restricted age range, these measures are not extremely reliable.

	possibility modals	necessity modals
anne	0.35	-0.46
aran	0.22	0.18
becky	0.17	-0.12
carl	0.42	0.15
domin	0.25	-0.54
gail	0.18	0.53
joel	0.14	0.08
liz	-0.18	0.03
nic	0.24	-0.40
warr	-0.34	0.36
ALL	0.18	-0.09

[87] In typical studies that use such measures of change (see in particular Hoff, 2003; Rowe, 2008, who test the effect of various input measures/ on child vocabulary development), the child/mother dyads are recorded at two different times points, in a more controlled way. For instance, Hoff (2003) tested 63 dyads, 10 weeks apart; Rowe (2008) tested 47 dyads, first recorded at age 2:6 and then at age 3:6.

[88] We excluded Nic from the analysis for necessity modals because the slope of her learning curve is probably not representative of an actual drop in performance, but more likely due to the reduced age range sampling problem already described.

Figure 4.13 Relation between age of production of possibility (in red) and necessity (in blue) modals for each child. *r* coefficients are reported in **Table 4.8.**

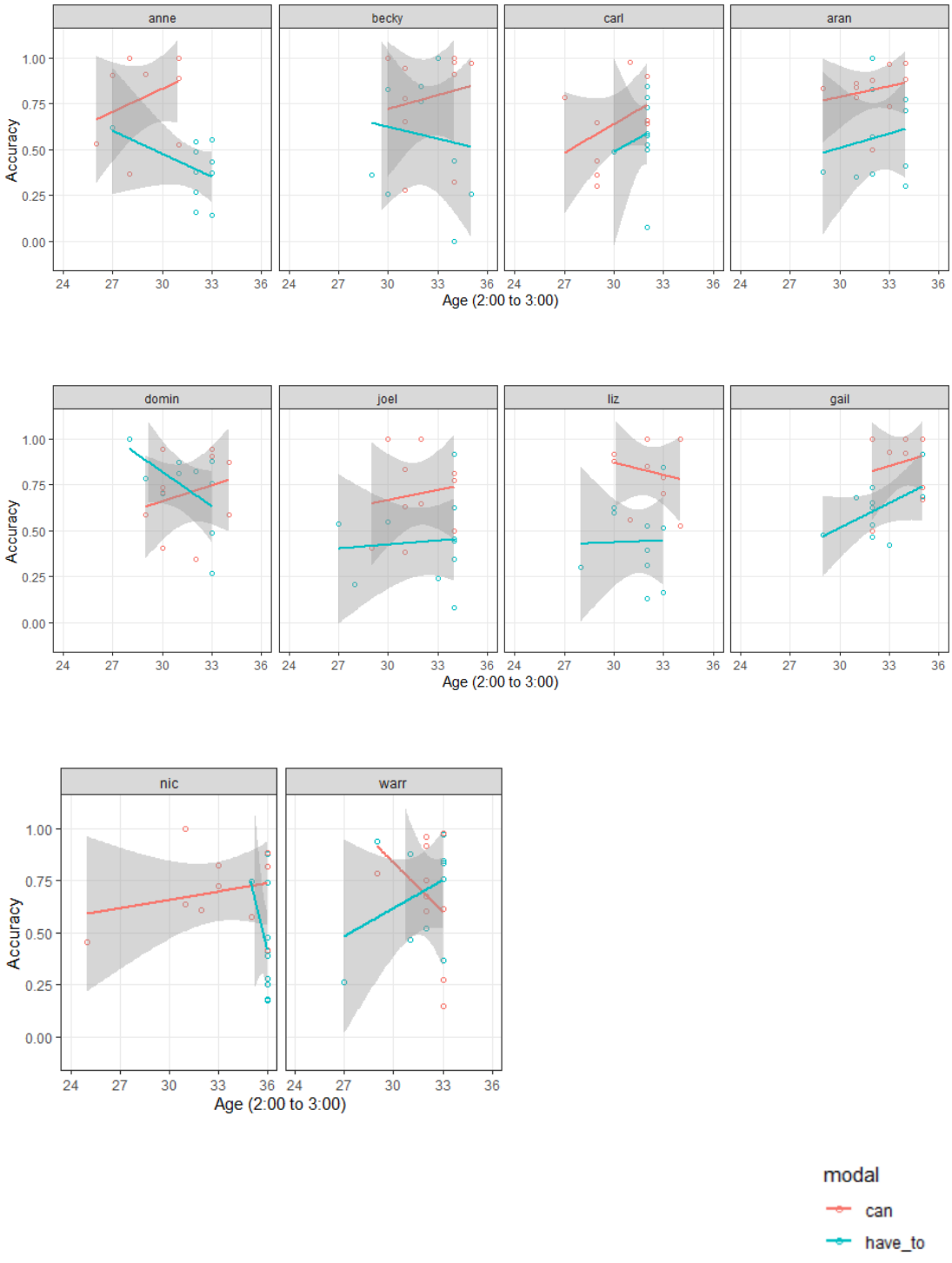


Table 4.10 Effect of input measures on change in mastery: Results of correlation analysis (Pearson's r) between input factors and the slopes of the curves, as reported in Table 4.9.

			Possibility modals *(can/able)*	**Necessity modals *(have to)***
Input measures	**Quantitative measures**	'Overall' frequency	r=-0.48, t (8)= -1.5, p=0.16 (NS) 95% CI: [-0.85; 0.21]	r=0. 50, t (7)= 1.5, p=0.17 (NS) 95% CI: [-0.25 ; 0.87]
		'Relative' frequency (*can* vs. *have to*)	r=-0.22, t (8)= -0.62, p=0.55 (NS) 95% CI: [-0.74; 0.48]	r=0.25, t (7)= 0.68, p=0.52 (NS) 95% CI: [-0.50; 0.78]
	Qualitative general measures	Mothers' MLU	r=-0.29 , t(8)= -0.85, p=0.42 (NS) 95% CI: [-0.78 ; 0.42]	r=0.42, t (7)= 1.2, p=0. 25 (NS) 95% CI: [-0.33; 0.85]
		Mothers' talkativeness	r=0.47 , t(8)= 1.5 , p=0.17 (NS) 95% CI: [-0.23; 0.85]	r=-0.50, t (7)= -1.5, p=0. 17 (NS) 95% CI: [-0.87; 0.25]
	Qualitative specific measures	Mothers' modal talk	r=-0.71 , t(8)= -2.8 , p=0.02 95% CI: [-0.92;-0.15]	r=0.46, t (7)= 1.36, p=0.22 (NS) 95% CI: [-0.30 ; 0.86]
		Mothers' necessity modal talk	r=-0.49 , t(8)=-1.6 , p=0.15 (NS) 95% CI: [-0.85; 0.20]	r=0.35, t (7)= 0.98, p=0.36 (NS) 95% CI: [-0.41; 0.82]
		Negation	r=0.13 , t(8)=0.38 , p=0.72 (NS) 95% CI: [-0.54; 0.70]	r=-0.25, t (7)= -0.67, p=0.52 (NS) 95% CI: [-0.78;0.50]
		Desirability score	*Not coded.*	r=0.58, t (7)=1.9, p=0.10 (NS) 95% CI: [-0.14; 0.90]
		Undesirability score		r=0.13, t (7)= 0.34, p=0.74 (NS) 95% CI: [-0.59 ; 0.73]

4.5 General discussion

The goal of this study is to start addressing the question of how children learn modal force *in practice*, by relating characteristics of their linguistic input to their mastery of modals. We focused on root modals (*can* vs. *have to*), since they are the first children produce and as a consequence, the only ones we could test at such an early age range (2- to 3-year-old).

Let's summarize our results. First, we replicate our findings from our first GF experiment on children's productions, looking at a larger sample (2*100 contexts). Again, we find a clear asymmetry in children's mastery of modal force: children seem to use possibility modals in an adult-like way, but not necessity modals. Looking more closely at individual patterns, we find that almost all children start producing *have to* after they produce *can*, with on average a 4-month difference. The experiment shows that for children's possibility modals, participants perform significantly above chance for all of the 10 children we could test, whereas for necessity modals, they are at chance for 4 of them, perform significantly above chance for 5, and below chance for 1 (Anne). We find that 7 children are not adult-like for necessity modals (vs. 4 for possibility modals). It is difficult to draw strong conclusions given the small size of the sample (10 children), but these results suggest that children might not all face the same difficulties with necessity modals: some (Warren and Dominic) may master them before 3-years-old.

Importantly, these results also show that no child clearly treats *have to* as a possibility modal: participants are at chance, but don't overguess possibility modals as they might do if children were using *have to* in the same situations as *can*. If children's struggles with necessity modals are partially semantic in nature, this seems to support the 'weak' version of the hypothesis (according to which children are just unsure about the meaning of necessity modals), rather than its 'strong' version (which claims they are mistaken, and think necessity modals mean possible).

We find no clear effect of age, but this might be because of the narrow age range, which may not be large enough to see an effect (presumably, age matters

eventually, since we find a difference between children and adults). Looking at variation between older children seems a promising avenue for future research.

Turning to the main goal of this study—to see what factors, in children's linguistic input, influence their modal mastery—, we find an effect of negation on children's performance for *have to*: children more exposed to *don't have to* in their input seem to master its force earlier. This is an important result, but might seem puzzling at first, given our discussion from Chapter 2. The fact that these cases are extremely rare is not the main problem: if children are able to use these highly informative cases, they might be able to learn even with few instances. But the issue is that it seems that in most cases, they are not used in a 'logically' informative way by adults, but rather, to convey prohibitions. How do children interpret these uses, and how could they exploit them when learning? The existing acquisition literature suggests that children are sensitive to the speech acts speakers performs quite early on (see e.g. Bloom, 2002; Clark & Amaral, 2010, a.o.), potentially more than they are able to objectively label a situation as possibility or necessity. So, is the effect of negation we find here due to children 'logical' use of negative environments to infer the force of *have to*, following Gualmini & Schwarz (2009)? Or is negation useful in another way? Alternatively, negation could help learners via prosody, or because it allows to put focus on necessity modals—i.e., for a very different reason than what Gualmini and Schwarz's (2009) suggest.

Even if it does not tell the nature of the mechanism at stake, this result suggests that children rely on negative environments to learn the force of *have to*. A question to address in the future is thus the extent to which this effect of negation generalizes to

other necessity modals, especially necessity modals that systematically outscope negation like *must* or *should.* As we discussed in Chapter 2, for these necessity modals, using negation could be misleading for children if they are to reason logically: do we find that the frequency of use of negation in their input leads to lower performance for *must*? Here, we didn't run the experiment on children's *must* (condition ROOT-AFF-1 in our first GF experiment on children's productions) because of how infrequent they are (only 4 children produce more than 10 *must* utterances in the whole corpus), but this could be an interesting follow up, maybe using a different corpus.

Note that we also proposed that negation could help to figure out possibility modals, as uses of *can't* are frequent in input, and context is particularly informative about their force. Here, we find no effect of negation for *can*, but this might be because children's performance is almost at ceiling: even if negation was helpful for possibility modals, we might not be able to detect it because we are looking at children that are already too old.

Our second main result is that we find that, while mere quantity of exposure to particular lexemes does not seem to affect children's mastery, there is an effect of modal talk in their input: children more exposed to modal talk in general seem to master *have to* earlier. This effect could come from modal talk making possibilities and necessities more salient, maybe showing that they are notions that can be talked about, and deserves to be explored further. It may also relate to the question of how knowing a dual 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 on modal learning).

We find no significant effect of frequency of exposure (i.e., the quantity of input) on children's mastery, be it relative to all mothers' utterances, or relative to the other modals. There is a small trend in the expected direction for necessity modals, which would suggest that children who hear them more often master them earlier, but it is not significant. This might suggest that what matters most is not quantity of exposure so much as quality of exposure, in line with the literature on other cases of word learning reviewed at the beginning.

Last, we find no evidence that the strength of the 'undesirability' signal in mothers' speech matters (i.e., how frequently mothers use necessity modals with undesirable prejacents). However, this (absence of) result is as yet tentative, as for reasons of time, we didn't run an actual experiment to assess desirability, as in Chapter 2, and coded desirability for necessity modals.[89]

Let's come back to what consequences these results might have regarding our different possible explanations for children's struggles with force. Why do children struggle specifically with necessity modals, and use them (and maybe comprehend them) in situations where adults would prefer using possibility modals? As discussed in Chapter 3, there might be several explanations for children's difficulties. Our results from this study are still compatible with all of these explanations: children's use of modals could be symptomatic of deployment issues, rather than wrong representations.

[89] Ideally, we would want to test more generally whether children do make use of contextual cues. To probe this, we intend to measure whether the general informativity of the situational context is predictive of children's modal mastery, by seeing whether children who master necessity modals early have mothers who use highly guessable modals from context, using the HSP paradigm but testing directly differences between mothers. We plan on investigating this in future research.

An interesting follow-up that might allow us to assess more directly the question of whether children know the underlying force of their modals, will be to see whether we find a relation between desirability and force in children's own modal productions: this would suggest that children know the underlying meaning, and that their struggles are in deploying them in the right situations.

Even though this study had intrinsic limitations, due to the low sample size (12 children) and restricted age range (2- to 3-year-old), it also can serve as a proof of concept that this methodology allows to identify factors that matter in the learning process. In future work, we plan to expand to a wider age range, and explore further cues that might matter. A first step in this direction will be to use a measure for each parent of how "good" their modal productions were, rather than focusing on the strength of the desirability in their speech.

Chapter 5. Conclusion

When, and how, do children figure out the force of modals? How do they learn that words like *can, may* or *might* express possibility, whereas words like *must, should* or *have to* have a stronger meaning, and express necessity?

In **Chapter 1**, we saw that what might make this mapping particularly challenging is that necessity entails possibility, which creates a Subset Problem for necessity modals. We considered three "solutions" to this Subset Problem. The first one was for learners to have a bias towards strong (necessity) meanings, in the spirit of Berwick (1985). The second one was for them to rely on downward-entailing environments, which reverse patterns of entailment (Gualmini and Schwarz 2009). The third one was that children use cues from the conversational context, if speakers use possibility and necessity modals in distinct situations, in ways that can be informative to children. Our goal was to assess the viability of each of these solutions, by studying in detail the linguistic input of English-learning children.

In **Chapter 2**, we thus explored in depth the speech young English-learning children get exposed to, on the basis of a detailed corpus study of their linguistic input and three experiments based on the corpus. How often do children hear possibility and necessity modals? Are they often used with negation? How informative is the conversational context about their force? The first two experiments (*Input Experiment 1* and *Input Experiment 2*) allowed us to assess the general informativity of conversational contexts, by asking adult participants to guess a modal blanked out from an adult's sentence in dialogue extracted from the corpus, using a variant of the Human

Simulation Paradigm (HSP) (Gillette et al. 1999). In the first experiment, the blanked modal statement is presented in context; in the second experiment, it is presented without context. The last experiment (*Input Experiment 3*) explored further one specific situational cue for root modals, namely the desirability of the event described by the prejacent.

Our results show that the conversational context is highly informative as to both forces: speakers use possibility and necessity modals in distinct situations, and adults are able to recover a modal's force solely on the basis of this conversational context. This means that, if children are sensitive to the same conversational cues as adults, they can in principle use them to figure out modal force. The nature of these conversational cues could be quite diverse, and might vary with modal flavor. In *Input Experiment 3*, we focused on the hypothesis that (un)desirability might matter to figure out root necessity modals. The results show that such a cue is indeed available in the input: root necessity modals are more often used with predicates that describe events perceived as undesirable than their possibility counterparts. If children expect necessities to correlate with undesirable events, they might use desirability to figure out the force of root modals.

The second main take-away from our input study is that using evidence from negative environments cannot be a general solution to the Subset Problem. Uses of negated necessity modals are either, depending on the modal, potentially informative but extremely rare (e.g., *don't have to*), or potentially misleading because of the scope irregularities between necessity modals and negation (e.g. *mustn't ~ can't*): the question of how children figure out the right scope relations for negated modal

sentences, and the potential polarity restrictions of modals, is an open one. Evidence from negative environments might be helpful for children to figure out the force of possibility modals: negated possibility modals (e.g., *can't*) are frequent in the input, and context is highly informative about their force.

Our input study results show that in principle, the conversational context may be sufficient to pick up on modal force, and thus, that learners may not need to rely on negative environments, nor on a necessity bias. But how do children actually figure out modal force? Are they able to use cues from the conversational context or negation?

To address this question, we first asked when children figure out modal force, and explored in **Chapter 3** the spontaneous modal productions of 2- to 3-year-old English children, again using a combination of corpus analysis and experimental methods. How early do children use possibility and necessity modals? And do they use them appropriately?

We find what seems to be a "Necessity Gap". Children master possibility modals early: at age 2, they use them frequently and productively (both with and without negation), and in an adult-like way (they do not use them in necessity situations). However, they seem to struggle with necessity modals: they produce these later on, much less frequently, hardly ever with negation, and often, in a non adult-like way: our GF experiment shows that they use them in situations where adults would prefer using possibility modals, and when negated, in situations where adults would prefer negated possibility modals. These results cast a new light on prior results from comprehension experiments. If this difficulty with necessity modals persists into the preschool years, it could help understand, maybe explain, the reported struggles with

both forces: both children's tendency to accept possibility modals in necessity contexts (as they may lack a relevant stronger alternative), and necessity modals in possibility contexts (as they may not be sure that these modals express necessity).

There are several possible explanations for the nature of children's difficulties with necessity modals. It remains unclear whether they stem from not knowing their underlying force, or whether children have successfully learned their force, but either have conceptual difficulty, pragmatic difficulty (deploying them in the right situations), or epistemic difficulty (judging what is, or can be said to be, necessary).

Our input study shows that a number of factors might combine to make the underlying force of necessity modals potentially more challenging to learn: first, their lower frequency in the input, compared to possibility modals; second, the way these modals interact with negation, especially for necessity modals like *must* which outscope negation; and finally, the logical Subset Problem, if it is a real one for learners.

How do children eventually figure out the force of necessity modals? In **Chapter 4**, we started addressing this question, by correlating differences in children's input to differences in their mastery of modals. Focusing on root modals *can* and *have to*, we found that two input factors are linked to children's mastery, and thus may play a role in the learning process. First, children who hear *have to* with negation more frequently appear to master it earlier. This is an interesting result, given our discussion of how children might solve the Subset Problem using downward-entailing environments. And this finding may at first appear paradoxical: we saw in Chapter 2 that the informativity of the average context of *don't have to* was low, and even more problematically, that most of these uses seem to correspond to cases which convey

prohibitions. How do children exploit negation? Do they reason 'logically', by seeing it used in non-necessity contexts? Or is a different mechanism at stake, involving for instance prosody? This is an open question, to be explored in the future.

Second, we found that while hearing more *have to* does not affect its mastery, exposure to modal talk in general matters in the learning: children exposed to more modal talk in general seem to master *have to* earlier. This effect deserves further exploration as well, and could come from modal talk making possibilities and necessities more salient and showing that they are notions that can be talked about, or because of the contrast with 'duals' helping figuring out the force of a single modal.

English has a rich modal inventory, where necessity modals are overall less frequent than possibility modals. From there, an important question is to see how these results generalize to other languages than English. How general are the difficulties we find with necessity modals? Can we really speak of a Necessity *Gap*, or is it just a natural consequence of other factors? In the future, I intend to look at other languages like French, which has a more restricted modal inventory than English, but where necessity modals interact differently with negation, and might be used differently by speakers. The small amount of reported corpus data suggests that in French, necessity modals might be more frequent in parents' productions than they are in English (Jeretič, 2020). This would allow us to assess, in particular, the role of frequency of exposure on children's modal proficiency.

Another important issue is how children learn the difference between 'weak' and 'strong' necessity modals. As mentioned in Chapter 1, necessity modals are often split into strong (*must*) and weak (*should*) necessity modals (von Fintel and Iatridou

2008); nouns (*slight possibility*) and adjectives (*likely*) can even encode finer-grained strength distinctions. Here, I focused on the main contrast between possibility and necessity modals, but the question of how children learn these finer-grainer contrasts needs to be explored.

To conclude, there are several reasons why learning the force of necessity modals might be challenging for learners. Some are specific to English—in particular, their low frequency. Some are more general, and might arise for all learners: the logical Subset Problem, and the way they interact with negation for some modals, since it is not only in English that we find necessity modals that outscope negation. Before we end, I will highlight a particular additional issue for learning necessity modals, that comes from the flavor variability of modals.

As discussed in Chapter 1, the flavor variability of English modals could in principle make the learners' task easier, in particular, if they expect a modal to always express the same force. For instance, having figured out that *must* expresses deontic necessity could allow them to infer, by extension, that it also expresses necessity when used to express epistemic flavor. If the contrast in force is easier to grasp for some flavors (e.g. deontic) than for others, this might be particularly helpful.

However, for the learner, the salience of ability interpretations of modals might create an additional issue to learn force. This is another kind of Subset Problem, but this one is not a matter of pure truth-conditional logic. The problem is not that necessity *logically* entails possibility (i.e., if 'you must$_{\text{deontic}}$', then, 'you can$_{\text{deontic}}$'); it is that in

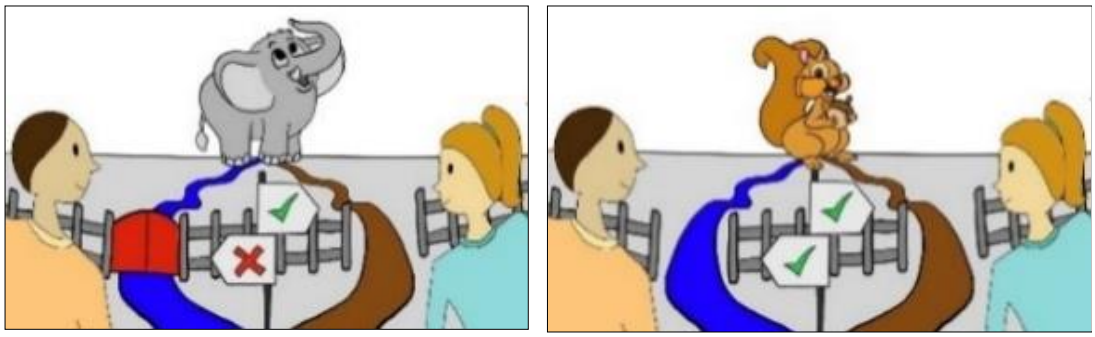
most cases when a necessity root modal is used (be it deontic or teleological), an ability statement would hold as well: when "you must$_{deontic}$", then "you can$_{ability}$".[90]

One argument that supports this additional possibility for why children may struggle figuring out the meaning of their necessity modals comes from a Novel Word learning experiment conducted with adults (Dieuleveut, Cournane & Hacquard, 2020). In this experiment, (adult) participants were asked to learn various novel modals for different 'flavors' of modals: teleological (goal-based) versus epistemic (knowledge-based) (e.g. "We **sig** go down the blue road"; "The keys **gleeb** be in the blue box"), in various situations of possibility and necessity, as illustrated in **Figures 5.1a-b** and **5.2a-b**. Results show that, both when tested in epistemic and teleological scenarios, adults behave as expected when learning these new modals in 'possibility' situations (**5.1b**/**5.2b**): they accept these novel modals when tested in 'necessity' situations (**5.1a**/**5.2a**). However, when they learn new modals in 'necessity' situations, they successfully learn the force (i.e., they reject them when tested in 'possibility' situations) for epistemic scenarios only; with teleological scenarios, they accept them in 'possibility' situations.

[90] See Kant (1781): "The action to which the "ought" applies must indeed be possible under natural conditions."

Figure 5.1 Visual stimuli and sentence frames used in Dieuleveut et al. (2020), for teleological condition, by situation type:

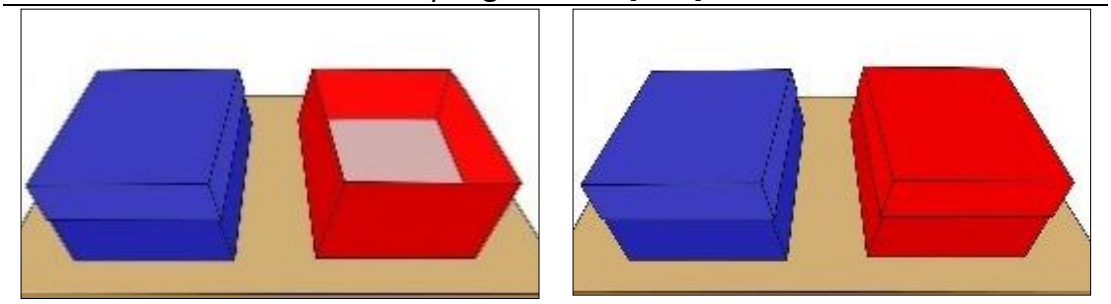
'We **sig** go down the [blue] road.'



5.1.a NECESSITY situation **5.1.b** POSSIBILITY situation

Figure 5.2 Visual stimuli and sentence frames used in Dieuleveut et al. (2020), for epistemic condition, by situation type:

'The keys **sig** be in the [blue] box.'



5.2.a NECESSITY situation **5.2.b** POSSIBILITY situation

These results are particularly interesting for us, as they open up a new possibility for what might make the meaning of necessity modals potentially challenging for children. Scenarios like the one illustrated in **Figure 5.1.a** make an ability interpretation salient: the question of whether it is 'possible or not' to go down the yellow road might be more relevant than whether it is 'possible or necessary' to use this road to get to their goal. In epistemic scenarios, the same problem may not arise (at least in the scenarios used in this experiment), as competition with an ability

interpretation is less likely; but as we saw, epistemic modals are rare in children's actual input, especially necessity ones. Adult learners' behavior in the teleological condition might be explained by differences in perspectives between them and the experimenter. This overlap in modal flavor, specifically, this competition with an ability interpretation when speakers use root necessity modals, could also contribute to children's difficulties with necessity modals we have described: If children tend to interpret situations as ability situations where parents intend a teleological necessity statement, they could lexicalize a possibility meaning for necessity modals. How compatible with an ability interpretation are natural occurrences of epistemic and root necessity modals, in children's actual input? This is another question we intend to investigate further.

Appendix A: Corpus results

Table (i). Counts and percentages of modal uses by force and flavor, for adults and children (excluding tags and repetitions), breakdown by lemma.

	ADULT (n=18,853)				CHILD (n=4,800)			
	root		epistemic		root		epistemic	
ALL	**17187**	**91.2%**	**1665**	**8.8%**	**4686**	**97.6%**	**114**	**2.4%**
POSSIBILITY	**12175**	**90.2%**	**1324**	**9.8%**	**3705**	**97.6%**	**93**	**2.4%**
can	10742	99.7%	37	0.3%	3619	100%	1	0%
might	*NA*		1154	100%	*NA*		80	100%
could	1096	90.4%	117	9.6%	79	91.9%	7	8.1%
able	315	100%	*NA*		3	100%	*NA*	
may	22	57.9%	16	42.1%	4	44.4%	5	55.6%
NECESSITY	**5012**	**93.6%**	**341**	**6.4%**	**981**	**97.9%**	**21**	**2.1%**
have to	2392	99.7%	6	0.3%	351	99.7%	1	0.3%
got to	930	99.3%	7	0.7%	288	100%	0	0%
should	641	92.1%	55	7.9%	19	90.5%	2	9.5%
need to	493	100%	*NA*		217	100%	*NA*	
supposed	326	97.3%	9	2.7%	9	100%	0	0%
must	147	35.8%	264	64.2%	96	84.2%	18	15.8%
ought to	83	100%	*NA*		1	100%	*NA*	

Table (ii). Embedding under attitude predicates (adults). The most frequent embedding verbs are *think*, *see* and *know*. *think* is very frequently used to embed epistemic modals (24.7% of possibility epistemic modals, 11.7% of necessity epistemic modals). *see* never embeds epistemic modals, but is quite frequent with possibility root modals. In children, we found 35/4800 cases of modals embedded under attitude predicates (*think*: 22, *see*: 4, *know*: 3, *say*: 3; *ask*, *bet*, *wish*).

	ADULT (n=18,853)							
	root				epistemic			
	possibility		necessity		possibility		necessity	
no embedding	11218	92.1%	4548	90.7%	968	73.1%	297	87.1%
think	298	2.4%	383	7.6%	327	24.7%	40	11.7%
see	427	3.5%	3	0.1%	0	0%	0	0%
know	55	0.5%	32	0.6%	9	0.7%	0	0%
say	45	0.4%	11	0.2%	2	0.2%	0	0%
be sure	28	0.2%	9	0.2%	0	0%	3	0.9%
suppose	18	0.1%	7	0.1%	5	0.4%	0	0%
bet	28	0.2%	0	0%	0	0%	0	0%
tell	8	0.1%	6	0.1%	1	0.1%	0	0%
wonder	6	0%	0	0%	5	0.4%	0	0%

Else: *mean, show, hope, expect, look like, wish, ask, presume, use to, be better, be determined, be insistent, believe, happen, have a feeling, have a look, insist, keep, like, make sure, pretend, reckon, want*

Appendix B: Experimental material

Input Experiment 1
→ http://spellout.net/ibexexps/modsquad/HSP_FC_epiP/experiment.html

You will read short excerpts from real conversations between mothers and their two-year-old children. In these conversations, there will be one or more words missing, indicated by _____. Complete the sentence by choosing the best of the two options below the conversation.
Pick the option that seems the most likely to correspond to what the mother actually said!

Here is an example:

MOTHER: time for a game.
MOTHER: what are they playing with?
CHILD: sand.
MOTHER: do you like to play with sand ?
CHILD: yeah.
MOTHER: what's that baby doing ?
CHILD: taking all the sand out.
MOTHER: and where's he _____ the sand?

putting	giving

The correct answer is "putting".

Input Experiment 2
→ https://spellout.net/ibexexps/modforce/modforce_FC0cxt_epiP/experiment.html

You are going to see short sentences. In these sentences, there are one or more words missing, indicated by _____. Your goal is to complete the sentence, by choosing the best of the two options. Pick the option that sounds the best to you!

Here is an example:

and where's he _____ the sand?

putting	giving

The best answer is "putting".

Sometimes, you will also be asked to solve simple additions or subtractions.

1 + 1 = _____

2	3

The right answer is "2".

Input Experiment 3

→ https://spellout.net/ibexexps/modforce/modforce_hspdesF_rootP1/experiment.html

You will see activities that came up in conversations between two year old children and their mothers. For each, say whether the activity sounds fun or not. Sometimes it might be hard to tell, but give your best guess.

Here is an example:

Doing a puzzle

Does this sound fun?

no	yes

Sometimes, you will also be asked to solve simple additions or subtractions.

1 + 1 = _____

2	3

The right answer is "2".

GF Experiment 1&2

→ https://spellout.net/ibexexps/modsquad/HSP_FC_dilch_rootP2/experiment.html

You will read short excerpts from real conversations between mothers and their two-year-old children. In these conversations, there will be one or more words missing, indicated by _____. Complete the sentence by choosing the best of the two options below the conversation.
Please answer based on what makes sense in the given context. Consider what *you* find most natural to fill the blank.

Here is an example:

MOTHER: are you tired now?
CHILD: take that elastic band off her.
MOTHER: would you like to go to bed?
MOTHER: Aran.
CHILD: take that elastic band off.
MOTHER: try please.
CHILD: please.
CHILD: I've been _____ all day, Anna.

working	giving

The correct answer is "working".

Bibliography

Abbott, B. (2008). Issues in the semantics and pragmatics of definite descriptions in English. *Reference: Interdisciplinary Perspectives*, 61–72.

Acredolo, C., & Horobin, K. (1987). Development of relational reasoning and avoidance of premature closure. *Developmental Psychology*, *23*(1), 13.

Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking Quality and Quantity of Parental Linguistic Input to Child Language Skills: A Meta-Analysis. *Child Development*, *92*(2), 484–501.

Asplin, K. N. (2002). *Can complement frames help children learn the meaning of abstract verbs?* University of Massachusetts Amherst.

Astington, J. W. (1993). *The child's discovery of the mind* (Vol. 31). Harvard University Press.

Austin, J. L. (1975). *How to do things with words*. Oxford university press.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, *10*(4), 533–581.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*(1), 40–62.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford university press.

Bascelli, E., & Barbieri, M. S. (2002). Italian children's understanding of the epistemic and deontic modal verbs *dovere* (must) and *potere* (may). *Journal of Child Language*, *29*(1), 87–107. https://doi.org/10.1017/S0305000901004925

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

Bindra, D., Clarke, K. A., & Shultz, T. R. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent" causal" and" logical" problems. *Journal of Experimental Psychology: General*, *109*(4), 422.

Bloom, P. (2002). *How children learn the meanings of words*. MIT press.

Bochnak, M. R. (2015). Variable force modality in Washo. *Proceedings of the Forty-Fifth Annual Meeting of the North East Linguistic Society (NELS 45)*, 105–114.

Brockmann, S., McConnell, S., Hacquard, V., & Lidz, J. (2018). Children's comprehension of pronouns and definites. *ZAS Papers in Linguistics*, *60*, 259–276. https://doi.org/10.21248/zaspil.60.2018.466

Brown, M. A. (1988). On the logic of ability. *Journal of Philosophical Logic*, *17*(1), 1–26.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Brown, R. (1973). *A first language: The early stages*.

Brown, R., & Hanlon, C. (1970). *1970: Derivational complexity and the order of acquisition in child speech. In Hayes, JR, editor, Cognition and the development of language. New York: John Wiley & Sons*.

Byrnes, J. P., & Duff, M. A. (1989). Young children's comprehension of modal expressions. *Cognitive Development*, *4*(4), 369–387. https://doi.org/10.1016/S0885-2014(89)90049-X

Cariani, F., & Santorio, P. (2018). Will done better: Selection semantics, future credence, and indeterminacy. *Mind*, *127*(505), 129–165.

Carnap, R. (1947). *Meaning and necessity; a study in semantics and modal logic*.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263–1266.

Chemla, E., & Singh, R. (2014). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373–386.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings of the 25th Boston University Conference on Language Development*, 157–168.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science Semantics Volume 3*. de Gruyter.

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, *8*(3), 425–504.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Clark, E. V., & Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Language and Linguistics Compass*, *4*(7), 445–457.

Clark, E. V., & MacWhinney, B. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition*, 1–33.

Coates, J. (1988). The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language*, *15*(2), 425–434.

Cournane, A. (2015). Revisiting the epistemic gap: Evidence for a grammatical source. *Proceedings of the 39th Annual Boston University Conference on Language Development*.

Cournane, A. (2020). Learning modals: A grammatical perspective. *Language and Linguistics Compass*, *14*(10), 1–22. https://doi.org/10.1111/lnc3.12399

Cournane, A. (2021). Revisiting the epistemic gap: It's not the thought that counts. *Language Acquisition*, 1–26.

Cournane, A., & Pérez-Leroux, A. T. (2020). Leaving Obligations Behind: Epistemic Incrementation in Preschool English. *Language Learning and Development*, *16*(3), 270–291. https://doi.org/10.1080/15475441.2020.1738233

Crain, S. (2012). *The emergence of meaning*.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. *Perspectives on Sentence Processing*, 443–467.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar*. Cambridge, MA: MIT Press.

De Haan, F. (1997). *The interaction of modality and negation: A typological study*. Taylor & Francis.

De Haan, F. (2011). Disambiguating modals: Constructions and must. *Oracle, Ms*.

De Villiers, J. (2007). The interface of language and theory of mind. *Lingua*, *117*(11), 1858–1878.

de Villiers, J., & Roeper, T. (2016). The acquisition of complements. In *The Oxford handbook of developmental linguistics*.

De Villiers, P. A. (2005). The role of language in theory-of-mind development: What deaf children tell us. *Why Language Matters for Theory of Mind, Apr, 2002, University of Toronto, Toronto, ON, Canada; This Chapter Originated from the Aforementioned Conference.*

Deal, A. R. (2011). Modals Without Scales. *Language*, *87*(3), 559–585. https://doi.org/10.1353/lan.2011.0060

Dell, F. (1981). On the learnability of optional phonological rules. *Linguistic Inquiry*, *12*(1), 31–37.

Dieuleveut, A., van Dooren, A., Cournane, A., & Hacquard, V. (n.d.). *Learning modal force: Evidence from children's production and input*. 10.

Dieuleveut, A., van Dooren, A., Cournane, A., & Hacquard, V. (2019). Acquiring the Force of Modals: Sig You Guess What Sig Means? *Proceedings of the 42nd Annual to the Boston University Conference on Language Development (BUCLD42)*.

Drubig, H. B. (2001). On the syntactic form of epistemic modality. *Ms., University of Tübingen*.

Dudley, R. (2017). *The role of input in discovering presupposition triggers: Figuring out what everybody already knew* [PhD Thesis]. University of Maryland, College Park.

FAGAN, S. M., & GERMANIC, C. T. I. (n.d.). Coates, Jennifer. 1983. The Semantics of the Modal Auxiliaries. London & Canberra: Croom Helm. *Die Unterrichtspraxis*, *10*, 2–73.

Fauconnier, G. (1975). Polarity and the scale principle. *Proceedings of Chicago*.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (n.d.). *MacArthur-Bates Communicative Development Inventories*.

Gazdar, G. (1979). A solution to the projection problem. In *Presupposition* (pp. 57–89). Brill.

Giannakidou, A. (2001). The meaning of free choice. *Linguistics and Philosophy*, *24*(6), 659–735.

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

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Gualmini, A., & Moscati, V. (2009). The early steps of modal and negation interactions: Evidence from child Italian. In *Romance languages and linguistic theory* (pp. 131–144). John Benjamins.

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

Hackl, M. (1998). *On the Semantics of" Ability Attributions"*.

Hacquard, V. (n.d.-a). *Logic and the lexicon: Insights from modality*.

Hacquard, V. (n.d.-b). *The grammatical category of modality*. 8.

Hacquard, V. (2011). 58. Modality. In *Handbücher zur Sprach-und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science* (pp. 1484–1515). De Gruyter Mouton.

Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language*, *34*(1), 73–96. https://doi.org/10.1111/mila.12192

Harrigan, K., Hacquard, V., & Lidz, J. (2018). Three-year-olds' understanding of desire reports is robust to conflict. *Frontiers in Psychology*, *9*, 119.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Heizmann, T. (2006). Acquisition of Deontic and Epistemic Readings of must and müssen. *UMOP*, *34*.

Hickmann, M., & Bassano, D. (2016). Modality and mood in first language acquisition. In *The Oxford handbook of modality and mood*.

Hirschberg, J. L. B. (1985). *A theory of scalar implicature*. University of Pennsylvania.

Hirst, W., & Weil, J. (1982). Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language*, *9*(3), 659–666. https://doi.org/10.1017/S0305000900004967

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433.

Hoffmann, T. (1966). Past tense replacement and the English modal auxiliary system. *Harvard Computation Laboratory Report NSF*, *17*.

Horn, L. (1972). On the semantic properties of logical operators in English (Doctoral dissertation). *University of California, Los Angeles*.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *89*(6), e572–e593.

Howard, A. A., Mayeux, L., & Naigles, L. R. (2008). Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language*, *28*(4), 375–402.

Hsu, N., Hadley, P. A., & Rispoli, M. (2017). Diversity matters: Parent input predicts toddler verb production. *Journal of Child Language*, *44*(1), 63–86.

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

Jasbi, M. (2018). *Learning Disjunction*. Stanford University.

Karmiloff-Smith, A. (1979). Micro-and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science*, *3*(2), 91–117.

Kenny, A. (1976). Human abilities and dynamic modalities. In *Essays on explanation and understanding* (pp. 209–232). Springer.

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. https://doi.org/10.1007/s10936-018-9573-8

Kratzer, A. (1981). Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, *10*(2), 201–216.

Kratzer, A. (1991). *Modality. Semantics: An international handbook of contempo-rary research, ed. by Arnim von Stechow and Dieter Wunderlich, 639–50*. Berlin: De Gruyter.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford University Press.

Kripke, S. A. (1963). Semantical analysis of modal logic i normal modal propositional calculi. *Mathematical Logic Quarterly*, *9*(5–6), 67–96.

Kuczaj, S. A., & Maratsos, M. P. (1975). What children can say before they will. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(2), 89–111.

Ladusaw, W. A. (1979). *Polarity Sensitivity as Inherent Scope Relations.* The University of Texas at Austin.

Laks, D. R., Beckwith, L., Cohen, S. E., & Cohen, S. E. (1990). Mothers' use of personal pronouns when talking with toddlers. *The Journal of Genetic Psychology*, *151*(1), 25–32.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.*

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78.

Levinson, S. C., Stephen, C., & Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Lewis, D. (1973). Counterfactuals and comparative possibility. In *Ifs* (pp. 57–85). Springer.

Lin, J., Weerman, F., & Zeijlstra, H. (2015). Emerging NPIs: The acquisition of Dutch hoeven 'need.' *The Linguistic Review*, *32*(2), 333–374.

Lin, J., Weerman, F., & Zeijlstra, H. (2018). Acquisition of the Dutch NPI *Hoeven* 'Need': From Lexical Frames to Abstract Knowledge. *Language Acquisition*, *25*(2), 150–177. https://doi.org/10.1080/10489223.2017.1348097

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

Mandelkern, M. (2019). What 'must'adds. *Linguistics and Philosophy*, *42*(3), 225–266.

Mandelkern, M., Schultheis, G., & Boylan, D. (2015). *I believe I can φ*.

Mandelkern, M., Schultheis, G., & Boylan, D. (2017). Agentive modals. *The Philosophical Review*, *126*(3), 301–343.

Manzini, M. R., & Wexler, K. (1987). Parameters, binding theory, and learnability. *Linguistic Inquiry*, 413–444.

Matthewson, L. (2013). Gitksan modals. *International Journal of American Linguistics*, *79*(3), 349–394.

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.

Moscati, V., & Gualmini, A. (2007). More facts that Isomorphism cannot explain. *Semantics and Linguistic Theory*, *17*, 202–209.

Moscati, V., Romoli, J., Demarie, T. F., & Crain, S. (2016). Born in the USA: A comparison of modals and nominal quantifiers in child language. *Natural Language Semantics*, *24*(1), 79–115. https://doi.org/10.1007/s11050-015-9120-1

Moscati, V., Zhan, L., & Zhou, P. (2017). Children's on-line processing of epistemic modals. *Journal of Child Language*, *44*(5), 1025–1040. https://doi.org/10.1017/S0305000916000313

Musolino, J. (2006). On the Semantics of the Subset Principle. *Language Learning and Development*, *2*(3), 195–218. https://doi.org/10.1207/s15473341lld0203_3

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.

Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316027073

Noveck, I. 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

Noveck, I. A., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language*, *23*(3), 621–643.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Orita, N., McKeown, R., Feldman, N., Lidz, J., & Boyd-Graber, J. (2013). Discovering pronoun categories using discourse information. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).

Ozturk, O., & Papafragou, A. (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, A. (1998). The Acquisition of Modality: Implications for Theories of Semantic Representation. *Mind and Language*, *13*(3), 370–399. https://doi.org/10.1111/1468-0017.00082

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, *105*(1), 125–165.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8

Perkins, L. (n.d.). *HOW GRAMMARS GROW: ARGUMENT STRUCTURE AND THE ACQUISITION OF NON-BASIC SYNTAX*. 293.

Peterson, T. (n.d.). *The Ordering Source and Graded Modality in Gitksan Epistemic Modals*. 15.

Piantadosi, S. T. (n.d.). *Learning and the language of thought*. 191.

Piantadosi, S. T. (2011). *Learning and the language of thought* [PhD Thesis]. Massachusetts Institute of Technology.

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*.

Piéraut-Le Bonniec, G. (1980). *The Development of Modal Reasoning Genesis of Necessity and Possibility Notions*.

Pinker, S. (1979). Formal models of language learning. *Cognition*, *7*(3), 217–283.

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

Prior, A. N. (1957). *Time and modality*. *Oxford*. England Clarendon Press.

Quine, W., & Van, O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. (pp. xv, 294). John Wiley.

Rakoczy, H., Warneken, F., & Tomasello, M. (2007). "This way!","No! That way!"—3-year olds know that two people can have mutually incompatible desires. *Cognitive Development*, *22*(1), 47–68.

Rasin, E., & Aravind, A. (2021). The nature of the semantic stimulus: The acquisition of every as a case study. *Natural Language Semantics*, *29*(2), 339–375. https://doi.org/10.1007/s11050-020-09168-6

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*(1), 12.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, *2*, 201–223.

Ruffman, T., Aitken, J., Wilson, A., Puri, A., & Taumoepeau, M. (2018). A re-examination of the broccoli task: Implications for children's understanding of subjective desire. *Cognitive Development*, *46*, 79–85.

Rullmann, H., Matthewson, L., & Davis, H. (2008). Modals as distributive indefinites. *Natural Language Semantics*, *16*(4), 317–357. https://doi.org/10.1007/s11050-008-9036-0

Sauerland, U. (2012). The Computation of Scalar Implicatures: Pragmatic, Lexical or Grammatical?: Computation of Scalar Implicatures. *Language and Linguistics Compass*, *6*(1), 36–49. https://doi.org/10.1002/lnc3.321

Schaeffer, J., & Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural Language & Linguistic Theory*, *23*(1), 53–101.

Searle, J. R. (1975). Indirect speech acts. In *Speech acts* (pp. 59–82). Brill.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, *14*(3), 301–321. https://doi.org/10.1016/0010-0277(83)90008-2

Shatz, M., & Wilcox, S. A. (1991). Constraints on the acquisition of English modals. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on Language and Thought* (1st ed., pp. 319–353). Cambridge University Press. https://doi.org/10.1017/CBO9780511983689.010

Skordos, D., & Papafragou, A. (2014). Scalar inferences in 5-year-olds: The role of alternatives. *Proceedings from the 38th Annual Boston University Conference on Language Development*.

Snedeker, J. (2000). Cross-situational observation and the semantic bootstrapping hypothesis. *Proceedings of the Thirtieth Annual Child Language Research Forum. Stanford, CA: Center for the Study of Language and Information*.

Snedeker, J., Gleitman, L., & Brent, M. (1999). The successes and failures of word-to-world mapping. *In*.

Snyder, W. (2007). *Child language: The parametric approach*. Oxford University Press.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques* [PhD Thesis]. Paris 7.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Citeseer.

Stalnaker, R. (1991). The Problem of Logical Omniscience, I. *Synthese*, *89*(3), 425–440.

Stalnaker, R. C. (1968). A theory of conditionals. In *Ifs* (pp. 41–55). Springer.

Stephany, U. (1983). *The development of modality in language acquisition*. Institut für Sprachwissenschaft, Universität zu Köln.

Stromswold, K. J. (1990). *Learnability and the acquisition of auxiliaries* [PhD Thesis]. Massachusetts Institute of Technology.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (Vol. 54). Cambridge University Press.

Szabolcsi, A. (2004). Positive polarity–negative polarity. *Natural Language & Linguistic Theory*, *22*(2), 409–452.

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.

Thornton, R., & Tesan, G. (2013). Sentential negation in early child English1. *Journal of Linguistics*, *49*(2), 367–411.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2016). Children's Knowledge of Free Choice Inferences and Scalar Implicatures. *Journal of Semantics*, *33*(2), 269–298. https://doi.org/10.1093/jos/ffv001

Van der Auwera, J. (2001). On the typology of negative modals. *Perspectives on Negation and Polarity Items*, *23*, 48.

Van der Auwera, J., Ammann, A., & Kindt, S. (2005). Modal polyfunctionality and standard average European. *Modality: Studies in Form and Function*, 247–272.

Van der Wal, S. (1996). *Negative polarity items and negation: Tandem acquisition* [PhD Thesis]. Rijksuniversiteit Groningen.

Van der Wouden, T. (1994). Polarity and 'illogical negation.' *Dynamics, Polarity and Quantification*, *17*, 16–45.

Van Dooren, A., Dieuleveut, A., Cournane, A., & Hacquard, V. (2017). Learning what must and can must and can mean. *Proceedings of the 21st Amsterdam Colloquium*, 225–234.

van Dooren, A., Tulling, M., Cournane, A., & Hacquard, V. (2019). *Discovering Modal Polysemy: Lexical Aspect Might Help*.

Van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua*, *120*(8), 1973–1990.

Veselinović, D., & Cournane, A. (2020). The grammatical source of missing epistemic meanings from modal verbs in child BCS. *Formal Approaches to Slavic Linguistics*, *26*.

Von Fintel, K., & Heim, I. (2011). *Intensional semantics*. Citeseer.

Von Fintel, K., & Iatridou, S. (2008). How to say ought in foreign: The composition of weak necessity modals. In *Time and modality* (pp. 115–141). Springer.

von Wright, G. H. (1951). Carnap's Theory of Probability. *The Philosophical Review, 60*(3), 362–374.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*(3), 245–275.

Wells, G. (1979). Describing children's linguistic development at home and at school. *British Educational Research Journal, 5*(1), 75–98.

Wexler, K., & Manzini, M. R. (1987). Parameters and learnability in binding theory. In *Parameter setting* (pp. 41–76). Springer.

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

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yanovich, I. (2016). Old English *motan, variable-force modality, and the presupposition of inevitable actualization. *Language, 92*(3), 489–521. https://doi.org/10.1353/lan.2016.0045