Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Cognition 103 (2007) 253–299
COGNITION
www.elsevier.com/locate/COGNIT

Evidentiality in language and cognition ☆,☆☆

Anna Papafragou [a,*], Peggy Li [b], Youngon Choi [c], Chung-hye Han [d]

[a] *Department of Psychology, University of Delaware, 109 Wolf Hall, Newark, DE 19716, USA*
[b] *Department of Psychology, Harvard University, William James Hall, 33 Kirkland St., Cambridge, MA 02138, USA*
[c] *Department of Psychology, University of Pennsylvania, 3401 Walnut St., Suite 400, Philadelphia, PA 19104, USA*
[d] *Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, BC Canada, V5A 1S6*

Received 21 June 2005; revised 31 March 2006; accepted 1 April 2006

Abstract

What is the relation between language and thought? Specifically, how do linguistic and conceptual representations make contact during language learning? This paper addresses these questions by investigating the acquisition of evidentiality (the linguistic encoding of information source) and its relation to children's evidential reasoning. Previous studies have hypothesized that the acquisition of evidentiality is complicated by the subtleness and abstractness of the underlying concepts; other studies have suggested that learning a language which systematically (e.g. grammatically) marks evidential categories might serve as a pacesetter for early reasoning about sources of information. We conducted experimental

☆ This manuscript was accepted under the editorship of Jacques Mehler.

☆☆ We wish to thank Lila Gleitman, Henry Gleitman, Eleni Miltsakaki, Jesse Snedeker and the members of the CHEESE seminar at the University of Pennsylvania for discussion of this project. Thanks also to Seung-yun Yang for experimental help. This research was partly supported by NIH/NRSA Grants #F32MH065020 to Anna Papafragou and #1F32HD043532 to Peggy Li and by SSHRC Grant #410-2003-0544 to Chung-hye Han.

* Corresponding author.
E-mail addresses: papafragou@psych.udel.edu (A. Papafragou), pegs@wjh.harvard.edu (P. Li), youngonc@sas.upenn.edu (Y. Choi), chunghye@sfu.ca (C. Han).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.04.001

studies with children learning Korean (a language with evidential morphology) and English (a language without grammaticalized evidentiality) in order to test these hypotheses. Our experiments compared 3- and 4-year-old Korean children's knowledge of the semantics and discourse functions of evidential morphemes to their (non-linguistic) ability to recognize and report different types of evidential sources. They also compared Korean children's source monitoring abilities to the source monitoring abilities of English-speaking children of the same age. We found that Korean-speaking children have considerable success in producing evidential morphology but their comprehension of such morphology is very fragile. Nevertheless, young Korean speakers are able to reason successfully about sources of information in non-linguistic tasks; furthermore, their performance in these tasks is similar to that of English-speaking peers. These results support the conclusion that the acquisition of evidential expressions poses considerable problems for learners; however, these problems are not (necessarily) conceptual in nature. Our data also suggest that, contrary to relativistic expectations, children's ability to reason about sources of information proceeds along similar lines in diverse language-learning populations and is not tied to the acquisition of the linguistic markers of evidentiality in the exposure language. We discuss implications of our findings for the relationship between linguistic and conceptual representations during development.

Keywords: Evidentiality; Language acquisition; Language and thought; Theory of mind; Semantics; Pragmatics

1. Introduction

What is the relation between language and thought? More specifically, how do linguistic and conceptual representations make contact during language learning? Most commentators accept the view that (at least to some extent) language acquisition builds on antecedently available concepts – hence (part of) the learner's task is to map novel words in the input onto conceptual representations already in the mind. However, there is little agreement about how much of language acquisition is constrained by pre-linguistic concepts and in what ways.

According to one widely held view, the relationship between linguistic and non-linguistic categories in development is rather transparent: not only does language acquisition depend on conceptual development but it also reflects it to a rather precise degree. In other words, the rate of emergence of various linguistic expressions in child language more or less directly indexes the degree of their conceptual complexity (see, e.g., Huttenlocher, Smiley, & Charney, 1983). A different position suggests that language itself has the power to shape non-linguistic categories. This perspective, famously associated with the writings of Benjamin Whorf (Whorf, 1956), holds that the systematic encoding of certain conceptual distinctions in grammar may encourage (or force) speakers of the language to use these distinctions consistently in their non-linguistic thinking. Several commentators have recently revived this perspective, arguing that language structure may provide

the basis for an individual's "default conceptual representation" (Pederson et al., 1998, p. 586). From a learning standpoint, this view entails that language-specific encoding patterns can affect the salience (or even, the availability) of certain conceptual distinctions in the learner's mind: children learning different languages may develop different concepts at different timetables depending on properties of the exposure language (Bowerman & Levinson, 2001).

These two positions agree that language and thought are tightly and causally connected but seem to take different perspectives on the direction of causality (even though it is sometimes suggested that both positions can be true of different aspects of development; Bowerman & Choi, 2003; Gentner & Boroditksy, 2001). Currently much experimental work seeks to evaluate these two differing positions. Most of this work focuses on the relationship between language and readily testable, perceptually grounded cognitive areas such as object individuation, space and motion (see the papers in Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003).

Here we want to contribute to these experimental efforts by turning to a more abstract domain, the ability to monitor the origins of one's beliefs (*source monitoring*). Humans are typically able to reason about the sort of evidence that led them to believe something; in other words, we know whether we directly saw an event happen, whether someone told us, or whether we inferred that the event took place on the basis of available evidence. Knowing what type of event led to a belief plays an important role in belief evaluation and belief change or update: for instance, one is less likely to believe in rumors than in one's own eyes. Source monitoring builds on the understanding that people stand in different and variable informational relations to the world – hence their beliefs may vary and be modified or updated as new evidence becomes available. This understanding is part of the adult *theory of mind*, the ability to attribute to oneself and others mental states and to reason in terms of mental states in order to explain and predict behavior.

Source distinctions are encoded in language through a variety of *evidentiality* markers. In English, such evidential devices are mostly lexical. For instance, in (1a) and (1b) the speaker conveys that she had direct perceptual access to the event of John's singing, while in (1c) and (1d) the evidence is indirect (hearsay in (1c) or some unspecified source in (1d)):

(1) a. I **saw** John sing.
b. I **heard** John sing.
c. John was **allegedly** singing.
d. John was **apparently** singing.

Other languages grammaticalize evidentiality through specialized and often obligatory verbal affixes, particles or other devices, as shown in the following examples from Colombian Tuyuka (Barnes, 1984) and Peruvian Quechua (Weber, 1986), respectively:

(2) a. díiga apé-**wi** 'He played soccer (I saw him)'
b. díiga apé-**ti** 'He played soccer (I heard the game and him but didn't see it or him)'
c. díiga apé-**yi** 'He played soccer (I have seen evidence that he played but did not see him play)'
d. díiga apé-**yigi** 'He played soccer (I obtained the information from someone else)'
e. díiga apé-**hĩyi** 'He played soccer (It is reasonable to assume that he did)'
(3) a. wañu-nqa-paq-**mi** 'It will die (I assert)'
b. wañu-nqa-paq-**shi** 'It will die (I was told)'
c. wañu-nqa-paq-**chi** 'It will die (perhaps)'

Both the semantic content and the internal organization of linguistic evidentiality make contact with fundamental aspects of the human ability to reason about the origins, reliability and strength of our beliefs. Two major features of evidential systems are particularly relevant in this respect. First, despite the considerable variability of evidential systems cross-linguistically (Aikhenvald & Dixon, 2001; Anderson, 1986; Chafe & Nichols, 1986; Cinque, 1999; Delancey, 2002; Faller, 2002; Garrett, 2000; Givón, 1982; De Haan, 1998, 2001; Ifantidou, 2001; Izvorski, 1998; Johanson & Utas, 2000; Kratzer, 1991; Mayer, 1990; Mushin, 2001; Palmer, 1986; Papafragou, 2000; Speas, 2004; Willett, 1988), the semantics of evidential morphology seems to draw in systematic ways from a relatively restricted range of basic evidential concepts. According to Willett (1988), who surveyed data from 32 languages, there are three main types of source of information that are encoded grammatically: direct access (in particular, perception), reports from others, and reasoning (where the last two fall under indirect access).

(4) **Basic categories of evidentiality**
A. Direct access/perception
B. Indirect access
b1. Report from others
b2. Reasoning

When additional distinctions are found, these seem to arise from subdivisions of the three major notional categories (or from the interaction of these distinctions with other grammatical features such as tense and aspect). For instance, direct access may be subdivided into visual, auditory and other types of sensory perception; reported information can be secondhand, thirdhand or general hearsay; and reasoning can be based on concrete evidence or mere conjectures.[1] By contrast, several other conceivable and salient sources of information never surface in evidential morphemes (e.g.

[1] As the earlier examples in (2) and (3) show, Quechua follows the basic three-way distinction in grammaticizable evidentiality, while Tuyuka makes use of a more elaborate system of five distinctions.

divine revelation, legal edict, parental advice, heartfelt intuition or 'gut feeling', learned through trial and error). This points to a highly constrained grammaticalization system cross-linguistically (Speas, 2004).

Second, across languages, evidentials form a scale defined by the reliability of the relevant informational sources (which is itself determined on non-linguistic grounds). In its most simplified and general form, this scale ranks direct access (e.g. visual perception) higher than indirect access (e.g. hearsay or inference).[2] This is because perceptually grounded beliefs, although not necessarily more likely to be true, are normally assumed to be causally related to the structure of reality; they are thus considered to be our securest form of contact with the world around us (Dancy, 1985, p. 178). By contrast, an inference, although valid, may prove to have been based on incomplete or unreliable premises and may need to be revisited; similarly, the reliability of hearsay depends on the trustworthiness of the reporting source:

(5) **Evidentiality scale**
Direct access ≫ Indirect access

The evidentiality scale can give rise to pragmatic effects: assuming that the speaker is trying to be adequately informative, the use of an evidential encoding a concept lower in the scale typically gives rise to the inference that the speaker was not in a position to offer a higher ranked term (Horn, 1972; cf. Urmson, 1963). For instance, in English, a speaker who utters "I hear that it's raining" indicates that she has had no direct visual access to the event, i.e. she hasn't seen that it's raining. Similar interpretations arise cross-linguistically (cf. for instance Faller, 2001, p. 52 on Quechua).

Evidentiality offers a good testing ground for investigating the relationship between language and our 'conceptual/intentional systems' (Hauser, Chomsky, & Fitch, 2002). For reasons just explained, from a learning perspective, evidential meanings require grasp of abstract and unobservable source concepts (cf. (4)) and subtle reasoning about the reliability of different sources of information (cf. (5)). Furthermore, evidentiality is a novel arena for investigating potential language-on-thought effects – perhaps an especially promising one: according to some commentators, linguistic effects on cognition are more likely to be found in domains removed from perception, involving higher-level cognitive representations where human cognition appears to differ from other species (Spelke & Tsivkin, 2001). Here we take up both of these themes in a series of cross-linguistic experiments exploring the relation between grammatical evidentiality and non-linguistic source monitoring in children. To introduce the specific hypotheses driving the experimental part of our study, in

[2] There is disagreement about how the two types of indirect access should be ranked relative to each other (De Haan, 1998; Oswalt, 1986; Willett, 1988). We cannot see any a priori reason for such a ranking: an expert's inference about the origins of a wine may override the information provided by its seller, but the inference of a novice wine taster will not. A more productive approach may be to allow inference and hearsay to form separate internal hierarchies depending on the reliability, completeness, etc., of the premises for the inference and the trustworthiness, recency, etc., of hearsay (Faller, 2001).

the remainder of Section 1 we briefly review prior work on children's understanding of the sources of their beliefs (Section 1.1) and the acquisition of linguistic evidentiality (Section 1.2) before sketching our experimental prospectus (Section 1.3).

1.1. The development of source monitoring

Studies looking at children's ability to explicitly identify the evidence for their beliefs generally conclude that young children are unable to encode source information at the time of experiencing an event; furthermore, the problem is more specific than simple memory limitations (Gopnik & Graf, 1988; O'Neill & Chong, 2001; O'Neill & Gopnik, 1991; Pillow, 1989; Povinelli & de Blois, 1992; Wimmer, Hogrefe, & Perner, 1988; Wimmer, Hogrefe, & Sodian, 1988; Woolley & Bruell, 1996). Typically, in these tasks, children discover the contents of a container through a single type of source (e.g. they are allowed to see it, they are being told by the experimenter, etc.) and are then asked how they found out. Overall, 3-year-olds are much poorer in verbally reporting the source of their beliefs than 4- or 5-year-olds.

Similar findings emerge from research looking at children's ability to attribute knowledge to other agents based on the agents' access to information. Three-year-olds do not realize that a person who did not hear a particular statement is ignorant compared to someone who did (Mossler, Marvin, & Greenberg, 1976). In other tasks, 3-year-olds do not select the character who had visual access to an object hidden inside a box as the one who knows what is hidden inside the box over another character who simply lifted or pushed the box (Povinelli & de Blois, 1992; Wimmer et al., 1988).[3] Full understanding of inference as a source of information appears only at the age of six (Sodian & Wimmer, 1987; Wimmer & Hogrefe et al., 1988) and more subtle distinctions among inference types come much later (Pillow, 2002). Other work examining children's ability to recognize that certain kinds of knowledge can only be gained by specific information channels – e.g. texture from touching, color from seeing, etc. – finds that 3-year-olds, and quite a few 4-year-olds, have great difficulty linking specific kinds of knowledge with the appropriate sensory modality (O'Neill & Astington, 1990; O'Neill, Astington, & Flavell, 1992; Pillow, 1993; Robinson, Thomas, Parton, & Nye, 1997). Three-year-olds may also overestimate the knowledge to be gained by a sensory experience (e.g. seeing: Robinson et al., 1997; cf. Taylor, 1988).[4]

[3] An exception is a study by Pratt and Bryant (1990), who find that 3-year-olds are very successful at a version of this task. However, their version includes extensive training and their group of 3-year-olds is older than those of similar studies (for discussion, see Povinelli & de Blois, 1992).

[4] To be sure, young children encode the origins of mental representations to some extent. Three-year-olds perform better with some sources (e.g. seeing) than with others (e.g. being told). In fact, when asked to report whether their beliefs were due to either seeing or telling, 3-year-olds' performance is well above chance (O'Neill & Gopnik, 1991; Whitcombe & Robinson, 2000). Children also engage in source monitoring if they need to contrast and evaluate conflicting sources of information. For instance, they rightly trust their own visual perception more than conflicting verbal reports from others (Mitchell, Robinson, Nye, & Isaacs, 1996). Furthermore, 3- and 4-year-olds are more likely to believe what they are told by an adult who has had visual evidence over an adult who has not (Robinson, Champion, & Mitchell, 1998).

Given these early difficulties with source reasoning, it is of interest to investigate more closely the relation between the development of source monitoring and the acquisition of linguistic evidentiality. An intriguing possibility is that learners of languages with systematic (e.g. grammatical) markings of evidential distinctions may find such distinctions to be more salient than learners of languages where evidential distinctions are not encoded in the grammar. So far, however, the acquisition of evidentiality cross-linguistically has been the topic of only a few studies (Aksu-Koç, 1988; Choi, 1995), which have been conducted independently of work on children's source monitoring abilities. We turn to these studies next.

1.2. The acquisition of evidentiality

Most of the available evidence on the acquisition of grammaticalized evidentiality comes from Aksu-Koç's pioneering work on Turkish (Aksu-Koç, 1988, 2000; Aksu-Koç & Slobin, 1986). Turkish obligatorily marks all past tense events with one of two suffixes: ***-mI*** (indirect evidence: inference/hearsay) or ***-dI*** (direct evidence):

(6) a. Ahmet gel **-mi** 'Ahmet came (I heard/I guessed)'
b. Ahmet gel **-di** 'Ahmet came (I saw him)'

In one of her experiments, Aksu-Koç (1988) showed children from three years up acted out stories in which a target event (e.g. the popping of a balloon) was explicitly shown or had to be inferred from the perceived outcome (e.g. the popped balloon). When asked to relate the story, children appeared to prefer ***-dI*** for directly perceived events and ***-mI*** for inferred events consistently only after the age of four, even though they began using them already from age two. To test whether the use of evidentials was accompanied by genuine understanding, Aksu-Koç asked children to judge whether a doll who had reported an event using ***-mi/-dI*** had seen the event or had been told about it. She found that learners of Turkish were not able to use evidential morphology to make consistently correct inferences about source knowledge even by the age of six (even though they were better with the 'direct experience' marker than with the 'indirect experience' one). In more recent work, Aksu-Koç and Alici (2000) found that the relative certainty communicated by Turkish evidentials is not appreciated even by 6-year-olds (see also Lee & Law, 2000, on the late comprehension of epistemic particles in Cantonese).

Taken together with the cognitive developmental data reviewed in the previous section, these results raise two issues. First, to what extent does the emergence of evidentials during language learning depend on the nature of their cognitive prerequisites? In studies of early lexical development, it is widely held that much, perhaps most, of the difficulty in learning mentalistic/abstract vocabulary comes from mastering the relevant concepts. As other commentators have remarked, the acquisition of words which refer to mental states (e.g. *think* or *know*) crucially depends on the ability to represent and reason about mental contents in adult-like fashion. The fact that such words are rare or altogether absent in the speech of very young children

(unlike action words such as *run* or *throw*) has been taken to directly reflect their underlying conceptual complexity (Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Smiley & Huttenlocher, 1995). In a similar spirit, Aksu-Koç (1988) concludes that the emergence of evidentiality is delayed partly because of its abstract and complex conceptual presuppositions:

> Children's early lack of sensitivity to the distinction between direct and indirect experience suggests that they are more attentive to concrete, referential and objective characteristics of situations than to subjectively relevant distinctions such as the speaker's attitude to the proposition asserted. (p.195)

The hypothesis that insensitivity to the class of evidential concepts creates difficulties for the acquisition of evidential morphology is certainly plausible. However, in order to evaluate it properly, one should be able to show that evidential concepts present problems for young learners using independent, non-linguistic tasks (of the sort reviewed in the previous section). In the absence of such comparisons, the relative complexity of evidential concepts can be evaluated only indirectly.

Similarly, within the evidential class, the order of appearance of individual morphemes has been taken to reflect at least in part the child's developing abilities to handle various information sources. For instance, the fact that, in Turkish child speech, the indirect evidence morpheme appears later than the direct evidence morpheme has been attributed to "the further complexity of making an inference ... as compared to simply accessing an experienced event from memory" (Aksu-Koç & Slobin, 1986, p. 166). A related explanation is suggested for the fact that the hearsay meaning appears later than the inferential meaning of the indirect evidence morpheme. Again, in order to properly evaluate these hypotheses, one would need an independent assessment of the relative complexity of informational sources through non-linguistic tasks on source monitoring.[5]

A second question arising from the present findings is whether the linguistic marking of evidentiality could affect non-linguistic source reasoning in speakers of different languages. One might hypothesize that the systematic (e.g. grammaticalized) marking of evidential distinctions in languages such as Turkish could make such distinctions more salient in the mental life of their speakers. This possibility was suggested in a different context by Whorf himself, who pointed out that Hopi – unlike English – marks evidential distinctions grammatically and concluded that this grammatical feature was bound to make certain conceptual distinctions easier to draw for the Hopi speaker because of the force of habitual linguistic practices (Whorf, 1956):

[5] It is interesting to note, in this respect, that commentators differ in their evaluations of the relative conceptual difficulty of informational sources. For instance, Asku-Koç and Slobin suggest that hearsay might be more complex than inference; other authors have explicitly argued that inference is plausibly the most challenging aspect of the systems monitoring informational access (see Section 1.1). Perhaps these authors have in mind different kinds of inference (e.g. logical vs. physical/circumstantial) but these issues have not been systematically explored so far.

> Why, for instance, do we not, like the Hopi, use a different way of expressing the relation of channel of sensation (seeing) to result in consciousness, as between 'I see that it is red' and 'I see that it is new'? We fuse two quite different types of relationship into a vague sort of connection expressed by 'that', whereas the Hopi indicates that in the first case seeing presents a sensation 'red', and in the second that seeing presents unspecified evidence for which is drawn the inference of newness ... We even have to think, and boggle over the question for some time, or have it explained to us, before we can see the difference in the relationships expressed by 'that' in the above examples, whereas the Hopi discriminates these relationships with effortless ease, for the forms of his speech have accustomed him to doing so. (p.85)

From a developmental perspective, this view suggests that children who learn languages with grammaticalized evidential systems might be more advanced in their source reasoning than learners of languages without such systematic contrasts (cf. Section 1.1). In its most radical version, this relativistic view entails that linguistic evidentials themselves may serve as a source of information for the acquisition of evidential concepts (cf. Bowerman & Levinson, 2001). In her discussion of the linguistic and conceptual development of evidentiality, Aksu-Koç leaves this possibility open and concludes that "it is necessary to make comparative studies between languages with and without evidentiality contrasts" – presumably including independent non-linguistic tasks (1988, p.203).

1.3. Experimental prospectus: Evidentiality and the language/cognition interface

In the experiments described below, we investigate linguistic evidentiality and non-linguistic source monitoring in very young children. We also compare the source monitoring abilities of learners exposed to languages with different evidential systems. Our goal is twofold: we ask, first, whether the acquisition of linguistic evidentiality is complicated by the subtleness and abstractness of the underlying concepts; we also ask whether learning a language which systematically (e.g. grammatically) marks evidential contrasts might serve as a pacesetter for early reasoning about sources of information. Our approach brings together two strands of research which have until now been pursued separately by researchers working on language acquisition or cognitive development.

Our experimental efforts focus on evidential morphology in Korean. Korean encodes evidentiality grammatically as an inflectional morpheme on the main verb of the sentence. Evidential morphemes form a subclass of 'sentence-ending' (SE) morphemes that express for the most part which clause type (e.g. declarative, interrogative, etc.) the sentence belongs to. These SE morphemes obligatorily occur at the end of the verb, following a tense morpheme. For instance, a declarative sentence ending with ***-e*** indicates that the speaker has direct evidence for the statement, whereas a declarative sentence ending with ***-tay*** indicates that the grounds for the speaker making the statement is hearsay:

(7) Toli-ka mantwu-lul mek-ess-**e**.
Toli-Nom dumpling-Acc eat-Past-Decl
'Toli ate dumplings.'

(8) Toli-ka mantwu-lul mek-ess-**tay**.
Toli-Nom dumpling-Acc eat-Past-Decl
'(I heard that) Toli ate dumplings.'

Our starting point is a study by Choi (1995) who conducted a longitudinal study of the production of evidential morphemes in the speech of three Korean-speaking children. Among her three subjects, ***-e*** and ***-tay*** had both appeared in production by the age of two, and by three, her subjects were using these and other evidential morphemes productively. The early acquisition of evidential morphemes by Korean-speaking children is quite surprising and Choi provides several possible explanations for it (e.g. SE suffixes are obligatorily produced in adult speech; they appear in a salient environment; they encode some form of modal meaning, rather than encoding a combination of modal, tense or aspectual meanings at the same time, etc.). Especially for ***-e***, it is noted that it is a very frequent morpheme which is almost the default declarative marker.

If Choi's observations are accurate, children learning Korean acquire evidential morphology much earlier than the age at which English-speaking children pass non-linguistic evidential tasks and perhaps earlier than the age at which evidential morphology is understood in a language such as Turkish. The Korean data thus raise two questions of interest. First, one would want to know whether very young Korean children have assigned the correct (adult) semantics to evidential morphemes (since this cannot be established solely on the basis of observational data). Second, one would want to know whether the presence of grammaticalized evidentiality could encourage Korean children to use the relevant conceptual distinctions at an earlier age and with greater reliability than their English-speaking peers (whose language does not grammaticalize evidential distinctions). A potential cognitive advantage for the Korean-speaking population seems more likely to arise before age four since, according to the evidence reviewed earlier, English-speaking three-year-olds still have difficulties reasoning about and reporting on belief sources.

These questions form the core of our experimentation. We concentrate on the linguistic distinction between the Korean morphemes ***-e*** and ***-tay*** (or direct evidence vs. hearsay) and its non-linguistic counterpart the distinction between visual perception and verbal report (seeing vs. telling). Specifically, in our first study we test Korean-speaking 3- and 4-year-olds' comprehension of the semantics and pragmatics of evidential morphemes and relate their linguistic scores to the very same children's performance in non-linguistic tasks of source reasoning (Experiment 1). We then probe further into both the comprehension (Experiment 2) and production (Experiment 3) of evidential morphology by Korean learners and compare the results to non-linguis-

tic source monitoring. Finally, we test non-linguistic source reasoning in English-speaking children of the same age and compare the data to those from Korean learners (Experiment 4).

2. Experiment 1

In this first experiment, we investigate Korean children's understanding of the semantic and pragmatic properties of evidential morphology and compare it to the source monitoring abilities of the same children in non-linguistic reasoning tasks.[6] Inspired by one of Aksu-Koç's (1988) tasks, we designed a Semantic task which tested children's comprehension of the morphemes ***-e*** and ***-tay***. We asked whether children could attribute a sentence marked with ***-e*** (e.g. "There is a puppy in the box-***e***") to the character who looked inside a container and a sentence marked with ***-tay*** (e.g. "There is a puppy in the box-***tay***") to the character who was verbally informed about its content.

Additionally, we asked whether children know the discourse functions of the evidential morphemes (in accordance with the evidentiality hierarchy; cf. (5)). Namely, do children know that the speakers' choice of morphemes imparts information about the speaker's level of certainty? Our Pragmatic task pits two statements whose contents and evidential morphemes differ and asks which one children tend to believe.

Our non-linguistic tests of source monitoring were modeled after the set of studies surveyed in Section 1.1. Our design included two source monitoring tasks which differed primarily in whether monitoring of knowledge in *oneself* or *others* was involved. For the 'Self' source monitoring task, children had to report how they found out about the content of a container. For the 'Others' source monitoring task, children had to identify which of two puppets had gained informational access to its content.

2.1. Method

2.1.1. Participants

A total of 32 3-year-old (mean age: 3;3, ranging from 2;11 to 3;9) and 32 4-year-old (mean age: 4;4, ranging from 3;11 to 4;9) Korean-speaking children participated. The children came from upper-middle-class families. They were recruited from several preschools in Seoul, Korea, and tested individually in a quiet room outside the children's classroom. Additionally, a control group of 8 Korean-speaking adults living in Seoul was recruited and tested in the linguistic tasks in the same way as the children.

[6] We should point out that by 'non-linguistic tasks' we mean source reasoning tasks that do not involve use of evidential morphology (even though they do involve the use of language).

2.1.2. Stimuli and procedure

2.1.2.1. Linguistic comprehension tasks. Stimuli were presented on the screen of a laptop computer. Materials involved animated scenarios created by the Macromedia Director program. The digitized audio for the animations was recorded from voices of three native Korean speakers and partly edited with the Jam-it software for special effects.

For the Semantic trials, children were introduced to a character named Billy who owned many boxes. Billy would be playing the "Who said that?" game and help the child uncover the things hidden inside the boxes. In each animation trial, Billy introduced two new boxes (one to his left and one to his right) and two new friends (also to his left and right) and asked the child to pay attention. Then he said: "Now one of my friends is going to look inside the box." The friend on the left opened and looked inside the box and then closed the box and returned to his initial place. After that, Billy announced: "Now I am going to tell one of my friends what is inside the box", while motioning for his friend on the right to come closer. After his friend moved next to him, he began whispering "Inside the box is..." and intentionally trailed off his voice. Afterwards, that friend also returned to his initial place. For all trials, the character on the left always looks and the character on the right always listens. Having an action consistently tied to a location reduces the possibility that children forget what happened.[7] To ensure that the children were paying attention, the experimenter would also probe the children with the questions "Who looked inside the box?" and "Who was told about the box?". These questions were fairly easy for the children. If they responded incorrectly, they would be shown the looking and telling events again.

Following the presentation of the two events, two narrow strips of curtain were lowered to cover just the two friends (leaving the boxes in full view). Then Billy told children: "Now one of my friends is going to say something, so listen carefully". The computer played the test sentence in a voice different from Billy's. The sentence involved either ***-e*** or ***-tay*** (e.g. "There is a balloon inside the box **-e/tay**"[8]). Afterwards the curtains were lifted to uncover the two friends. Billy then asked children: "Who said that?". The pairs of friends were matched such that the voice uttering the test sentence could belong to either friend. If children hesitated, the experimenter repeated the test sentence and the question. After children made a choice, the computer displayed the contents of both boxes. Both boxes always contained the same thing (what was described in the test sentence).

Each child was administered two ***-e*** trials and two ***-tay*** trials. The ***-e*** and ***-tay*** trials were blocked and their presentation order was counterbalanced across children so that half of the children received the two ***-e*** trials first and half received the two ***-tay*** trials first.

Half of the children (16 3-year-olds and 16 4-year-olds) also received two training trials prior to the ***-e*** and ***-tay*** test trials. Training trials were embedded in stories that

[7] A potential problem is that, if children had a particular bias (e.g. picking the left side), we would not know whether it was a side-bias or a perceptual source bias. Fortunately no such pattern emerged when we examined the data.

[8] Sangca aney phwungsen-i iss-e/iss-tay
box inside balloon-Nom be-e/be-tay
'There is a balloon inside the box -e/-tay.'

had the same structure as the test stories but involved the open class words 'look' (*po*) and 'tell' (*malhaycwu*). There were two such trials, one with 'look' (''I looked inside the box'') and one with 'tell' (''Billy told me what is inside the box''), always presented in that order. These trials served as baseline comparisons for the ***-e*** and ***-tay*** trials. We expected children to perform equally well, if not better, on these open class word trials because they transparently stated the speaker's access to information. One might also expect that, after first practicing with 'look' and 'tell', children might improve their performance on the ***-e*** and ***-tay*** trials by paying more attention to the linguistic cues distinguishing the two sources.

For the Pragmatic task, a narrator named Zowie asked the child to join her in a game called ''What animal is behind the curtain?''. She emphasized that there was only one animal behind the curtain and asked the child to listen to her friends to figure out which animal that was. There were two test trials. Zowie stood in the middle of the screen, in front of a curtain. For each trial a different pair of friends was present. One friend stood to the left and one to the right of Zowie. The two friends took turns uttering a different sentence each (''There is a cat behind the curtain-**e**''[9] or ''There is a puppy behind the curtain **-tay**''[10]). Who spoke first was randomly determined; however for one trial, the character on the left uttered the ***-e*** sentences while on the other trial, the sentence was uttered by the character on the right.

The order of the Semantic and Pragmatic tasks was counterbalanced across children of both ages.

2.1.2.2. Non-linguistic source monitoring tasks. A paper dollhouse served as the stage for the source monitoring tasks. The dollhouse consisted of places (e.g. drawer, refrigerator) in which items (e.g. plate, slippers) could be hidden. A total of eight different items were hidden, each in a different location, prior to the arrival of child participants. We also selected two puppets familiar to children (Mickey and Minnie) who would discover the hidden objects.

For a typical experimental session, children were first introduced to the dollhouse and told that they would play a treasure hunt game to reveal items hidden in secret places. To engage children and encourage them to speak, the experimenter asked them to name objects, furniture, and potential hiding places in the room during a warm-up period. After children were comfortable answering questions, the experimenter administered the two source monitoring tasks (the Self and Others tasks). The presentation order of the two tasks was counterbalanced across each age group.

The Self task, modified from O'Neill and Gopnik (1991), involves having children discover the contents of secret hiding places within the dollhouse. For each trial, children either saw for themselves or were told about the content of a new hiding place,

[9] Khethun twyey koyangi-ka han mali-ka iss-e
curtain behind cat-Nom one classifier-Nom be-e
'There is a cat behind the curtain -e'.

[10] Khethun twyey kangaci-ka han mali-ka iss-tay
curtain behind puppy-Nom one classifier-Nom be-tay
'There is a puppy behind the curtain -tay'.

and then had to report how they found out. The experimenter would point out: ''There is something hidden inside. Do you want to know what is inside?''. The experimenter would next instruct children, depending on the trial, to either have a look inside or to let her tell them what is inside. To check whether children were paying attention, the experimenter then probed them for the identity of the item (e.g. ''What is in the cabinet?''). Immediately after replying, children were asked: ''How did you know? Did you look? Did I tell you?''. Altogether children received four trials, two involving looking and two involving telling. The two types of trials (looking vs. telling) were blocked, and the order of presentation was counterbalanced such that half of the children received the looking questions before the telling questions and the other half received the telling questions before the looking questions.

For the Others task, participants had to choose the more knowledgeable of two characters (Mickey or Minnie). To establish that children knew the names of the two characters and were willing to choose between the two when given a forced choice question, we added a warm-up phase. During this phase, children had to watch each of the two characters perform a certain action (Mickey wash his hands, Minnie brush her hair) and then answer a question about who performed the action. Once they willingly and correctly answered these warm-up questions, the experimenter continued to the test questions.

For each trial, both characters engaged in some action in the scene. However, only one of the two characters performed an action that enabled him or her to discover the content of a container. The question is whether children could reason about the knowledge state of the two characters based on what the characters had done.

Again, there were four questions, two of which involved looking and two telling. In the looking trials, one of the characters would look into a secret hiding place (e.g. the cabinet) and one would kick or tap it. To ensure that children were paying attention, the experimenter would ask questions similar to those asked during the warm-up phase (e.g. ''Who looked inside the cabinet?'' ''Who kicked the cabinet?''). These questions were fairly easy for the children. If they responded incorrectly, the experimenter would reenact the scene with the puppets and correct the children's response. After determining children knew who did what, the experimenter then asked the test questions: ''Who knows what is in the cabinet? Mickey or Minnie?''.

In the telling trials, the experimenter spoke to one of the two characters and indicated her intentions to convey the contents of a secret hiding place by beginning an utterance (e.g. ''In the closet, there is a . . .''). However, instead of finishing the message out loud, she pretended to whisper the rest of the message into the character's ear. The experimenter performed some irrelevant action (e.g. kissing or hugging) with the other character. For example, the experimenter would also begin with an utterance, ''I am going to give you a kiss . . .'', then proceeded to kiss the character. As in the looking condition, children had to determine which character knew what was hidden inside the secret place (''Who knows what is in the closet? Mickey or Minnie?'').

In half of the trials, Mickey was the more knowledgeable character and in the other half, Minnie was. Again, the two types of trials (looking vs. telling) were blocked, and the order of presentation was counterbalanced. The non-linguistic

tasks preceded the linguistic tasks and a short break intervened. (In what follows, we reverse this order for ease of presentation.)

2.2. Results

2.2.1. Linguistic comprehension results

Korean adults were 100% correct on all questions of the linguistic task. For the Semantic task, they attributed the ***-e*** sentences to the character who looked and the ***-tay*** sentences to the character who was verbally informed. For the Pragmatic task, they chose to believe what was conveyed by the ***-e*** sentence over the ***-tay*** sentence.

Children's performance on the Semantic and Pragmatic tasks is presented in Fig. 1. Starting with the Semantic task, we analyzed children's performance on the two types of test questions (***-e*** and ***-tay***). We submitted the percentage of correct responses to a 2 (Morphology: ***-e***, ***-tay***) × 2 (Group with Open Class Trials: Yes, No) × 2 (Age: 3, 4) ANOVA[11], with Morphology as a within-subjects factor. The analysis yielded no main effect of Morphology: ***-e*** trials (61% correct) did not differ from ***-tay*** trials (52% correct). Furthermore, there was no Age effect or any interactions with that term: 3-year-olds' performance (59% of correct responses) was not significantly different from 4-year-olds' (54% of correct responses). Finally, there was no main effect of Group with Open Class Trials and no interaction effects with that term: children who received the open class word trials scored 59% correct and children who did not scored 55%. Next, we tested each group plotted in Fig. 1a against chance performance (50% correct), and found that only 4-year-olds' comprehension of ***-e*** was above chance ($t(31) = 2.06$, $p = .048$); in all other conditions performance was at chance ($p > .15$).

The children who received open class word trials scored 66% on the open class trials. Analysis of the results from the open class items showed that 4-year-olds almost scored significantly better than 3-year-olds by a two-tailed *t*-test (78% correct vs. 53%, $t(30) = 1.78$, $p = .085$). Testing against chance by a two-tailed *t*-test indicated that 4-year-olds were significantly above chance ($t(15) = 3.09$, $p < .01$), while 3-year-olds were not. Thus, importantly, the 4-year-old children were starting to pass the open-class word trials while still failing at the closed class evidential morphology trials. Furthermore, of the 17 subjects who correctly answered both open class word questions, only 6 (38.9%) scored 75% or more correct on the Semantic task. Hence being able to answer the open class questions for verbs of perception and communication does not necessarily entail higher likelihood of passing test questions for evidential morphology (even though the presentation conditions in the training and test tasks were similar).[12]

[11] We will not include the order of presenting the ***-e*** and ***-tay*** sentences nor the order in which the Semantic and Pragmatic task were administered because these factors yielded no significant main effects.

[12] Inversely, of the 15 children who failed at the open class questions, only 6 passed the Semantic task. Thus, the number of children who tend to pass the Semantic task and fail at the open-class trials is not reliably higher than the number of children who pass the open-class trials but fail on the Semantics task, $\chi^2(1, N = 32) = .07$, $p > .05$.

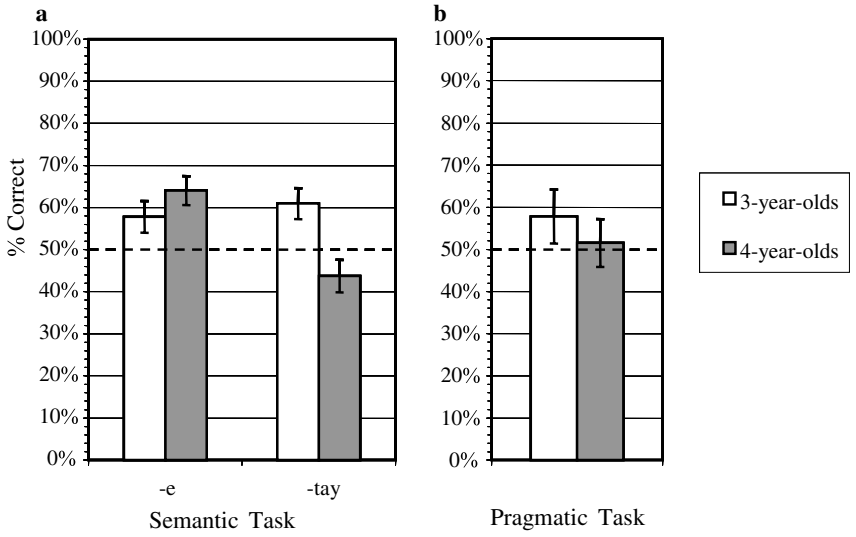

Fig. 1. Korean-speaking children's performance on linguistic tasks (Experiment 1). (a) Semantic Task; (b) Pragmatic Task.

Children's performance on the Pragmatic task was as poor as their performance on the Semantic task. A one-way ANOVA on the percent of correct responses showed that 3- and 4-year-olds did not differ significantly in performance (58% vs. 52% of correct responses respectively, $F(1,62) = .531$, $p = .469$) and their performance was not different from chance (3-year-olds: $t(31) = 1.22$, $p = .28$; 4-year-olds: $t(31) = .273$, $p = .79$).[13]

2.2.2. Source monitoring results

Korean children's source monitoring performance is presented in Fig. 2. Using the percentage of questions answered correctly as the dependent variable, we conducted a 2 (Age: 3, 4) × 2 (Perspective: Self, Others) × 2 (Source: Look, Tell) ANOVA. (We did not include factors pertinent to the presentation order of the questions because preliminary analysis indicated that order contributed no main or interaction effects). We found a main effect of Age ($F(1,62) = 11.09$, $p < .001$): 3-year-old children on average scored 73% while 4-year-olds scored significantly better at 91%. There was additionally a main effect of Perspective ($F(1,62) = 17.75$, $p < .001$): performance on the Self questions was significantly better than performance on the Others questions (89% vs. 75%, respectively). There were no other significant main effects or interactions.[14]

[13] A regression of the Pragmatic score on the Semantic score showed no relation between the two tasks. Most of the two scores cluster around 50% correct, and hence provide no useful correlation.

[14] Robinson et al. (1998) showed that being able to explicitly reflect on "how" and "why" someone knows is more difficult than implicitly judging who knows. This is relevant for the present results: our Self task involves explicit reflection of "how" children found out, while the Others task only involves implicit judgment.

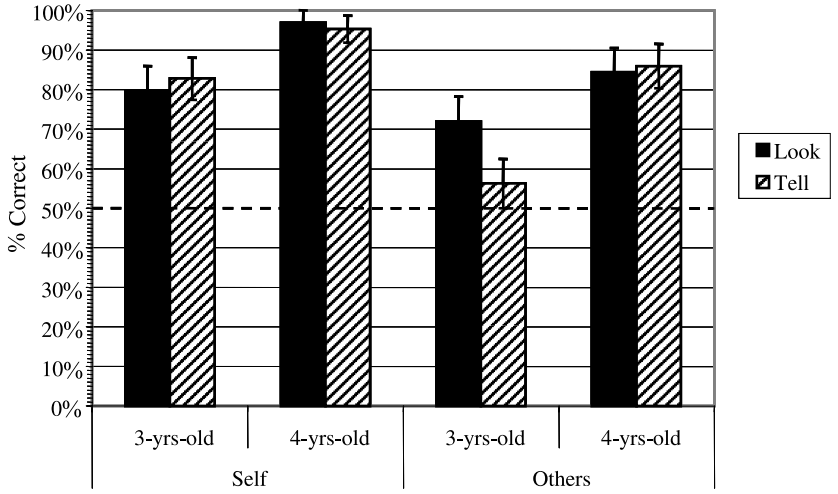

Fig. 2. Korean-speaking children's performance on source monitoring tasks (Experiment 1).

Separate 2 (Perspective) × 2 (Source) ANOVAs for each age group confirmed the effect of Perspective as the only significant main effect for each age group (3-year-olds: $F(1,31) = 11.471$, $p = .002$; 4-year-olds: $F(1,31) = 6.356$, $p = .02$).

Both 4- and 3-year-olds were significantly above chance (4-year-olds: Self: $M = 96\%$, $t(31) = 14.4$, $p < .001$; Others: $M = 85\%$, $t(31) = 6.9$, $p < .001$; 3-year-olds: Self: $M = 81\%$, $t(31) = 7.2$, $p < .001$; Others: $M = 64\%$, $t(31) = 2.8$, $p < .01$).

2.2.3. Comparison between linguistic comprehension and source monitoring

In our final analysis, we compared the linguistic results of this section with the non-linguistic understanding of information source in Korean children. The non-linguistic score was calculated as a percentage correct out of eight source monitoring questions (4 Self and 4 Other questions) and the linguistic score was the cumulative score out of 6 total questions (2 Semantic ***-e*** questions, 2 Semantic ***-tay*** questions, and 2 Pragmatic questions). The performance on the non-linguistic score did not correlate with the linguistic score ($r = -.17$, $p = .18$, $n = 64$) due to the fact that children were failing the linguistics tasks. In fact, regardless of which non-linguistic subscore (e.g. Self, Others, Look, etc.) was compared with which linguistic subscore (e.g. Semantic total, Semantic ***-e*** questions, etc.), the non-linguistic scores did not correlate with the linguistic scores.

We also submitted the percentage correct to a 2 (Task: Non-linguistic, Linguistic) × 2 (Age: 3, 4) ANOVA, with Task as a within-subjects factor and Age as a between-subjects factor. We found an almost significant effect of Age ($F(1,62) = 3.357$, $p = .072$), with 4-year-olds averaging 72% correct and 3 year-olds averaging 66% correct. Most importantly for present purposes, we found a strong effect of Task ($F(1,62) = 46.46$, $p < .001$): the Non-linguistic score (82%) was higher than the Linguistic score (46%). Finally, the analysis revealed a Task × Age effect ($F(1,62) = 9.914$, $p < .003$). As Fig. 3 indicates, this effect is due to a larger Age effect

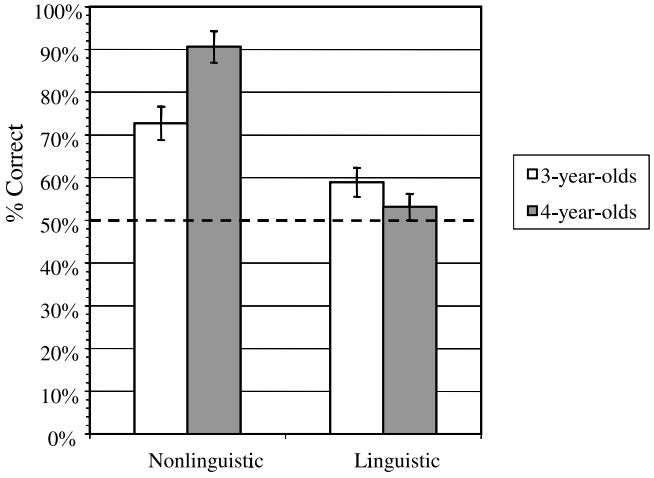

Fig. 3. Comparison of Korean-speaking children's performance on non-linguistic and linguistic evidentiality tasks (Experiment 1).

on the Non-linguistic task than the Linguistic task.[15] Separate ANOVAs comparing the Age for each task confirmed that there was no Age effect for the Linguistic task but a large Age effect for the Non-linguistic task ($F(1,62) = 11.09$, $p < .001$).

2.3. Discussion

Our experimental results extend what is known in the literature about the acquisition of evidential systems. Our linguistic tasks indicate that young Korean learners cannot attribute an evidentially marked utterance to the appropriate speaker and cannot judge that hearsay statements, other things being equal, carry reduced levels of speaker certainty. These results are similar to findings from Turkish, according to which children do not exhibit knowledge of evidential morphemes in comprehension tasks until after the age of five (Aksu-Koç, 1988) and the pragmatic effects of these morphemes until even later (Aksu-Koç & Alici, 2000).

Korean children's difficulties in our linguistic tasks contrast sharply with results from our non-linguistic tasks which show that the same children successfully engage in source monitoring. Even 3-year-olds who had a more difficult time compared to 4-year-olds explicitly reporting the appropriate sensory experience that led to their knowledge performed above chance in the source monitoring tasks. The difference is remarkable given that the structure of the tasks is quite similar: For instance, in the Others task children had to attribute knowledge to a character based on his evidential access to information (perception or communication), whereas in the

[15] Again, the basic results do not change regardless of which linguistic score (or subscore) is compared to which non-linguistic score (or subscore). There is always an effect of Task (Non-linguistic better than Linguistic), and always an effect of Task × Age (with a larger Age effect for Non-linguistic than Linguistic tasks).

Semantic task, children had to attribute an utterance marked for evidential access to one of two characters.[16] Overall, then, it seems that mental-state attribution in the source monitoring tasks is easier than utterance attribution in the Semantic task (a point we return to in Section 6).

One might object that certain aspects of our linguistic tasks were inherently difficult for children. For instance, it might be argued that the Pragmatic task involved evaluation of the connection between evidential source and certainty. However, it is worth pointing out that, using contrastive tasks very similar to our Pragmatic task, other researchers have shown that children can make inferences about speaker certainty from the use of epistemic modals (e.g. *may vs. should*) or mental verbs (e.g. *know vs. think*) in English by the age of four (Hirst & Weil, 1982; Moore, Bryant, & Furrow, 1989; Moore & Furrow, 1991; Papafragou, 1998 for a review; but see Bascelli & Barbieri, 2002).[17]

Still, it is possible that the present findings underestimate what Korean children know about linguistic evidentials. Specifically, the first of our linguistic tasks requires children to reason explicitly about the meaning of the morphemes ***-e*** and ***-tay*** in order to infer who might have produced an utterance containing them. Furthermore, the very process of attributing an utterance to a potential speaker could have complicated the task in ways irrelevant to children's knowledge of evidentiality: children had to remember multiple events and characters in order to match up an utterance with a prior event involving the speaker. These objections are mitigated somewhat by two facts. First, children could in fact recall which character looked inside the box and which character was told when explicitly asked. Second, and more crucially, 4-year-olds were above chance at attributing utterances marked with open class words explicitly specifying information source, even though these same children could not subsequently select the appropriate speaker on the basis of the ***-e*** and ***-tay*** morphemes. Nevertheless, we decided to administer a variation of our comprehension task that might better tap into Korean children's understanding of these evidential morphemes to a new group of Korean preschoolers.

3. Experiment 2

To further examine 3- to 4-year-old Korean children's understanding of evidentiality, we adopted a modified version of the Truth-Value Judgment task (Crain &

[16] We note that the tasks under comparison, even though similar, are not identical. In the Others task, the child needs only to contrast an event that leads to information (e.g. looking or being told) to an uninformative event (e.g. tapping or being kissed). But for the Semantic task, the child needs to contrast two different modalities of knowledge acquisition (i.e. looking vs. being told) and match one to the linguistic stimulus. As a result, it is possible that the Semantic task may be more difficult. That said, children did not have any difficulty identifying who looked or who was told when probed by the experimenter.

[17] In fact, using the same paradigm and animations from the Pragmatic task we asked 3- and 4-year-old English speakers which animal they would pick when hearing "I know there is cat behind the curtain" and "I think there is a puppy behind the curtain." Four-year-olds performed above chance (78% correct, $t(15) = 4.39$, $p < .001$), selecting the animal associated with *know* rather than *think* while 3-year-olds were at chance (59% correct, $t(15) = 1.14$, $p = .27$).

Thornton, 1998: TVJ task heretofore). Using this paradigm, we presented children with a puppet who had either directly witnessed an event or had been told about the event by an event participant. The puppet produced a report sentence about the event using either ***-e*** or ***-tay***. We then asked the child whether the puppet's report about the event was acceptable or just silly. By presenting children with a single event paired with a single sentence, we reduced the demands of remembering multiple events and characters, which might have contributed to the complexity and difficulty of the Semantic task in Experiment 1. Furthermore, children were no longer required to reason explicitly about the meanings of evidential morphemes (e.g. by considering who could have used the particular morpheme in a given situation).

3.1. Method

3.1.1. Participants

Eleven Korean 3-year-olds (mean age = 3;4, ranging from 3;0 to 3;10) and 11 4-year-olds (mean age = 4;5, ranging from 4;0 to 4;11) participated in this study. The children were recruited from the city of Suwon in the vicinity of Seoul and all came from upper-middle-class families. Of these, one 3-year-old and three 4-year-olds were excluded due to their response biases ('yes' responses throughout the entire session). To these, we added 2 4-year-old participants who were recruited from the vicinity of Philadelphia, US, since their level of Korean and their demographics were comparable to those of Korean children recruited in Korea. Therefore, 10 children in each age group entered the analyses. Ten adult native speakers of Korean also participated in the study.

3.1.2. Stimuli and procedure

Two characters, Jin (a female human actor) and Tim (an elf-looking puppet wearing a large red hat), were first introduced to the children via a photo display on a computer screen. Then Tim physically appeared at the experimental scene and stood next to the child. The child was told that Tim and Jin are close friends and they played together the day before, and that they would watch what they actually did the day before through a computer display. The child was also told that, after viewing the event, Tim would tell him/her about the event they viewed. But Tim was unreliable, so sometimes he was being good but other times he was being silly in his story telling. The child was asked to decide whether the way Tim told the story was correct/good or wrong/silly. And if Tim was good, the child was to reward him by patting his hat while saying "Good!". If the child thought that Tim was silly or wrong, then s/he was asked to punish Tim by pinching his nose while saying "Silly!". Thus, in each trial, the child watched a video-clip describing a single event, and then heard Tim's report about the event, followed by the question whether Tim's statement was good or silly.

Two types of event were video-taped and presented on the computer screen: one in which Tim is watching while Jin carries out a certain action (e.g. eats an apple or takes pictures), and another in which Jin reports an event to the puppet (e.g. "I went to the market with Mom yesterday"). We labeled the first type of event a 'Looking

event' and the second a 'Telling event'. There were a total of 4 events in each type, thus yielding 8 total videotaped events (see Appendix A for detailed descriptions of each event). For each video clip, there was a corresponding statement made by Tim, the puppet, about the event. Tim's statements ended in either ***-e*** or ***-tay***. Two of the Looking events were paired with ***-e*** statements while the other two were paired with ***-tay*** statements. Similarly, two of the Telling events were paired with ***-tay*** statements whereas the other two were followed by ***-e*** statements. Thus, there were a total of 8 trials (see Appendix A), of which half contained properly matched event-statement pairs (direct perception – direct evidence or indirect access – hearsay) and the other half mismatched pairs (direct perception – hearsay or indirect access – direct evidence). Since in all the test trials the base propositions (minus the sentence-ending evidential morphemes) were true, we chose to exclude those children who gave affirmative judgments about the statements throughout the trials because of the possibility that they might have paid attention only to the content of the base sentences while ignoring the sentence-ending morphemes. The test trials began with a Looking event and the order of presentation was fixed (see Appendix A).

3.2. Results

For purposes of analysis, we obtained an average proportion of acceptance of the statements per individual split by event (Looking/Telling) and morpheme type (***-e***/***-tay***).

As shown in Fig. 4a, when the event was of the Looking type (e.g. Tim watches while Jin takes photographs), adults accepted the puppet's reports that ended with ***-e*** 100% of the time, but those with ***-tay*** only 15% of the time. By comparison, when the event was of the Telling type (e.g. Jin tells Tim that she went to the market with her Mom), reports with ***-tay*** were accepted 100% of the time (see Fig. 4b). Interestingly, for the Telling events, the acceptance rate of ***-e*** statements was pretty high (70%) among adults. Indeed, this type of response was viable if the observer assumes that Tim trusted what Jin told him to be true, and that Jin is truthful. Under these circumstances, ***-e*** is interpreted as a declarative marker and not a marker of direct evidence. This reasoning was confirmed when we conducted post-experiment interviews with the participants who accepted the ***-e*** statements for the Telling events. The percentage of acceptance was submitted to a 2 (Event: Looking, Telling) × 2 (Morpheme: ***-e***, ***-tay***) ANOVA. The analysis yielded a main effect of Event ($F(1,9) = 8.4$, $p < .05$) and a main effect of Morpheme ($F(1,9) = 8.4$, $p < .05$). The main effect of Event corroborated the observation that adults more often accepted either type of utterance as appropriate for the Telling events (85%) than for the Looking events (57.5%). The main effect of Morpheme indicated higher acceptance of ***-e*** (85%) than ***-tay*** statements (57.5%). As predicted, there was also a significant Event × Morpheme interaction ($F(1,9) = 39.3$, $p < .001$). Planned pair-wise comparisons of the Morpheme for each Event type showed that for the Looking events, adults accepted the ***-e*** statements more often than the ***-tay*** statements ($F(1,9) = 63.4$, $p < .001$). The reverse, with higher acceptance of ***-tay*** than ***-e*** statements, was not significant for the Telling events ($F(1,9) = 3.86$, $p = .08$). Subsequent

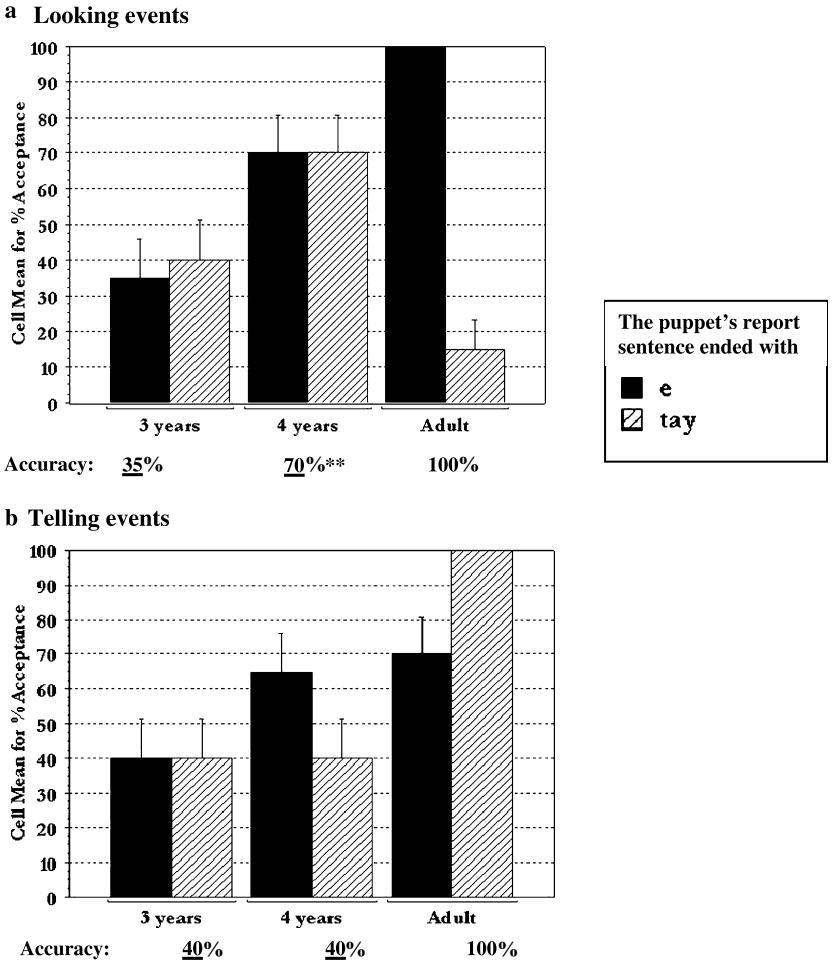

Fig. 4. Korean 3- and 4-year olds' and adults' mean proportion of acceptance of the puppet's sentences for two different event types: (a) looking events and (b) telling events. $^{**}p < .05$ (tests against chance).

t-tests against chance (50% acceptance rate) showed that the percentage of acceptance for the ***-e*** statements in the Telling condition (70%) did not differ from chance ($t(9) = 1.31$, $p = .22$), whereas ***-tay*** was accepted below chance for the Looking condition ($t(9) = -3.28$, $p = .01$).

In contrast to the adults, 3- and 4-year-olds were equally likely to accept the statements with ***-e*** or ***-tay*** for the Looking events (Fig. 4a). In other words, they were equally likely to say that ***-tay*** statements are as good as ***-e*** statements even when Tim directly perceived the events. In the Telling condition, 3-year-olds showed a similar pattern; that is, they did not differ in terms of accepting or rejecting the ***-e/-tay*** statements (Fig. 4b). Four-year-olds, accepting slightly more ***-e*** statements than ***-tay***

statements, oddly showed the opposite trend than the adults. However, submitting the percentage of acceptance into a 2 (Event: Looking, Telling) × 2 (Morpheme: -e, -tay) × 2 (Age: 3s, 4s) ANOVA with the last factor as a between-subjects factor yielded only a significant effect of Age ($F(1,18) = 5.05$, $p = .037$, and $p > .15$ for all other effects). The 4-year-olds were overall more willing to accept the statements than the 3-year-olds (61.3% vs. 38.8%), although it is unclear why this should be the case. Importantly though, regardless of the children's baseline acceptance rates for the utterances, children did not differ in their acceptance rates for ***-e*** and ***-tay*** statements for each type of event.

In sum, the pattern of our results suggests that Korean children by the age of 4 have not yet fully acquired the meaning of these evidential morphemes: they do not reliably distinguish between the meanings of ***-e*** and ***-tay*** – hence their understanding of these morphemes is not yet in place and must still be developing.

3.3. Discussion

In this experiment, using a different comprehension (TVJ) task, we found similar results to those found in Experiment 1: specifically, we showed that 3-to-4-year-old Korean children have not yet fully acquired the meaning of the evidential morphemes ***-e*** (direct evidence) and ***-tay*** (hearsay). Unlike our previous Semantic and Pragmatic tasks, children in the TVJ task did not have to keep track of 'who did what' (e.g. who looked into the box versus who was informed by someone else), nor did they have to reason about who could have produced a certain utterance under the experimental circumstances. Despite the fact that the task demands were not as high as in the previous experiment, we still did not find evidence that Korean children by the age of 4 understand the meanings of these evidential morphemes.

Somewhat unexpectedly, adults allowed some room for variation in the use of these evidential morphemes. Particularly, in the Telling events, the puppet's reports with ***-e*** were highly acceptable given that the puppet trusted its informant. Also, in the Looking events, reports with ***-tay*** were not entirely bad for adults since the purpose of ***-tay*** might have been to indicate that the puppet was simply trying to be indirect, rather than to mark hearsay. Nevertheless, the overall pattern of adult responses differed significantly from the children's judgments.

The total set of comprehension tests so far, then, points to the conclusion that Korean learners by age four have limited understanding of the semantics of evidential morphemes. This conclusion is further supported by other versions of evidential comprehension tasks we devised and administered. In one of these tasks, 3- and 4-year-old Korean children were presented with two characters and each character offered a statement about the contents of a container. The statements were identical except for a sentence-ending marker (e.g. "There's a parrot in the box ***-e/tay***"). The child then had to say which of the two characters saw what was in the box (or, alternatively, which one was told by the experimenter). The idea was that, if participants could use the semantic information encoded in the morphemes as a basis for their response, they should decide that the '***-e***' character saw what was in the box and

the '***-tay***' character was verbally informed about its contents (a group of adult controls behaved in just this way). However, neither age group performed significantly above chance in this task.

Despite the convergence of our linguistic tasks, it remains an open possibility that Korean children's own production of sentence-final morphology might reveal some knowledge of evidential ***-e*** and ***-tay*** which does not surface in explicit comprehension tasks. As mentioned in Section 1, there is some evidence that Korean children produce evidential morphology correctly already from the age of two (Choi, 1995). As a final probe into the acquisition of evidentiality, we turn to young Korean children's use of evidential morphemes in their own production and compare such use to their source monitoring abilities.

4. Experiment 3

In this experiment, we attempted to elicit children's production of the evidential morphemes for direct evidence (***-e***) and hearsay (***-tay***) in two different conversational situations. For ***-e***, the child had to correct a puppet's silly hearsay statement contradicting the child's own description of a previously experienced event. For ***-tay***, the child first learned about two puppets' activities by hearing their descriptions, and then had to inform another person about what the puppets said. In addition to the production task, we administered the source monitoring task used in Experiment 1 to the children who participated in the production study. In this experiment, we also included children at the age of 5 in order to draw more detailed developmental comparisons, especially about the acquisition of evidential morphology.

4.1. Method

4.1.1. Participants

A total of 45 monolingual Korean children were recruited for this study: sixteen 3-year-olds (age range: 3;0–3;11, mean = 3;6), 15 4-year-olds (age range: 4;0–4;11, mean = 4;7), and 16 5-year-olds (age range: 5;0–5;11, mean = 5;3). All of them were from middle to upper middle class families living in Suwon, Korea. Due to difficulties in scheduling children for multiple sessions, not all children participated in all three tasks administered for the study. All children, participated in both the hearsay (***-tay***) elicitation task and the source monitoring task. However, one 3-year-old did not complete the source monitoring task. In the direct evidence (***-e***) elicitation task, 14 three-year-olds, 7 four-year-olds, and 6 five-year-olds participated.

4.1.2. Stimuli and procedure

4.1.2.1. Elicited production task. For elicitation of the direct evidence morpheme (***-e***), a puppet named Elmo (enacted by the experimenter) was first introduced to the child, and the child was told that Elmo wanted to talk about what the two of them did the day before. The child was also informed that Elmo was known to lie about many things and that the child would want to watch out and correct what Elmo said.

Once the situation was set up, Elmo started out by telling the child what he himself had done and then asked the child what s/he did the day before (e.g. ''I ate strawberries yesterday. What about you, Suzie?''). After the child answered (e.g. ''I ate ice cream''), the experimenter asked Elmo what the child did. Elmo always produced a silly report about what the child just said by changing the verb, such as ''Suzie kicked ice cream yesterday, hehehe,'' and Elmo's report always ended with ***-tay*** (hearsay). Then the experimenter asked the child whether the puppet lied or not, and encouraged the child to correct Elmo's silly report. Crucially, in order to correct Elmo, children would have to state what they really did, changing the verb and turning the reportative marker ***-tay*** in Elmo's misleading statement back into ***-e*** (first-hand experience). This kind of conversational exchange was repeated four times so that each child was given four opportunities to correct the puppet's statements. The verbs and statements provided by Elmo varied by trials and by individual child since the conversation with the puppet was put together on the fly.

For purposes of eliciting the hearsay morpheme (***-tay***), two puppets (a raccoon in a trash can and a duck) were used for the experimental set-up. The child's task was to report to the experimenter what the puppets did the day before after hearing each puppet's report of their activities. The experimenter told the child that she does not understand what the puppets are saying but is very curious to learn what they did through the child's report. The puppets' reports on their activities were pre-recorded, spoken by two adult Korean speakers (we used a female voice for the duck and a male voice for the raccoon). Once the child became comfortable with the puppets and the experimenter, the experimenter encouraged the child to ask each puppet in turn what they did the day before, and then tell the experimenter. When the child asked each puppet, their answer was played back through a speaker hidden under each puppet to create the effect of real conversation. The puppets' reports always ended with ***-e*** (e.g. the duck said: ''I fought with my friends yesterday-***e***.''). After the child listened to each puppet's answer, the experimenter encouraged the child to tell her about what s/he heard – initially by simply looking curious and expecting an answer. If the child hesitated, the experimenter would attempt to elicit the child's response by using a phrase such as ''duck-nun?'' ('duck-topic marker'; here roughly, 'what about the duck?'). If the child was still unwilling to answer, then the experimenter used specific questions such as ''What did the duck say she did yesterday?'' or ''What did the duck do yesterday?''. The child was given a total of 8 opportunities (4 for each puppet) to report the puppets' activities. Again, in order to pass this task successfully, the child had to transform the puppet's statement into a report by turning the sentence-final morpheme ***-e*** into the hearsay marker ***-tay***.[18]

4.1.2.2. Source monitoring task. For the source monitoring task, the same materials (the paper dollhouse with Mickey and Minnie) and procedures were used as in Experiment 1, including the Self/Others and Look/Tell conditions (with 2 trials in

[18] Notice that the structure of the two elicitation tasks is not identical (e.g. only the first one involves deception/lying; only the first one requires a correction of both the main verb and the evidential marker). These differences were necessary given the semantic content of the target morphemes.

each condition). Approximately half of the children in each age group started with the Self task followed by the Others task, and the other half followed the reverse order.

4.2. Results

4.2.1. Elicited production results

Fig. 5 shows the proportion of each different sentence-ending (SE) morpheme the children produced in the direct evidence (5a) and hearsay (5b) elicitation conditions, respectively. The direct evidence production task elicited, alongside ***-e***, a small proportion of tokens containing the SE morpheme ***-nuntey***. This morpheme indicates an incomplete thought or prompts confirmation from the listener (Lee, 1997). Additionally, one utterance contained ***-tako***, a complementizer specifically used to mark the embedded declarative clause as being quoted, often in combination with the verb 'say' (i.e. ''...tako hata'', 'say that...'; Suh, 2000). In the context of our experiment, the child appeared to use the ***-tako*** complementizer to quote his own remark while omitting the verb 'say' (e.g. ''(I said) that I ate ice cream''). Finally, in some cases children answered that Elmo had not lied ('True!' responses) even though he had.

None of the children in all three age groups simply reiterated the puppet's silly report in its original form. To begin with, none of the children reproduced the ***-tay*** morpheme following the puppet's report: the majority of children in all age groups corrected the puppet's silly report by marking it with the ***-e*** morpheme. This pattern could be generously interpreted as showing that children from the age of 3 years are sensitive to the fact that the morpheme ***-tay*** is not appropriate for talking about one's own direct experience. However, given that ***-e*** is the most frequent SE morpheme in speech, it could alternatively be the case that children are simply inserting the ***-e*** marker as a default. The data from direct evidence elicitation alone are insufficient to adjudicate between these two alternative explanations; we need to supplement them by data from the hearsay elicitation task.

In the hearsay production, aside from no responses and noun phrase responses without a predicate (e.g. ''strawberries''; labeled as 'no SE'), all three age groups used either ***-e*** or ***-tay*** in their reports to the experimenter. Four children also used the ***-tako*** morpheme instead of ***-tay*** in five of the production tokens.

The use of ***-tay*** turned out to be variable across age groups. As illustrated in Fig. 5b, 5-year-olds were highly productive (84%) in using ***-tay*** in their reports. Four-year-olds spontaneously marked their reports with ***-tay*** 74% of the time. In contrast to older children, 3-year-olds marked their utterances with ***-e*** about 38% of the time and with ***-tay*** only 57% of the time. The proportion of utterances marked with ***-tay*** in each child was positively correlated with the child's age in months ($r = .427$, $p = .003$, $n = 47$), indicating a developmental pattern in which children become more reliable in marking hearsay with ***-tay*** as they grow older.

Further examination of individual patterns by dividing the children into groups based on the number of times they marked hearsay with ***-tay*** supported this conclusion. As Table 1 indicates, the majority of 4- and 5-year-old children, in contrast to 3-year-olds, almost always correctly marked their utterances with ***-tay*** and rarely

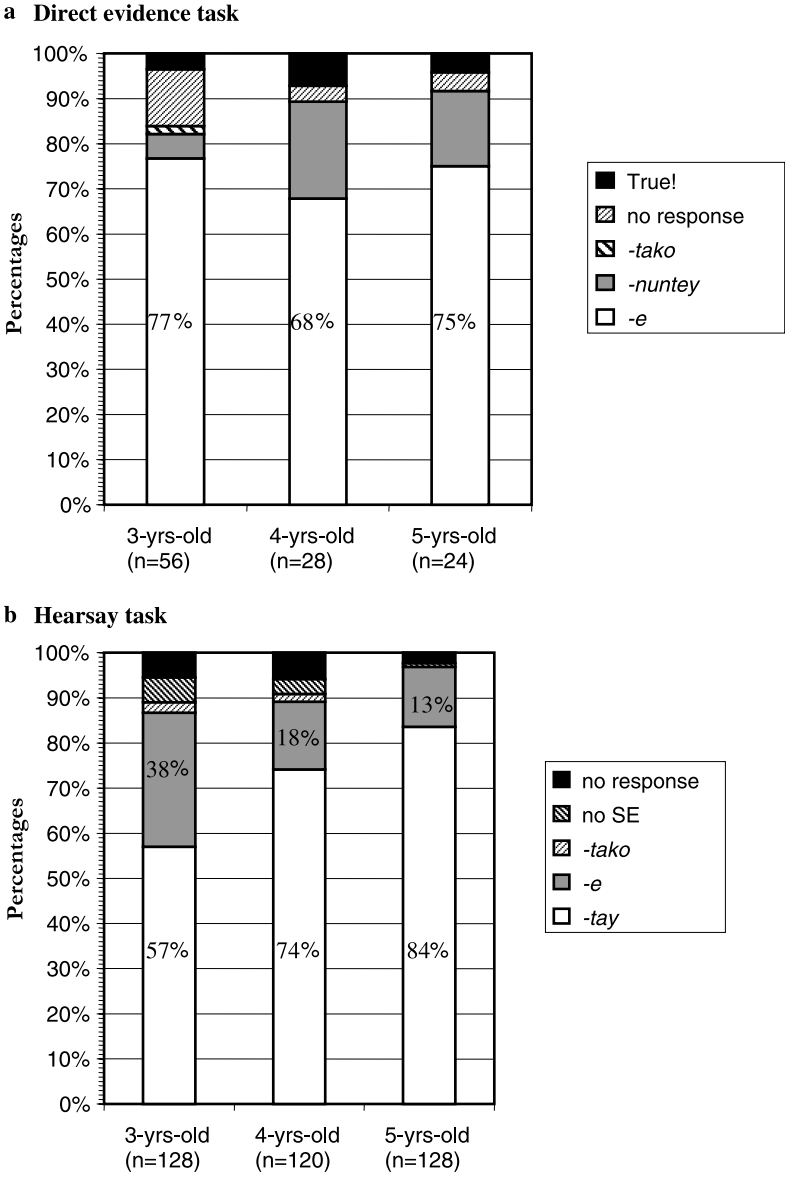

Fig. 5. Percentages of each type of sentence-ending morpheme in (a) the direct evidence elicitation and (b) the hearsay elicitation task. (Note: the number in parentheses next to age group refers to the total number of production tokens in each age group.)

incorrectly with ***-e***. Furthermore, the 3-year-old children's tendency to mark hearsay with ***-tay*** did not all center around the mean of 57%. Instead, the group also included children who rarely or never marked their utterances with ***-tay*** and others who frequently did so. Five of sixteen children marked ***-tay*** 25% or less of the time and 7 of

Table 1
Number of children divided by proportion of utterances marked with ***-tay*** (Experiment 3)

Age group	# of utterances (out of 8) marked with ***-tay***			
	0–1	2–3	4–5	6–8
3-years				
# of children	4	1	4	7
Ave. % utt. w/-tay	6.3%	25.0%	59.4%	89.3%
Ave. % utt. w/-e	71.9%	75.0%	21.9%	3.6%
Ave. age	3;4	3;2	3;6	3;8
4-years				
# of children	2	1	1	11
Ave. % utt. w/-tay	0.0%	37.5%	50.0%	93.2%
Ave. % utt. w/-e	75.0%	25.0%	25.0%	2.3%
Ave. age	4;6	4;7	4;3	4;9
5-years				
# of children	1	0	2	13
Ave. % utt. w/-tay	0%	–	62.5%	93.3%
Ave. % utt. w/-e	100%	–	31.3%	3.8%
Ave. age	5;0	–	5;2	5;4

16 marked ***-tay*** 75% or more of the time. Children who consistently produced ***-tay*** 75% or more of the time were also older than those who did not (mean age 3;8 vs. 3;4, $t(14) = 2.3$, $p = .035$). Hence, the data suggest that between 3 and 4 years of age children learn to productively use ***-tay*** to mark hearsay.

4.2.2. Source monitoring task results

Fig. 6 shows the mean percentage of questions answered accurately on the source monitoring task by those Korean children who also participated in the production study. A 3 (Age: 3, 4, 5) × 2 (Perspective: Self, Others) × 2 (Source: Look, Tell) ANOVA was conducted with both perspective and source as within-subject factors. As in Experiment 1, the main effect of Perspective was significant $F(1,43) = 3.95$, $p = .05$, with the Self (85%) task being much easier than the Others task (75%) of correct responses.

Unlike Experiment 1, the main effect of Age was not significant, but it was in the right direction with increasing percentage correct for older children (3-years: 73% (SD = 24%), 4-years: 81% (SD = 19%), and 5-years: 86% (SD = 16%)).[19] The lack of significance is probably due to the smaller number of subjects and larger age range sampled for each age group in this experiment relative to Experiment 1. However, there was a significant interaction effect between the Age and Source, $F(2,43) = 5.45$, $p = .01$. Fig. 6 suggests that the interaction is probably due to the much higher percentage correct for the Look condition relative to the Tell condition with just the 3-year-olds. Post hoc comparisons (Tukey) revealed that 3-year-olds

[19] The non-linguistic score was marginally significantly correlated with more sensitive measures of children's age in months ($r = .24$, $p = .064$, $n = 46$) or days ($r = .28$, $p = .057$, $n = 46$).

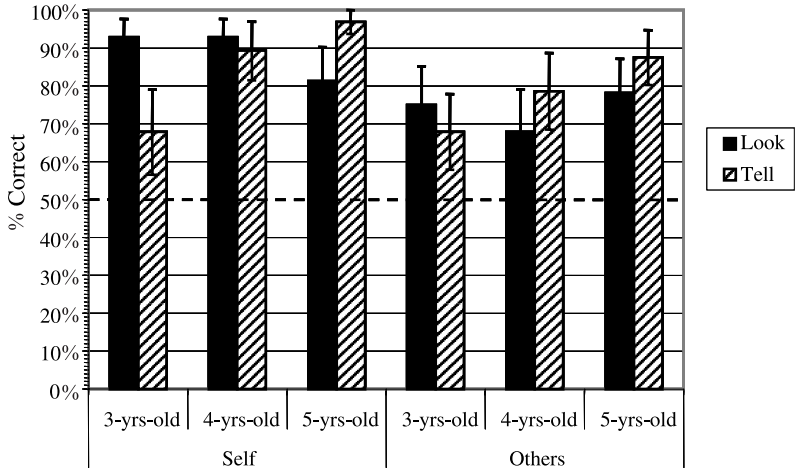

Fig. 6. Three- to five-year-old Korean children's mean accuracy in the source monitoring task (Experiment 3).

were significantly less accurate than the 5-year-olds in Tell events (65% vs. 92%, $p < .05$) while they did not differ from older children in the Look events (82% vs. 80% (4-years) and 80% (5-years)). Four-year-olds did not significantly differ from either 3- or 5-year-olds in the Tell condition. Children's performance was well above chance with the exception of the 3-year-olds in the Tell condition ($p = .08$).

Overall, even though data from the youngest (3-year-old) children tend to be more variable, these source monitoring results are consistent with our previous findings from Experiment 1: between 3 and 4 years children become better at source monitoring. This convergence was statistically confirmed: an ANOVA comparing 3- and 4-year-olds in Experiment 1 and the present experiment yielded no main effect of Experiment.

4.2.3. Comparison between production and source monitoring tasks

The proportion of utterances marked with ***-e*** produced by each child in the direct evidence elicitation task is not correlated with the same child's score for the source monitoring task ($r = -.12$, $p = .57$)[20] due to the low variance in the proportion of ***-e*** statements produced (across all three age groups, the proportion of children marking ***-e*** is similar and relatively high; Fig. 5a). Henceforth, we focus on the production of ***-tay*** in the hearsay elicitation task and its relation to non-linguistic source monitoring ability. An analysis of correlation showed a significant positive association between the proportion of utterances marked with ***-tay*** in the hearsay production task and the mean accuracy in the non-linguistic source monitoring task, $r = .49$, $p = .001$ ($n = 47$). This correlation holds even when controlling for the effect of

[20] In fact, the proportion of utterances marked with ***-e*** did not correlate with any of the more specific subscores of the non-linguistic source monitoring task (e.g. scores on Look or Tell trials; $p > .30$).

Table 2
Korean children divided by age, proportion of utterances marked with ***-tay*** for the hearsay elicitation task, and performance in non-linguistic source monitoring tasks (Experiment 3)

Group	# participants	Ave. Age	% of utterances marked w. ***-tay***	% correct on non-linguistic task
3-year-olds	Young ($N = 8$)	3;3	43.8	71.4
	Old ($N = 8$)[a]	3;9	70.3	75.0
4-year-olds	Young ($N = 7$)	4;5	66.1	78.6
	Old ($N = 8$)	4;10	81.3	82.8
5-year-olds	Young ($N = 8$)	5;2	78.1	84.4
	Old ($N = 8$)	5;5	89.1	87.5

[a] Missing 1 young participant in non-linguistic task.

children's age in months, $r = .43$, $p = .003$. In particular, the performance in the Tell source condition of the Others task was positively associated with the proportion of utterances with ***-tay*** in the hearsay production task, $r = .54$, $p < .001$. This suggests that the more hearsay morphology a child produced, the more accurate the performance of that child was when evaluating the knowledge that others obtained from communication.

Table 2 displays the mean accuracy score on the source monitoring task for the various groups of children subdivided by their age and their ***-tay*** score. The table shows how linguistic and non-linguistic performance with evidentiality grows with age; interestingly, for the youngest age group (young 3-year-olds), performance in the source monitoring tasks is much higher than in the linguistic production task (71.4% vs. 43.8% of correct responses).

4.3. Discussion

Experiment 3 yielded two particularly noteworthy findings. First, in contrast to Experiments 1 and 2, in which we found little evidence that young Korean children understand the meaning of evidential morphemes, we saw quite widespread productive use of such morphemes in the speech of young children. Specifically, we found quite accurate usage of the direct evidence morpheme ***-e*** from the age of three years. Interestingly, we also observed quite productive use of the hearsay morpheme ***-tay*** among children between the ages of four and five in the particular conversational setting we provided. Although 3-year-olds were not as productive as older children in using the hearsay morpheme, the observed proportion of utterances marked with ***-tay*** (56.7%) was quite high, considering that the task did not provide exemplars for children to choose from and children had to come up with an appropriate SE morpheme spontaneously. These data confirm and empirically extend earlier reports of spontaneous production of evidentials in young Korean children (Choi, 1995). Furthermore, by eliciting production under controlled circumstances, the present experiment excludes the possibility that these early uses of evidentials are based on unanalyzed memorized verb phrases (as could be the case with at least some occurrences of evidentials in naturalistic speech; cf. Aksu-Koç, 1988; Papafragou, 1998):

the present results demonstrate truly creative uses of evidential morphology with a variety of different verbs.

A second important result is that linguistic competence with evidentiality seems to proceed hand in hand with the conceptual understanding of informational sources. Confirming data from Experiment 1, we show that Korean children between the ages of three and four become capable of monitoring evidential sources quite successfully – and at the same time, their production of linguistic evidentials (especially, the acquisition of the distinction between markers of direct evidence vs. hearsay) becomes more stable. This pattern of results, even though unexpected on the basis of the large asymmetries between linguistic comprehension and non-linguistic evidential reasoning in our prior experiments, nevertheless seems to correspond more faithfully to linguistic and cognitive developments around this time.

These novel findings in turn raise some crucial questions. Perhaps the most pressing question is what is responsible for the difference between the current production data and the comprehension data obtained in our earlier studies. Recall that, in one of our tasks (Experiment 1), children were asked to attribute an utterance marked by ***-e*** or ***-tay*** to a character who had direct (visual) or indirect (hearsay) access to the contents of a container. In our Truth-Value Judgment task (Experiment 2), children needed to reject a statement containing an inappropriate evidential suffix (i.e. a suffix which did not semantically correspond to the source of the proposition). Children were poor in both of these comprehension measures. The fact that children of the same ages (especially 4-year-olds) succeed in producing these evidential morphemes in their speech seems to reverse the usual pattern where comprehension precedes production of linguistic forms during language development. The production-comprehension asymmetry appears all the more puzzling given that linguistic evidentiality needs to be acquired from the input, hence to be understood by young children, before it can be appropriately produced.

Even though unexpected, we suggest that the production-comprehension asymmetry is by no means mysterious for at least two reasons. First, the comprehension and production tasks required children to take different perspectives. In the production tasks, children were asked to evaluate the situation from their own point of view. That is, it was the children themselves who had an experience or heard information that was to be reported to a third person. By comparison, the comprehension tasks always required children to consider the situation from another's point of view (i.e. evaluate a statement in terms of how well it corresponded to a puppet's informational access).

Second, as we have already pointed out, our battery of comprehension tasks might have more of a metalinguistic flavor than the ordinary demands of language comprehension. It may well be that Korean preschoolers in our experiments had some knowledge of evidential semantics but were not able to reason on the basis of this knowledge in order to offer appropriate linguistic judgments. A related possibility is that, in both of our comprehension experiments, children may have failed to see the importance of the evidential suffix in formulating their response and may instead have focused on the base proposition. This strategy would predict the observed pattern of responses: in Experiment 1, where participants had to choose

one of two characters who could have uttered a sentence on the basis of its evidential morphology, responses were random; in Experiment 2, where children had to correct a statement on the basis of the appropriateness of the evidential morphology, they overwhelmingly accepted the statement as long as the base proposition was true. In sum, these tasks may have collectively masked an implicit understanding of 'evidential stance' by Korean children by focusing on explicit understanding of linguistic evidentiality (cf. also Aksu-Koç, 1988).[21]

There are other production-comprehension asymmetries in language acquisition which may have similar explanations. For instance, in sentences such as *Big Bird is touching him*, children are willing to accept that the NP 'Big Bird' and the pronoun 'him' can be co-referential even though they rarely make such errors in production (de Villiers, Cahillane, & Altreuter, in press). To take another example, third person singular /s/ in English verb morphology is produced correctly from the third year, but children cannot use it as a cue in comprehension until much later (de Villiers & Johnson, 2005). It is possible that metalinguistic awareness is involved in explicitly judging, e.g. that a verb marked with third-person morphology needs to have a singular, rather than a plural, subject, especially when the subject itself does not offer transparent cues (see de Villiers & Johnson, 2005, for such a proposal).

A second question which emerges from the present findings is how exactly the acquisition of evidentiality is related to developments in evidential reasoning. Is it possible that the gradual acquisition of the grammatical encoding of evidential distinctions contributes to the development of language-independent abilities to reason about and evaluate evidential sources? This question cannot be answered by looking within the Korean child population, where linguistic and conceptual developments in evidentiality seem to be tightly interlocked. Therefore, in our last experiment, we turn to English-speaking children and test their ability to reason about and remember sources of information in themselves and others. If learning the grammatical encoding of evidentiality confers a cognitive advantage onto Korean learners, then they should be better in source monitoring tasks than their English-speaking peers.

5. Experiment 4

This experiment replicates our earlier source monitoring tasks with a population of English-speaking 3- and 4-year-olds. The data are then compared to Korean data from our prior non-linguistic experiments.

[21] The role of the implicit/explicit distinction has been repeatedly pointed out in different contexts in the developmental literature. For instance, Zabrucky and Ratner (1986) report that both sixth and third graders read incongruent sentences more slowly than congruent sentences, but third graders detect far fewer such inconsistencies in their verbal reports. In the area of theory of mind, we know that 3- to 4-year-olds correctly evaluate who they should believe when two sources of information conflict but may be unable to explicitly justify their preference (Robinson & Whitcombe, 2003; Whitcombe & Robinson, 2000). And in false belief tasks, 3-year-old children may correctly look to the location where a character thinks an object is after the object has been secretly moved; however, children give the wrong answer if asked where the character will look for the object (Clements & Perner, 1994; Garnham & Ruffman, 2001).

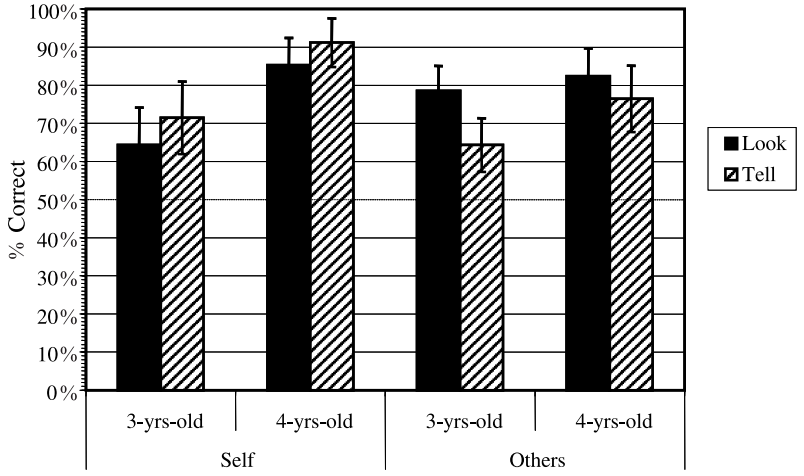

Fig. 7. English-speaking children's performance on source monitoring tasks (Experiment 4).

5.1. Method

5.1.1. Participants

Thirty-three English-speaking children were recruited from three preschools in the greater Boston area, and five English-speaking children were recruited from a preschool in Newark, Delaware. Like Korean children, these children were from upper middle class families. Each child was tested individually in a quiet room within the preschool. All children completed the study. A total of 21 3-year-olds (mean age: 3;5, ranging from 3;0 to 3;11) and 17 4-year-olds (mean age: 4;4, ranging from 4;0 to 4;11) English speaking children participated.

These children were compared to the Korean children in Experiments 1 and 3 of the same age range (3;0–4;11). These included 47 3-year-olds (mean age: 3;4, ranging from 3;0 to 3;11) and 46 4-year-olds (mean age: 4;5, ranging from 4;0 to 4;11).[22]

5.1.2. Stimuli and procedure

Children were presented with the source monitoring tasks used in Experiments 1 and 3. Half of the children in each age group started with the Self task followed by the Others task, and the other half of the children followed the reverse order.

5.2. Results

5.2.1. Source monitoring in English-speaking children

Beginning with separate within-language analyses of the English data (summarized in Fig. 7), we first conducted a 2 (Perspective: Self, Others) × 2 (Source: Look,

[22] When the age in days of the children was subjected to *t*-tests, the Korean- and English-speaking children's ages did not differ (3-year-olds: $t(66) = 1.12$, $p = .27$; 4-year-olds: $t(61) = 1.04$, $p = .30$).

Tell) × 2 (Age: 3, 4) ANOVA with Perspective and Source as within-subjects factors and Age as a between-subjects factor. (As before, we did not include the order of presenting the Perspective and Source questions as factors in this analysis because no order effects or interaction with the order terms were found in a preliminary analysis.) The results yielded a main effect of Age ($F(1, 36) = 4.60$, $p = .04$): 4-year-old children scored significantly better than 3-year-olds ($M = 84\%$ correct versus $M = 70\%$, respectively). There were no other main or interaction effects.

We also conducted two separate 2 (Perspective: Self, Others) × 2 (Source: Look, Tell) ANOVAs for each age group in the English data. The analysis for the 3-year-olds yielded no significant main or interaction effects. This analysis shows that the responses for the Self questions ($M = 68\%$) were not significantly different from the Others questions ($M = 72\%$). The analysis for the 4-year-olds, however, yielded a significant main effect of Perspective ($F(1, 20) = 4.2$, $p = .05$) and no other effects. Therefore, the Self questions were significantly easier ($M = 88\%$) than the Others questions ($M = 79\%$) for the 4-year-olds. Both 3-year-olds (Self: $t(20) = 3.42$, $p < .01$; Others: $t(20) = 3.87$, $p < .001$), and 4-year-olds (Self: $t(16) = 7.88$, $p < .001$; Others: $t(16) = 5.49$, $p < .001$) were above chance on both types of Perspective question.

5.2.2. Korean–English comparison in source monitoring tasks

Fig. 8 shows the average percentage of source monitoring questions answered correctly by each age and language group. As the figure suggests, performance improved with age and the two language groups did not differ from each other. To compare English and Korean speakers on their source monitoring statistically, we submitted the percentage of questions answered correctly into a grand 2 (Perspective: Self, Others) × 2 (Source: Look, Tell) × 2 (Age: 3, 4) × 2 (Language: English, Korean) ANOVA with language and age as the between-subjects factors. (We did not include

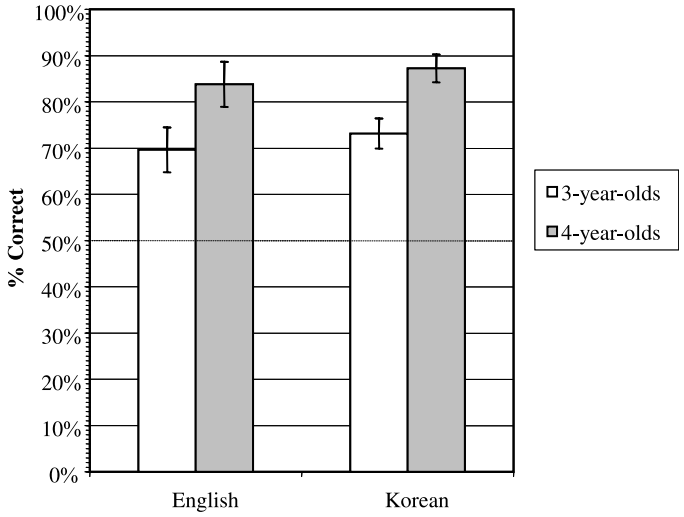

Fig. 8. Comparison of Korean- and English-speaking children's performance on source monitoring tasks (Experiments 1, 3, 4).

factors pertinent to the presentation order of the questions because preliminary analysis indicated that order contributed no main or interaction effects.) We found a main effect of Age ($F(1,125) = 11.71$, $p < .001$) with 3-year-olds performing worse than 4-year-olds (86% vs. 72% correct). Crucially, there was no main effect of Language, indicating that the English speakers performed just as well as the Korean speakers (77% vs. 81%, respectively). There was also no effect of Age × Language. We found a main effect of Perspective ($F(1,125) = 8.73$, $p < .01$), with the Self condition being easier than the Others condition (83% vs. 75% correct). We also found a marginal interaction of Perspective × Language ($F(1,125) = 4.03$, $p = .047$) and no other interaction. Post hoc pair-wise comparison with Bonferroni corrections for each language indicated that only the Korean children were contributing to the effect of Perspective (Self: 87% correct vs. Others: 74%, $F(1,125) = 21.06$, $p < .001$), while English speaking children were not (Self: 78% vs. Others: 75%, $p = .57$).[23]

5.3. Discussion

In this experiment, we looked at English-speaking children's source monitoring abilities and compared them to prior data from Korean-speaking children. Both linguistic populations yielded very similar patterns of results. Children in the Korean and English group performed equally well on identifying looking and being told as sources of information. Most importantly, 3-year-olds consistently had a more difficult time than 4-year-olds explicitly reporting the appropriate sensory experience that led to their knowledge. Our English results are broadly in accord with previous findings on source monitoring in young children, especially those findings that contrasted directly the informational role of looking vs. telling (O'Neill & Gopnik, 1991; Povinelli & de Blois, 1992; Wimmer et al., 1988; Woolley & Bruell, 1996; but cf. Gopnik & Graf, 1988; Whitcombe & Robinson, 2000).

Taken together, our data suggest that the improvement in Korean-speaking children's ability to monitor the sources of their beliefs follows the same timetable as in English-speaking children. This is important because it offers a novel perspective onto the correlation between non-linguistic source monitoring and production of evidential morphology observed in the Korean data of Experiment 3. More concretely, the comparison with English data suggests that language acquisition (specifically, the acquisition of evidentiality) cannot be the reason why Korean children become progressively better at thinking about information sources between the ages of three and four – English-speaking preschoolers are just as good at source monitoring even

[23] One may be tempted to analyze this interaction as evidence that Korean children gain earlier awareness of sources, perhaps as a result of the grammatical marking of evidentiality in their language. But as shown in Table 2, across age groups, Korean children are more advanced in their understanding of sources than in knowledge of linguistic evidentials (evidenced by their production of these morphemes). Especially within the group of 3-year-olds, we see clear evidence for development in the linguistic domain (i.e. increased uses of ***-tay***) while performance on non-linguistic source monitoring remains stable (and relatively high – over 70% of success rate). We take this as evidence that linguistic progress is not driving non-linguistic understanding of sources.

though English does not mark evidentiality grammatically. Rather what seems to underlie the developmental improvement in source monitoring skills in both child populations is cognitive (theory of mind) development which seems to proceed uniformly in children growing up in different linguistic communities.

Other related studies on Korean children's development of theory of mind, specifically their ability to reason about false beliefs, also converge on a similar conclusion. In a classic version of a false belief task, two characters, Sally and Anne, hide an object in a certain place. Anne then leaves the scene and Sally moves the object to a new location. Anne comes back and the child is asked where she is going to look for the object. Notice that, in this situation, Anne's knowledge state is incorrect and different from Sally's (and the child's) due to differential access to information. Therefore, passing the task requires being able to track informational access (e.g. Sally saw and Anne did not) and link informational access to knowledge state (i.e. Sally knows and Anne does not) in ways similar to our Others task. Korean-speaking children behave just like English-speaking children in such false belief reasoning scenarios: 3-year-olds typically fail and 4 year-olds typically pass (Wellman, Cross, & Watson, 2001).

6. General discussion

In this paper we examined the acquisition of linguistic evidentiality and its relation to children's source monitoring abilities in order to address two broad hypotheses about early linguistic and conceptual development. On the first of these hypotheses, the cross-linguistic timetable of language learning is predicted rather straightforwardly from the underlying complexity of the encoded concepts; on the second, the emergence and use of conceptual distinctions can be subject to language-specific effects. To test the first hypothesis, we investigated 3- and 4-year-old Korean learners' comprehension and production of evidential morphology and compared them to their (non-linguistic) ability to recognize and report different types of evidential sources. To test the second hypothesis, we compared Korean children's source monitoring abilities to those of English-speaking children of the same age in order to see whether learning a language with vs. without grammaticalized evidentiality has possible effects on source reasoning.

Our results point to two major descriptive conclusions. First, young Korean children's comprehension of evidentiality is not very stable. However, even 3-year-olds have productive command of morphemes encoding both direct access and hearsay; furthermore, the acquisition of these linguistic markers of evidentiality proceeds hand in hand with children's developing knowledge about sources of information.

Second, Korean and English learners behave identically in non-linguistic source reasoning tasks. Specifically, 4-year-olds in both groups can remember and report the origins of their beliefs (seeing or telling). Children of this age are also able to recognize that visual access or verbal report can create knowledge for others: they understand that a character who has looked inside a container or has been verbally informed will know its contents, while another character who hasn't had visual or verbal information won't know. Younger children, although poorer in these source

reasoning tasks, are still above chance in their performance. These results are broadly in accord with previous findings on source monitoring in young children, especially those which contrasted directly the informational role of seeing vs. telling, and point to a cross-linguistically consistent timetable for the acquisition of evidential distinctions (O'Neill & Gopnik, 1991; Povinelli & de Blois, 1992; Whitcombe & Robinson, 2000; Wimmer et al., 1988; Woolley & Bruell, 1996).

Taken together, these results have a number of specific implications for both the acquisition of linguistic evidentiality and the relation between the linguistic marking of evidentiality and non-linguistic source reasoning. We take up these issues next.

6.1. The acquisition of evidentiality

One of the most striking findings of the present studies is the fact that Korean 3- and especially 4-year-olds successfully produce appropriate direct evidence or hearsay evidentials. Taken together with other cross-linguistic data (Aksu-Koç, 1988; Choi, 1995; see also Matsui, Yamamoto, & McCagg, 2004 on Japanese), these results strongly argue against the assumption that young children are not able to express certain mentalistic concepts verbally because they cannot fully entertain these concepts (an issue we return to later in the discussion).

Putting aside the issue of potential conceptual difficulty, these early successes are remarkable given the many challenges evidentiality poses for the young language learner. Evidential relations are abstract and unobservable; they cannot be learned through ostension and are even less directly connected to circumstances in the world (or event structure) than aspect or tense distinctions. Clues to the meanings of evidentials are buried in speakers' minds; little in the extra-linguistic environment can guide language learners towards the meanings of *allegedly* or *apparently*. To the extent that the discovery of the meaning of a linguistic stimulus requires extracting commonalities across learning environments in which the stimulus occurs, evidential markers are particularly challenging: their meanings are related to very subtle, and often non-existent, cross-situational concomitants. As a result, even if learners are sophisticated enough in their source reasoning, they face considerable *mapping* problems in trying to pair bits of linguistic strings in their input to specific source concepts.[24]

[24] Such mapping problems have been originally pointed out and discussed in the context of learning procedures for mental verbs (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Papafragou, Cassidy, Gleitman, & Hulbert, 2004; Snedeker & Gleitman, 2004). As these authors point out, observational cues to the meaning of mental verbs such as *think* or *believe* are few and impoverished compared to those for action verbs such as *catch* or *eat*. On at least some occasions, a child might hear the verb *eat* while an eating event is unfolding or about to begin and therefore infer that the two are related (Brown, 1957; Tomasello & Kruger, 1986). However, it is unlikely that a child will find observational information just as useful in linking *think* to occurrences of thinking. Since observational cues are among the very first tools the child can bring to bear on the language learning task, this approach predicts that the first words to be learned will have 'concrete' or observable referents (objects or events); words with more 'abstract' denotations, such as mental verbs, are expected to arise later, once linguistic support for their acquisition has been mastered by learners. Linguistic evidentials raise many of the same mapping challenges for learners (albeit in the domain of closed-class items).

In the specific case of Korean, evidential morphemes are verbal suffixes but take scope over the whole proposition (relating that proposition to its evidentiary sources). Their acquisition presupposes the ability to correctly parse the sentence, identify the morphemes and their distributional properties (e.g. verb-final placement), and perform a mapping between their syntactic (sentence-final) and semantic (high-scope, broadly modal) features.

How do Korean learners overcome these obstacles and manage to exhibit productive use of evidential morphology already before the age of five? One possibility is that hypotheses about the meaning of evidentials are constrained by the fact that grammaticalized evidentiality draws on a restricted inventory of evidential distinctions cross-linguistically. As mentioned in Section 1, these distinctions include direct (sensory) and indirect (inference or hearsay) access, plus elaborations and combinations of those; more specific sources of information (e.g. divine intervention, gossip, etc.) are not grammaticalized despite the fact that they can be salient in the linguistic and cultural environment of the speakers. To the extent that young learners' hypotheses are constrained by what can serve as a reasonable basis for a grammaticalized category, they should be able to home in on the meanings of evidential morphemes in a relatively consistent manner.

Once the meanings of evidential morphemes are in place, children can start exploiting evidential scales (i.e. the ranking of information sources in terms of reliability) for pragmatic effects. The ability to rank evidential sources may be within the capacities of 3- and 4-year-olds (Whitcombe & Robinson, 2000), but the derivation of pragmatic effects from the use of different evidential markers is more complex and requires subtle reasoning about the communicative intentions of the speaker. There is independent evidence that pragmatic effects from linguistic scales may be difficult for preschool children (Noveck, 2001; Papafragou & Musolino, 2003). If so, we would expect difficulties with the pragmatic effects of evidential scales to extend well beyond the acquisition of the lexical/grammatical meaning of specific evidentials.

Also at later stages, we expect children to acquire the ability to draw inferences from the use of evidential morphology and make linguistic judgments. Adults are able to attribute an evidentially marked statement to a character based on the character's access to the event (cf. Experiment 1) and reject statements which contain inappropriate evidential suffixes (e.g. suffixes which do not match the evidential relation between the speaker and a proposition; cf. Experiment 2). It would be interesting to use our methods to discover when Korean-learning children acquire these abilities (evidence from Turkish suggests that these emerge only after the fifth year; Aksu-Koç, 1988).

6.2. The cognitive roots of evidentiality

As mentioned in Section 1, much of the literature on language development assumes that the learning difficulty posed by evidential and other mental terms is mostly conceptual: since mental terms denote abstract and complex concepts, their acquisition has to await the relevant cognitive advances in the growing child. Our

results from Korean point to a different picture: 4-year-old and – to a certain degree – 3-year-old Korean children can reason successfully about sources of knowledge in others and can remember the sources of their own beliefs. Hence young Korean children are not cognitively limited to contemplating only the concrete and the 'objective' but can entertain abstract and fleeing notions of evidential sources and their causal relation to knowledge. Together with our production data, these results offer convergent evidence that young children can handle source concepts early on.

On a methodological level, our findings underscore the usefulness of non-linguistic tasks as an independent metric of cognitive development when studying language acquisition and its cognitive prerequisites. Consider as a concrete example the acquisition of hearsay morphemes. Recall that in Turkish, according to Aksu-Koç and Slobin (1986), the hearsay interpretation of the 'indirect' evidential ***-mI§*** is acquired later than its inferential interpretation and, in any case, is not available in 3-year-old children. The authors tentatively attribute this ordering to the cognitive demands of hearsay interpretations, which presuppose the integration of someone else's utterance as a basis for one's assertion. They conclude: ''In any case, the hearsay function is not primary in development'' (ibid., p. 166). Our own linguistic comprehension data show that 3- and 4-year-old Korean children also do not understand the meaning of the Korean hearsay morpheme in comprehension tasks. However, the reason for the delay cannot be conceptual, since the same children can both report communication as a knowledge source for themselves and link communication to knowledge in others in tasks which do not involve evidential morphology.

In fact, a closer comparison between linguistic comprehension and non-linguistic conceptual tasks reveals that linguistic tasks may place heavier cognitive demands on participants than formally very similar source monitoring tasks. Recall that one of our source monitoring tasks (the 'Others' task) involved attributing knowledge to one of two characters who had access to sensory or communicated information but not to another character who lacked such access. A second task (the 'Self' task) required children to report how they came to know something by choosing perception ('Did you see it?') or communication ('Did I tell you?') as the appropriate source. Both 3- and 4-year-olds seemed to be very good at these kinds of task (even though there was considerable improvement between the two age groups). In one of our linguistic comprehension tasks (Experiment 1), children were asked to attribute an utterance marked by ***-e*** or ***-tay*** to a character who had direct (visual) or indirect (hearsay) access to the contents of a container. In our Truth-Value Judgment task (Experiment 2), children needed to reject a statement containing an inappropriate evidential suffix (i.e. a suffix encoding hearsay in case the speaker had directly witnessed an event). Children were poor in both of these linguistic tasks. As already discussed, the comprehension tasks have a metalinguistic flavor and may require more sophisticated understanding of the underlying semantic distinctions than comprehension of evidential morphology during ordinary semantic processing. It remains an interesting fact, however, that the attribution of knowledge on the basis of information access (a metacognitive task) should be easier than the attribution of an utterance encoding information access to the appropriate speaker (a metacommunicative task).

6.3. Language-on-thought effects?

We finally turn to what we consider as a particularly interesting aspect of our findings. Recall that some authors have expressed the expectation that the grammatical marking of evidentiality may make evidential distinctions more salient in the thought processes of the speakers (e.g. Whorf, 1956). Other authors have left open the possibility that children learning languages with different evidential systems might diverge in their source reasoning (Aksu-Koç, 1988; Choi, 1995). The present paper adopted a comparative approach focusing on Korean, a language with grammaticalized and frequently used (Choi, 1995) evidentials, and English, a language where the expression of evidentiality is mostly lexical and thus variable. If the presence of grammaticalized evidentiality can encourage sensitivity to informational source, one might expect that young Korean learners would have an advantage in source reasoning tasks over their English-learning peers. As the comparison between the English and Korean non-linguistic data shows, however, children's ability to reason about sources of information is not affected by whether such sources are grammaticalized in the language children are exposed to. To the contrary, the development of non-linguistic evidential reasoning (at least with respect to the two sources we have focused on) seems to proceed uniformly in distinct language-learning populations.

To be sure, the absence of language-driven effects in our Korean and English data does not necessarily preclude the existence of cognitive asymmetries between speakers of other languages with different evidential systems. After all, even though English and Korean differ in their formal means of marking evidentiality, they both possess the resources to refer to sources of information and a not-so-awkward translation is possible between utterances of the two languages. It could be claimed that given a pair of sufficiently distinct languages, one would be more likely to uncover linguistic effects on source reasoning (e.g. memory for sources or estimation of source reliability). In this sense, it would be illuminating to compare English to a language with a more elaborate and subtle evidential paradigm such as Tuyuka or Quechua (cf. Section 1), or pursue a comparison between languages with obligatory vs. optional encoding of evidentiality. We are currently pursuing further cross-linguistic studies of linguistic and non-linguistic aspects of evidentiality with a diverse array of tasks and age groups (cf. Ozturk & Papafragou, in preparation).

The present data bear on a stronger hypothesis relating language and source monitoring abilities. One might argue that evidential concepts themselves are formulated 'under linguistic guidance' – specifically during the process of figuring out semantics for novel expressions during language learning (as it has been proposed, e.g. for the domain of space: Bowerman & Choi, 2003). The present findings cast doubt on such a potential link, demonstrating instead the independence of language from source monitoring abilities. In fact, a closer inspection of evidential systems cross-linguistically offers good reasons to assume that the fundamentals of source reasoning precede (and structure) language acquisition. Recall that evidential paradigms draw on a small set of fundamental semantic distinctions between types of information access (direct/sensory access vs. indirect access, where the latter is further subdivided

into hearsay and inference); furthermore, individual languages combine and refine these distinctions in relatively predictable ways (cf. Section 1). As noted by Bloom (2000), the view that languages build on non-linguistic primitives is plausible only if one does not have to posit a new set of non-linguistic basic concepts for every language we look at: "the variation that exists should be highly constrained" (p. 402). The fact that evidential systems cross-linguistically converge on the same narrow set of distinctions points to the conceptual basicness of these distinctions prior to the emergence of language.

Independently of these linguistic facts, empirical evidence from two distinct sources supports the conclusion that source cognition is organized independently of linguistic influence. First, source reasoning, together with broader mindreading abilities, is selectively impaired in autistic individuals. Autistic children, for instance, have trouble recognizing that seeing leads to knowing (Leslie & Frith, 1988; Hogrefe, Wimmer, & Perner, 1986; Perner, Frith, Leslie, & Leekham, 1989).[25] This breakdown pattern in theory-of-mind capacities has been taken as evidence for a modular, probably innate mechanism responsible for attributing mental states to others and reasoning about them. For present purposes, the same facts can be interpreted as demonstrating the internal organization of a conceptual system independently of external (e.g. linguistic) influences.

Second, recent experimental results suggest that human source monitoring abilities may be shared in part with other primates. According to Hare, Call, and Tomasello (2001), chimpanzees seem to understand some things about what others do and do not see (and hence know or do not know) by monitoring the gaze direction of conspecifics in competitive situations. Their observations revise earlier findings which suggested that chimpanzees may not be able to recognize the causal link between visual access and knowledge (e.g. Povinelli, Nelson, & Boysen, 1990) and open the way to recognizing some causal reasoning about mental states in non-humans (for discussion, see Povinelli & Vonk, 2003; Tomasello, Call, & Hare, 2003). To the extent that source reasoning can emerge in non-linguistic species, the claim that source reasoning is introduced by linguistic elements becomes much less plausible.

Summarizing, these observations, together with our overall findings, can be given a unified interpretation if we assume that source monitoring abilities in humans emerge independently of, and prior to, aspects of linguistic systems encoding source information. In short, linguistic evidentiality is not, and could not, be a pacesetter for cognition in the strict sense.[26] Rather, human beings are cognitively predisposed to monitor the origins of information they come to possess and to compare and

[25] Interestingly, it seems that causally linking seeing to knowing is a slightly easier task than false belief attribution, since there are autistic individuals who can succeed in the former but not the latter (Hogrefe et al., 1986).

[26] Even though the claim is frequently made, it is not at all clear how language learning could introduce truly novel concepts. For instance, how could the acquisition of evidential morphology produce the conceptual resources to think about informational access if these resources were not antecedently available to the child? (cf. Fodor, 1975, and for discussion Gleitman & Papafragou, 2005).

evaluate such sources (even though the development of these abilities takes considerable time and world knowledge; cf. Section 1). Individual languages select from a basic inventory of source distinctions when organizing their evidential paradigms. Children learning languages which lexicalize or grammaticalize sources of information (or inferentially suggest them, even when not explicitly encoding them) draw from and are guided by this set of initial distinctions. Finally, differences in the linguistic encoding of evidentiality do not materially affect young learners' reasoning about informational access.

This picture coheres with a rather traditional, distinctly universalist view of the relationship between language and thought, according to which language acquisition requires and builds on prior – quite plausibly universal – conceptual distinctions. Children learning individual languages face the problem of selecting from among the concepts already available to them the one that the adult speaker could have in mind and is talking about (cf. Chomsky, 2000; Gleitman, 1990; Pinker, 1984; Slobin, 1973). We find it unlikely that abstract aspects of language (such as the semantic distinctions underlying evidentiality) could be learned otherwise.

Appendix A. Materials for Experiment 2

Event description [Type]	Tim's report about the event
Tim watches as Jin eats an apple [**Looking**]	ecey cininun sakwalul mekesse yesterday Jin-top apple-acc eat-pst-e Yesterday, Jin ate an apple-***e***
Tim watches as Jin takes pictures [**Looking**]	ecey cininun sacinul ccikesse yesterday Jin-top photo-acc take-pst-e Yesterday, Jin took picture(s)-***e***
Tim watches as Jin sings a song [**Looking**]	ecey cininun nolaylul pwullesstay yesterday Jin-top song-acc sing-pst-tay Yesterday, Jin sang a song-***tay***
Tim watches as Jin puts on a hat [**Looking**]	ecey cininun mocalul ssesstay yesterday Jin-top hat-acc put on-pst-tay Yesterday, Jin put on a hat-***tay***
Jin tells Tim that she kicked a puppy the day before [**Telling**]	ecey cininun kangacilul pallochasstay yesterday Jin-top puppy-acc kick-pst-tay Yesterday, Jin kicked (a) puppy-***tay***

Appendix A (*continued*)

Event description [Type]	Tim's report about the event
Jin tells Tim that she had an ice cream the day before [**Telling**]	ecey cininun aisukhulimul mekesstay yesterday Jin-top ice cream-acc eat-pst-tay Yesterday, Jin ate ice cream-***tay***
Jin tells Tim that she talked with grandmother over the phone the day before [**Telling**]	ecey cininun halmenilang cenhwahaysse yesterday Jin-top granny-and phone-pst-e Yesterday, Jin called grandmother-***e***
Jin tells Tim that she went to the market with her Mom the day before [**Telling**]	ecey cininun emmalang sicangkasse yesterday Jin-top mom-and go-to-market-pst-e Yesterday, Jin went to market with Mom-***e***

Note: nom, nominative; acc, accusative; top, topic; loc, locative; pst, past tense; comp, complementizer; rel, relative.

References

Aikhenvald, A., & Dixon, R. (Eds.). (2001). *Studies in evidentiality*. Amsterdam: Benjamin.

Aksu-Koç, A., & Slobin, D. (1986). A psychological account of the development and use of evidentials in Turkish. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 159–167). Norwood, NJ: Ablex.

Aksu-Koç, A., & Alici, D. (2000). Understanding sources of beliefs and marking of uncertainty. In E. Clark (Ed.), *Proceedings of the 30th Annual Child Language Research Forum* (pp. 123–130). Stanford: CSLI.

Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge: Cambridge University Press.

Aksu-Koç, A. (2000). Some aspects of the acquisition of evidentials in Turkish. In L. Johanson & B. Utas (Eds.). *Evidentials: Turkic, Iranian and neighbouring languages* (pp. 15–28). Berlin: de Gruyter.

Anderson, L. (1986). Evidentials, paths of change and mental maps: typologically regular asymmetries. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 273–312). Norwood, NJ: Ablex.

Bascelli, E., & Barbieri, M. (2002). Italian children's understanding of the epistemic and deontic modal verbs *dovere* (must) and *potere* (may). *Journal of Child Language, 29*, 87–107.

Barnes, J. (1984). Evidentials in the Tuyuka verb. *International Journal of American Linguistics, 50*, 255–271.

Bartsch, R., & Wellman, H. (1995). *Children talk about the mind*. Oxford: Oxford University Press.

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Bowerman, M., & Choi, S. (2003). Space under construction: language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 389–427). Cambridge, MA: MIT Press.

Bowerman, M., & Levinson, S. C (2001). Introduction. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.

Brown, R. (1957). Linguistic determinism and the parts of speech. *Journal of Abnormal and Social Psychology, 55*, 1–5.

Chafe, W., & Nichols, J. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.

Choi, S. (1995). The development of epistemic sentence-ending modal forms and functions in Korean children. In J. Bybee & S. Fleischman (Eds.), *Modality in grammar and discourse* (pp. 165–204). Amsterdam: Benjamin.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

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

Clements, W., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.

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

Dancy, J. (1985). *An introduction to contemporary epistemology*. Oxford: Blackwell.

Delancey, S. (2002). The mirative and evidentiality. *Journal of Pragmatics, 33*, 369–382.

de Villiers, J., Cahillane, J., & Altreuter, E. (in press). What can production reveal about Principle B? *Proceedings of Generative Approaches to Language Acquisition in North America* (*GALANA*) 1.

de Villiers, J. & Johnson, V. (2005). The information in third-person /s/: Acquisition across dialects of American English. Submitted for publication.

De Haan, F. (1998). The category of evidentiality. Unpublished ms., University of New Mexico.

De Haan, F. (2001). The place of inference within the evidential system. *International Journal of American Linguistics, 67*, 193–219.

Faller, M. (2001). Remarks on evidential hierarchies. In D. Beaver, S. Kaufmann, B. Clark, & L. Casillas (Eds.), *Proceedings of the 'Semfest'* (pp. 37–59). Stanford: CSLI Publications.

Faller, M. (2002). Semantics and pragmatics of evidentials in Cuzco Quechua. Stanford University doctoral diss.

Fodor, J. (1975). *The language of thought*. New York: Crowell.

Garnham, W., & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding of belief?. *Developmental Science 4*, 94–100.

Garrett, E. (2000). Evidentiality and assertion in Tibetan. UCLA doctoral diss.

Gentner, D., & Boroditksy, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge: Cambridge University Press.

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

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

Givón, T. (1982). Evidentiality and epistemic space. *Studies in Language, 6*(1), 23–49.

Gleitman, L. (1990). The structural sources of verb meaning. *Language Acquisition, 1*, 1–55.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & R. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 633–661). Cambridge: Cambridge University Press.

Gopnik, A., & Graf, P. (1988). Knowing how you know: young children's ability to identify and remember the sources of their beliefs. *Child Development, 59*, 1366–1371.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139–151.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: what it is, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hirst, W., & Weil, J. (1982). Acquisition of epistemic and deontic meanings of modals. *Journal of Child Language, 9*, 659–666.

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: a developmental lag in attribution of epistemic states. *Child Development, 57*, 567–582.

Horn, L. (1972). *On the semantic properties of logical operators in English*. Bloomington, IN: IULC.

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: evidence from verb meanings. *Psychological Review, 90*, 72–93.

Ifantidou, E. (2001). *Evidentials and relevance*. Amsterdam: Benjamin.

Izvorski, R. (1998). The present perfect as an epistemic modal. In A. Lawson & E. Cho (Eds.), *Proceedings of SALT 7*. Cornell University: CLC Publications.

Johanson, L., & Utas, B. (Eds.). (2000). *Evidentials: Turkic, Iranian and neighbouring languages*. Berlin: de Gruyter.

Kratzer, A. (1991). Modality. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research* (pp. 639–650). Berlin: de Gruyter.

Lee, C. (1997). The acquisition of mood and modals in Korean. In *Linguistic theory and the acquisition of Korean Semantics and Syntax I* (pp. 73–101). Seoul, Korea: Minum Publisher.

Lee, T., & Law, A. (2000). Evidential final particles in child Cantonese. In E. Clark (Ed.), *Proceedings of the 30th Annual Child Language Research Forum* (pp. 131–138). Stanford: CSLI.

Leslie, A., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology, 6*, 315–324.

Matsui, T., Yamamoto, T., & McCagg, P. (2004). Who can you trust? A closer look at preschoolers' developing sensitivity to epistemic expressions. Talk delivered at the 29th Annual Boston University Conference on Language Development, Boston, November 5–7.

Mayer, R. (1990). Abstraction, context and perspectivization – evidentials in Discourse Semantics. *Theoretical Linguistics, 16*, 101–163.

Mitchell, P., Robinson, E., Nye, R., & Isaacs, J. (1996). Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition, 59*, 1–21.

Moore, C., & Furrow, D. (1991). The development of the language of belief: the expression of relative certainty. In D. Frye & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding* (pp. 173–193). Hillsdale, NJ: Erlbaum.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*, 167–171.

Mossler, D., Marvin, R., & Greenberg, M. (1976). Conceptual perspective taking in 2- to 6-year-old children. *Developmental Psychology, 12*, 85–86.

Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling*. Amsterdam: Benjamin.

Noveck, I. (2001). Why children are more logical than adults: an experimental investigation of scalar implicature. *Cognition, 78*, 165–188.

O'Neill, D., & Astington, J. (1990). Preschooolers' developing understanding of the role sensory experiences play in knowledge acquisition. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April.

O'Neill, D., Astington, J., & Flavell, J. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development, 63*, 474–490.

O'Neill, D., & Chong, S. (2001). Preschool children's difficulty understanding the types of information obtained through the five senses. *Child Development, 72*, 803–815.

O'Neill, D., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology, 27*, 390–397.

Oswalt, R. (1986). The evidential system of Kashaya. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic encoding of epistemology* (pp. 29–45). Norwood, NJ: Ablex.

Ozturk, O., & Papafragou, A. (in preparation). The acquisition of evidentiality in Turkish.

Palmer, F. (1986). *Mood and modality*. Cambridge: Cambridge University Press.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*, 153–182.

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

Papafragou, A. (2000). *Modality: Issues at the semantics-pragmatics interface*. Oxford: Elsevier.

Papafragou, A., Cassidy, K., Gleitman, L., & Hulbert, J. (2004). How children acquire mental verbs. Talk delivered at the 29th Annual Boston University Conference on Language Development, Boston, November 5–7.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language, 74*, 557–589.

Perner, J., Frith, U., Leslie, A., & Leekham, S. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Development, 60*, 689–700.

Pillow, B. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116–129.

Pillow, B. (1993). Preschool children's understanding of the relationship between modality of perceptual access and knowledge of perceptual properties. *British Journal of Developmental Psychology, 11*, 371–389.

Pillow, B. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development, 73*(3), 779–792.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Povinelli, D., & de Blois, S. (1992). Young children's (*Homo sapiens*) understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology, 106*(3), 228–238.

Povinelli, D., Nelson, K., & Boysen, S. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 108*, 74–80.

Povinelli, D., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Science 7*(4), 157–160.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61*, 973–982.

Robinson, E., Champion, H., & Mitchell, P. (1998). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology, 35*, 535–546.

Robinson, E., Thomas, G., Parton, A., & Nye, R. (1997). Children's overestimation of the knowledge to be gained from seeing. *British Journal of Developmental Psychology, 15*, 257–273.

Robinson, E., & Whitcombe, E. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development, 74*, 48–62.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 175–178). New York: Holt, Rinehart and Winston.

Smiley, P., & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects and persons. In M. Tomasello & W. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 21–61). Hillsdale, NJ: Erlbaum.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In G. Hall & S. Waxman (Eds.), *Weaving a lexicon*. Cambridge, MA: MIT Press.

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*, 424–433.

Speas, M. (2004). Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua, 114*, 255–276.

Spelke, E., & Tsivkin, S. (2001). Initial knowledge and conceptual change: space and number. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 70–100). Cambridge: Cambridge University Press.

Suh, J. (2000). *Modern Korean grammar*. Seoul, Korea: Hanyang University Press.

Taylor, M. (1988). The development of children's understanding of the seeing-knowing distinction. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 207–225). New York: Cambridge University Press.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Science, 7*(4), 153–156.

Tomasello, M., & Kruger, A. (1986). Joint attention on action: acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language, 19*, 311–333.

Urmson, O. (1963). Parenthetical verbs. In C. Caton (Ed.), *Philosophy and ordinary language* (pp. 220–246). Urbana: University of Illinois Press.

Weber, D. (1986). Information perspective, profile, and patterns in Quechua. In W. Chafe & J. Nichold (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 137–155). Norwood, NJ: Ablex.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of Theory-of-Mind development: the truth about false belief. *Child Development, 72*(3), 655–684.

Whitcombe, E., & Robinson, E. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development, 15*, 329–346.

Whorf, B. L. (1956). In J. Carroll (Ed.), *Language thought and reality*. Cambridge, MA: MIT Press.

Willett, T. (1988). A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language, 12*, 51–97.

Wimmer, H., Hogrefe, G.-J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development, 59*, 386–396.

Wimmer, H., Hogrefe, G.-J., & Sodian, B. (1988). A second stage in children's conception of mental life: understanding informational access as origins of knowledge and belief. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 173–192). Cambridge: Cambridge University Press.

Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology, 32*, 335–346.

Zabrucky, K., & Ratner, H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development, 57*, 1401–1418.