

Testing theories of temporal inferences: Evidence from child language *

Alexandre Cremers
Universiteit van Amsterdam

Frances Kane
Ulster University

Lyn Tieu
*Western Sydney University &
Macquarie University*

Lynda Kennedy
Ulster University

Yasutada Sudo
UCL

Raffaella Folli
Ulster University

Jacopo Romoli
Ulster University

Abstract Sentences involving past tense verbs, such as *My dogs were on the carpet*, tend to give rise to the inference that the corresponding present tense version, *My dogs are on the carpet*, is false. This inference is often referred to as a ‘cessation’ or ‘temporal’ inference, and is generally analyzed as a type of implicature (Thomas 2012; Altshuler & Schwarzschild 2013; Musan 1995; Magri 2009; 2011; Sudo & Romoli 2017). In the literature, there are two main proposals for capturing this asymmetry: one assumes a difference in informativity between the past and present counterparts mentioned above (Altshuler & Schwarzschild 2013), while the other proposes a structural difference between the two (Thomas 2012). The two approaches are similar in terms of their empirical coverage, but they differ in the predictions they make for language acquisition. We used a novel animated picture selection paradigm (building on Katsos & Bishop 2011) to investigate the predictions of the two approaches to temporal inferences. Specifically, we compared the performance of a group of 4–6-year-old children and a group of adults on temporal inferences, the “not all” scalar implicature of the quantifier “some”, and inferences of adverbial modifiers under negation. The results of our experiment revealed that overall, children computed all three inference types at a lower rate than the adult controls, however they

* For invaluable discussion and suggestions, we thank Pranav Anand, Ryan Bochnak, Richard Breheny, Danny Fox, Anne Mucha, Anne Renans, Guillaume Thomas, and the audiences at Ulster University, Sinn und Bedeutung 22, and LAGB. This work has been supported by the British Academy/Leverhulme trust grant SG-153-238 and the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021). Cremers and Kane contributed equally to this work and share joint first-authorship.

were more adult-like on temporal inferences and on the inferences of adverbial modifiers than on the scalar implicature of “some”. We discuss the implications of the findings, both for a developmental alternatives-based hypothesis, which posits that children’s difficulties with certain implicatures arise from a difficulty in accessing the required lexical alternatives (e.g., [Barner et al. 2011](#); [Tieu et al. 2016; 2017](#); [Singh et al. 2016](#)), as well as theories of temporal inferences, arguing that the finding that children were more (and equally) adult-like on temporal inferences and adverbial modifiers supports a structural theory of temporal inferences along the lines of [Thomas \(2012\)](#).

Keywords: temporal inferences, cessation implicatures, scalar implicatures, adverbial modifiers, alternatives-based hypothesis, language acquisition, semantics, pragmatics

1 Introduction

Sentences involving past tense verbs like (1a) tend to give rise to the inference that the corresponding present tense sentence (2a) is false. By contrast, present tense sentences like (2a) do not suggest that the corresponding past tense version (1a) is false.¹

- (1) a. My dogs **were** in the basket.
 b. \rightsquigarrow *My dogs aren't in the basket*
- (2) a. My dogs **are** in the basket.
 b. $\not\rightarrow$ *My dogs weren't in the basket*

The inference in (1b) is generally referred to as a ‘cessation’ or ‘temporal’ inference. It is typically analyzed as a type of scalar implicature, arising through a comparison between (1a) and its alternative in (2a) ([Musan 1995](#); [Magri 2009; 2011](#); [Altshuler & Schwarzschild 2013](#); [Thomas 2012](#); [Sudo & Romoli 2017](#)).

There are two main proposals in the literature for capturing the asymmetry between the sentences in (1a) and (2a). The first is based on the assumption

¹ A similar situation arises in the case of future sentences like (i) and their present tense counterparts, i.e. (ii). In order to set aside the potential complications involving the contribution of *will* (see [Santorio & Romoli To appear](#) and references therein), we will focus on past tense sentences like (1).

- (i) a. My dogs will be in the basket.
 b. \rightsquigarrow *My dogs aren't in the basket (yet)*
- (ii) a. My dogs are in the basket.
 b. $\not\rightarrow$ *My dogs will not be in the basket*

that the two sentences differ in their informativity; we will refer to this approach as the SEMANTIC approach. This approach argues that a present tense sentence like (2a) entails its past counterpart in (1a). This, in turn, predicts that when (2a) is uttered, its alternative, (1a) cannot be negated as it is entailed by the assertion itself (Altshuler & Schwarzschild 2013). An alternative analysis of the asymmetry above, which we will refer to as the STRUCTURAL approach, assumes a structural difference between a present tense sentence and its past counterpart. On this latter approach, a sentence like (1) is logically independent from a sentence like (2), but the former is more complex – and in fact structurally contains – the latter (Thomas 2012). This assumption, in combination with a structural theory of alternatives (Katzir 2007; Fox & Katzir 2011; Trinh & Haida 2015), results in (2) being an alternative of (1) but not vice versa. This in turn correctly predicts that (2) cannot have the negation of (1) as an inference.

The two approaches share assumptions regarding the semantics of the present and past tense sentences, as well as the idea that temporal inferences are a type of implicature; they thus have similar empirical coverage, for example capturing the asymmetry between (1a) and (2a). One crucial point of divergence between the two analyses, however, is that they make different developmental predictions. In particular, as we argue below, only the structural approach, in combination with a recent alternatives-based hypothesis concerning the acquisition of implicatures, predicts that children should be more adult-like on temporal inferences compared to more traditional implicatures.

The present study tests this prediction by comparing the performance of 4–6-year-old children and adults on sentences like (1). We developed a novel ‘group’ selection task based on Katsos & Bishop (2011: Exp. 3) and tested children and adults on temporal inferences, scalar implicatures, and inferences arising from adverbial modification in negative sentences. The main finding was that children were more adult-like in their derivation of temporal inferences and inferences arising from adverbial modification in negative sentences than in their derivation of the *not all* implicature of “some”. We argue that the observed pattern of performance is predicted by Thomas’s (2012) theory of temporal inferences, combined with a developmental alternatives-based hypothesis for scalar inferences (such as the Restricted Alternatives Hypothesis proposed in Tieu et al. 2016).

The remainder of this paper is organised as follows. In Section 2, we present the two approaches to temporal inferences in more detail. In Section 3, we introduce the lexical alternatives-based hypothesis concerning the acquisition of implicatures. In Section 4 we discuss the case of negated adverbial modifiers and in Section 5, we outline the experimental predictions for temporal inferences and inferences arising from negated adverbial modifiers. In Section 6, we present our experiment and in Section 7 we discuss the results in the context of the theories and their predictions.

2 Two theories of temporal inferences

As outlined in Section 1, there are two main proposals for explaining the asymmetry between past tense sentences like (1a) and present tense sentences like (2a). The SEMANTIC approach is based on informativity, while the STRUCTURAL approach is based on structural differences between the two sentences. We outline some common assumptions of the two approaches below before discussing each in greater detail.

2.1 Common assumptions

Under both approaches, a semantics for present and past sentences along the lines of (3) and (4) is assumed (where **time**(*c*) represents the time of utterance in the given context):²

- (3) $\llbracket \text{My dogs were in the basket} \rrbracket^c = \exists t [t < \mathbf{time}(c) \wedge \mathbf{in-the-basket}(\mathbf{my-dogs})(t)]$
- (4) $\llbracket \text{My dogs are in the basket} \rrbracket^c = \exists t [\mathbf{time}(c) \in t \wedge \mathbf{in-the-basket}(\mathbf{my-dogs})(t)]$

In the absence of additional assumptions, the meanings in (3) and (4) make the two sentences logically independent: the first simply tells us that there is a time prior to the utterance time in which my dogs were in the basket. This leaves open whether they are in the basket at a time that includes the utterance time, which corresponds to the conditions under which (4) is true. Conversely, (4) merely says that there is a time that includes the utterance time in which my dogs are in the basket. It leaves open whether there was a time beforehand in which they were in the basket.

In addition, under both approaches, the temporal inference arising from (3) should be analyzed as a type of implicature. The main argument for this approach to temporal inferences comes from the fact that, like standard scalar implicatures, temporal inferences are easily cancellable. The scalar implicature of “some” in (5), for example, is easily cancelled:

- (5) John met some of his students yesterday, in fact he met all of them.

² This is the semantics assumed in Sudo & Romoli (2017). The semantics assumed by Thomas (2012) and Altschuler & Schwarzschild (2013) differ slightly but not in a crucial way for our purposes. The most important difference is with respect to the assumption about stative predicates that we describe below, which changes the entailment relation between (3) and (4). We return to this in the next subsection.

The temporal inference arising from (6) can likewise be cancelled when followed by "...and they still are", without incurring a contradiction.

- (6) My dogs were in the basket.
 (7) My dogs are in the basket.

Let us briefly sketch how the implicature from (6) to the negation of (7) is derived. We will illustrate this using a simplified Gricean algorithm for scalar implicatures, acknowledging however that any theory of scalar implicatures that includes the characteristics below will be compatible with the implicature approaches we have mentioned.³ The basic idea is that rational interactions between conversational participants are guided by general principles of co-operation (see Grice 1975 and much subsequent work). In particular, the assumption is that upon hearing an utterance, the hearer will reason about what the speaker might have said instead, with various assumptions about what led the speaker to say what she said rather than something else. Among these assumptions, the one that is relevant here is the assumption that the speaker is being as informative as is required. The fact that the speaker chose to assert what they did and not something else (among a set of restricted relevant competitors) leads the hearer to conclude that the competitors that are not already entailed by the assertion must be false.⁴ Therefore in the case of temporal inferences, assuming that (6) and (7) are competitors, upon hearing the sentence in (6) the hearer will conclude that the speaker must think that (7) is false, thereby deriving the desired temporal inference that *my dogs are not in the basket*. This follows the simplified gricean algorithm in (8):

- (8) The speaker said A.
 The speaker could have said B instead (but didn't).
 It is false that B.

A problem arises, however, given the assumed literal meanings provided in (3) and (4) and the reasoning provided above, namely, we should expect to derive an implicature from (7) that is the negation of (6). That is, we would incorrectly predict from (7) the inference that *my dogs were not in the basket (before)*. Yet, as we have observed, only (6) gives rise to a temporal inference involving the negation of (7). This is where the two theoretical approaches diverge.

³ Notice in particular that at least for Thomas (2012), the theory has to be one in which implicatures are derived not only on the basis of stronger alternatives but also on logically independent ones.

⁴ More precisely, the hearer will conclude that it is not true that the speaker believes that the stronger competitors are true, which, given other assumptions, can be strengthened to the speaker's belief that the competitors are false, and even further to the competitors actually being false. See Gamut (1991) for discussion.

Under the SEMANTIC approach, the asymmetry between (6) and (7) is derived by assuming differences in the semantic relationships between the two sentence types. On the other hand, the STRUCTURAL approach derives the asymmetry by making additional assumptions regarding the syntactic structures of the two sentences, in combination with a theory of scalar competitors that is sensitive to the syntactic structure. In the next subsections, we outline the two approaches and their differences in greater detail. We then turn to the different developmental predictions they make, which constitute the motivation for the experiment we present in Section 6.

2.2 The semantic approach

As discussed above, without additional assumptions, the literal meanings of (6) and (7) are logically independent. Altshuler & Schwarzschild (2013) argue, however, that (7) actually entails (6), once the following assumption is made:

- (9) **Temporal profiles of statives:** If a tenseless stative clause ϕ is true at moment m , then there is a moment m' preceding m at which that ϕ is true.

And indeed, it is easy to see that if we assume (9), then it follows that (7) is stronger than (6): this is because if there is a time which includes the utterance time at which my dogs are in the basket, then on the basis of (9), there must be a moment prior to that time (and henceforth prior to the utterance time), in which my dogs were in the basket. This automatically makes (6) true. The opposite does not hold, of course: there can be a time prior to the utterance time in which my dogs were in the basket, without there being a time including the utterance time in which they still are in the basket. In sum, given the assumption in (9), Altshuler & Schwarzschild (2013) argue that (7) asymmetrically entails (6). This, in turn, in combination with a theory of implicature as sketched above, correctly predicts that the former cannot have the negation of the latter as an implicature, because the latter is entailed by the former.

In other words, for Altshuler & Schwarzschild (2013), the asymmetry between past and present sentences is based on in the entailment relation between them. In the next subsection we turn to an alternative approach proposed by Thomas (2012), which involves a very different way of obtaining the asymmetry between present and past sentences.

2.3 The structural approach

There exists an alternative to the semantic approach discussed above that does not involve the assumption in (9). According to [Thomas \(2012\)](#), a present sentence and its past counterpart remain logically independent and cannot be distinguished by informativity as per the semantic approach. This, in combination with a theory of implicatures that allows implicatures to arise from alternatives that are logically independent from the assertion, would correctly predict the target temporal inference from the past sentence to the negation of the present counterpart, but it would also incorrectly derive a corresponding temporal inference from the present tense sentence to the negation of the past counterpart. To address this issue, [Thomas \(2012\)](#), building on an observation in [Dowty \(1979\)](#), argues that the two sentences are structurally distinct. In particular, a present tense sentence would have the LF in (10a) (where the T head only contains a pointer to the time of utterance N), while the past tense counterpart would have the LF in (10b), involving additional covert temporal operators.

- (10) a. $[[T \ N] \ [\text{my dogs are in the basket}]]$
 b. $[[T \ \text{ONCE} \ [\ \text{PAST N}] \ [\text{my dogs are in the basket}]]$

Importantly, (10b) is more complex and in fact contains (10a). This, in combination with a structural theory of alternatives ([Katzir 2007](#); [Fox & Katzir 2011](#); [Trinh & Haida 2015](#)) immediately predicts that (10a) can be an alternative of (10b), but not vice versa. This is because under such approaches to alternatives, a sentence can have its subparts as alternatives but not vice versa. In other words, this structural asymmetry ensures that the present tense sentence is an alternative of the past tense one, but the latter is not an alternative of the former. This, in turn, correctly predicts that (10b) (i.e. (6)) can have the negation of (10a) (i.e. (7)) as an inference (given our description of the derivation of inferences above), while (10a)/(7) cannot give rise to the negation of (10b)/(6) as an inference.

In sum, the two approaches outlined above derive the asymmetry between a present tense sentence and its past counterpart with respect to temporal inferences in quite different ways. We turn next to the developmental predictions of these approaches, which diverge given recent hypotheses regarding children's performance on scalar implicatures.

3 A developmental alternatives-based hypothesis

A fairly robust finding in the developmental literature on scalar implicatures is that 4–6-year-old children tend to differ from adults in how they respond to underinformative scalar sentences. For example, several earlier studies observed that children tend to respond to underinformative sentences containing scalar terms like disjunction and the existential quantifier “some” on the basis of a *literal* interpretation, rather than one that includes the relevant scalar implicature (see Noveck 2001; Chierchia et al. 2001; Papafragou & Musolino 2003; Guasti et al. 2005, among many others). In one of the earliest studies, Noveck (2001) presented participants with sentences such as (11), which gives rise to the scalar implicature that *not all giraffes have long necks*. This implicature is falsified by common knowledge, and so participants were expected to reject the sentence if they computed the implicature.

(11) Some giraffes have long necks.

Noveck observed that child participants rejected statements like (11) less often than adults did, and took this as evidence that preschool-aged children derive fewer scalar inferences than adults. This behavioral difference between children and adults has since been replicated.

More recent studies have revealed that there are certain inferences that children can compute at adult-like rates, in contrast to their performance on standard cases like the *some-not-all* implicature. One such example is the so-called ‘free choice’ inference, investigated by Tieu et al. (2016). Consider the sentence in (12a) and its associated inference in (12b):

(12) a. Jack can have cake or ice cream.
 b. \rightsquigarrow *Jack can have cake and Jack can have ice cream*

The inference in (12b) is traditionally referred to as a ‘free choice inference’, the intuition being that the sentence in (12a) grants Jack free choice between the two options of cake and ice cream. Such inferences have received a scalar inference analysis in the formal semantics literature (Kratzer & Shimoyama 2002; Alonso Ovalle 2005; Fox 2007; Klinedinst 2007; Chemla 2009; Santorio & Romoli To appear), which we return to below. Tieu et al. (2016) investigated children’s interpretation of modal statements containing disjunction (in Mandarin) and ‘any’ (in English). Both sentence types give rise to free choice inferences. Tieu et al. observed that children computed free choice inferences around 90% of the time, whereas they computed standard implicatures involving plain disjunction

("or"/"and") and modals ("may"/"must") at typically low rates. In addition to free choice inferences, children have been reported to compute a handful of other inferences at adult-like rates, including the *exactly-n* inference of numerals (Barner & Bachrach 2010; Papafragou & Musolino 2003), ad hoc implicatures (Barner et al. 2011; Stiller et al. 2015), ignorance inferences (Hochstein et al. 2016), and various inferences of simple and embedded disjunctions (Singh et al. 2016; Tieu et al. 2017).

The seeming variability in children's performance on implicatures (i.e. their relatively poorer performance on implicatures involving plain disjunction, the quantifier "some", and the weak modal "may"/"might", compared to their strong performance on free choice, ad hoc, and numeral implicatures) has led to much recent discussion in the developmental literature. One recent developmental proposal aims to capture children's variable performance by appealing to the nature of the *alternatives* required to compute the relevant inferences. Specifically, the inferences on which children perform at non-adult-like levels tend to involve *lexical alternatives* (e.g., "some" vs. "all", "or" vs. "and", "may" vs. "must"), while the inferences that children are able to compute at adult-like levels involve alternatives that can be retrieved from the context or the sentence itself (see the 'Restricted Alternatives Hypothesis' proposed in Tieu et al. 2016 as well as discussion in Barner et al. 2011; Singh et al. 2016; Tieu et al. 2017, among others).

Under this approach, children differ from adults in computing scalar implicatures because they struggle to access the alternatives required for the computation of the relevant implicatures. This is supported by the finding that children in fact readily compute inferences for which the relevant alternatives are easily accessible or salient in the context (Barner et al. 2011; Tieu et al. 2016; Singh et al. 2016; Tieu et al. 2017, among others; see also discussion in Skordos & Papafragou 2016 for the role of the *relevance* of the required alternatives). Take the case of the free choice inference. On the implicature approach to free choice inferences, the inference in (13b) is derived through the negation of the 'substring' alternatives in (13c) and (13d) (if it's false that Jack is restricted to cake and it's false that Jack is restricted to ice cream, it follows that he can have the cake and he can have the ice cream).

- (13) a. Jack can have cake or ice cream.
 b. \rightsquigarrow *Jack can have cake and Jack can have ice cream*
 c. Jack can only have cake.
 d. Jack can only have ice cream.

Given the alternatives are provided as substrings of the assertion, children are not required to have lexicalized the scalar alternatives, nor do they have to re-

trieve scalar alternatives from the lexicon. In contrast to the free choice inference, consider a traditional implicature like (14b) from (14a). The *not-all* implicature requires that the child access the stronger scalar term “all” upon hearing “some”, a term that is not contained within the original assertion.

- (14) a. Some of my dogs jumped on the bed.
 b. \rightsquigarrow *Not all of my dogs jumped on the bed*

Summarizing, the alternatives-based hypothesis locates the source of children’s variable behavior in the nature of the alternatives involved in implicature computation. For our purposes, the most relevant prediction of this approach is that children should perform at adult-like levels on scalar implicatures for which the relevant alternatives are structurally contained within the assertion. As we will see in the next subsection, assuming the alternatives-based hypothesis, the semantic and structural approaches to temporal inferences described above make divergent predictions for how children should perform on temporal inferences.

4 The case of negated adverbial modifiers: A baseline

Sentences containing negated adverbial modifiers, such as (15a), tend to give rise to the inference that the corresponding unmodified positive sentence is true; for instance, (15a) suggests quite robustly that my dogs did jump (Simons 2001; Katzir 2007; Schlenker 2008).

- (15) a. My dogs didn’t jump high.
 b. \rightsquigarrow *My dogs jumped*

The inference in (15b) is typically analyzed as a scalar implicature arising from the negation of the simpler alternative in (16), i.e. the sentence without the adverbial modifier (Katzir 2007). As (16) is more informative than (15a), the hearer will reason along the lines above and conclude that (16) is false, thereby concluding (15b).

- (16) My dogs didn’t jump.

This kind of explanation also extends to analogous inferences arising in other downward-entailing contexts, as shown by Katzir (2007). What is crucial for our purposes is that the inference of the negated adverbial modifier arises from an alternative that is structurally contained within the assertion, as shown in (17a) and (17b).

- (17) a. [not[my dogs [jumped [high]]]]
 b. [not[my dogs [jumped]]]

This inference therefore presents a straightforward test of the alternatives-based hypothesis. Children are expected to compute this inference at an adult-like rate, since the required alternative is easily retrievable from the assertion itself. The inference of negated adverbial modifiers will therefore provide a reasonable baseline for comparison with temporal inferences, on the one hand, and the standard lexical scale-based implicature of “some”, on the other hand.

5 Predictions

The two theories of temporal inferences that we have discussed both manage to capture the relevant temporal inference that arises from past tense sentences like (6). When we turn to child language, however, we see that the two theories make divergent predictions for how children should perform on temporal inferences.

Let us consider first the **STRUCTURAL** approach, which assumes that a present tense sentence is structurally contained within its corresponding past tense one. This assumption, in combination with a developmental alternatives-based hypothesis leads to the prediction in (18) that children should be more adult-like in their computation of temporal inferences, compared to how they perform on classical scalar implicatures.

- (18) **Prediction of the STRUCTURAL approach:** Children will display more adult-like behavior on temporal inferences than on standard scalar implicatures involving lexical replacement.

Under the **SEMANTIC** approach, on the other hand, temporal inferences arise as a regular implicature. Crucially, this theory does not make any assumptions about the structural relationship between the past tense sentence and its present counterpart. There is, therefore, no expectation that children should be more or less adult-like on temporal inferences compared to other implicatures. If anything, without further assumptions about the syntax of these sentences, this approach instead predicts that children should display similar performance on temporal inferences and classical scalar implicatures, or at least that they might differ from adults in the same way on the two inferences (see [Tieu et al. Under review](#) and [Renans et al. 2017](#) for discussion of a similar ‘uniformity’ prediction in the domain of plurality inferences). That is, if children differ from adults, they should differ to the same degree for temporal inferences and standard implicatures; there is no

specific theoretical reason to expect that they should fare better (i.e. be more adult-like) on one inference compared to another:⁵

- (19) **Prediction of the SEMANTIC approach:** The difference that is observed between children and adults on standard implicatures, if any, should also be observed for temporal inferences.

In sum, comparing temporal inferences to classical scalar implicatures, the STRUCTURAL approach predicts an interaction between group and inference type, with children performing more like adults on temporal inferences than on standard implicatures (that require lexical replacement). On the other hand, the SEMANTIC account predicts no such interaction.

To test the predictions above, we designed an experiment to test children and adults on temporal inferences, the standard *not all* implicature of the quantifier “some”, and the inferences of negated adverbial modifiers. The case of “some” provides a baseline inference on which children are expected to differ from adults, as the inference requires the lexical replacement of alternatives. The case of the negated adverbial modifier involves alternatives that are contained within the assertion, and therefore provides a baseline inference for which we expect to observe more adult-like behavior from children, since no lexical replacement of alternatives is required.

To summarize, we have two inferences that will serve as a baseline against which temporal inferences will be compared, one involving replacement of lexical alternatives, as in (20), and one involving non-replacement alternatives, as in (21).

- (20) a. **Some** of my dogs jumped.
 b. **All** of my dogs jumped. REPLACEMENT ALTERNATIVE
 c. \rightsquigarrow *Not all of my dogs jumped*
- (21) a. My dogs didn't jump high.
 b. My dogs didn't jump. NON-REPLACEMENT ALTERNATIVE
 c. \rightsquigarrow *My dogs jumped*

⁵ As an anonymous reviewer points out, one could supplement the SEMANTIC approach with an assumption about saliency of alternatives, for instance that temporal alternatives are simply more salient to children than the alternative of *some*. This has been proposed in the case of numerals, where children's performance is typically more adult-like compared to their performance on other scalar terms (Papafragou & Musolino 2003, among others). Unlike the case of numerals, however, where children are explicitly taught the numeral scale from an early age, and therefore might plausibly be more familiar with the alternatives, it is not clear to us why temporal alternatives should be more salient for children than scalar quantifier alternatives. Without independent justification for distinguishing among these alternatives through saliency, we therefore identify the main prediction of the SEMANTIC approach as that in (19).

Given this three-way comparison, the `STRUCTURAL` approach, in combination with the developmental alternatives-based hypothesis, gives rise to the prediction in (22).

- (22) **Prediction of structural approach in combination with alternatives-based hypothesis:** Children’s performance on temporal inferences and negated adverbial modifiers should be more adult-like than their performance on the *not all* implicature of “some”.

We turn now to our experiment, which tested the prediction in (22).

6 Experiment

We tested the predictions outlined in Section 5 by comparing the performance of a group of 4–6-year-old children with a group of adults on temporal inferences, the classical scalar implicature of “some”, and negative sentences involving adverbial modification.

6.1 Methods

6.1.1 Participants

38 English-speaking children (4;02–5;11, $M=5;04$) recruited from preschools in Belfast and 40 English-speaking adults recruited through Amazon Mechanical Turk (age range 21–55) participated in the experiment. One child and four adults were excluded from analysis because their error rate on control and filler items exceeded 25%, leaving a total of 37 children and 36 adults for analysis.

6.1.2 Procedure

We developed a novel task based on the sentence-to-picture-matching task employed in [Katsos & Bishop \(2011: Exp. 3\)](#), which involved matching a spoken sentence to one of three picture alternatives featured in an animated video sequence.⁶ Participants watched a series of animated videos, each involving three

⁶ A pilot study using a truth value judgment task showed that adults were unwilling to reject sentences like (1) when its literal meaning was true but its temporal inference was false. We thus moved to a selection task, which intuitively made the reading with the implicature more attractive, but at the time were not aware of the results reported in [Katsos & Bishop \(2011: Exp. 3\)](#). As we will discuss further in the Discussion, our experiment did not replicate Katsos & Bishop’s findings for the quantifier “some”: even with the selection task, the children we tested were not adult-like at deriving the *not all* implicature.

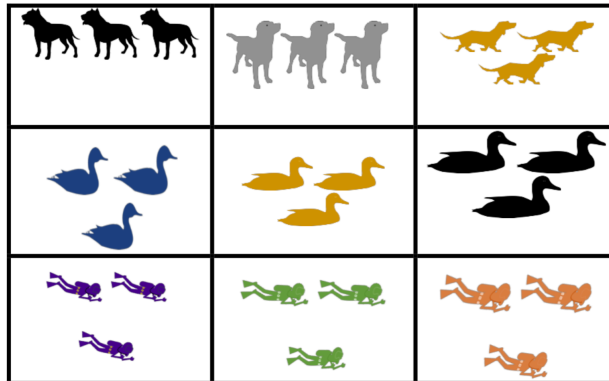


Figure 1: Sample answer sheet for collecting responses (children).

groups of characters (differently colored dogs, birds, divers, or ducks). The participants' task was to 'guess' which colour characters belonged to Raffie based on a clue that the puppet would give at the end of each video sequence. At the beginning of each video sequence, the narrator introduced the characters on the screen. A sample introduction is given below ⁷

- (23) Narrator: *There are three dogs playing in the room. See! Black dogs, grey dogs and brown dogs. One of the three groups of dogs belongs to Raffie, your job is to guess which colour dogs are his. Listen for the clue.*

Once the characters had completed their animated actions, a bell would sound, indicating that the puppet was about to provide a clue. The puppet then provided a clue relating to the action. A sample clue from the adverbial modifier condition is provided in (24).

- (24) Puppet: *My dogs didn't jump high.*

Participants were then asked to provide a response to the question 'Which dogs belong to Raffie?' Child participants provided their responses by placing stickers on multiple choice answer sheets that corresponded to the characters on the screen (see Figure 1). Adult participants provided their responses by clicking on one of three response buttons indicating the colour associated with the group of dogs they wanted to choose as their answer (e.g., 'Black', 'Grey', 'Brown').

⁷ Testing with child participants was done face to face with the narrations carried out by the individual conducting the experiment. For adult participants, we developed an online version of the experiment. An anonymized online version of the experiment can be found at <http://spellout.net/ibexexps/AnonymizedExps/SuB>.

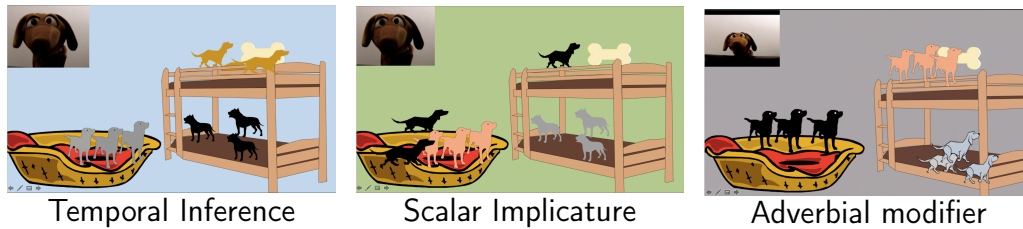


Figure 2: Screen captures of target items in each condition.

6.1.3 Materials

The three groups of characters (*Literal*, *Target*, *False*) differed in whether they made the literal interpretation and the implicature of the sentence true or false. The *Literal* group satisfied the literal interpretation of the puppet's clue, but not its implicature; the *Target* group was consistent with both the literal meaning and the implicature of the puppet's sentence. Deriving the appropriate inference was thus necessary in order to distinguish between these two groups. The third group, the *False* group, was a distractor that failed to satisfy the literal meaning of the sentence; the inclusion of these distractors allowed us to check whether children had correctly understood the task.

The target sentences and their respective implicatures are provided in (25)–(27). Figure 2 provides sample screen shots of the target items for the TI, SI, and AM conditions. In Table 1 we describe the animations for each character group. Figure 3 provides an example of the visual display prior to and after the relevant action for the TI condition.

- (25) My dogs were in the basket.
 ~> *My dogs aren't in the basket* TEMPORAL INFERENCE (TI)
- (26) Some of my dogs are in the basket.
 ~> *Not all of my dogs are in the basket* SCALAR IMPLICATURE (SI)
- (27) My dogs didn't jump high.
 ~> *My dogs jumped* ADVERBIAL MODIFIER UNDER NEGATION (AM)

Each participant received 23 trials in total: 4 repetitions of each target type, 6 control items that contained a Target Group and two distinct False Groups (2 for each target type), 2 present tense controls, and 3 fillers. Items were presented in one of two pseudo-randomized orders, which ensured, among other things, that the present tense controls never appeared before a TI target item (since this would make the alternative contextually salient for the temporal inference). To make it clear that the puppet's use of the past tense could not refer to a time coinciding

	Target Group	Literal Group	False Group
Literal meaning:	True	True	False
Inference:	True	False	False
TI (25)	in basket, then move	in basket throughout	never in basket
SI (26)	some on the bed, some not	all on the bed	none on the bed
AM (27)	jumped low	didn't jump	jumped high

Table 1: Location of Target, Literal, and False groups during animation, for each condition.

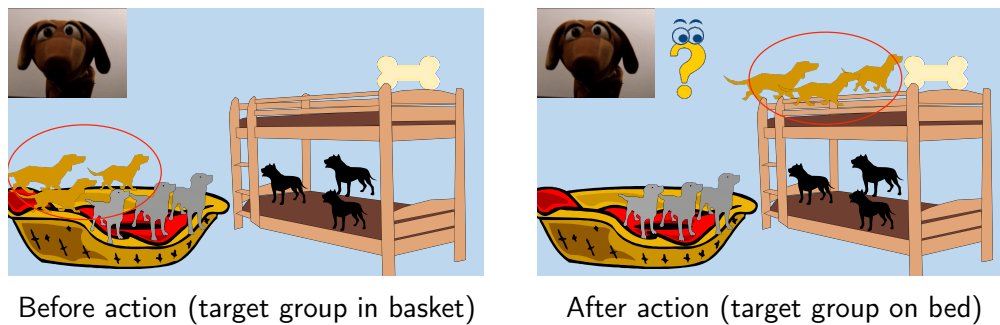


Figure 3: Example stimuli for the TI item “My dogs were in the basket”.

with the end state, the characters remained animated on the screen until a response was provided.

6.2 Results

The mean error rate on control and filler items before exclusion of any participants was 12% for adults and 6% for children, suggesting that the task did not pose any particular difficulty. We then excluded four adult participants and one child whose error rate was above 25% error rate. Figure 4 presents the proportion of Target, Literal, and False group selections for the target items, by adults and children.

For statistical analysis, we excluded the False responses (under the assumption that they reflected difficulty in understanding the stimuli/situation and not the target sentence) and focused on Target and Literal responses on the target trials. The dependent variable was therefore a two-level categorical variable. Figure 5 provides a boxplot displaying each participant’s behavior once False responses are excluded. A mixed-effects binomial logistic regression model was fitted to the data predicting the probability of a Target response as a function of Condition (TI, SI or AM, treatment-coded with TI as a baseline), Group (child vs. adult,

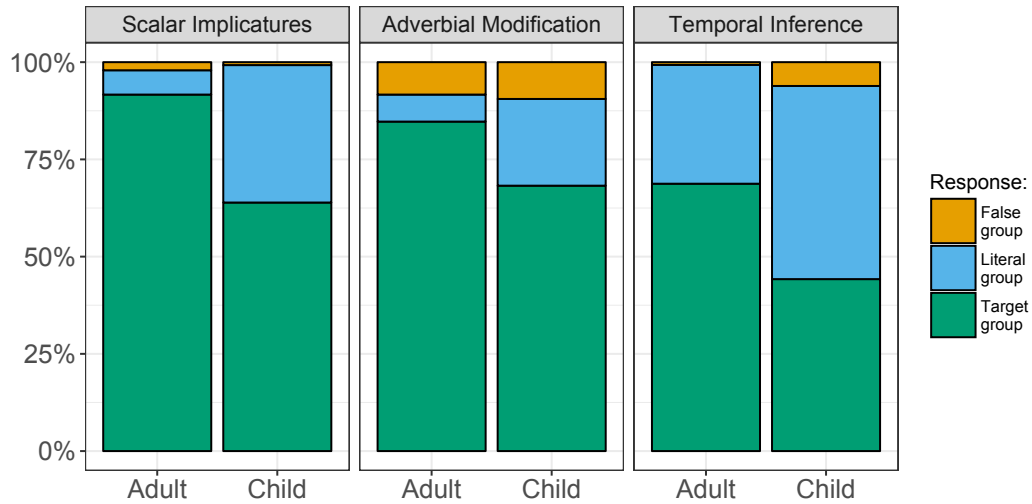


Figure 4: Percentages of group selections per condition.

sum-coded), and their interaction. We included random slopes for Condition and further simplified the random effects structure following the recommendations of Bates et al. (2015).⁸ The detailed results are provided in Table 2.

	Estimate	<i>z</i> -value	<i>p</i> -value	
(Intercept)	0.6	1.6	0.102	
[Condition:AM]	1.5	3.4	<.001	***
[Condition:SI]	2.4	4.0	<.001	***
[Group]	-1.5	-2.4	0.015	*
[Condition:AM] × [Group]	-.06	-0.1	0.920	
[Condition:SI] × [Group]	-2.2	-2.3	0.022	*

Table 2: Results of binomial logistic regression model on participants' responses to targets (excluding False responses)..

The model revealed a main effect of Condition, indicating that participants overall selected the Target more often in the AM and SI conditions than in the TI condition. There was also a main effect of Group, showing that adults selected the

⁸ More concretely, after fitting a maximal model in the sense of Barr et al. (2013), we ran a principal component analysis on the random effects and refitted the model, keeping only components which explained at least 5% of the variance explained by the main component. The selection of components for Subject and Item random effects was done independently, as the random effects for Item tend to be much smaller. The final 'parsimonious' model included two Subject random effects and one Item random effect.

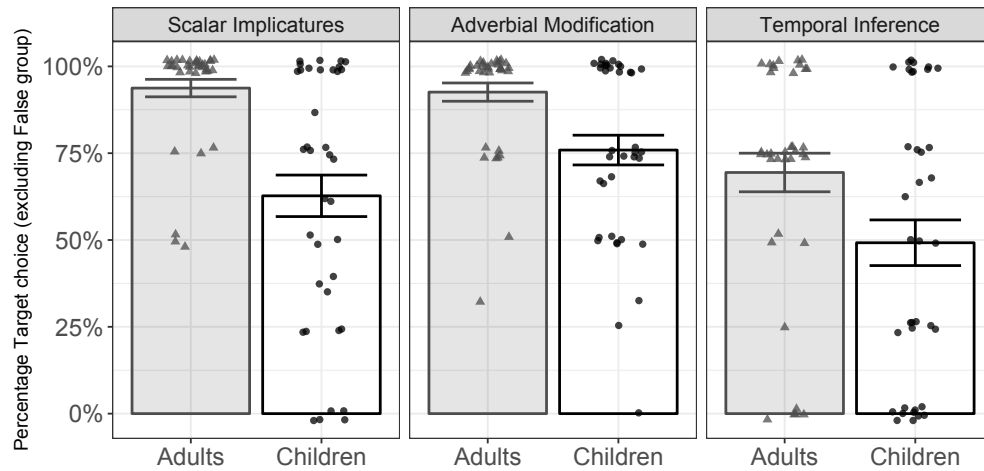


Figure 5: Percent Target group selection against Literal group selection for individual participants. Error bars correspond to standard errors.

Target group more often than children in the TI condition. Crucially, there was a significant interaction between Group and Condition when comparing the TI and SI conditions, but no such interaction when comparing the TI and AM conditions, indicating that children were equally adult-like on the AM and TI conditions, but were significantly less adult-like in the SI condition.

7 Discussion

The current study employed a novel animated selection paradigm building on [Katsos & Bishop \(2011\)](#). We investigated the developmental predictions of the STRUCTURAL and SEMANTIC approaches to temporal inferences by comparing the performance of a group of 4–6-year-old children and a group of adults on temporal inferences like in (28), the classical *not all* implicature of the quantifier “some” in (29), and the inference of negated adverbial modifiers, like (30).

- (28) My dogs were in the basket.
 ↪ *My dogs aren't in the basket*
- (29) Some of my dogs are in the basket.
 ↪ *Not all of my dogs are in the basket*
- (30) My dogs didn't jump high.
 ↪ *My dogs jumped*

The results of our experiment reveal that overall, children computed all three inference types at a lower rate than the adult controls. Importantly, however, while children differed from adults across inference types, they were more adult-like on temporal inferences and the inferences of adverbial modifiers than on scalar implicatures. In particular, we found an interaction between group and inference type when comparing temporal inferences and the implicature of *some*, but no such interaction when comparing temporal inferences with the inference of adverbial modifiers. The children's data overall supports the alternatives-based hypothesis we discussed in Section 3: children were relatively more adult-like on adverbial modification, which does not require any lexical replacement of alternatives, while they performed relatively worse on the implicature of "some" which requires the lexical replacement of "some" with "all". Moreover, the finding that children were equally adult-like on adverbial modification and temporal inferences lends support to a structural approach to temporal inferences along the lines of [Thomas \(2012\)](#). Under such an approach, the present tense sentence is structurally contained within its corresponding past tense one, and no lexical replacement of alternatives is required to derive the inference.

A further finding of the current study is that both children and adults computed fewer temporal inferences overall than either scalar implicatures or adverbial inferences. This observed variability across inference types is somewhat reminiscent of recent work indicating that adults compute different scalar implicatures at varying rates ([van Tiel et al. 2016](#)). However, as discussed below, our forced-choice selection task should minimize such variation because it rewards the derivation of any relevant inference. While we do not have a definitive explanation for the observed difference, we suspect it may come from uncertainty as to which time counts as 'present' in the context of the task; despite our efforts, some participants may have anchored the whole story in the past, and this would have made the temporal inference irrelevant. The lower rate of Target responses in adults further complicates the interpretation of our results. First, with varying baselines, the statistical estimates for interactions become dependent on the choice of link function in the binomial model. We verified that our results held with probit, Cauchy CDF, and complementary log-log links, so we are confident that the observed pattern of interaction is not an artifact of the logit link function. Second, and more problematic, we cannot rule out that the low rates of target choices in adults and children have different sources. For instance, one might imagine that adults are more likely than children to anchor the story in the past, while the low rate of Target responses in children mostly reflects difficulty with the derivation of the temporal inference. As a consequence, our conclusion relies on the assumption that whatever caused a low rate of Target responses in adults had the same effect on children, such that

the interaction with the SI and AM conditions only measures children's specific difficulty with the derivation of the implicature.

Notice that the finding that children computed fewer implicatures from "some" than adults diverges from the results reported in [Katsos & Bishop \(2011\)](#), who found children to be adult-like on "some" in their selection paradigm. Our finding of non-adult-like behavior is more consistent with previous studies that have used the more traditional binary truth value judgment task (e.g., [Papafragou & Musolino 2003](#)). One potential reason for the difference in findings might be a matter of sample size: we tested 38 children, while Katsos and Bishop tested 15 children in their selection paradigm. A more interesting reason for the discrepancy in findings could be that [Katsos & Bishop \(2011\)](#) included "all" controls, while we were careful not to include the scalar alternatives required for the derivation of the target inferences. While these factors may not entirely explain the divergent findings, they could partially explain why children appeared to be relatively more adult-like in Katsos & Bishop's experiment compared to ours.

Finally, let us consider the use of the selection paradigm in the context of an existing explanation for children's performance on implicatures. As we have already observed, the selection paradigm seems to be well-suited for capturing inferences that are not strong enough to trigger pure rejection in a standard binary truth value judgment task. In fact, [Katsos & Bishop \(2011\)](#) propose that a binary judgment task might *underestimate* children's ability to compute implicatures. This is because, as they propose, children are more *pragmatically tolerant* of underinformative statements than adults are, leading them to accept an implicature-violating sentence in a binary truth value judgment task, even if they are able to derive the implicature. While the selection paradigm does seem to be better suited at capturing weaker inferences like the temporal inference, the children we tested nevertheless displayed difficulties with implicatures that may not be fully explained by the notion of pragmatic tolerance. In principle, the selection task should help to circumvent this issue because it strongly encourages the derivation of an inference, since the literal meaning is not sufficient to narrow down the choice to a single answer. By contrast, the truth value judgment task requires participants to make a binary choice between accepting an underinformative utterance or rejecting it on the basis of pragmatic considerations. Similarly, the covered-box paradigm (see [Huang et al. 2013](#) and [Pearson et al. 2011](#)), which has also been used to test for the derivation of inferences, also requires participants to make a binary choice between accepting a sub-optimal picture or opting for the unknown covered one (participants must choose the visible picture if it is compatible with the target sentence, and the covered picture otherwise). In our selection paradigm, the choice is much easier: after accessing the literal reading of the puppet's statement, participants can rule out the False group, but would

not at that point be able to make a choice between the two remaining groups. Stopping here would require them to make a random choice between the Literal and Target groups (or base their choice on other, non-linguistic considerations), and this might then push them to derive any inference that could help them to decide amongst the pictured groups. In a sense, this task translates the pragmatic violation that ensues from the use of an underinformative statement into a very concrete communication failure in the absence of the implicature, penalizing participants who do not derive an implicature by requiring them to find other cues to make a choice. Nevertheless, we observed differences between children and adults on the target inferences, suggesting pragmatic tolerance might not be the entire explanation.

8 Conclusion

The current study employed a novel animated selection paradigm to investigate the developmental predictions of the STRUCTURAL and SEMANTIC approaches to temporal inferences. The results of our experiment reveal that overall, children computed all three inference types at a lower rate than the adult controls, but they were more adult-like on temporal inferences and the inferences of adverbial modifiers than on the scalar implicature of “some”, which requires the lexical replacement of “some” with “all”. We discussed the implications of the findings, both for a developmental alternatives-based hypothesis, which posits that children’s difficulties with certain implicatures arise from a difficulty in accessing the required lexical alternatives, as well as for theories of temporal inferences, arguing that the finding that children were more (and equally) adult-like on temporal inferences and adverbial modifiers supports a structural theory of temporal inferences along the lines of Thomas (2012).

References

- Alonso Ovalle, Luis. 2005. *Disjunction in alternative semantics*: UMASS dissertation.
- Altshuler, Daniel & Roger Schwarzschild. 2013. Moment of change, cessation implicatures and simultaneous readings. In *Proceeding of Sinn und Bedeutung* 17.
- Barner, David & Asaf Bachrach. 2010. Inference and exact numerical representation in early language development. *Cognitive Psychology* 60. 40–62.
- Barner, David, N. Brooks & Alan Bale. 2011. Accessing the unsaid: the role of scalar alternatives in children’s pragmatic inferences. *Cognition* 188. 87–96.

- Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967* .
- Chemla, Emmanuel. 2009. Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics* 2(2).
- Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini & Luisa Meroni. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th boston university conference on language development*. 157–168. Cascadilla Press Somerville, MA.
- Dowty, David R. 1979. *Word meaning and Montague grammar*, vol. 7. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. Palgrave.
- Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107.
- Gamut. 1991. *Logic, language and meaning*. University of Chicago Press.
- Guasti, Maria Teresa, Gennaro Chierchia, Stephen Crain, Francesca Foppolo, Andrea Gualmini & Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes* 20(5). 667–696.
- Hochstein, Lara, Alan Bale, Danny Fox & David Barner. 2016. Ignorance and inference: Do problems with Gricean epistemic reasoning explain children's difficulty with scalar implicature? *Journal of Semantics* 33(1). 107–135.
- Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What exactly do numbers mean? *Language Learning and Development* 9. 105–129.
- Katsos, Napoleon & Dorothy VM Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81.
- Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistic and Philosophy* 30(6). 669–690.
- Klinedinst, Nathan. 2007. *Plurality and possibility*: UCLA dissertation.
- Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Yukio Otsu (ed.), *Proceedings of the Tokyo conference on psycholinguistics*, vol. 3. 1–25. Tokyo: Hituzi Syobo.
- Magri, Giorgio. 2009. *A theory of individual-level predicates based on blind mandatory scalar implicatures*: Massachusetts Institute of Technology dissertation.

- Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in DE environments. In David Nan Li & Lutz (ed.), *Semantics and Linguistic Theory (SALT) 20*. Vancouver, British Columbia.
- Musan, Renate. 1995. *On the temporal interpretation of noun phrases*. Cambridge: Massachusetts Institute of Technology dissertation.
- Noveck, Ira. 2001. When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* 78(8). 165–188.
- Papafragou, Anna & Julien Musolino. 2003. Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* 86. 253–282.
- Pearson, Hazel, Manizeh Khan & Jesse Snedeker. 2011. Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of scalar implicature. In David Lutz & Nan Li (eds.), *Semantics and Linguistic Theory (SALT) 20*. Cornell University Ithaca, NY: CLC Publications.
- Renans, Agata, Jacopo Romoli, Maria-Margarita Makri, Lyn Tieu, Hanna de Vries, Raffaella Folli & George Tsoulas. 2017. Abundance inference of pluralised mass nouns is an implicature: Evidence from Greek. Ms. Ulster University and University of York.
- Santorio, Paolo & Jacopo Romoli. To appear. Probability and implicatures: a unified account of the scalar effects of disjunction under modals. *Semantics & Pragmatics* .
- Schlenker, Philippe. 2008. Presupposition projection: the new debate. *SALT proceedings* .
- Simons, Mandy. 2001. On the conversational basis of some presuppositions. In Rachel Hastings, Brendan Jackson & Zsafia Zvolenszky (eds.), *Semantics and Linguistic Theory (SALT) 11*. 431–448.
- Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar & Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24(4). 305–352.
- Skordos, Dimitrios & Anna Papafragou. 2016. Children’s derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18.
- Stiller, A., N. Goodman & M. Frank. 2015. Ad-hoc implicature in preschool children. *Language Learning and Development* 11(2). 176–190.
- Sudo, Yasutada & Jacopo Romoli. 2017. Lifetime effects as presuppositional scalar strengthening. In *Proceedings of SALT 27*.
- Thomas, Guillaume. 2012. *Temporal implicatures*: MIT dissertation.
- Tieu, Lyn, Cory Bill, Jacopo Romoli & Stephen Crain. Under review. Testing theories of plural meanings: Insights from child language.
- Tieu, Lyn, Jacopo Romoli, Peng Zhou & Stephen Crain. 2016. Children’s knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*

- 33(2). 269–298.
- Tieu, Lyn, Kazuko Yatsushiro, Alexandre Cremers, Jacopo Romoli, Uli Sauerland & Emmanuel Chemla. 2017. On the role of alternatives in the acquisition of simple and complex disjunctions in French and Japanese. *Journal of Semantics* 34(1). 127–152.
- Trinh, Tue & Andreas Haida. 2015. Building alternatives. *Journal of Semantics* .
- van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175.