Ailís Cournane & Valentine Hacquard

5 Adapting acquisition methodologies to study modality in underdescribed languages

Abstract: This chapter focuses on methods used to test modality in child language acquisition. Acquisition and fieldwork approaches are united by the goal to understand the representational systems of the grammars under inquiry (developing or adult). They differ in what is known and unknown. In language acquisition we typically know a lot about language specific target constructions, but not the development of full competency, while in fieldwork we don't know as much about those targets, but we can trust that adult speakers have full competency. When we design child studies, we create replicable sets of carefully controlled contexts and stimuli. Focusing on a few methodological paradigms that have been successful for deepening our understanding of modal development, we consider the advantages and challenges associated with adapting these methods to study underdescribed languages. We speculate on how fieldworkers might be able to make the best use of these methods, in a way that complements existing methods.

1 Introduction

Fieldworkers and acquisitionists studying modality both aim to accurately describe and explain modal systems: for their syntax, semantics and pragmatics. A common ultimate goal is to learn how much variation and similarity exists in

Ailís Cournane, New York University, 10 Washington Pl., New York, NY, 10003, USA,
e-mail: cournane@nyu.edu
Valentine Hacquard, University of Maryland, College Park, 1401 Marie Mount Hall, College Park, MD, 20742, USA, e-mail: hacquard@umd.edu

cross-linguistic modal systems. We ask, how can this notionally defined area of language get packaged into grammars? What is yet unknown is quite different, however, between fieldwork and acquisition work. In fieldwork on underdescribed languages we don't yet know the (full) modal system for adult grammatical competence. In acquisition work, we typically know the modal system of the target language very well (as for e.g., English or Spanish), but don't yet know the developmental path children take to arrive at adult grammatical competency. Given the particular challenges of working with children, and isolating and characterizing grammatical competence in development (as distinct from conceptual and processing development), we have to be creative and careful to arrive at useful methodologies. Our successes may translate well to understudied languages, when considering the common goals of fieldworkers and acquisitionists, so long as we bear in mind the differing unknowns (target grammar vs. developmental grammar). In this paper, we: (a) share our carefully controlled first language acquisition materials for modal language, and the insights we've learned about modal development from using these materials, (b) relate and compare our methods to existing semantic fieldwork methods, (c) offer suggestions about adapting our materials for research on understudied languages, noting that our materials are suitable for working with adults or children, and (d) advocate for an increased back-and-forth between our two subfields: fieldwork on underdescribed languages has helped us understand better what children entertain as possible for modal language systems (see Cournane 2020), and acquisition work on modals helps us understand how learning shapes adult modal systems.

In fieldwork, the question of how modal concepts are grammatically expressed is addressed by seeing what patterns are actually attested in the language under study. For the fieldworker, the main goal is describing a language whose modal properties are not yet known from a linguistics perspective. The question is: how is modality expressed in the language, and how does that fit into the known typology of modal systems? We have learned more about the kinds of modals and modal systems languages may have from this work, especially on understudied languages (Bochnak 2015; Deal 2011; Matthewson 2010; Peterson 2010; Rullmann, Matthewson and Davis 2008; Rullmann and Matthewson 2018; Vander Klok 2012, i.a.). In acquisition work, the question of how modal concepts are grammatically expressed is addressed by seeing what hypotheses children entertain or fail to entertain along their learning path. For the acquisitionist, the main goal is describing when and how children acquire the target system (usually well-known), and understanding aspects of the input and changes within the child that drive development. The question is: how is the target system learned? Acquisitionists must be attuned to social, conceptual and other non-linguistic developments that may affect children's modal language use or their ability to

perform certain tasks (i.e. "task effects"). Contrast that to the fieldworker, who can trust that speakers have developed adult competence in linguistic and non-linguistic domains, that there are no conceptual or socio-pragmatic deficits due to immature development (though there may be sociocultural differences which can affect tasks).

In our projects on modal acquisition, we have primarily worked with English-learning children (our "convenience sample" when working with monolingual populations in North American cities, though working with small children is rarely convenient), but have also done experiments with Bosnian/Croatian/Serbian-learning children (Veselinović 2019; Cournane and Veselinović *accepted*), and worked on corpora in Dutch (van Dooren et al. 2019), French (Cournane and Tailleur, 2021), and Bosnian/Croatian/Serbian (Veselinović and Cournane 2020). None of these languages is underdescribed, but in acquisition that is often seen as an advantage. Knowing the language well is helpful when characterizing the target language, the input the child receives, and the learning path more generally. There are fewer unknowns. However, this real advantage for understanding the dynamics of development has contributed to acquisition work being even more narrowly centered on a few well-studied languages than linguistics research is more broadly.

Adapting methods specifically designed for one population and language will always involve creativity and effort to make sure the method maintains its integrity for addressing the desired research questions and hypotheses. That said, adapting from acquisition to fieldwork is a reasonably good match, despite on-the-surface major differences between our populations. We share many of the same challenges, many that researchers working with mature speakers of well-studied and widely spoken language do not.

First, few speakers and small sample size is common in both fields (see Bochnak and Matthewson 2015: 3–5). In acquisition work it is often difficult to get participants who fit the eligibility requirements of a study. Even working on widely spoken languages, recruiting and testing children is much more challenging than with adult participants from the same communities. Children cannot consent for themselves, so there are many more steps and individuals (i.e., caregivers, teachers, daycare directors, etc.) involved in the consent and assent[1] process. The process involves

---

[1] Children cannot give informed consent to participate in a research study (Parental/Guardian informed consent is required), so to adhere to legal and ethical requirements and best practices, we instead need them to give oral assent before we run an experiment with them. This involves a short conversation where we explain, in an age-appropriate way, that we want to learn more about how children talk and that to do that we're asking for their help. We ask them if they are willing to play the language game with us, and if they agree, we continue.

heavy recruiting effort for each participant that is successfully recruited. This leads to the problem of a "small n", that is, samples which lack power for statistical analysis. In brief, we usually do not have the resources (time, personnel, money) to devote to collecting large ns, before sharing our work and bringing it to publication (contrast this with studies conducted with adults online, where hundreds of participants are collected within a matter of days and costs are relatively low). And, adding to the small n challenge, many participants' data need to be excluded from analysis because children have a higher likelihood of being non-compliant participants. Work on understudied languages also usually faces issues related to limited data. Because of small ns, methods and analysis may in some ways be a closer fit from acquisition work than from psycholinguistics more broadly.

Second, in both fields we need to be extremely careful about making assumptions from what we know as linguists or speakers. This comes down to not being a member of the population. Many linguists are not native speakers of the understudied languages under inquiry, and we are never children. Compare this to when linguists work with mature speakers of a language that they are a native speaker of – their implicit biases about the phenomena are much more likely to be inconsequential to a clear view of facts, because they are a member of the broad population of study (abstracting away from sociolinguistic variation). In acquisition and fieldwork, we run a high risk of tacitly imposing our knowledge as linguists or speakers into the materials, data, or analysis. This issue can be particularly insidious in acquisition, as researchers often are native speakers of the language they are studying in development: it is very easy to assume that if a child learning our own language behaves a certain way, they are using the same knowledge or abilities that mature speakers like us do. Or if a child gives a somewhat opaque response, we may overinterpret it, filling in missing information from our own knowledge (and ironically, can end up measuring ourselves rather than the participant). The first author often teaches her students to pretend that child English (or whatever language we are looking at) is a different language or dialect from what they speak, to gain a healthy distance from making these kinds of tacit assumptions that can muddy a clear view on the facts.

And, lastly, looking at semantic and pragmatic areas of languages, controlling the context for acceptability and felicity is essential, because "[u]tterances are only true or false, felicitous or infelicitous, in context." (Bochnak and Matthewson 2020: 262; see also Bochnak and Matthewson 2015; Burton and Matthewson 2015; Cover 2015; Ferreira and Müller, this volume, §3.1; Vander Klok 2014, 2019). Here too, we cannot assume that because the context we created supports our own interpretation for e.g., an epistemic possibility modal, that participants or speakers will have the same interpretation of that context. To combat these challenges, as in fieldwork, we try to use multiple methods on the same phenomenon

to triangulate to the truth about the phenomenon under inquiry. And, we carefully monitor participants throughout testing, and probe for qualitative data and explanation where possible, to shed further light on responses to our materials. And, during development, we extensively pilot our materials out on others, both adults and children in person, to hone the materials to where we need them to be so we can be as confident as possible about the data they will garner. These practices help mitigate both the small n issue and the insidious effects of context interpretation. We will discuss some further, more specific, ways we deal with these challenges for the specific methodologies we cover.

We will not cover the literature on modal development (see Papafragou 1998; Hickmann and Bassano 2016; Cournane 2020 for overviews). Instead, we highlight some key questions that motivate acquisition studies and their potential relevance for fieldwork, and provide references for the interested reader here (1–5). In Section 2, we use our materials to showcase some new findings our methods have revealed even for relatively well-studied English learners. There are several aspects of modality that makes it particularly interesting from an acquisition standpoint:

1. Modal expressions are used to describe abstract concepts, which children may or may not grasp innately or early in development (possibility, necessity, desire, knowledge, etc.) (Leahy and Carey 2020; Shtulman and Phillips 2018; i.a.).
2. Modal expressions are abstract vocabulary or constructions with no obvious physical correlates, whose acquisition may thus need to rely heavily on syntactic and pragmatic cues (Dieuleveut et al. 2019; van Dooren et al. 2017; i.a.; see also Gleitman 1990; Hacquard and Lidz 2018).
3. The same modal words can express different flavors (in about ¼ of the world's languages according to van der Auwera and Amman 2005,[2] including those languages where acquisition has been most extensively studied), raising the question of how children figure out this one-to-many mapping (Cournane 2015; Papafragou 1998; van Dooren et al. 2017; i.a.).
4. Modals (at least in those languages where acquisition has been extensively studied) are often used in pragmatically enriched ways: they can be used to perform indirect speech acts like requests, and with scalar implicatures. These pragmatically enriched uses both raise the question of how children disentangle semantic and pragmatic contributions of modal statements,

---

[2] Variable-force modals were not counted in van der Auwera and Ammann (2005), therefore the one-fourth of the typological report reflects only flavour-variability, underestimating meaning-variability more generally (Matthewson 2013).

and provide a rich testing ground to probe children's pragmatic abilities (Dieuleveut et al. 2019, 2022; Noveck 2001; Ozturk and Papafragou 2015; i.a.).
5. Modals don't all behave in a uniform way in how they scope relative to negation (Iatridou and Zeijlstra 2013), and, controversially, to tense and aspect (see Hacquard 2009; Klecha 2016; Rullmann and Matthewson 2018, i.a.), raising the question of when and how children figure out these scopal constraints, and how these scopal constraints affect modal acquisition (Jeretič 2018; Koring, Meroni and Moscati 2018; Moscati and Crain 2014; Dieuleveut et al. 2022; i.a.).

Because the fieldworker can rely on informants having mature conceptual and pragmatic abilities, some of the acquisition studies designed to test children's conceptual and pragmatic competence with modals may be less relevant. However, the fieldworker faces some of the issues that make both the acquisition of modals *and* its investigation by acquisitionists challenging: can the same modal expressions be used to express different forces or flavors? How do they interact with elements like negation, tense, aspect or evidentiality? How does one tease apart the semantic and pragmatic contributions of a modal utterance? We will look at three methods that have helped us understand more about modal development, and consider how these relate to existing fieldwork methodology, and how they may be adapted for work on underdescribed languages.

Sharing our successes continues a long tradition of adapting acquisition methods for use in fieldwork contexts: notably, truth-value judgment tasks (Crain and Thornton 1998; Gordon 1998) and frog stories (Berman and Slobin 1994, using Mercer Mayer's "Frog, Where Are You?"). Here we aim to offer more materials and methods, specifically about modality, to the important and pressing enterprise of learning more about the semantics and pragmatics in understudied languages.

2 Case studies: Acquisition methodologies for modal language

We'll present three studies: one production study to elicit modals from participants, and two comprehension studies, to test understanding of certain modal expressions, one focused on flavour, the other on force. We begin by explaining our methods, and then compare and contrast them to existing fieldwork methods, exploring similarities and differences. For each method, we lay out potential gain from adapting these methods to the field, while acknowledging

several challenges that fieldworkers may face in adapting our materials to use on underdescribed languages. Where possible, we make suggestions for adaptations and applications. We provide comparisons, comments and suggestions that apply to all three studies we present in the General Discussion (§3). We make our materials available at OSF (https://osf.io/v9ure/) and welcome inquiries for further information or resources to the NYU Child Language Lab, www.childlanguagelab.com.

2.1 A sentence-repair task: Modal production task

Cournane, Hirzel and Hacquard (submitted) use a sentence repair task (see also Cournane 2014) to see what lexical preferences speakers have for expressing modal meanings to match particular carefully controlled situational contexts, and how children's preferences differ from adults' preferences. This method elicits modal productions for a 2x2 set of contexts, crossing modal flavour (teleological (=root), epistemic) and force (possibility, necessity) giving 4 unique situational combinations (Figure 1). This acquisition study was informed by the work on understudied languages. Namely, our 'modal meaning space' is inspired by Nauze (2008) and Vander Klok (2012), and our research questions are directly inspired by the variable-force modal literature (Bochnak 2015; Deal 2011; Peterson 2010; Rullmann, Matthewson and Davis 2008; Yanovich 2013). We explore the possibility that our English-learning children may entertain non-English, but cross-linguistically available, semantic representations for modals in their input. Acquisition research prior to this study had taken for granted that modal verbs are either possibility or necessity, probably because both possibility and necessity modals exist in English and other common languages which have provided the vast majority of evidence for child linguistic development. However, the comprehension literature for modal force shows non-adult behaviours through early school age (Byrnes and Duff 1989; Hirst and Weil 1982; Noveck 2001; Noveck, Ho and Sera 1996; Ozturk and Papafragou 2015, i.a.), which have been attributed to conceptual or pragmatic immaturity, without also considering that learners may hypothesize variable-force modals, especially if they fail to identify clear necessity modals (Dieuleveut et al. 2022).

In our study, children (3- and 4-year-olds, n= 46) and adults (n= 24) heard stories about going to stores via different coloured roads (teleological) and hiding in different coloured boxes (epistemic) (similar to Ozturk and Papafragou 2015's Experiment 1). Each flavour-force pairing is maximally similar to the others, with variation (e.g., boxes vs. roads) of only the critical type to clearly support flavour and force distinctions. These scenarios thus provide us with confidence that our

scenes will be interpreted as intended, so that we can trust that participants' uses of modal lexemes are in keeping with how they can grammatically encode each of the modal force-flavor combinations of interest. There were 4 items per condition (each cell in Figure 1). The stories provided a natural and supportive narrative for all four types of test sentences. Note that when there are two equally salient possibilities, it is pragmatically strange to only highlight one (*She can go down the red road*), so we mention the other possibility before the target.

Figure 1: Sample trials in each condition crossing FORCE (POSSIBILITY, NECESSITY) and FLAVOUR (EPISTEMIC, TELEOLOGICAL). Arrows indicate changes from one scene to the next on a tablet. (Drawings by Mina Hirzel; Stories by Ailís Cournane, Mina Hirzel & Valentine Hacquard. 2018. https://osf.io/v9ure/).

The task was to repeat story sentences with obscured modals to a shy snail puppet called Mr. Drooly, who was listening alongside the participant, so he could hear them. Pink noise blocked the modal but preserved the syntactic frame in which the modal occurs. This frame contains aspectual cues consistent with the intended flavour: eventive *go* for teleological (1a) and stative *be* for epistemic (1b) in the prejacent (see Condoravdi 2002; Portner 2009). Participants corrected the glitch with a modal of their choosing. This method works well for giving a choice to the speaker for how to repair the sentence, because multiple distinct modals can be used in the same slot in the English sentence, between the subject and bare verb complement: e.g., *can*, *must*, *might*, *have to*, *should*. This methodology is innovative for testing modal development because it targets both force and flavour together as the dependent variable, and prompts participants to produce modals of their own choosing but in an experimenter-constrained manner. In

short, it allows us to get snapshots of how individuals can use modal forms to express their 'modal meaning space'.

(1) a. Kat   <<noise>> go down the red path (given goal   TELEOLOGICAL
            to get to the bakery)
     b. Nick <<noise>> be hiding in the red box (given   EPISTEMIC
            evidence the other is empty)

Our results show that adults behave as expected for English, using different lexical items for the force dimension: in epistemic contexts, they primarily produced *could* for possibility, and *must* for necessity. In teleological contexts adults produced mostly *could* for possibility and *should* or *have to* for necessity. Thus, in necessity contexts adults tended to differentiate by both force and flavour. And, adult uses showed flavour-variability, especially for *could* across teleological and epistemic possibility conditions. Children produced fewer modal sentence-repairs than adults did (36% vs. 99%). This is a risk with any production task, as participants may respond in non-compliant or unexpected ways, and children usually yield higher rates of non-target data. However, children's non-modal and non-frame-compliant material (i.e., modal, but not fitting the syntactic frame in the prompt) were usually also informative about modal language knowledge and preferences. For example, if a child produced "*Maybe* Nick is hiding in the red box" this supplies a semantically appropriate modal repair (*maybe*) but also alters the frame (*be* > *is*). Child modal results are provided in Figure 2. We've shaded modal forms as follows: possibility modals in light grey, necessity in dark grey, and future in medium grey. Children tend to produce possibility modals for all 4 conditions, but appear to differentiate by flavour: they use more *might* in epistemic contexts, but more *can* in teleological contexts. Some children used *have to*, but they used it for both possibility and necessity conditions. Children also use many future modals, but similarly across flavours and forces. One interesting finding for our purposes here is that English-learning children appear to use particular modals for both forces (cf. variable-force modals), contra English-speaking adults.

Results are informative for child modal systems, as children generally use the same modal words as adults, but this study reveals how these map differently to the 2x2 modal meaning space we set-up. These results suggest children may not yet have the adult linguistic representations for the modal verbs of English. Child productions of possibility modals are more in line with adult productions than their uses of necessity modals, consistent with similar asymmetries in spontaneous corpus data (Dieuleveut et al. 2019, 2022), and experimental comprehension data (Ozturk and Papafragou 2015; Cournane et al. *in prep*). Children tend
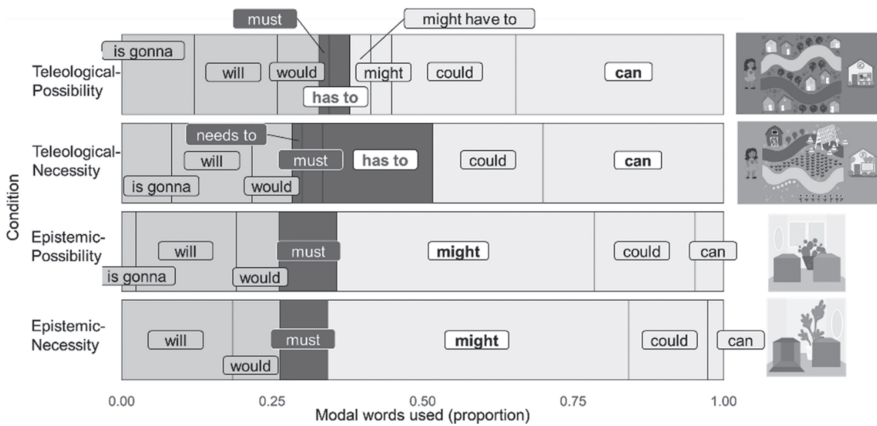
Figure 2: Children's modal word responses by condition. Possibility = Light Grey. Necessity = Dark Grey. Future = Medium Grey.

to prefer using different modals for different flavours, but, unlike their adult counterparts, they do not as readily distinguish by force, using the same modals across both possibility and necessity conditions (a pattern remarkably like variable-force modals, at least on the surface, but seen in 4-year-old English learners).

2.1.1 Comparison to existing methods, potential added value, and adaptation challenges

These materials may not be culturally or age-appropriate for everyone. Possible adaptations include superficial changes like recasting characters, locations and hiding locales to be more culturally or age relevant, or changing what noise obscures the portion of audio (e.g., we used pink noise and Cournane (2014) used a dog's bark, but a bird call or car honk should be equally effective). For older children and adults, the shy snail may be a fun childhood throwback or could be infantilizing or confusing. We explicitly tell adult controls who do our child studies (typically college students) that the materials are designed to work with young children. Easy ways to mature the materials are to change the signal to a radio transmission that gets obscured at some points, or if available, to use writing and have smudges or blanks for key words or morphemes (an option mentioned in Vander Klok 2014 as well). This could add to the fun of the task – perhaps building a backstory where someone's notebook was found but with water damage and we need to decode the content. The recorded components could be scrapped in favour of live narration if the fieldworker is confident with

elicited production prompts; This change would be especially useful for underdescribed languages with a primarily oral culture.

This study is a type of targeted elicited production task (for use in semantic fieldwork, see e.g., *The BowPed Topological Relations Picture Series,* Bowerman and Pederson 1992, discussed in Bochnak and Matthewson 2020), using contextualizing stimuli to constrain speaker productions to a desired area of language. More specifically, this is a sentence-repair/completion task. The basic method involves obscuring a portion of an utterance (or text) to elicit production of a piece of language, of the speaker's choosing, to fit a context of the researcher's choosing. With this methodology, many questions related to productive possibilities and preferences can be addressed, although adapting the method to linguistic properties of individual languages may require some adaptability and creativity.

This methodology also has elements of existing storyboarding fieldwork methods for semantic research (see Burton and Matthewson 2015). Storyboards provide a narrative series of images (or videos) to contextually (conceptually, semantically and pragmatically) support a targeted language use (see e.g., Matthewson 2013; Vander Klok 2014, 2019; Ferreira and Müller this volume, §3.2; and Kolagar and Vander Klok this volume, for storyboards targeting modal meanings). They "combine the advantages of spontaneous speech with the benefit of being able to test hypotheses about particular linguistic elements or constructions." (Burton and Matthewson 2015: 135). The images encourage speakers to talk about particular topics and distinctions. Deconstructing the items of a child behavioural experiment of the type described here essentially gives us multiple, minimally different, short storyboards. Like storyboards, each experimental item has a short series of images and supporting storyline designed to semantically and pragmatically support the use of modal language along two major dimensions that languages are known to grammatically distinguish (force,[3] flavour). These scenarios, with richly controlled and maximally similar contexts across forces and flavour, may be a useful extension from storyboards because they provide multiple slight variations on a theme – in brief, our materials could be adapted into many similar storyboards.

Similar to storyboards, our materials are visually presented, and the visual scenes for each storyline are carefully controlled for force and flavor distinctions. Being visually supported, these scenarios can thus serve as non-translation-based materials for targeted modal fieldwork (important in semantic fieldwork, see Matthewson 2004; Zhornik and Pokrovskaya 2018). The shared language can be used

---

3 We are sketching force as possibility or necessity, but bear in mind that modals may also be gradable (e.g., Lassiter 2010).

to help set up the stories and context, with a reduced risk of priming for the test language due to using modal expressions in the set-up (this is equally important to avoid in child experimental work, so we have made the set-up modal-free for English[4]). And, unlike contexts that are presented orally or via text, the visual scenes reduce working memory load. This is critical for work with small children to reduce unwanted task effects. With fieldwork, this is also helpful, as reducing memory load or room-for-imagination/enrichment allows greater trust that the language facts collected are actually of the type the researcher sought to collect (Bochnak and Matthewson 2020: 276).

Different from storyboards, these do not incorporate a fun twist to make the experience more enjoyable for the consultant (common though not essential for storyboards, Burton and Matthewson 2015: 145), though they are happy, child-friendly, and relatively short (for children's attention spans and to get multiple items into a short period of time). Also, if we treat every item as a mini-storyboard, then each story lacks a resolution. Because we repeat multiple items within participants, we need to avoid learning or expectations affecting perception of the task goals, so we don't provide resolution (e.g., imagine we revealed which box Nick was hiding in in an epistemic possibility scenario, and it was red, children might then expect Nick to like red boxes, and reframe the study as a guessing game rather than a scene-describing game). Instead, at the end of the story we provide a wrap-up of the whole experience. Likewise, at the beginning of the experiment we provide a general intro for the epistemic hiding scenes and the around town shopping scenes. The overall storyline arc (see Louie 2015) thus "bookends" the multiple story items per teleological and epistemic experimental block.

For direct adaptation of this sentence-repair method to understudied languages – including the trick of obscuring a portion of the target utterance to elicit very targeted productions – substantial challenges may arise because of linguistic differences. English is a language with many free modal words (auxiliaries and semi-auxiliaries) that fit the same linear slot between the subject and bare verb, so this method is a good fit for English because speakers can choose from among many modals to repair the glitch in the sentence, thus revealing something about the meanings they have for those modals through how they use them. If in the target language, modal meaning is carried by a very short duration morpheme or by suprasegmentals, or when the language's modal vocabulary is

---

4 The potential exception is the use of future *going to* in setting up the teleological stories, but this was necessary to make the scenario clear as to the future goal (teleological modality).

not concentrated in the same linear position or morphosyntactic category,[5] this method may be more difficult to adapt, or indeed not possible to adapt in its full form. However, it may be possible to obscure a larger portion of the utterance that contains the modal variation of interest (e.g., perhaps the entire verb or verb phrase for mood variation). If the construction of interest has the modal element (or covarying element, e.g., one could have constant modals in the test sentence and obscure tense marking, to test what tense is compatible with what modal uses) at the end of the utterance, the fieldworker can start the sentence and ask the consultant to complete it as they think best. This can be done with a hanging sort of prosody, as in "The girl went to the bakery to pick up *a* . . .", feeding the sentence for the participant to complete. In brief, one may find it useful to only partially prompt the test sentence.

Our scenarios may be also be adopted partially, rather than with the full experimental design or linguistic manipulations. When deconstructed, the scenarios we created are similar to components of the Modal Questionnaire (Vander Klok 2014; see also the revised version by Vander Klok, this volume), and could be used in a similar way to explore the modal expressions of a language. For example, a fieldworker may hypothesize that a particular modal element is a possibility modal from hearing it used in contexts compatible with that meaning, but be uncertain as yet as to whether it is flavour-variable or force-variable (under certain conditions). Furthermore, since necessity entails possibility, it may also be a necessity modal. The fieldworker could use our visually supported scenarios to test out which flavours and forces the modal is compatible with, particularly to try to see which scenarios its use is not acceptable for (if any). The consultant may offer alternatives as they tune-in to the dimensions of variation being probed – like whether there is more than one open possibility, or the temporal orientation of the scene (achieving the goal of reaching the location in the near future, or being in a hiding space in the present). As discussed in the modal fieldwork methodology literature, targeted follow-ups are an important way to learn more from a particular task (e.g., Bochnak and Matthewson 2020; Vander Klok 2019); If the participant offers one modal construction in response, ask them "are there other ways you could say this?".

In these sentence-repair studies (Cournane 2014; Cournane, Hirzel and Hacquard, submitted) participants often recast epistemic necessity as a simple declarative with *is*. And for root modals, they sometimes recast as imperatives ("Go down

---

5 English also has modal meanings expressed by many different syntactic categories, but crucially for this method there is a critical mass of modal verbs that express different forces and flavors and all occur in the same position.

the red road!"). Knowing the landscape of English, it is straightforward to see how these unanticipated strategies for responding to the task items (unanticipated because they don't use the provided syntactic frame) nonetheless capture the relevant meanings of the scenarios. In a child experimental study these kinds of responses cannot be included in the quantitative assessment (though are considered qualitatively), which could be considered a disadvantage. But unanticipated strategies may also be an advantage in fieldwork in the case that they reveal alternative grammatical ways of expressing the "same" meanings (see Vander Klok 2019, §4.3). Targeted follow-ups may be useful here as well. They may get consultants or participants to provide alternative ways of saying things, and perhaps talk about subtle differences in acceptability or felicity or usage patterns, that will help better understand the modal expressions (and related elements) of the language.

Finally, a putative drawback of production task is that they don't directly provide negative data: our study allows us to learn what linguistic expressions speakers prefer to use in the contexts we set-up, but not about which expressions they cannot use in those contexts. This could be seen as a problem, as often noted in discussion of experimental design practices for linguistics, "the methodology must allow the researcher to probe for negative data: contexts where a well-formed utterance is not acceptable." (Bochnak and Matthewson 2020: 3). However, the method also allows us to probe for speaker's own preferences and avoids certain disadvantages of comprehension tasks – there is little room for a *yes*-bias or guessing strategies, as in most types of comprehension task (e.g., forced-choice or judgment tasks). So, while responses are less predictable and attest only positive data, they are rich in their own right, and get at questions and hypotheses about language production that comprehension-based methods do not. Production tasks are an important tool for our methodological toolbox, and are especially useful when methods are compared against other types of methods, to help triangulate to the facts about the phenomenon.

2.2 Comprehension task 1: Modal force study

Regarding modal force, children need to figure out both what the underlying force of a modal is (e.g., possibility vs. necessity), whether their language has modal expressions for both possibility and necessity, and when the use of one might be more appropriate than the other (e.g., necessity modals are often more appropriate in necessity contexts than their possibility counterparts, even if the latter are logically true in these contexts). One challenge is that

necessity entails possibility. On the semantic end of things, what prevents an English-learning child from treating *must* or *have to* as encoding possibility, if a possibility meaning is true whenever *must p* or *have to p* is uttered? Perhaps all children need is to observe these modals in downward entailing environments (Gualmini and Schwarz 2009), since they reverse entailment patterns. However, in speech to children, necessity modals rarely appear with negation, let alone other downward entailing environments (Dieuleveut et al. 2019, 2022). Moreover, some necessity modals like *must* outscope negation, while others, like *have to*, scope under it (Zeijlstra and Iatridou 2013), making it difficult to use negation to figure out force (see Jeretič 2018; Dieuleveut et al. 2019, 2022). On the pragmatic end of things, how do children figure out that a modal like *might* is true, but underinformative in necessity situations? To pick up on the underinformativity of *might*, one must be aware of the existence of *must*, but as we saw, acquiring necessity modals faces its own challenges. Moreover, not all languages have modals with scalemates (e.g., Nez Perce, Deal 2011), so children cannot bank on their language having duals. Learning the semantics and pragmatics of modal force involves overcoming many learning challenges, analogous to the fieldwork challenges of discovering the modal force facts of an underdescribed language (for further discussion of the learnability of modal force, and its possible resolution, see Dieuleveut et al. 2019, 2022).

Prior studies on modal force development have primarily used epistemic paradigms, involving a hidden object or character to test child interpretations (as in the hiding-in-boxes scenarios in Figure 1) (for overview, see Ozturk and Papafragou 2015). Earliest work (Hirst and Weil 1982; see also Bascelli and Barbieri 2002) also included a deontic scenario but children performed much more poorly than in the epistemic task (guessing the location of a peanut), likely due to task complexity effects (selecting between two puppet teachers, who gave differing orders to a puppet student for no clear reason) (Hirst and Weil 1982). Children tend to overaccept underinformative modal uses (*can* when *have to* is felicitous), apparently making logical judgements rather than pragmatic ones. This type of non-adult behaviour is consistent with children having pragmatic difficulties with generating scalar implicatures (e.g., Noveck 2001). However, under the right circumstances, adults can also be made to behave logically and accept possibility modals in necessity contexts. Moreover, when the alternatives are provided or made very salient children perform better (see Ozturk and Papafragou 2015). Together these findings suggest that children's difficulties may reside more with knowing which alternatives are relevant in context, rather than with the pragmatic reasoning itself (see also Barner, Brooks and Bale 2011; Skordos and Papafragou 2016).

For example, Experiment 1 in Ozturk and Papafragou (2015) involved characters hiding in boxes onstage (hiding while curtains were drawn), followed by the curtains opening on a test scenario (four distinct scenarios: (a) a single closed box, (b) two closed boxes, (c) one closed and one open box and the sentence targets the closed box, (d) one closed and one open box and the sentence targets the open box). An experimenter produced statements (e.g., The cow *may*/*has to* be in the red box) or questions (e.g., *Can* the cow be in the red box? vs. Does the cow *have to* be in the red box?) for participants to judge or answer. Results showed that both adults and children accepted *can*/*may* when underinformative (in scenarios (a) and (c)), both groups apparently not computing scalar implicatures under these circumstances. More surprisingly, children also accepted uses of necessity *have to* in possibility scenarios (scenario (b)) about half of the time, unlike adults who judged these statements false or responded "no" to the questions. Ozturk and Papafragou (2015) argued for a conceptual explanation, essentially that children struggle with reasoning about more than one open possibility at the same time (Acredolo and Horobin 1987; Piéraut-Le Bonniec 1980). In possibility scenarios (two closed boxes) children randomly commit to one of the possibilities to rapidly resolve the uncertainty, and this results in *have to* being true about half the time (essentially children flip a coin on which open possibility to choose; though see Moscati et al., 2017).

Cournane, Dieuleveut, Repetti-Ludlow and Hacquard (in prep) extended the methods from Ozturk and Papafragou (2015) to test teleological scenarios, adapting the materials from Hirzel, Hacquard, and Cournane (submitted). Do children perform similarly for force with a root modality, which they have more experience with for *can* and *have to* from the input (see Shatz and Wilcox 1991; van Dooren et al. 2017)? We chose to test force in teleological scenarios because they are a subtype of root modality that is more readily imageable than deontic (permission, obligation), and allows us to maintain the same experimental structure as the epistemic hidden-box tasks, using open and closed roads. We are testing 3-to-4-year-olds and adult controls in this study, which involves a protagonist (Kat) going to different shops to prepare for a friend's birthday party. A pre-recorded narrator describes onscreen pictures (Figure 3: A, B), and a puppet (Logan) says the test sentence (Figure 3: C). The narrator then prompts the child to judge whether Logan is right or wrong. We ran (or are running) three experiments: Experiment 1 tests *can* and *have to* between subjects, Experiment 2 tests *can, can't* and *have to, doesn't have to* between subjects, and Experiment 3 tests *can* and *have to* within subjects. To make Logan's negative statements felicitous in Experiment 2, we systematically added a line to the Narrator's set-up, asking Logan a question (e.g., Narrator: "Logan, to get to the pizzeria, can Kat go down the green road?", Logan: "No, Kat can't . . .").
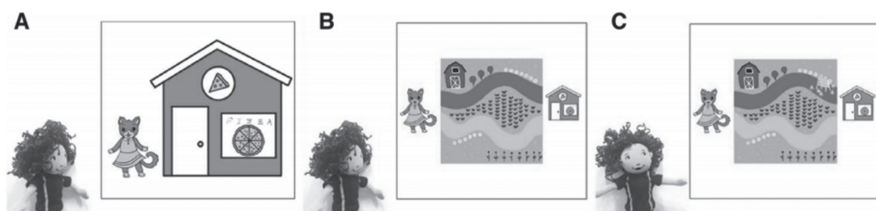
A Narrator: *Cat is going to the pizzeria to get a pizza for the party.*
B Narrator: *There are two ways to get to the pizzeria: the green road and the yellow road.*
[Animation – construction drops on] *But uh oh! The green road is blocked!*
C Logan [Animation – mouth moves] *To get to the pizzeria, Cat has to go down the yellow road.*
Prompt (Narrator): *Is Logan right?*

Figure 3: Sample item for Modal Force Study (Experiment 1), illustrating a *has to* use in a necessity scenario. (Drawings by Mina Hirzel & Chiara Repetti-Ludlow; Stories by Ailís Cournane, Anouk Dieuleveut, Chiara Repetti-Ludlow & Valentine Hacquard. 2019. https://osf.io/v9ure/).

Our results on Experiment 1 (*can* and *have to* tested between subjects) show that both children (n= 24, 12 per modal) and adults (n= 20, 10 per modal) accept *can* in both possibility and necessity contexts, apparently not computing scalar implicatures in the necessity contexts (just as with uses of *may/can* in epistemic necessity scenarios in Ozturk and Papafragou 2015). Most children uniformly accept *have to* in possibility scenarios, contra adults who reject it in these scenarios, and contra Ozturk and Papafragou (2015) where children were at chance. For Experiment 2, adults behave as we expected – the same as Experiment 1 for the positive items, and the expected reversed judgements for the negated items. Child data collection is ongoing (at time of writing, n= 17). So far, children behave more or less as expected with *can/can't*, but not with *have to/doesn't have to*, although this task shows more noise[6] than Experiment 1. Our method of supporting negation adds an unanticipated challenge for children: they appear to have especial difficulty when the target is that Logan is wrong for saying "yes", or right for saying "no". This mismatch appears to confuse some children. We have not yet begun data collection on Experiment 3 at time of writing. Our overall results so far suggest preschoolers are adult-like for *can*, but they tend to treat *have to* like *can*. This suggests that children may have a possibility meaning for *have to*.

The teleological task results for *have to* in Experiment 1 differ from the parallel epistemic results from Ozturk and Papafragou (2015). The children in the teleological study (Cournane et al. *in prep*) are about a year younger on average

---

6 In reference to experimental results, *noise* refers to more variance in responses among individuals, including from extraneous sources (factors that were not controlled for, or not under study).

than those in the epistemic study (Ozturk and Papafragou 2015). In teleological possibility contexts children accept *have to* the majority of the time, while in epistemic possibility contexts they accept *have to* only about half the time. One possible explanation for this discrepancy is that in the epistemic task children may interpret *have to* as root (consistent with how *have to* is predominantly used in the input, as root). Consider that if one hears 'the cow has to be in the red box' in a scene with two closed boxes (red and blue), the adult-like response is 'no (because the cow could also be in the blue box)'. But, on a root reading of *have to* (~ 'The cow *is obliged to* be in the red box'), children's acceptance of necessity in possibility situations is force-appropriate. That is, children may not always interpret the flavour of the modal in line with researchers' intent, unlike adults who are more accommodating and savvier. Another possible explanation is that children initially think that *have to* is a possibility modal, consistent with our teleological results for 4-year-olds. Then by age 5, some children may have learned that *have to* is a necessity modal, consistent with the mixed results for *have to* with Ozturk and Papafragou's epistemic results. The differing results between epistemic and teleological paradigms for testing modal force, and the inherent overlap of possible flavours in particular situations, underscore that flavour construals can affect judgments about the truth and felicity of the modal force.

2.2.1 Comparison to existing methods, potential added value, and adaptation challenges

First, relatively straightforward possible adaptations to fieldwork include superficial changes like recasting characters, locations and story topics to be more age or culturally appropriate, perhaps changing the birthday party preparation to preparing for a wedding, or to travel to another city or country. The roles of the narrator and the puppet can be carried out live with two researchers (or one, perhaps using a translation language for the set-up and then asking for judgement on target language test sentences to draw the contrast between set-up and judgement).

This task is a variation on the classic truth-value judgment task (Crain and Thornton 1998; Gordon 1998), which has already been widely adopted from child language research to other populations, including to understudied languages. In these tasks we set-up careful scenarios that support more than one possible interpretation, and ask participants to judge whether something a puppet says is true or false relative to the scenario. The participant's judgement sheds light on how they interpret the test sentences. Ideally, the design should make the critical sentences false, as a rejection is stronger than an acceptance, and the follow-up

of "Why not?" or "What really happened?" works well to get the reasoning behind the rejection. For example, adults who reject *have to* sentences (e.g., "To get to the Pizzeria, Kat has to go down the red road") in the possibility scenarios with two open roads, say things like, "No, she doesn't have to because the yellow road is open too." In fieldwork, truth-value judgments can be done singly, but in work with children we try to have at least 4 items of the same type (e.g., judgments of *can* with necessity scenarios).

Results from this study complement those from our Modal Repair Task (§2.1). These studies cover roughly the same age group, and with children from similar communities (English-speaking children from the New York City and Washington, D.C. areas). The Modal Repair Task gives production data, and we saw that children use the same modals across both forces (e.g., *can* and *have to* show up for both teleological possibility and necessity contexts). In this Modal Force Study, using a truth-value judgment task, we can test acceptability of those same modals with the same population of children and adult controls to see what patterns hold and improve our understanding of children's modal representations. This is parallel to how fieldworkers use multiple methodological strategies for the same phenomenon, with the same consultant(s), to best understand the semantics and pragmatics of the expression (see Bochnak and Matthewson 2015).

These materials were designed to test force interpretation for particular English modals (*can*, *have to*), for which we know the semantic and pragmatic facts. And, relevant for our Experiment 2, we know the relative scope of the modals with and without negation (*can't* = not > possible; *doesn't have to* = not > necessary) which can vary by modal (in particular for necessity modals, see Iatridou and Zeijlstra 2013), which allows us to interpret non-adult child patterns for negative test sentences as possibly relating to different relative scope. For example, a couple children tested on the *have to/doesn't have to* condition for Experiment 2 rejected both positive and negative sentences. These children may have a necessity semantics for *have to* because they behave like adults in necessity contexts (contra the majority of children tested thus far), but they may erroneously assume *have to* outscopes negation (necessary > not; akin to *mustn't*), causing them to reject *doesn't have to* in possibility contexts. These challenges may make this task (at least Experiment 2) too "late-stage" for some fieldwork settings, if certain facts about the modals under study are still unknown. However, these materials may be used in a partial way (i.e., adopting some components and not others), to help gather more information, which can be compared against other targeted inquiry. As with other methods, we often probe participants with follow-up questions after they have made their judgement, e.g., "Why is Logan wrong?" and these responses can help interpret participants' quantitative responses.

Multiple items per test condition allows multiple repetitions of the same kinds of scenarios (we have 4 items per condition here). This may be useful for seeing patterns of acceptance/rejection for variable-force modal items (cf. patterns of picture-selection for the Modal Flavour Task in Cournane and Pérez-Leroux 2020, for English variable-flavour modal *must*, covered in §2.3). And, repeating the same kind of scenario could potentially be helpful to diagnose these variable-force expressions in carefully constructed usage contexts. That said, pragmatic factors and the entailment relationship – e.g., *may* is true wherever *must* is – complicate modal force inquiry, so targeted diagnostics are also necessary (see Bochnak 2015; Deal 2011; Peterson 2010; Rullmann, Matthewson and Davis 2008; Yanovich 2013).

In both acquisition and fieldwork, one has to keep in mind that participants/consultants may not necessarily hone in on the intended force/flavor, given the frequent overlap in both force (possibility true in necessity contexts), and flavor (ability/teleological). For example, what prevents participants from interpreting *have to* in the epistemic contexts of Ozturk and Papafragou (2015) as deontic instead? Or, in our contexts, what prevents participants from interpreting *can* as about Kat's ability as opposed to about the possible roads to get to the pizzeria? We tried to mitigate this issue by making the question under discussion (see e.g., Roberts 2012) explicit in the overall set-up (Kat needs to get to different stores to prepare for a birthday party), and within each item by having the puppet repeat "To get to the pizzeria, . . ." before every test sentence (note that we varied shops for each item).

In this task, the test sentence is judged as a whole, so it would be easy to manipulate the test sentence as needed. For example, stories can be reused with different modal elements (e.g., this study could be re-run with no substantive changes just by swapping *have to* out for *must*; this would give the same judgement predictions except for when negated, as *have to* scopes below negation but *must* above) or with modalized sentences using different kinds of modal elements (particles, adverbs, etc.) as the whole sentence is judged by the participant/speaker. It is important to maintain the set-up to support the test sentence usage and make as clear as possible the intended flavour of the scenario. To ensure this, consider how we mention all possibilities in the scene to set up both possibility and necessity readings, how we used a polar question before the test sentence to license the use of negation, and how we explicitly state the goal and attempt to control the question under discussion.

Having a consultant judge another speaker, in this classic child-directed version of the truth-value judgment task, could be problematic. For children this is essential to avoid introspection and to reduce a type of child yes-bias where they are inclined to agree with adult teacher-figures, but more ok with telling puppets they are "silly", but for adults who are capable of metalinguistic introspection

this may invite vagueness into the judgements – perhaps there is some variation among speakers, and the consultant has heard someone say something like what the puppet said, but would not say it themselves. The first author recalls this happening when doing fieldwork in Inuktituk (Inuit) (Consultant: Oleekie Etungat), where the consultant sometimes said sentences were fine, but then divulged that they were how people in a different region of Baffin Island spoke, not how she spoke in her regional variety. The puppet can be eliminated in favour of having more direct judgements, either spoken by the fieldworker or written down.

2.3 Comprehension study 2: Flavour preference task

This method involves a Picture Choice Task targeting deontic (root) and epistemic interpretations of modalized sentences. It is from Cournane (2015) and Cournane and Pérez-Leroux (2020), and adapted to Bosnian/Croatian/Serbian in Veselinović (2019) and Veselinović and Cournane (*accepted*) (see also Fond 2005; Heizmann 2006). We will focus on the English study, but also note how it was adapted to grammatical properties of Bosnian/Croatian/Serbian. In English (and Bosnian/Croatian/Serbian) a modal verb's interpretation differs by syntactic construction. In English, the availability of root and epistemic flavour is constrained by interaction with aspect (e.g., Hacquard 2009; Ramchand 2018). Specifically, when a modal auxiliary verb like *must* is followed by a bare eventive verb like *vote* (2a), both a root and epistemic interpretation are possible. Note that the epistemic interpretation with a bare eventive verb has a habitual construal. With grammatical aspect marking (perfect or progressive), *must* has an epistemic interpretation (2b, c).[7] Given the syntactic construction (*Modal + Bare Eventive Verb* or *Modal + Grammatical Aspect*), what interpretation will child and adult speakers prefer – deontic or epistemic? The participant chooses between two pictures (the dependent variable), one portraying a deontic interpretation and one an epistemic (see Figure 4). This method allows the researcher to test what interpretations are available for specific modal constructions, and what preferences speakers have when both interpretations are available (Cournane and Pérez-Leroux 2020, use this method in part to study change-in-progress for modal *must*).

---

7 A root interpretation is possible, but requires making the sentence future-in-the-past, as by adding the adverbial "by tonight" to (2c). This is related to how root modalities are restricted to future-oriented prejacents (see Condoravdi 2002; Rullmann and Matthewson 2018; Werner 2006); perfect and progressive marking render prejacents past and present-oriented, respectively, so root modality is ruled out without additional temporal operators or a strong context to the same effect.

(2)   a.   *Annemarie **must** vote . . .*
         . . . because it's her civic duty.   (Root)
         . . . because she's really interested in politics.   (Epistemic)

     b.   *Anouk **must** be voting . . .*
         . . . because she left the house with her I.D.   (Epistemic)

     c.   *Anouk **must** have voted . . .*
         . . . because she is wearing an "I voted" sticker   (Epistemic)

Figure 4 summarizes the materials and design, using a sample story, and our Appendix provides the full list of materials for the English study. We introduced participants to a Penguin telling stories from a book (presented pre-recorded on a laptop). Penguin sets the scene for each mini-story with an initial picture and comment, then turns the book back to himself, turns the page (we play a page-turning noise), and says the test sentence. Then, after saying the test sentence, Penguin turns the book back towards the participant and says, "See look!", showing two possible pictures (we counterbalanced which side the epistemic and deontic pictures appeared on). The participant is then prompted to select the picture "Penguin was looking at" when he said the test sentence. We ran 54 children (aged 3–6) and 19 adults, all speakers of the local dialect in Toronto, Canada. In Figure 4 we also see the English test sentences, our independent variable: MODAL ONLY with the auxiliary modal *must* followed by a bare eventive verb, and MODAL ASPECT, with *must* followed by grammatical aspect marking (perfect, progressive). We expected adults to prefer epistemic interpretations for the modal aspect sentences, and deontic interpretations for the modal only sentences (these are ambiguous, but the mini-stories don't explicitly support a habitual reading of the bare verb to give an epistemic interpretation of *must*). We expected children to begin with a (a) deontic bias (in line with root biases observed in child spontaneous production data, e.g., Papafragou 1998; van Dooren et al. 2017, 2019), (b) to gradually become more adult-like (differentiate interpretations by construction), and (c) to show more epistemic interpretations for modal only sentences than adults (in line with language change-in-progress, see Cournane and Pérez-Leroux 2020, for details).

   Results are given in Figure 5. Adults behaved as expected, strongly preferring epistemic interpretations for MODAL ASPECT sentences, and variably selecting both deontic and epistemic pictures for MODAL ONLY sentences, but with a preference for deontic interpretations (adult responses are mostly in the d-quadrant in Figure 5). Adults differentiated for interpretations by construction. Small children had a slight overall deontic bias (3-year-old responses were mostly in the c-quad-
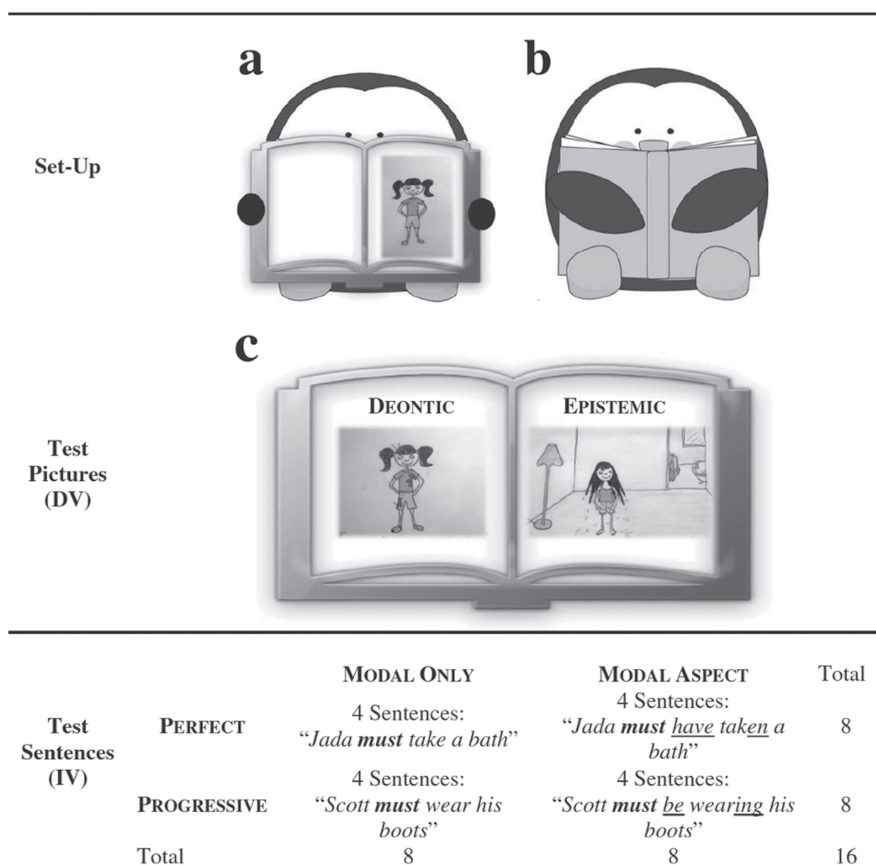
Set-Up (a) (b)

Test Pictures (DV) (c) DEONTIC EPISTEMIC

		MODAL ONLY	MODAL ASPECT	Total
Test Sentences (IV)	**PERFECT**	4 Sentences: *"Jada **must** take a bath"*	4 Sentences: *"Jada **must** have tak<u>en</u> a bath"*	8
	PROGRESSIVE	4 Sentences: *"Scott **must** wear his boots"*	4 Sentences: *"Scott **must** <u>be</u> wear<u>ing</u> his boots"*	8
	Total	8	8	16

Figure 4: Flavor preference task, sample stimuli, and design. Narration for each image:
(a) Penguin: This is Jada. She likes to play in the mud, but she also likes to be clean. (b) [Page turning noise]. Penguin: Oh! [Test Sentence]. (c) Penguin: See look! Experimenter: Penguin said, [repeat test sentence]. Which picture was penguin looking at? Each participant saw eight stories with MODAL ONLY sentences and eight MODAL ASPECT sentences, four each of which were progressive and perfect. The participant was prompted to pick either the deontic or the epistemic picture. (Drawings by Ailís Cournane; Stories by Ailís Cournane & Ana Teresa Pérez-Leroux. 2013. https://osf.io/v9ure/).

rant), but were otherwise close to chance. No child group differentiated for interpretations by construction, unlike adults. And, 5-year-olds show a strongly significant epistemic bias (their responses are mostly in the b-quadrant). This epistemic picture-selection bias is adult-like for the modal aspect sentences, but not for the modal-only sentences (it is in line with the child-driven change-in-progress hypothesis). This study allowed us to see what interpretations speakers have for
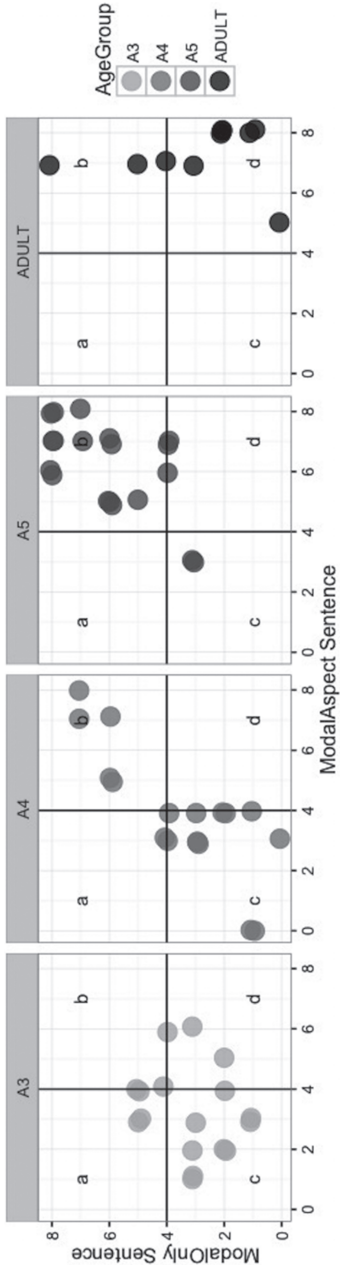
Figure 5: Results for all age groups (A3 = 3-year-olds, A4 = 4-year-olds, A5 = 5-year-olds). Individual speakers are plotted by the number of epistemic picture choices given to the MODAL ONLY (y-axis) and MODAL ASPECT (x-axis) sentences.

particular modal sentences, in a constrained way. Because of the 8 items per construction condition (modal only, modal aspect (4 each perfect and progressive) it also allowed us to assess individual interpretive preferences for variable-flavour modal *must*, and how those pattern within our age groups.

In Bosnian/Croatian/Serbian, modal verbs like *morati* 'must' (citation form) are also variable-flavour, but the syntactic properties of root vs. epistemic uses of the modal verbs differ from English. Briefly, Bosnian/Croatian/Serbian has categorically different constructions for root and epistemic uses of the modal verbs (Veselinović 2017, 2019). When the modal appears in a biclausal structure, with the modal verb in default agreement form, and the embedded main verb in imperfective the interpretation is epistemic (3a). When the modal appears in a monoclausal structure, with the modal verb agreeing with the subject, and the main verb in perfective form, the interpretation is root (3b) (see Veselinović 2019 for details and experimental results confirming this distinction for adult speakers of Bosnian/Croatian/Serbian).[8]

(3) a. ***Mora*-∅ *da se*[9] *mede*   *kupa-ju***
         must-3SG DA SE little.bear.PL bathe-3PL
         'The bear cubs must be bathing'   *Root, ✓Epistemic

     b. *Mede   **mora**-ju da se o-kupa-ju*
         little.bear.PL must-3PL DA SE PFV-bathe-3PL
         'The bear cubs must bathe'   ✓Root, *Epistemic
                                                      (Veselinović 2019: 190)

Some story items from the English study were straightforward to translate into Bosnian/Croatian/Serbian. Others were problematic because the main verb aspect was not of the right base aspectual class in Bosnian/Croatian/Serbian. New verbs (and corresponding stories) had to be selected, according to the following criteria: (i) the aspectual pair of verbs had to exist, and (ii) the imperfective had to be of a form termed 'simple' in the BCS literature, while the perfective had to be formed via prefixation from the imperfective. Also, the default agreement for the modal verb *morati* is only clearly present when the subject is plural, so all subjects were made plural. We used *mede* "little bears/bear cubs" throughout. The materials were all re-created so they would be by the same

---

8 The following abbreviations will be used here: 3 third person; SG singular; PL plural; PFV perfective.

9 The functional morphemes *da* and *se* have differing analyses in the literature, see Veselinović (2019) for overview.

hand, have uniform verb types, and have consistent subjects. The scenarios and deontic and epistemic pictures were then experimentally normed with Bosnian/ Croatian/Serbian native speakers prior to conducting the actual study. Norming of this kind is done in order to see if the scenarios we'd created were good enough renditions of deontic and epistemic scenarios to use in the main study. For details and results see Veselinović (2019). Results were very like those in English, despite input and syntactic differences between *must* and *morati* (see Veselinović 2019; Cournane and Veselinović *accepted*)

2.3.1 Comparison to existing methods, potential added value, and adaptation challenges

First, possible adaptations to the fieldwork setting include superficial changes like recasting characters, locations and story topics to be more culturally appropriate, and changing the book reading arrangement to instead involve looking at photographs, videos or drawings, or even looking at scenes out the window. What is critical is that the scene is out of view at the time the participant hears the test sentence, and only later do they see the options. The assumption is that participants will arrive at an interpretation prior to seeing the scenes, and that interpretation will guide their selection. Another adaptation would be to add more possible interpretations. We opted for only two because it was reasonable for *must*-interpretations to get at the broad, syntactically-mediated distinction between root and epistemic, and to reduce memory and attention load of stories for small children.

This is a type of forced choice task. Forced choice tasks are common in L1A and more broadly in psycholinguistics (see e.g., Ambridge and Rowland 2013), and are also used in fieldwork (e.g., which sentences is better for a given context? Or, which context is better for a given sentences? See e.g., Vander Klok 2014). More specifically, this is a picture choice task aimed at assessing interpretative possibilities or preferences (dependent variable, the choice between pictures) for single sentences (independent variable). There is some similarity here to acceptability judgment tasks used in fieldwork (see Bochnak and Matthewson 2020: 263–265 for an overview), but the pairings are expanded beyond a one-to-one match-up. In this study we offer two possible interpretations (operationalized as pictures) for one sentence. We used only two pictures to limit visual load and task complexity, but providing up to 4 pictures is fairly common for forced choice tasks, and increases the power associated with each selection type (chance behaviour becomes ¼ instead of ½, so consistent).

There is considerable flexibility to this method as pictures can be reused with different modal elements (e.g., the English study could be re-run with very little change just but swapping *must* out for *have to*). If picture adaptations are required for the language (as happened for Bosnian/Croatian/Serbian, due to verbal aspectual classes), we recommend norming the new pictures independently (i.e., running them online on MTurk or similar for their feasibility as good examples of e.g., an epistemic scene) so one can trust that the pictures are good examples of the desired interpretation (see Cournane 2015; Veselinović 2019; for norming processes). Each predicate brings its own slight idiosyncrasies to bear, making this norming especially important.

The multiple items per test condition allows repetitions. This may be useful for seeing patterns of preference by construction for variable-flavour modal items, and could be helpful to diagnose these items. However, one should be aware that the implicature that a deontic obligation will get carried out (i.e., if something must be done, it will be; see Traugott and Dasher 2002) may allow participants to select epistemic pictures with deontic modal sentences. To diagnose this and also to generally garner more insights into responses, it is helpful to ask follow-up prompts (as we did in the child studies), such as, "How did you know it was that picture?" (after the participant chooses a picture). Qualitative data that arises spontaneously during testing is often very informative, and can be increased by explicitly asking follow-up questions about test items at least some of the time (we aim for at least one prompt per unique condition). For example, children who picked the epistemic picture for the MODAL ONLY sentences regularly followed-up this prompt with a brief discussion of the evidence, e.g., how Jada's hair is wet or how she's in her pajamas ready for bed. Because fieldworkers also work one-on-one with consultants, these follow-up prompts can be an integral part of using this method, as they are in acquisition.

These materials were designed to test flavour interpretation for particular necessity modals, and so the scenarios make a contrast not only between deontic and epistemic, but between deontic obligation and epistemic necessity. This means that possibility modals could be tested using this method, but the materials would need to be altered to support open possibilities in root and epistemic flavours (see the sample scenarios in Figure 1, for Cournane, Hirzel, and Hacquard, submitted). The picture selection could be an array of all four major flavour-force combinations (Figure 1) to assess which are possible interpretations. A mix-and-match approach to the methods presented here may help solve some adaptation issues.

These materials could be deconstructed into storyboards, as mentioned for the other methods we have presented as well. The forced-choice picture prompts could be integrated into broader stories, creating a combination storyboard and

picture choice task. For example, one could set up a story like Vander Klok (2019), and then provide a test sentence to the speaker and give them two possible continuation pictures for the story. The question would be, which picture better illustrates the test sentence? This could be useful to do targeted work on constructions to rule out close alternative interpretations, or to learn about meaning variability or ambiguity.

Along the same lines, this method can easily be "flipped" to a sentence choice task (another type of forced choice task), where the researcher provides just one picture context (independent variable), and then two minimally different sentences that the researcher should select between on semantic or pragmatic grounds (dependent variable). Instead of addressing the question, "which pictures depicts your interpretation of the test sentence?" (picture choice task) this method addresses the question, "which construction (or form) better expresses what is happening in the picture?". The set-up and prompt question can help target either semantic (e.g., Which sentence is right?) or pragmatic (e.g., Who said it better?) interpretations, and the opportunity for follow-up discussion is rich, as there are minimally different sentences to discuss in reference to a picture scenario.

This sentence choice method was used in two other studies with *must* reported in Cournane and Pérez-Leroux (2020), comparing *must* sentences to unmodalized counterparts in deontic (children, n= 52; adults, n= 10) and epistemic scenarios (children, n= 35; adults, n= 9). In these tasks, using the epistemic version as an example, participants heard short introductions the same as for the picture choice task, but then they saw just one continuation picture (either epistemic, showing indirect evidence for the prejacent, or actual, showing direct evidence for the prejacent). For every test item they heard two sentences, one with *must* (e.g., *Scott must be wearing his rainboots*) and one without (e.g., *Scott is wearing his rainboots*), spoken by two puppets. Participants were trained to pick the puppet who was "paying closer attention to the story" and who "said it better". For further example experiments and discussion of judging one sentence versus selecting between two for semantic and pragmatic meaning with modals see Ozturk and Papafragou (2015). More broadly for different methods in pragmatic tasks, especially for scalar implicatures, see Skordos and Papafragou (2016).

3 General discussion

We discussed three distinct methods we have used to learn more about child modal language development: an elicited production task involving sentence repairs, and two comprehension tasks: a truth-value judgment task and a picture selection

task, to probe children's grasp of modal flavour and force, and usage preferences. All of these studies test predictions about modal language and different methodological choices are fitted to the properties of the target language and the conceptual domain. Methods and materials can also be combined as researchers see fit, to address questions differing from those addressed in each of the studies described above. Just as fieldworkers must triangulate to semantic representations and pragmatic felicity patterns from using multiple different methods (Bochnak and Matthewson 2015), so must acquisitionists. This "the more the merrier" approach is required even to address basic questions about the domain of inquiry because every method invites different task effects, and children are developing multiple cognitive and linguistic abilities and knowledge in parallel, making interpretation challenging, even for well-controlled experiments. Each methodology is just a tool for addressing research questions in a controlled way, taking into consideration the subtleties of the linguistic domain, the strengths and limitations of the method, and the particular challenges of the population (e.g., working memory limitations with children).

Our methodologies may help contribute to the general push for more experimental methods in field linguistics and work on understudied languages (e.g., Clemens et al. 2015; Li et al. 2011; Rech et al. this volume; Tollan, Massam and Heller 2019; Whalen and McDonough 2015). This work can help increase understanding of the languages under study, and contribute to psycholinguistic and sociolinguistic theories, which have been overwhelmingly built on English and other commonly studied languages. And, because our child-friendly experiments are designed to rely neither on reflection nor metalinguistic ability, they are suitable for work with all ages, including small children (for acquisition work on underderscribed languages see Courtney 2008; Demuth, Moloi and Machobane 2010; de Villiers et al. 2009; Eisenbeiß 2006; Gagliardi and Lidz 2014; Lima 2014; Pye 2017; Pye and Pfeiler 2014; Viau and Lidz 2011; among others).

All of our studies are experimental in nature. Wholesale adaptation of experimental work from well-studied languages to understudied languages is often thought to be untenable because of the necessity of multiple participants (see Bochnak and Matthewson 2015, Introduction, for discussion). The languages we have worked on for exploring modal development have millions of native speakers, and are being acquired by children. For many understudied languages, unfortunately these markers of an unendangered language are not true: often there are few native speakers, or the language is no longer being transmitted to children. In these cases, either standard experimental methods, child acquisition research, or both, are not possible. Psycholinguistics work, and experimental methods for human behaviour more broadly, rely on population sampling: getting enough data for statistical power, to infer findings from the sample to the population

more generally. That said, child language work has notoriously "small ns" in the wider psycholinguistics and developmental psychology world, and researchers have had to be particularly savvy to develop best practices for working with small datasets. For this reason, while acknowledging additional challenges, we suggest fieldworkers interested in doing experimental work, but with constraints on how many participants they can test, may rely more readily on child studies for ways to get the most from small datasets. For example, as suggested above for each of our studies, we also gather qualitative data from participants, to try to better understand their selection or judgment behaviours.

One pillar of experimental design is repetition (for statistical power) and reproducibility. Our methods for child studies involve running multiple trials of the same condition (e.g., four trials of the epistemic possibility condition for the sentence repair task) with multiple children, in order to sample enough from the population to do inferential statistics. While this kind of experimental method may not be feasible with many understudied languages, the many similar scenarios, with minor variations on the same recipe, may be useful for repetitions with the same speaker in the same or different sessions. Intra-speaker reproducibility of this type is a practicable alternative to population (re)sampling methods of reproducibility (Bochnak and Matthewson 2015: 5). With many similar items (i.e., all the stories we develop for each condition of each of our studies) portions of our tasks can be repeated at different intervals over time without repeating identical items.

Retesting participants or consultants on the same items risks them memorizing their way of responding to the task item(s), inviting task effects which obscure the facts of interest. We try to reduce this in experimental work by introducing variation that we are confident will be inconsequential. When one is doing intra-speaker replication, there is the possibility that speakers remember how they responded last time and are primed by themselves (see also Burton and Matthewson 2015: 145). This could be problematic because they may not be tapping in to their grammar directly, but into the task memory and this may reduce variation in responses, which perhaps the language can allow. For example, in English speakers may give *have to* to all necessity items in our Modal Repair task because they self-prime, when in fact their grammar has a many-to-one form-to-meaning mapping for those items (allowing also *should, must, need to*, etc.). This self-priming can also be lessened by inconsequential variations in stimuli, in addition to employing targeted follow-ups and multiple methodologies. In sum, acquisitionists and fieldworkers are both interested in grammatical competence of individuals, but must work with less-than-ideal circumstances from the standpoint of best practices of experimental design and statistical sampling. That said, they work on particularly interesting and important populations, and so must find creative ways to do sound science.

Semantic judgements can be subtle and are highly context-dependent, notably so for variable-flavour or variable-force modals, so carefully controlled stories with vetted conditions (e.g., workshopped, piloted, normed) are important (Bochnak and Matthewson 2015). All of our studies involve visual scenes, which provide the simplest, most direct way, to set up reliable contexts (videos or animations are likewise good, and sometimes even better, see Bar-El 2015; Bochnak and Matthewson 2020: 271–272). For our visual scenes and story narratives, we workshop the details of our materials extensively in our lab groups and ask around to linguists and/or psychologists working in related areas for their expertise, making adjustments as needed. Then we pilot our tasks on adults and children, making further adjustments as needed before entering the actual data collection phase. We can be confident that adults interpret them as intended, as the English-speaking adult participants behaved as expected for well-studied English modals in our studies – adult controls often serve as both the baseline and the proof of utility of the methods in child studies. Since we are confident about many aspects of the adult patterns for English, if adults perform as expected we can be confident that our task is working the way we intended it to.

It bears stressing that for all L1A experiments it is always best practice to run adults on the same exact studies as children – even in a well-studied language like English and even when the researchers speak the language – as a way to verify that the task is testing what we expect it to, and that the linguistic description is characteristic of the sample population. For example, in running Cournane and Pérez-Leroux (2020), the first author learned that working with a different variety of Canadian English than her own, speakers had different preferences for *must* vs. *have to* in epistemic contexts. In other words, her judgements when creating the materials were not the same as those of the test population, even though she speaks an only minimally different dialect. Working on underdescribed languages without the benefits of (as much) sociolinguistic and change knowledge, it may be harder to work out the nature of observed intra- or inter-speaker variations, and between children and adults (Are they due to developmental changes? Changes in progress? Or, stable semantic or pragmatic distinctions?). Running multiple adults on the same studies, and running adults on the same studies as children may help to disentangle (or perhaps discover?) distinct contributions to variation, like interspeaker variation (related to dialect differences and language change) vs. intraspeaker meaning variability.

Furthermore, another advantage of adapting child experimental work, rather than much other psycholinguistic work with adult speakers, is related to this fact that the "Fieldworker-consultant relationship is not fully parallel to the investigator-subject relationship" (Bochnak and Matthewson 2015: 5). The fieldwork-consultant relationship is more parallel with investigators and *child* subjects, as we

nearly always test one-on-one in person, and we check in with the child regularly. This is much less often the case with adult participant studies, where we either set them up and leave them to it in the lab, or we simply provide written instructions for online tasks. With child participants, we also sit "face-to-face and assess the situation in real time" (Bochnak and Matthewson 2015: 5) and thus garner extra qualitative data to help us make decisions about which data to exclude (i.e., if the child was clearly distracted or said something to make us think they were not doing the task as intended, such as repeatedly telling us they really like Frogs when the task involves choosing between two puppets, one a frog). Adapting to challenges in advance and on-the-fly is also something both fieldworkers and child acquisitionists surely know better than most other sub-types of linguist, as each fieldwork situation is unique and each phenomenon studied with children has added challenges. With child participants, unlike with consultants, we rarely have a longstanding relationship so individual characteristics are not as readily assimilated into qualitative interpretations as with typical fieldwork.

Finally, one constant concern we have in interpreting our results is whether children differ from adults (and the researcher's intentions for the stimuli) by how they interpret aspects of our scenarios, or how they pragmatically enrich our stories, rather than how they use modal verbs. For example, in our modal force task, could children – like adults – know that *have to* a necessity modal, but be pragmatically enriching possibility contexts to make it felicitous? This issue is also present in fieldwork (and indeed any study using constructed stimuli to explore interpretation or felicity), and has been extensively discussed in the semantics fieldwork methods literature (e.g., Bochnak and Matthewson 2020; Matthewson 2004; Vander Klok 2019). Our strategies to try to mitigate these risks has been similar: (a) workshop, pilot and norm materials in advance, (b) run studies on control groups to improve confidence that the materials are working as intended (or to know when to go back to the drawing board), (c) gather qualitative responses to supplement selection or judgment data and provide evidence either that participants are or are not interpreting materials as intended, and (d) remain mindful of implicit biases with respect to materials interpretation – we want to measure the participants' interpretations of the materials not our own.

4 Conclusion

In this chapter we have discussed three methodological paradigms used in first language acquisition studies on modality: a sentence-repair modal production task, a picture preference forced-choice modal flavour comprehension task, and

a truth-value judgment modal force comprehension task. These methods have helped us better understand child modal interpretations and have proven useful for shedding new light on modal development (see Cournane 2020 for a recent and more language acquisition centered discussion). For all three paradigms outlined, if adaptations for underdescribed languages maintain the overall procedure, it would be possible to work both with adults and with children because these tasks are designed to not require meta-knowledge nor meta-judgements. We hope that the methodological challenges we have faced and found solutions for can be helpful not just in the usual developmental work on well-studied languages, but also in work on speakers of underdescribed languages. Developmental work on underdescribed languages is of particularly great value, though involves both the challenges of working on underdescribed languages and working with children. We hope that sharing our materials and experiences can contribute to facilitating at least some of the developmental research aspects of the tasks.

Acquisitionists and fieldworkers working on modality share a common goal of describing and explaining possible modal systems. Extant modals and modal systems are all examples of how this conceptual domain can be encoded into language. So, as increased cross-linguistic fieldwork on modality has helped us better understand linguistic modality, it has in turn helped us better consider hypotheses that child learners may entertain during acquisition. For example, foundational work on modal force development (overwhelmingly on well-studied Indo-European languages, primarily English) predates the semantic description of languages with variable-force modals. Going forward, we now know that languages can have variable-force modals, and so we can now consider this as a viable possibility for development (even if not present in the target input language). Can the tools we have used to learn more about modal development for a particular target language be flipped, to help learn more about adult modal (=target, in child studies) grammars? Building on a long tradition of adapting acquisition materials to the field, we think there is promise and that sharing our materials and insights from modal acquisition may be a way to give back to the linguistic subfield that has helped us expand and clarify our hypothesis space for child learners.

References

Ambridge, Ben & Rowland, Caroline F. 2013. Experimental methods in studying child language acquisition. *WIREs Cognitive Science* 4 (2). 149–168.
Acredolo, Curt & Horobin, Karen. 1987. The child's relational reasoning and tendency toward premature closure. *Developmental Psychology* 23 (1). 13–21.

Bar-El, Leora. 2015. Documenting and classifying aspectual classes across languages. In M. Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in semantic fieldwork*, 75–109. Oxford, UK: Oxford University Press.

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

Bascelli, Elisabetta & Maria Silvia Barbieri. 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.

Bochnak, M. Ryan. 2015. Variable force modality in Washo. In Thuy Bui & Denis Özyildiz (eds.), *Proceedings of 45th North-East Linguistics Society*, 105–114. (NELS45). Cambridge, MA: Massachusetts Institute of Technology.

Bochnak, M. Ryan & Lisa Matthewson (eds.). 2015. *Methodologies in semantic fieldwork*. Oxford, UK: Oxford University Press.

Bochnak, M. Ryan & Lisa Matthewson. 2020. Techniques in complex semantic fieldwork. *Annual Review of Linguistics* 6 (1). 261–283.

Bowerman, Melissa & Erik Pederson. 1992. Topological relations picture series. In Stephen C. Levinson (ed.), *Space Stimuli Kit 1.2: November 1992,* 51. Nijmegen, the Netherlands: Max Planck Institute for Psycholinguists.

Burton, Strang & Lisa Matthewson. 2015. Targeted construction storyboards in semantic fieldwork. In M. Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in Semantic Fieldwork*, 135–156. Oxford, UK: Oxford University Press.

Byrnes, James P. & Michelle A. Duff. 1989. Young children's comprehension of modal expressions. *Cognitive Development* 4 (4). 369–387.

Clemens, Lauren Eby, Jessica Coon, Pedro Mateo Pedro, Adam Milton Morgan, Maria Polinsky, Gabrielle Tandet & Matthew Wagers. 2015. Ergativity and the complexity of extraction: A view from Mayan. *Natural Language and Linguistic Theory* 33 (2). 417–467.

Condoravdi, Cleo. 2002. Temporal interpretation of modals: Modals for the present and the past. In David Beaver, Luis Casillas Martinez, Brady Clark & Stefan Kaufmann (eds.), *The construction of meaning*, 59–88. Stanford, CA: CSLI Publications.

Cournane, Ailís. 2015. *Modal development: Input-divergent L1 acquisition in the direction of diachronic reanalysis*. Toronto, ON: University of Toronto dissertation.

Cournane, Ailís. 2020. Learning modals: A grammatical perspective. *Language and Linguistics Compass* 14 (10). 1–22.

Cournane, Ailís, Mina Hirzel & Valentine Hacquard. submitted. Mapping modal verbs to meanings: An elicited production study on 'force' and 'flavor' with young preschoolers.

Cournane, Ailís & Ana Teresa Pérez-Leroux. 2020. Leaving obligations behind: Epistemic incrementation in preschool English. *Language Learning and Development* 16 (3). 270–291.

Cournane, Ailís & Sandrine Tailleur. 2021. La production épistémique chez l'enfant francophone: Une étude comparative. [Epistemic production in francophone children: A comparative study]. *Arborescences* 10. 47–72.

Cournane, Ailís & Dunja Veselinović. accepted. If they must, they will: Children overcommit to likeliness inferences from deontic modals. *Glossa*. https://doi.org/10.16995/glossa.5802

Courtney, Ellen H. 2008. Child production of Quechua evidential morphemes in conversations and story retellings. El Paso, TX: University of Texas manuscript.

Cover, Rebecca Tamar. 2015. Semantic fieldwork on TAM. In M. Ryan Bochnak & Lisa Matthewson, *Methodologies in semantic fieldwork*, 233–268. Oxford, UK: Oxford University Press.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in universal grammar*. Cambridge, MA: The MIT Press.

De Villiers, Jill G., Jay L. Garfield, Harper Gernet-Girard, Tom Roeper & Margaret Speas. 2009. Evidentials in Tibetan: Acquisition, semantics, and cognitive development. *New Directions for Child and Adolescent Development* 2009 (125). 29–47.

Deal, Amy Rose. 2011. Modals without scales. *Language* 87 (3). 559–585.

Demuth, Katherine, Francina Moloi & Malillo Machobane. 2010. 3-Year-olds' comprehension, production, and generalization of Sesotho passives. *Cognition* 115 (2). 238–251.

Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane & Valentine Hacquard. 2019. Acquiring the force of modals: Sig you guess what sig means? In Megan M. Brown & Brady Dailey (eds.), *Proceedings of the 43nd Annual to the Boston University Conference on Language Development,* 189–202. (*BUCLD* 43). Somerville, MA: Cascadilla Press.

Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane & Valentine Hacquard. 2022. Finding out modal force: evidence from children's production and input. *Natural Language Semantics*. https://doi.org/10.1007/s11050-022-09196-4

Eisenbeiß, Sonja. 2006. Documenting child language. *Language Documentation and Description* 3. 106–140.

Fond, Marissa Joanne. 2003. *Deontic and epistemic modal expression: Theory and acquisition in English and Spanish*. Northampton, MA: Smith College BA thesis.

Gagliardi, Annie C. & Jeffrey Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90 (1). 58–59.

Gordon, Peter. 1998. The truth-value judgment task. In Dana McDaniel, Cecile McKee & Helen Smith Cairns (eds.), *Methods for Assessing Children's Syntax*, 211–235. Cambridge, MA: The MIT Press.

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

Hacquard, Valentine. 2009. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy* 32 (3). 279–315.

Heizmann, Tanja. 2006. Acquisition of deontic and epistemic readings of *must* and *müssen*. In Tanja Heizmann (ed.), *University of Massachusetts occasional papers in linguistics* 34. 21–50. Amherst, MA: GLSA, University of Massachusetts.

Hirst, William & Joyce Weil. 1982. Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language* 9 (3). 659–666.

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

Jeretič, Paloma. 2018. Evidence for children's dispreference for weakness: A corpus study. New York, NY: New York University qualifying paper.

Klecha, Peter. 2016. Modality and embedded temporal operators. *Semantics and Pragmatics* 9. 1–55.

Koring, Loes, Luisa Meroni & Vincenzo Moscati. 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.

Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. *Semantics and Linguistic Theory* 20. 197–215.

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

Li, Peggy, Linda Abarbanell, Lila Gleitman & Anna Papafragou. 2011. Spatial reasoning in Tenejapan Mayans. *Cognition* 120 (1). 33–53.

Lima, Suzi. 2014. All notional mass nouns are count nouns in Yudja. *Semantics and Linguistic Theory* 24. 534–554.

Louie, Meagan. 2015. The problem with no-nonsense elicitation plans for semantic fieldwork. In M. Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in semantic fieldwork*, 47–71. Oxford: Oxford University Press.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70 (4). 369–415.

Matthewson, Lisa. 2010. On apparently non-modal evidentials. *Empirical Issues in Syntax and Semantics* 8. 333–357.

Matthewson, Lisa. 2013. On how (not) to uncover cross-linguistic variation. In Stefan Keine & Shayne Sloggett (Eds.), *Proceedings of the North East Linguistic Society* 42, 323–342.

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

Moscati, Vicenzo, Likan Zhan & Peng Zhou. 2017. Children's on-line processing of epistemic modals. *Journal of Child Language* 44 (5). 1025–1040.

Nauze, Fabrice Dominique. 2008. *Modality in typological perspective*. Amsterdam, NL: Institute for Logic, Language and Computation: Universiteit van Amsterdam dissertation.

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

Noveck, Ira A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78 (2). 165–188.

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

Papafragou, Anna. 1998. The acquisition of modality: Implications for theories of semantic representation. *Mind and Language* 13 (3). 370–399.

Peterson, Tyler Roy Gösta. 2010. *Epistemic modality and evidentiality in Gitksan at the semantics-pragmatics interface*. Vancouver, BC: University of British Columbia dissertation.

Piéraut-Le Bonniec, Gilberte. 1980. *The development of modal reasoning*. New York, NY: Academic Press.

Portner, Paul. 2009. *Modality*. Oxford, UK: Oxford University Press.

Pye, Clifton. 2017. *The comparative method of language acquisition research*. University of Chicago Press.

Pye, Clifton & Barbara Pfeiler. 2014. The comparative method of language acquisition research: A Mayan case study. *Journal of Child Language* 41 (2). 382–415.

Ramchand, Gillian. 2018. *Situations and syntactic structures*. Cambridge, MA: The MIT Press.

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 1–69.

Rullmann, Hotze & Lisa Matthewson. 2018. Towards a theory of modal-temporal interaction. *Language* 94 (2). 281–331.

Rullmann, Hotze, Lisa Matthewson & Henry Davis. 2008. Modals as distributive indefinites. *Natural Language Semantics* 16 (4). 317–357.

Shatz, Marilyn, and Sharon A. Wilcox. 1991. Constraints on the acquisition of English modals. In Susan A. Gelman & James P. Byrnes (eds.), *Perspectives on language and thought: Interrelations in Development,* 319–353. Cambridge, UK: Cambridge University Press.

Shtulman, Andrew & Jonathan Phillips. 2018. Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology* 165. 161–182.

Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18.

Tollan, Rebecca, Diane Massam & Daphna Heller. 2019. Effects of case and transitivity on processing dependencies: Evidence from Niuean. *Cognitive Science* 43 (6). e12736.

Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge, UK: Cambridge University Press.

van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane & Valentine Hacquard. 2017. Learning what *must* and *can* must and can mean. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *Proceedings of the 21st Amsterdam Colloquium*, 225–234. Amsterdam, the Netherlands: Institute for Logic, Language and Computation (ILLC), University of Amsterdam.

van Dooren, Annemarie, Maxime A. Tulling, Ailís Cournane & Valentine Hacquard. 2019. Lexical Aspect and Modal Flavor in Dutch. In Megan M. Brown and Brady Dailey (eds.), *Proceedings from the 43nd Annual Boston University Conference on Language Development*, 203–216. (*BUCLD* 43). Somerville, MA: Cascadilla Press.

Vander Klok, Jozina. 2012. *Tense, aspect, and modality in Paciran Javanese*. Montreal, QC: McGill University dissertation.

Vander Klok, Jozina. 2014. On the use of questionnaires in semantic fieldwork: A case study on modality. *Proceedings of Language Documentation and Linguistic Theory* 4. 1–11.

Vander Klok, Jozina. 2019. Exploring modality and temporality interactions through the storyboard Bill vs. The weather. *Semantic Field Methods* 1 (1). 1–29.

Veselinović, Dunja. 2019. *The syntax and acquisition of modal verb flavors*. New York, NY: New York University dissertation.

Veselinović, Dunja & Ailís Cournane. 2020. The grammatical source of missing epistemic meanings from modal verbs in child BCS. In Tania Ionin & Jonathan Eric MacDonald (eds.), *Formal approaches to Slavic linguistics* 26, 417–436. Ann Arbor, MI: Michigan Slavic Publications.

Viau, Joshua & Jeffrey Lidz. 2011. Selective learning in the acquisition of Kannada ditransitives. *Language* 87 (4). 679–714.

Werner, Tom. 2006. Future and non-future modal sentences. *Natural Language Semantics* 14 (3). 235–255.

Whalen, Doug & Joyce McDonough. 2015. Taking the laboratory into the field. *Annual Review of Linguistics* 1 (1). 395–415.

Yanovich, Igor. 2013. *Four pieces for modality, context and usage.* Cambridge, MA: Massachusetts Institute of Technology dissertation.

Zhornik, Daria & Sophie Pokrovskaya. 2018. Modelling visual stimuli for descriptive fieldwork among the Upper Lozva Mansi: Metalanguage vs. target language. Paper presented at *Fieldwork: Methods and Theory" Workshop,* University of Gothenburg, Sweden, December 13–14, 2018.

Appendix: Flavor preference task stimuli, English (Cournane & Pérez-Leroux 2020)

(Drawings by Ailís Cournane; Stories by Ailís Cournane & Ana Teresa Pérez-Leroux. 2013. https://osf.io/v9ure/)

1–8 Progressive stories, 9–16 Perfect stories

1. INTRO: *This is Iryna; she doesn't like to be dirty.*
    MODAL-ONLY: *Iryna must take a bath*
    MODAL-ASPECT: *Iryna must be taking a bath*

    Intro   Picture Choice (shown: deontic left, epistemic right)

2. INTRO: *This is Joanna and her cat, Slushie. Slushie loves to hunt mice. Joanna is scared of mice!*
    MODAL-ONLY: *Slushie must hunt a mouse*
    MODAL-ASPECT: *Slushie must be hunting a mouse*

    Intro   Picture Choice (shown: deontic left, epistemic right)

3. INTRO: *Alex wants to make snowmen in the snow, but it's very very cold outside!*
   MODAL-ONLY: *Alex must be wearing her warm winter clothes*
   MODAL-ASPECT: *Alex is wearing her warm winter clothes*

   Intro   Picture Choice (shown: deontic left, epistemic right)

4. INTRO: *Michelle doesn't know how to swim. Her mom wants her to learn*
   MODAL-ONLY: *Michelle must swim*
   MODAL-ASPECT: *Michelle must be swimming*

   Intro   Picture Choice (shown: deontic left, epistemic right)

5. INTRO: *Baby Chipmunk wants to eat candy. Her mom wants her to eat healthy acorns!*
   MODAL-ONLY: *Baby Chipmunk must eat her acorns*
   MODAL-ASPECT: *Baby Chipmunk must be eating her acorns*

   Intro   Picture Choice (shown: deontic left, epistemic right)

6. INTRO: *Scott wants to go play in the rain; but he doesn't want to wear any clothes!*
    MODAL-ONLY: *Scott must wear his rain boots*
    MODAL-ASPECT: *Scott must be wearing his rain boots*

    Intro   Picture Choice (shown: deontic left, epistemic right)

7. INTRO: *Matt was mean to his sister. Their mom is mad.*
    MODAL-ONLY: *Matt must say sorry to his sister*
    MODAL-ASPECT: *Matt must be saying sorry to his sister*

    Intro   Picture Choice (shown: deontic left, epistemic right)

8. INTRO: *Becky has never ridden a bike before. Her mom wants her to learn safely.*
    MODAL-ONLY: *Becky must ride a bike with training wheels*
    MODAL-ASPECT: *Becky must be riding a bike with training wheels*

    Intro   Picture Choice (shown: deontic left, epistemic right)

9. INTRO: *Mark is in his painting class. His teacher shows him what to paint*
    MODAL-ONLY: *Mark must paint a flower*
    MODAL-ASPECT: *Mark must have painted a flower*

    Intro   Picture Choice (shown: deontic left, epistemic right)

10. INTRO: *This is Jada. She likes to play in the mud, but she also likes to be clean*
    MODAL-ONLY: *Jada must take a bath*
    MODAL-ASPECT: *Jada must have taken a bath*

    Intro   Picture Choice (shown: deontic left, epistemic right)

11. INTRO: *Dan's father is on the phone with his grandma. Grandma hates Dan's long, messy hair*
    MODAL-ONLY: *Dan must get a haircut*
    MODAL-ASPECT: *Dan must have gotten a haircut*

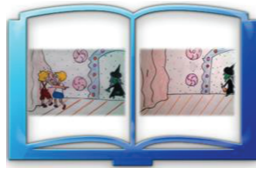   Intro   Picture Choice (shown: deontic left, epistemic right)

12. INTRO: *Hansel and Gretel are in the witch's house; the witch is mean, especially to children!*
    MODAL-ONLY: *Hansel and Gretel must hide behind the curtains*
    MODAL-ASPECT: *Hansel and Gretel must have hidden behind the curtains*

    Intro   Picture Choice (shown: deontic left, epistemic right)

13. INTRO: *Michael is on his way to school, but it's raining outside!*
    MODAL-ONLY: *Michael must use an umbrella*
    MODAL-ASPECT: *Michael must have used an umbrella*

    Intro   Picture Choice (shown: deontic left, epistemic right)

14. INTRO: *Chris and Doggy are hungry; there's pizza for Chris and dogfood for Doggy.*
    MODAL-ONLY: *Doggy must eat his dogfood*
    MODAL-ASPECT: *Doggy must have eaten his dogfood*

    Intro   Picture Choice (shown: deontic left, epistemic right)

15. INTRO: *Ross sees a rabbit and a fox – the fox is hungry and likes to eat rabbits!*

    MODAL-ONLY: *The rabbit must jump the fence*

    MODAL-ASPECT: *The rabbit must have jumped the fence*

    Intro   Picture Choice (shown: deontic left, epistemic right)

16. INTRO: *Sarah was really naughty – she painted all over her wall!*

    MODAL-ONLY: *Sarah must clean the wall*

    MODAL-ASPECT: *Sarah must have cleaned the wall*

    Intro   Picture Choice (shown: deontic left, epistemic right)