

## Children's comprehension of syntactically-encoded evidentiality\*

Lauren Winans,<sup>1</sup> Nina Hyams,<sup>1</sup> Jessica Rett<sup>1</sup> & Laura Kalin<sup>2</sup>

<sup>1</sup>University of California, Los Angeles & <sup>2</sup>University of Connecticut

### 1. Introduction

This paper examines children's acquisition of English copy-raising constructions (henceforth CRCs). These constructions are of particular interest because they encode evidentiality (Asudeh & Toivonen 2012, Rett et al. 2013), which is the implicit citing of the speaker's source of evidence for an at-issue proposition. A classic example of evidentiality is found in Cuzco Quechua and exemplified below in (1)-(3):

- (1) Para-shan-n-**mi**. (Faller 2002, 2006)  
rain-PROG-3-mi  
p = 'it is raining'  
ev = speaker sees that p
- (2) Para-shan-n-**si**.  
rain-PROG-3-si  
p = 'it is raining'  
ev = speaker was told that p
- (3) Para-shan-n-**chá**.  
rain-PROG-3-chá  
p = 'it is raining'  
ev = speaker conjectures that p

---

\*Our thanks to Anya Mancillas, Akane Nakagaki, Allison Echeveria, and Robyn Orfitelli for helpful discussions; Katie Sloss and Oliver Northrup for help with the pictures; Jesse Harris and Adam Chong for help with the statistics; and to all the children that participated as well as the staff at the UCLA Fernald Center, University Center, Bright Horizons, and Children's Corner. We're grateful to audiences at NELS 45, the 89th LSA Annual Meeting, and the UCSD Developmental Lunch for comments. This research was funded by a 2011-2012 COR grant from the UCLA Faculty Senate to N. Hyams and J. Rett.

The sentences in (1)-(3) all have the same at-issue content (i.e., that it is raining), but differ in their commitments regarding how the speaker came to know that information (Faller 2002, 2006). In Cuzco Quechua, evidentiality is encoded morphologically and obligatorily. In contrast, there are languages like English that optionally encode evidentiality. In both types of language, the evidential component is variously described as encoding not-at-issue or speech-act-level content (Faller 2002, 2006, Murray 2010). That is, the meaning contributed by the evidential is backgrounded and is not directly challengeable.

Previous studies on the acquisition of evidentiality have focused on languages in which evidentiality is encoded morphologically and obligatorily, including Turkish (Aksu-Koç and Slobin 1986, Aksu-Koç 1988, Ozturk and Papafragou 2007, 2008, Aksu-Koç et al. 2009), Korean (Choi 1995, Papafragou et al. 2007), Bulgarian (Fitneva 2008), Tibetan (de Villiers et al. 2009), and Quechua (Courtney 2008).<sup>1</sup> In this paper we examine the acquisition of evidentiality in English, a language which, along with Swedish, has been claimed to encode evidentiality syntactically (and optionally) in CRCs (Asudeh and Toivnen 2012).<sup>2</sup> An example is in (4).

- (4) a. Ernie looks like he got sick.  
b. It looks like Ernie got sick.

The sentences in (4) differ in their evidential contribution. The raised sentence, (4a), is acceptable only when the speaker has direct evidence, that is, in situations where the speaker has directly perceived Ernie. In contrast, the unraised version, (4b), is acceptable in both direct-evidence situations (as above) and indirect-evidence situations, those in which the speaker has not seen Ernie but instead has inferred that Ernie is sick from some other evidence (e.g., a doctor's note). The unraised form is thus unspecified with respect to evidentiality. This evidential pattern, which was confirmed for the adult grammar in an on-line felicity judgment task described in Rett et al. 2013,<sup>3</sup> is summarized in Table 1.

(5) *Evidential properties of English CRCs*

Evidential Source	Raised (4a)	Unraised (4b)
<i>Direct</i>	✓	✓
<i>Indirect</i>	✗	✓

In the following sections we report and compare the results of two acquisition studies on CRCs. Section 2 reviews our study of children's naturalistic production of CRCs (Rett et al. 2013, Rett & Hyams 2013). In Section 3 we present the results of a new comprehension study of CRCs. A comparison across these two studies shows a strong production-comprehension asymmetry: in production children have an adult-like distribution of CRCs by age two or three, while performance on the comprehension task is not adult-like even at age six. In Section 4, we provide an overview of various cross-

<sup>1</sup>See Koring and de Mulder (2011) for a study of Dutch evidential verbs.

<sup>2</sup>A theory-neutral term for this construction is 'perception verb similatives' (Rett & Hyams 2013).

<sup>3</sup>This study included 90 participants and asked speakers to rate the appropriateness of raised and unraised CRCs in a variety of contexts.

linguistic findings on the acquisition of evidentiality in languages that encode evidentiality morphologically. These studies consistently find an asymmetric pattern similar to what we observe in English. In Section 5, we discuss this asymmetry generally and propose possible explanations. In addition to the well-known difficulty associated with testing the acquisition of not-at-issue content, we also point out problems that are specific to testing evidential meanings. Section 6 concludes.

**2. Naturalistic production study (Rett et al. 2013)**

In this section we briefly review the findings of a naturalistic production study of English CRCs. The study was based on several corpora in the CHILDES database, listed in (6), of English-speaking children between the ages of two and six (MacWhinney and Snow 1985). (For more details of this study see Rett et al. 2013 and Rett & Hyams 2013.)

- (6) Bates, Bernstein-Ratner, Bliss, Bloom 1970, Bloom 1973, Bohannon, Brent, Brown, Clark, Cornell, Demetras Trevor, Demetras Working, Evans, Feldman, Garvey, Gathercole, Gleason, Haggerty, Hall, Higginson, HSLLD, Kuczaj, MacWhinney, McCune, Morisset, Nelson, New England, Peter Wilson, Post, Providence, Sachs, Snow, Soderstrom, Suppes, Tardif, Valian, Van Houten, Van Kleek, Warren-Leubecker, Weist

In total we found 70 utterances of declarative CRCs that contained the verbs *look*, *sound*, and *seem*. These utterances were coded for syntax (raised or unraised) and evidence type (direct or indirect). Fifty-four utterances were unambiguous for these two factors. The distribution of the children's CRCs is given in (7).

- (7) *Copy-raising and evidence source in English-speaking children (Rett et al. 2013)*

Evidence source	Syntax		Total
	Raised	Unraised	
<i>Direct</i>	21 (51%)	20 (49%)	41
<i>Indirect</i>	0	14 (100%)	14
Total	21	34	55

The pattern in (7) is perfectly aligned with the adult system as schematized in (5). In direct-evidence situations, children produce both raised and unraised sentences in roughly equal proportions. In marked contrast, in indirect-evidence situations, children only produce unraised sentences. That is, we found no illicit raised sentences in indirect-evidence situations. This pattern was observed even in the youngest children, ages two and three. In short, we found that English-speaking children correlate the evidence they have with the syntax of copy-raising from their earliest productions of the construction.

Moreover, despite the non-obligatory, non-grammaticized nature of evidentiality in English, the children in this study show no production delay relative to children acquiring evidential languages such as Turkish in which evidentiality is morphologically-encoded and obligatory (as we discuss in section 4). This goes against the neo-Whorfian view that language-specific properties influence the age at which children acquire the conceptual understanding of evidential source (and hence the semantics of evidentiality). In other

words, early attention to evidential source does not depend on having a language with obligatory, morphologically-encoded evidentiality. These results are in line with proposals by Papafragou et al. (2004), Gleitman & Papafragou (2005), and Koring & de Mulder (2011), for example, who argue that the conceptual framework for marking linguistic evidentiality is in place at a relatively young age and not subject to much language-specific variation.

### 3. Comprehension study

In this section, we discuss a comprehension task on English CRC constructions carried out with four to six year-olds and adult controls.

#### 3.1 Methods

Subjects were tested in a ‘felicity judgment task’ in which they were asked if a puppet’s description of a picture is “good” or “silly”. They received the instructions in (8).

- (8) “This is our friend Frank (*pointing to puppet*). Frank is learning how to describe pictures. He is pretty good, but sometimes he says things that are a little bit silly. Do you think you can help Frank and tell him if his description is good or silly?”

The experiment had three phases: training, familiarization, and testing. The testing phase was comprised of twenty test items and eight fillers. Examples of test items are provided in (9) and (10). The test items varied on two dimensions: the syntax (raised (9a) or unraised (9b)), and the context/picture (indirect evidence (10a) or direct evidence (10b)).

- (9) a. Ernie looks like he got sick.  
b. It looks like Ernie got sick.

- (10) a. *Indirect evidence*



- b. *Direct evidence*



There were a total of ten predicates that could occur in any of the four conditions, yielding forty test items; see Appendix for a complete list of test sentences. Each subject saw only half of the experimental sentences, that is, no subject saw the same picture or heard the same sentence twice. As each picture appeared, an experimenter would discuss the picture with the subject. For example, when (10a) appeared the experimenter would say that the picture is of Ernie’s room, pointing out the tissues, cough syrup, and the inscription on the balloon saying “get well soon”. Next the experimenter would ask the

puppet to describe the picture. The puppet would then utter either a raised or an unraised CRC, as in (9), and would then ask the child, "Is that a good way of saying what's going on in the picture, or is it a silly way of saying what's going on in the picture?"

Prior to the testing, children were trained on giving felicity judgments. The infelicitous training items, as well as six of the eight fillers, consisted of cases of presupposition failure, under-informative descriptions, and generally infelicitous descriptions. For example, in one picture Ernie is dressed like a pirate, and the puppet says, "Ernie's pirate ship is big." Since the picture does not show a pirate ship, the sentence is not a good description of the picture. During the training, the experimenter emphasized that the goal was to help the puppet, Frank, give the best description possible. In designing the training and fillers, we used insights from the literature on the acquisition of scalar implicatures (Papafragou & Musolino 2003, Guasti et al. 2007, Foppolo et al. 2012). During the training phase, subjects received feedback that emphasized the difference between the picture being silly and Frank's utterance being silly, as well as the difference between a true description and a felicitous one. After the training phase, no further feedback was provided.

The familiarization phase contained only two pictures and exposed the complete paradigm by providing a raised and unraised form of a single sentence. For each of the two pictures, the puppet would say the first sentence (e.g., *Ernie looks like he is a pirate*) and give the subject a chance to respond. He would then say that he had another sentence for the picture and say the opposite form (e.g., *It looks like Ernie is a pirate*). The order of raised and unraised sentences for this phase was varied between subjects.

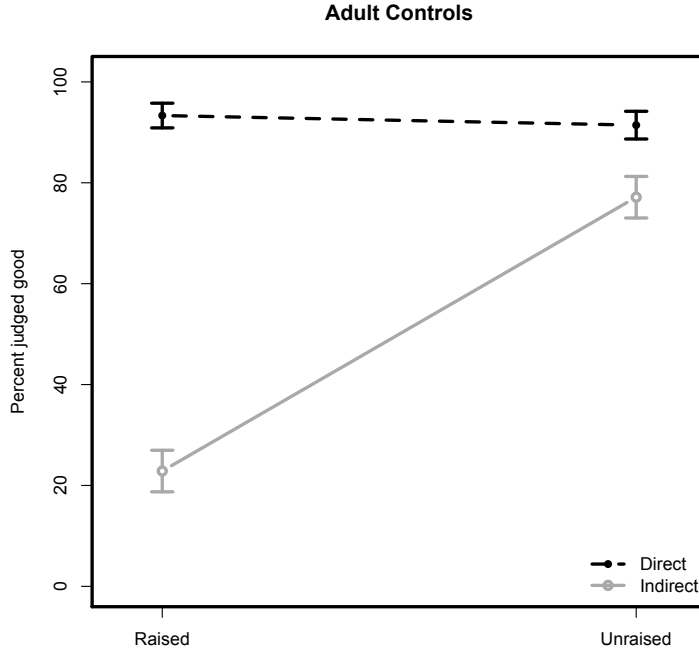
### **3.2 Subjects**

We tested forty-two children and twenty-one adult controls (UCLA undergraduates). The children were divided into three groups: four year-olds (N=13, mean age=4;5), five year-olds (N=13, mean age=5;5), and six year-olds (N=16, mean age=6;7). The data from eleven other children were excluded from the analysis because they either responded "good" to all items or responded "silly" to all items; most of the excluded children were four year-olds. The experiment was conducted at either the UCLA Language Acquisition Lab or schools in the Los Angeles area.

### **3.3 Results**

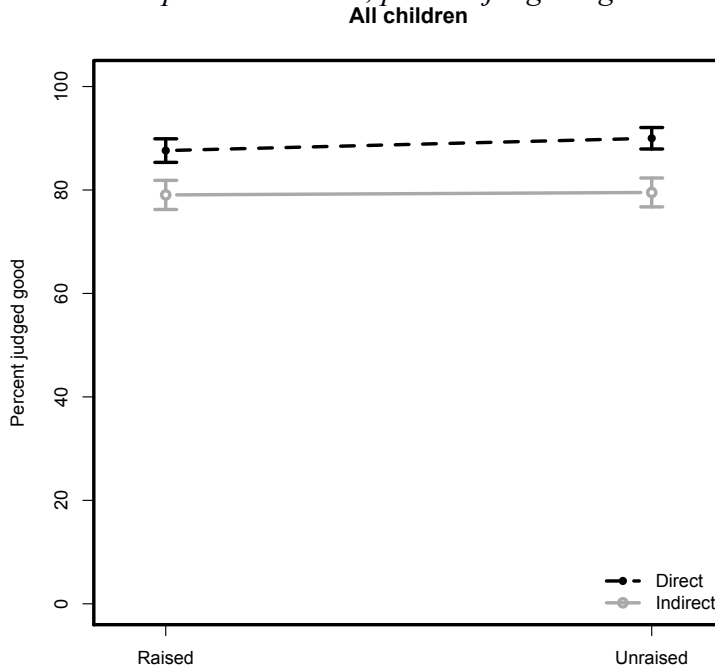
The adult data showed a significant effect for evidence type: overall, sentences were accepted more frequently in direct evidence scenarios than in indirect ones. Unraised sentences were accepted with direct-evidence scenarios 91% of the time and with indirect-evidence scenarios 77% of the time. Crucially, the adult data also showed a significant interaction between raising and evidence type: raised sentences were more likely to be accepted with direct evidence scenarios than indirect ones. (These are the results of a mixed-effects logistic regression,  $p < 0.001$ , Bates et al. 2014.) The figure in (11) shows the crucial adult results: raised sentences were accepted with indirect pictures 23% of the time, whereas they were accepted with direct pictures 93% of the time.

(11) *Adult responses to CRCs, percent judged "good"*



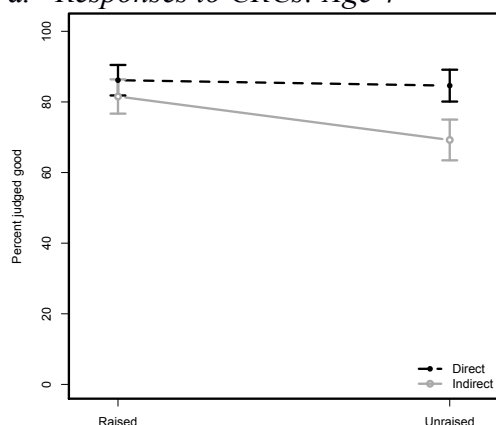
By contrast, the child data showed no significant effect at any age. Children were equally likely to accept raised and unraised sentences. And most importantly, they were equally likely to accept a raised sentence with an indirect picture as with a direct picture. The figure in (12) shows the child data collapsed across all age groups. The figures in (13) show each age group separately. While the percentages do change slightly across the age groups, the differences are not statistically significant.

(12) *All child responses to CRCs, percent judged "good"*

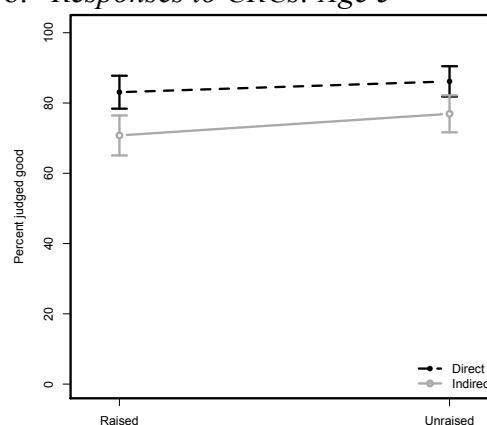


## Children's comprehension of evidentiality

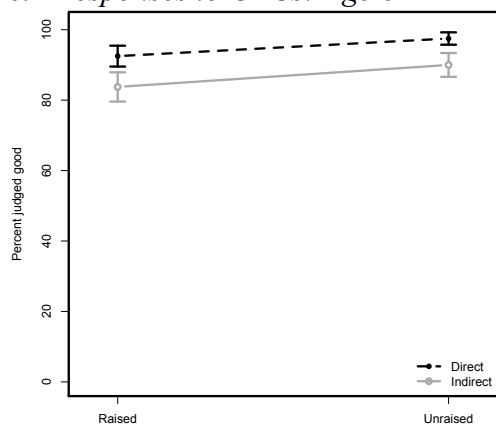
(13) a. Responses to CRCs: Age 4



b. Responses to CRCs: Age 5



c. Responses to CRCs: Age 6



Thus, in contrast to the production results discussed earlier, children seem unable to correlate evidence type (direct or indirect evidence) with the syntax of copy-raising sentences (raised or unraised) in our comprehension task, even at the oldest age tested.

### 4. Comparison with other languages

In Section 2 we showed that children as young as two or three produce CRCs and do so in an adult-like way, producing raised CRCs only when they have direct evidence. These results point to a strikingly early acquisition of the evidential component of CRCs. In contrast, the comprehension study does not find evidence of this knowledge, suggesting that children as old as six have not yet acquired the evidential component of CRCs.

Interestingly, this same asymmetric performance in production vs. comprehension is found in other, typologically distinct languages including Korean (Choi 1995; Papafragou et al. 2007) and Turkish (Aksu-Koç 1988, Aksu-Koç et al. 2009).<sup>4</sup> In these languages, children spontaneously produce evidential markers by age two or three, while their

---

<sup>4</sup>De Villiers et al. (2009) report that young Tibetan-speaking children also fail in comprehension tests of evidentiality, achieving adult performance at about eight to nine years old. They also discuss children's spontaneous production of evidential morphology in a way that suggests production of evidential markers in Tibetan, too, is significantly earlier than comprehension.



performance on comprehension tasks lags behind their production by several years. Thus, the asymmetry holds regardless of the type of evidential strategy a language uses, whether morphological or syntactic, obligatory or optional. Further, English-speaking children show no advance or delay (in either production or comprehension) relative to children acquiring morphologically and obligatorily-marked evidentiality.

It is also important to note that the production/comprehension asymmetry is found regardless of the particular experimental procedures used to test comprehension. For example, Aksu-Koç (1988), Papafragou et al. (2007), and Ozturk & Papafragou (2007, 2008) all used variations on the ‘speaker matching task’. In this kind of task the child hears a story in which two characters have different types of evidence for the at-issue proposition, one direct (visual), the other indirect (hearsay or inference). The child then hears an evidentially-marked sentence—either direct or indirect—and must decide which of the two characters uttered the sentence.

Two other types of tasks that have been used less commonly include the ‘identify the source’ task (de Villiers et al. 2009), in which the child hears an utterance containing an evidential marker and is then asked how the speaker came to know that information, and the ‘speaker reliability task’ (Fitneva 2008, Ozturk & Papafragou 2007), in which two puppets each make a statement and the child is asked who s/he believes more. The idea here is that if the child understands direct vs. indirect evidential marking s/he will find the person/puppet with direct evidence more credible.<sup>5</sup> In all these tasks, as in our felicity judgment task (see also Papafragou et al. 2007), children show the same non-adult performance.<sup>6</sup>

If children have acquired the semantics of evidentiality, as suggested by spontaneous production, what is it about these tasks that would mask this ability? Aksu-Koç and colleagues have attributed the comprehension lag to the difficulty of the tasks themselves: “[They] pose additional demands on the children’s working memory, their role taking ability, and their ability to coordinate temporal and informational perspectives” (Aksu-Koç et al. 2009: 19). The tasks require the child to keep in mind the form of the utterance that is used and evaluate it against the evidence provided. The tasks also require the child to consider the perspective of another individual.

The study presented in Section 3 was designed to minimize some of these concerns. There is only one puppet in our study, and the puppet and the child always share the same information, so as to lessen the burden of considering a different perspective, or multiple perspectives. In the next section, we elaborate on how the evidentials used in these experiments require knowledge beyond what is required to use and understand evidentials in natural conversation.

---

<sup>5</sup>Fitneva (2008) reports that at age six the children “almost reach significance.” As in the Tibetan study (fn. 4), the age at which Bulgarian-speaking children spontaneously produce evidential markers is not reported.

<sup>6</sup>Several studies also include elicited production tasks in which children are asked to describe animated scenarios (e.g. Aksu-Koç 1988, Papafragou et al. 2007, Ozturk & Papafragou 2007, 2008). Overall, children’s performance in these tasks was also delayed relative to their spontaneous production, but not as significantly as in comprehension tasks. For example, Aksu-Koç et al. (2009) report spontaneous production of Turkish direct and indirect evidentials around age 3 and good performance in elicited production tasks around age 4.



## **5. The production/comprehension asymmetry**

Previous work in several languages has found that children's comprehension of evidentials lags significantly behind their production. Our study replicates this effect for optional, syntactically-encoded evidentiality in English. We now turn to the question of why this asymmetry is so consistent across evidential strategies and comprehension tasks.

Clearly, it cannot be the case that the production results and the comprehension results both accurately reflect the child's linguistic knowledge. Thus there are two possible explanations for the asymmetry, outlined in (14). The first takes the comprehension results as the true measure of children's acquisition of evidential meaning, whereas the second takes children's naturalistic production to be the true indicator of their abilities.

- (14) a. *Explanation 1*: Comprehension results reflect actual development; the early production does not have true evidential meaning.
- b. *Explanation 2*: Production results reflect actual linguistic development; the comprehension results are artifacts of the experiments.

There are clear arguments against Explanation 1 in our data. First, in the production study, children as young as two or three produced the (marked) raised sentences and did so exclusively in direct evidence situations. If children at this age did not have the semantics (and syntax) of evidentiality, we would not expect their production to be perfectly aligned with the adult grammar. Second, in the comprehension study, children's justifications demonstrated knowledge of evidentiality. When children who rejected raised sentences in indirect situations were asked *why* the description was "silly", they would respond with adult-like justifications: "Ernie isn't in the picture" or "You can't see Ernie." These kinds of justifications were not provided for other conditions. This suggests that children are rejecting raised sentences in indirect contexts for the same reason adults are; they just are not rejecting them at the same rate as adults.

That leaves us with Explanation 2. But what exactly is causing the poor performance on these comprehension tasks? One possible explanation is that children have difficulty judging the appropriateness of not-at-issue content generally (as opposed to the truth or falsity of at-issue content). A second explanation might be that the comprehension tasks do not appropriately restrict the interpretation of the evidential. Finally, children could be avoiding a (difficult) felicity judgment by ignoring the matrix subject and defaulting to the unmarked form. We will discuss each of these possible explanations in turn.

### **5.1 Difficulty making felicity judgments**

Previous studies on the acquisition of not-at-issue content have shown that the methodologies used have a huge effect on how children perform on these tasks. Noveck (2001) presented evidence that children fail to compute implicatures in experimental tasks until even the oldest ages tested (9 or 10 years old). But subsequent studies provided evidence of earlier comprehension when the children were trained to give felicity judgments (Papafragou & Musolino 2003) or the tasks were modified in some

way to be more natural (Papafragou & Tantalou 2004, Guasti et al. 2005, Foppolo et al. 2012). Likewise, Syrett et al. (2009) and Dudley et al. (2013) have shown that children can demonstrate knowledge of presuppositions when they are not asked explicitly about felicity. Syrett et al. (2009) introduce the Presupposition Assessment Task to test the presuppositions of *the*. In this task, a puppet makes a request (e.g., “Please give me the red rod”). The subject then responds by either giving the red rod if there is one (and only one) or by rejecting the infelicitous request (if there is more than one or no red rod).

It is likely therefore that part of the comprehension lag can be attributed to the fact that evidentials encode not-at-issue content. While a truth-value judgment requires that a subject relate semantic content to a model, a felicity judgment requires that a subject relate semantic content to a model and a context of utterance. But this difference alone does not account for children’s behavior in the present comprehension task. Recall that the fillers in our experiment tested other not-at-issue content. We found that children had much more difficulty with infelicitous evidentials than with other types of infelicity. The relevant data are in (15). (See Appendix for a complete list of filler sentences.)

(15) *Fillers in comprehension study, percent rejected*

	Infelicitous fillers	Infelicitous evidentials
Adults	94%	77%
4 year-olds	63%	18%
5 year-olds	65%	29%
6 year-olds	80%	16%

The results in (15) are noteworthy in several respects. First, both the children (at all ages) and the adult controls reject infelicitous evidentials at a lower rate than the infelicitous fillers. This suggests that both children and adults have a harder time judging the infelicity of evidentials as compared to other types of infelicity. Second, by age six, children show improved performance in judging infelicitous fillers, but they do not improve in judging infelicitous evidentials. Third, children in all age groups reject the infelicitous evidentials substantially less than the adults do. These findings suggest some additional difficulty in judging the felicity of evidentials, which persists even after children improve on other types of infelicity. We therefore conclude that the comprehension lag in the acquisition of evidentials cannot be fully explained by their not-at-issue status. A final point worth noting is that the children are willing to reject the infelicitous fillers, so the pattern of responses to the evidentials does not simply reflect a ‘yes’ bias in our subjects. The next subsection looks more closely at why the comprehension of evidentials may be particularly hard to test in experimental tasks.

## 5.2 Failure to question the speaker or restrict the evidential interpretation

There are a couple of ways that the tasks used to test the comprehension of evidentiality are different from the way evidentials are used in natural discourse. First, while presuppositions are generally associated with discourse-old information, evidentials arguably encode discourse-new information (Murray 2010). This might make evidential content susceptible to the same sincerity conditions as asserted at-issue content; in

particular, it might mean that the utterer of a sentence with a speaker-oriented evidential marker is generally taken to be an authority on their own evidence. In the context of the felicity judgment task described in Section 3, this would predict that the child is more likely to defer to the puppet as an authority on his own evidence.

Second, the experimental tasks that have been used to test evidentiality all rely on the subject interpreting the evidential relative to the immediate context only (i.e., with respect to all and only the information depicted in the picture). However, evidential information is generally not discourse-bound; evidentials do not require evidence to be contextually salient (Murray 2014). The comprehension tasks described here rely on the assumption that the child will interpret the evidence provided in the experimental context as exhaustive. But there is in principle nothing stopping a charitable child from allowing that the puppet could have additional evidence: the child is told what evidence the puppet has, but not what evidence he does not have.

Adults are arguably able to overcome these artificial aspects of the experiment: they can in principle question the speaker's source of evidence and restrict the speaker's evidence to the immediate context. Children, in contrast, may not be willing or able to adjust their behavior to these artificial aspects of the experimental setting.

### **5.3 Non-interpretation of evidence source**

A final possibility is that children avoid making a felicity judgment by ignoring the matrix subject. As noted earlier, felicity judgments are more complex than truth-value judgments. Faced with this complexity, children might assign to both the raised (16a) and unraised (16b) a representation like (16c), in which the matrix subject is omitted. (16c) is unspecified with respect to evidentiality: it is acceptable with both direct and indirect evidence; as long as there is some visual evidence that supports the at-issue proposition (a requirement of the perception verb *look*), the utterance is true and felicitous.

- |      |                                  |                    |
|------|----------------------------------|--------------------|
| (16) | a. Ernie looks like he got sick. | <i>raised</i>      |
|      | b. It looks like he got sick.    | <i>unraised</i>    |
|      | c. Looks like he got sick.       | <i>pro-dropped</i> |

The sentence in (16c) is fully grammatical as an instance of 'expletive drop' or 'diary drop' (Haegemann 1990) and indeed, we found many such examples from the children in our CHILDES production study (Section 2), as well as in the adult input to these children in the files. (For discussion of these sentences, see Rett & Hyams 2013.) It is possible to interpret the existence of these constructions as evidence that children are ignoring the evidential component of the stimuli in our evidential comprehension task.<sup>7</sup>

---

<sup>7</sup>The 'diary drop' account would obviously not explain the comprehension lag in languages with morphologically-encoded evidentiality. However, it is possible that 'diary drop' is an instance of a more general tendency for children to "ignore" the evidential component in these tasks. Evidentiality is often encoded in a polysemous morpheme (see Rett & Hyams 2013 for discussion); in such languages, children might assign the morpheme only its aspectual or temporal meaning and "ignore" the evidential component as a way to avoid the felicity issue. This is consistent with claims that Turkish children, for example, master the temporal function of the direct evidential morpheme before its evidential function (Papafragou et al. 2007, Ozturk & Papafragou 2007, Aksu-Koç et al. 2009).

The various explanations of the evidential comprehension lag just outlined differ in their theoretical underpinnings and assumptions, but they all predict a similar pattern of responses in the comprehension experiment. If children fail to question the speaker's evidence or fail to restrict the evidential to the immediate context, they will judge the evidential to be felicitous. If they ignore the matrix subject, the sentence is interpreted as unmarked with respect to evidentiality. In either case, we predict that children will overaccept evidential stimuli relative to adults' performance, either because they are interpreting the evidential requirement as less restrictive than adults do, or because they are failing to represent the evidential component at all. These explanations both predict that children will judge the evidential sentences to be relatively felicitous across evidence conditions, which is in fact what we see, (12)-(13).

## 6 Conclusion

The naturalist production study of Rett et al. (2013) and Rett & Hyams (2013) shows that children acquiring syntactically-encoded evidentiality show no production delay relative to children acquiring morphologically-encoded evidentiality, and are in fact adult-like in their production of evidentials relatively early (at age two or three). But there seems to be a conflict between naturalistic production and comprehension, as tested experimentally, across languages and evidential strategies: While children produce evidentials very early, their performance on the comprehension task lags behind significantly. We have argued that this delay is most likely an effect of the experimental methods used to test comprehension, rather than a delay the acquisition of evidentiality. Children seem to find it relatively difficult to provide adult-like felicity judgments in experimental settings, but their performance on felicity tasks involving evidentiality is hindered above and beyond their ability to evaluate other types of not-at-issue content.

Our conclusions here suggest that, in future work, evidential comprehension tasks should avoid confounds about speaker authority and avoid relying on children's interpretation of evidentials as discourse-bound. One way to do this would be to rely on the child's evidence, rather than the experimenter's or puppet's. This could be achieved, for example, using speech acts in which evidentials become addressee-oriented (rather than speaker-oriented), e.g. in imperatives like "Point to the picture where Ernie looks like he's sick." The use of imperative sentences should also be integrated with an experimental design that tests knowledge of the evidential implicitly, rather than explicitly, adopting something like Syrett et al's (2009) Presupposition Assessment Task.

### Appendix: Items in the CRC comprehension task

#### (17) *Test items*

- a. Ernie/It looks like he/Ernie got sick.
- b. Ernie/It looks like he/Ernie rode his bike to school.
- c. Ernie/It looks like he/Ernie painted a picture.
- d. Ernie/It looks like he/Ernie played in the mud.
- e. Ernie/It looks like he/Ernie won the race.
- f. Ernie/It looks like he/Ernie went to Disneyland.
- g. Ernie/It looks like he/Ernie got a haircut.

### *Children's comprehension of evidentiality*

- h. Ernie/It looks like he/Ernie brought his train to school.
- i. Ernie/It looks like he/Ernie made breakfast.
- j. Ernie/It looks like he/Ernie built a tower.

#### (18) *Fillers*

- a. Picture: Ernie and Bert are playing basketball.  
Sentence: Ernie is playing basketball.
- b. Picture: Ernie is standing next to his dog. The dog has multiple black spots.  
Sentence: Ernie's dog has one black spot.
- c. Picture: Ernie is reading to Big Bird, Bert, and another student.  
Sentence: Big Bird is listening to the story.
- d. Picture: Ernie is reading to Big Bird and Bert. Big Bird is (as usual) shirtless.  
Sentence: Big Bird's shirt is red.
- e. Picture: There are two books on Ernie's bed.  
Sentence: There's a book on Ernie's bed.
- f. Picture: Ernie's dog is on his front lawn.  
Sentence: Ernie's cat is fluffy.
- g. Picture: Ernie's train is in pieces on the floor. Ernie is crying.  
Sentence: Ernie's train is broken.

### **References**

- Aksu-Koç, A. 1988. *The acquisition of aspect and modality*. Cambridge, Mass.: Cambridge University Press.
- Aksu-Koç, A., and D. Slobin. 1986. In *Evidentiality: the linguistic coding of epistemology*, ed. W. Chafe and J. Nichols, 159–167. Norwood, NJ: Ablex.
- Aksu-Koç, A., H. Ögel Balaban, & I.E. Alp. 2009. Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development* 125:13–28.
- Asudeh, A. & I. Toivonen. 2012. Copy raising and perception. *Natural Language and Linguistic Theory* 30:2, 321–380.
- Choi, S. 1995. The development of epistemic sentence-ending modal forms and functions in Korean children. In *Modality in Grammar and Discourse*, eds. J. Bybe and S. Fleischman, 165–204. Amsterdam: Benjamins.
- Courtney. 2008. *Child production of Quechua evidential morphemes in conversations and story retellings*. Submitted ms.
- Dudley, R., N. Orita, V. Hacquard, & J. Lidz. 2013. Three-year-olds' understanding of know and think. In *Experimental perspectives on presupposition*, ed. F. Schwarz. Springer.
- Faller, M. 2006. Evidentiality below and above speech acts. In *Functions of Language*, eds. L. Egberg and C. Paradis. John Benjamins.
- Faller, M. 2002. Semantics and pragmatics of evidentials in Cuzco Quechua. Doctoral dissertation, Stanford University.
- Fitneva, S. 2008. The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language* 35:845–868.

- Foppolo, F., M-T. Guasti, & G. Chierchia. 2012. Scalar implicatures in child language: Give children a chance. *Language learning and development* 8:365–394.
- Gleitman, L. & A. Papafragou. 2005. Language and thought. In *The Cambridge Handbook of Thinking and Reasoning*, eds. K. Holyoak & R.G. Morrison, 633–661. Cambridge, Mass.: Cambridge University Press.
- Guasti, MT., G. Chierchia, F. Foppolo, A. Gualmini, & L. Meroni. 2007. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20:667–696.
- Haegeman, L. (1990) Non-overt subjects in diary contexts. In *Grammar in Progress, GLOW essays for Henk van Riemsdijk*, eds. J. Mascaró & M. Nespó. Dordrecht: Foris. 167–174.
- Koring, L., & H. de Mulder. 2014. Understanding different sources of information: conceptual and linguistic development. *Journal of Child Language*
- MacWhinney, B., & C. Snow. 1985. The child language data exchange system. *Journal of Child Language* 12:271–295.
- Murray, S. 2010. Evidentiality and the structure of speech acts. Doctoral dissertation, Rutgers University.
- Murray, S. 2014. Varieties of update. *Semantics and Pragmatics* 7:1–53.
- Noveck, I. 2001. When children are more logical than adults: Experimental investigations of scalar implicatures. *Cognition* 78:165–188.
- Ozturk, O. & A. Papafragou, A., 2007. Children's acquisition of evidentiality. In *Proceedings from the 31st Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- Ozturk, O. & A. Papafragou. 2008. The acquisition of evidentiality and source monitoring. In *Proceedings from the 35th Annual Boston University Conference on Language Development*, 368–377. Somerville, MA: Cascadilla Press.
- Papafragou, A., P. Li, Y. Choi, & C. Han. 2007. Evidentiality in language and cognition. *Cognition* 103:253–299.
- Papafragou, A. & J. Musolino. 2003. Scalar implicatures at the semantic-pragmatic interface. *Cognition* 80:253–282.
- Papafragou, A., & N. Tantalou. 2004. Children's computation of implicatures. *Language Acquisition*, 12.1:71-82.
- Rett, J., N. Hyams, & L. Winans. 2013. The effects of syntax on the acquisition of evidentiality. In *Proceedings from the 37th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- Rett, J. & N. Hyams. 2013. The acquisition of syntactically encoded evidentiality. *Language Acquisition: A Journal of Developmental Linguistics* 21.2:173–198.
- Syrett, K., C. Kennedy, & and J. Lidz. 2009. Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics* 27:1–35.
- de Villiers, J., J. Garfield, H. Gernet-Girard, T. Roeper, & M. Speas. 2009. Evidentials in Tibetan: Acquisition, semantics, and cognitive development. *New Directions for Child and Adolescent Development* 125:29–47.