Functions of Evidentials in Turkish Child and Child-Directed Speech in Early Child-Caregiver Interactions

Berna A. Uzundag, Süleyman S. Tasci, Aylin C. Küntay, and Ayhan Aksu-Koç

About one quarter of languages of the world have grammaticized evidential markers to denote the type of evidence a speaker has for the information she asserts (Aikhenvald, 2004). In evidential languages, utterances express mode of access to information along with the informational content conveyed (Aksu-Koç, 2009).

Here we report a corpus analysis of the use of evidentials by Turkish-speaking caregivers and their young children on the cusp of language acquisition. We first explain how evidentiality is marked in Turkish and then present the current findings about the acquisition of evidentials.

1. Evidentiality in Turkish

In Turkish, evidentiality is grammaticized through multifunctional tense-aspect-modality suffixes on verbs and on non-verbal predicates. The basic distinction is made between direct and indirect experience, which are marked linguistically with the neutral perfective aspect/past tense marker –DI and the perfect aspect/evidential marker –mIş respectively.

Direct experience corresponds to situations where the speaker had experiential access to the asserted event. On the other hand, when knowledge is gained indirectly, such as (a) from the verbal report of another person (hearsay: Ali git-miş ‘(I heard that) Ali has left’, or (b) by inferring a process from the present physical evidence (inference: Ali git-miş ‘(I infer that) Ali has left’ (from the absence of his coat and bag)), the indirect experience marker is used. Besides the inference and hearsay functions that indicate a ‘non-witnessed process’, the –mIş inflection also has a narrative function indicating nonfactual events such as in folktales, myths, dreams, and jokes. Akin to this function, –mIş is used in the context of pretend play. It can also convey surprise, irony, or compliment, or more generally, new information (Aksu-Koç & Slobin, 1986).

* Berna A. Uzundag, Koç University, buzundag13@ku.edu.tr. Süleyman S. Tasci, Koç University. Aylin C. Küntay, Koç University & Utrecht University. Ayhan Aksu-Koç, Boğaziçi University.

2. Acquisition of Turkish Evidentials

Previous research via semi-naturalistic observations has indicated that Turkish-speaking children acquire the use of evidential forms between 1;6 and 3;0 years (Aksu-Koç, 1988). The direct experience marker –DI past is the first verb inflection that is encountered in children’s speech, emerging between 1;6 and 2;0 years. At first, it is used to comment on immediately preceding events. After a few months of using –DI, children start using the indirect experience marker –mIş, which is first used to comment on the existing state of objects in a joint attentional context with the caregiver, as in burda-y-miş, ‘(I see/realize) it is here’. Describing pictures and telling stories are also among the common situations where –mIş is used by the child. Children then start expressing the inferential function in the first half of their third year to talk about changes of state, where the process has not been witnessed. Finally the usage of the hearsay function emerges between 2 and 3 years, where children report the information conveyed verbally by others.

Semi-naturalistic data point to a much earlier successful usage of –mIş in children’s speech in comparison to experimental comprehension and production studies. In a production study, Aksu-Koç (1988) had 3- to 6-year-old children watch events acted out with toys. The children were tested to see whether they used the direct experience marker –DI to report events they witnessed in all its phases, and the indirect experience marker –mIş to report the events they witnessed only the beginning and the end result of (i.e. using –mIş in an inferential sense). Children aged between 3;0 and 3;6 could provide –DI as the correct response nearly 90 percent of the time when the event phases were entirely experienced, whereas correct usage of the indirect experience form –mIş with its inferential function reached a level of 70 percent between 3;6 and 4;0 years. The hearsay function was assessed with a cognitively demanding role-play task using dolls, where the speaker doll was acted out by the experimenter and the reporter doll by the child. Children displayed correct use of the hearsay function between 4 and 4;6 years. Although these data show the same acquisitional order of usage of the evidential functions in developmental time as observed in semi-naturalistic recordings, successful performance is attenuated by experimental task demands.

Aksu-Koç (1988) also conducted a comprehension study with picture-stories, where children were required to identify the speakers of utterances containing –DI or –mIş based on their knowledge of whether the speaker gained the information through a direct or an indirect perspective. Successful performance on the comprehension task was achieved at a later age than production with 50% correct use of –mIş by 5 years, probably indicating that children found it more difficult to provide an answer from another speaker’s perspective. This comprehension-production asymmetry, i.e. the fact that performance on the comprehension tasks lags behind performance on the production tasks, was also observed in other studies with children aged between 3 and 6 years (Ünal & Papafragou, 2016), and between 5 and 7 years (Ozturk &
Papafragou, 2008; 2015). In the production experiment of Ozturk and Papafragou, children in the youngest age group (64-72 mo) could produce –DI with no difficulty, but children in the oldest age group (85-96 mo) still made errors when they were expected to express hearsay and inference functions via using –mlş. Six-year-olds could only use the hearsay function, and seven-year-olds could make use of the inferential function only around chance level. These findings are in contrast with naturalistic observations of successful usage of inference before hearsay (Aksu-Koç, 1988). Ozturk and Papafragou (2015) suggested that either inference and hearsay functions posit different levels of complexity for the child, or inferential usage is less common than hearsay making it more difficult to use in required situations. To test the first suggestion, an experiment was conducted, where children either saw an event happening on the screen (See trial), or watched someone talk about something they did (Hear trial), or saw hints about what might have happened (e.g. a partially eaten cake) (Infer trial). After children reported what happened, they were asked about their source of information by asking them ‘How do you know? Did you see? Did you listen?’ (in the See vs. Hear condition) or ‘Did you see or did you see something and understand?’ (in the See vs. Infer condition). Children performed better in the See vs. Infer condition than in the See vs. Hear condition, not quite explaining the lower performance in the Infer trials in the production study. With the present corpus study, we had the possibility to directly assess Ozturk and Papafragou’s second hypothesis about the inferential usage being less frequent than the hearsay function.

Ozturk and Papafragou suggested that their results are in line with Wimmer, Hogrefe, and Perner’s (1988) approach about the development of source information. Namely, visual perception is the least complex form of source information, where inference is the most complex one with verbal information occupying a position in between. However, inference in their sense seems to be different than the inferential meaning that the –mlş particle conveys. Wimmer et al.’s understanding of inference depends on the findings of Sodian and Wimmer (1987), who tested children’s ability to make logical inferences in the name of another person. What the indirect experience marker in Turkish conveys is rather information that the speaker arrives at upon encountering an end state without witnessing the process that has led to it (e.g. bozul-muş, ‘(it) is broken’ upon seeing that a toy does not work). Ögel (2007; Ögel-Balaban, Aksu-Koç, & Alp, 2012) presented toys to children, and later showed the same toys with observable changes (e.g. getting wet, being broken). Children were expected to notice the changes and comment on them by using the indirect experience marker in its inferential function. In the production task assessing the usage of the hearsay function, children were told a chain of events by the experimenter, and asked to retell the events to a third party. It was found that children used the form for the inferential function more successfully than for the hearsay function.

As this brief literature review suggests, the ongoing debates regarding the comprehension and production of the indirect experience marker –mlş by Turkish children revolve around the following points of discussion: (1)
children’s performance on the experimental tasks lags behind their performance in their spontaneous speech, (2) children make more errors in comprehension than in production tasks, which might be related to methodological and/or conceptual factors, and (3) how source monitoring abilities (e.g. visual perception, linguistic report based on hearsay, inference) map onto successful usage of different functions of –mIş, and whether the usage/understanding of the inferential function precedes or follows the usage of the hearsay function.

3. Present Study

In this corpus study, we pursued the goals of (1) investigating the order of acquisition of different functions of the evidential marker –mIş by Turkish-speaking children by using longitudinally obtained corpus data relatively denser than used in earlier studies, (2) charting the distribution and frequency of different source-marking functions within child-directed speech and children’s speech, and (3) examining the relations between input and child’s developmental trajectory of –mIş utterances.

Owing to the exploratory nature of this study, we also addressed some further questions. Specifically, (4) we observed nonfactual functions of the –mIş particle that have either not been documented before or not received much attention. In addition, (5) we examined how evidential functions are employed in families with different socioeconomic status (SES).

4. Corpus and Methodology

We used the Koç University Longitudinal Language Development Database (Küntay, Koçbaş, & Taşçı, 2015; Ural, Yüret, Keteç, Koçbaş, & Küntay, 2009) that consists of longitudinally obtained video recordings of eight children between 8 and 36 months of age. Children were videotaped in their home environments while engaging in daily activities and interacting with caregivers. For each child, one-hour sessions were recorded twice a month. For the present study, we used data from six children since data collection for one of the children was terminated at 21 months of age, and the data for another child was not yet fully transcribed.

Table 1. Education and SES of children’s parents in the sample.

<table>
<thead>
<tr>
<th>Child</th>
<th>Highest Attained Education</th>
<th>Education in Years</th>
<th>SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Secondary School</td>
<td>8</td>
<td>low</td>
</tr>
<tr>
<td>C2</td>
<td>Secondary School</td>
<td>8</td>
<td>low</td>
</tr>
<tr>
<td>C3</td>
<td>Primary School</td>
<td>5</td>
<td>low</td>
</tr>
<tr>
<td>C4</td>
<td>University</td>
<td>15</td>
<td>high</td>
</tr>
<tr>
<td>C5</td>
<td>High School</td>
<td>11</td>
<td>high</td>
</tr>
<tr>
<td>C6</td>
<td>University</td>
<td>15</td>
<td>high</td>
</tr>
</tbody>
</table>
Three children came from families of low SES, with parents having attained less than or equal to 8 years of education. The other three children’s parents had high SES, with 11 or more years of education. The type and number of years of education as well as the SES information are summarized in Table 1.

5. Coding

We extracted the utterances that contained the indirect experience marker –mlṣ from the corpus. The instances in adult-to-adult and researcher-to-child speech were eliminated, leaving child’s speech and caregiver-to-child speech for analyses.

Table 2. Mean number of utterances per session and the percentage of utterances containing the evidential –mlṣ, for each child in CDS and CHI.

<table>
<thead>
<tr>
<th>Child</th>
<th># of sessions</th>
<th>Mean number of utterances per session</th>
<th>% of utterances containing the evidential –mlṣ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CDS</td>
<td>CHI</td>
</tr>
<tr>
<td>C1</td>
<td>57</td>
<td>114.3</td>
<td>176.5</td>
</tr>
<tr>
<td>C2</td>
<td>56</td>
<td>455.4</td>
<td>317.1</td>
</tr>
<tr>
<td>C3</td>
<td>41.75</td>
<td>379.8</td>
<td>338.7</td>
</tr>
<tr>
<td>C4</td>
<td>40.5</td>
<td>572.7</td>
<td>360.8</td>
</tr>
<tr>
<td>C5</td>
<td>51</td>
<td>386.9</td>
<td>264.1</td>
</tr>
<tr>
<td>C6</td>
<td>46</td>
<td>493.7</td>
<td>295.4</td>
</tr>
</tbody>
</table>

* Utterances where –mlṣ had an aspecital or participial function are not included.
† A thirty-minute recording was counted as a 0.5 transcription.

The corpus sample contained 83,580 child and 113,301 child-directed utterances in total with 4,759 child-directed (CDS) and 1,823 child (CHI) utterances containing the evidential –mlṣ. Table 2 shows the mean number of utterances for CDS and CHI for each child and the percentage of utterances containing the evidential –mlṣ.

Each utterance containing –mlṣ was coded in terms of (a) grammatical function, (b) source of information, and (c) pragmatic function, explained in more detail in the following subsections. The coding was conducted by the first and second authors, where each coded half of the instances and checked the coding of the other half. The third and the fourth authors were consulted for any disagreements. Figure 1 shows a summary of the coding scheme.

---

1 An R script was written to obtain the utterances from the sample that contained the allomorphs of –mlṣ (miş, miş, muş, müş), lengthier pronunciations (miş, miş, muş, müş), child-like pronunciations (miş, miş, muş, müş), and pronunciations in different accents (ms, mis, mus, mıs).
In addition to the evidential usage of –mIş, we annotated the grammatical contexts where –mIş does not express evidential modality. One such context is aspectual where –mIş precedes another tense-aspect-modality marker.

(1) *git–miş-ti*
    go-PERF-PAST
    (s)he had gone

In addition, there is a participle function of –mIş, where it is devoid of its evidential meaning (Aksu-Koç & Slobin, 1986; Göksel & Kerslake, 2005; Gürer, 2014). We removed these kinds of –mIş utterances from our data since they do not encode an evidential meaning.

(2) *kuru-muş çiçek*
    dry-PTPL flower
    dried flower

5.2. Source of Information

Each utterance that contained the evidential –mIş was coded with respect to its source of information. The source of information was classified into the following categories: (a) perceptual, (b) inference, (c) hearsay, or (d) nonfactual. An utterance was coded as perceptual, when the speaker makes an observation that does not require an explicit inference (e.g. *ayıcığın burada-y-miş ‘(I see that) your teddy bear is here’; çay sicak-miş ‘the tea is hot (now that I tasted it)’).
An inferential utterance contained a comment on the present state of an object, person or situation resulting from a past process that has not been witnessed (e.g. *pili bit-miş* ‘(apparently) its batteries died’, inferring this upon noticing that a toy does not work).

An utterance was coded as hearsay, if the speaker did not directly witness the event, but heard it from someone else (e.g. *Ayşe Ankara’ya gid-ecek-miş, Ali söyledi* ‘Ayşe is (reportedly) going to Ankara, Ali told me’).

Finally, the nonfactual coding category was used, if the utterance included (a) pretend play, (b) story-telling, or (c) nonfactual statements to regulate the behavior of the addressee.

Since caregivers sometimes also tended to use nonfactual reported speech, some utterances were coded as nonfactual-hearsay to denote this usage. In other words, the speakers reported an utterance that is not actually said by someone else, or that cannot be inferred from the situation (e.g. *baban sana çikolata getir-ecek-miş* ‘your father will bring you chocolate’). Examples for each category will be provided in Section 5.3 below.

Erroneous uses of the indirect experience marker were also noted. The most common error was using the evidential marker even if the speaker directly experienced the event in the past (e.g. *parka git-miş-iz*, literal: ‘we reportedly went to the park’, intended reading: ‘we went to the park’).

5.3. Pragmatic Function

Each utterance was further coded with respect to the pragmatic function it conveys if the evidential conveyed such a meaning.

Some of the utterances conveying nonfactual information (and especially child-directed ones) served another purpose not documented before, namely regulating the behavior of the child. Behavior regulation was used to influence the child’s action in the desired direction by diverting his attention, mentioning positive or negative outcomes of his behavior, and giving examples of behavior from other individuals.

In sum, based on the previous literature and our current observations, we defined the following six categories of pragmatic function: (a) narrative, (b) pretend play, (c) behavior regulation, (d) surprise, (e) irony, and (f) compliment. Whenever the source of information was coded as nonfactual, the pragmatic function belonged to one of the following categories: narrative, pretend play, or behavior regulation. The narrative category was used if the utterance contained story-telling or part of a children’s song (e.g. *bir kral ve kraliçe yaşar-r-miş* ‘there once lived a king and a queen’). The pretend play category was used if the utterance contained a nonfactual statement related to the play context (e.g.

---

2 Children are also reported to make the error of using the direct experience marker –DI instead of nonwitnessed –mIş (Aksu-Koç, 1988). We cannot provide comparative figures from the present corpus since the present analysis focused only on utterances with the –mIş form.
ayıcığın uykuşu gelmiş ‘teddy bear is sleepy’). Finally, the behavior regulation category was used whenever the utterance contained a statement to regulate the behavior of the addressee, where we defined three subcategories: (a) attention-getter, (b) positive consequence, or (c) negative consequence. Attention-getters were usually used to divert the child’s attention away from an undesired behavior (e.g. kuş gel-miş seni çağırıyor ‘the bird is here and calling for you’). Attention-getters may also be used with real stimuli in the environment (e.g. bak burda neler var-miş ‘look what you have here’), thus, some factual (e.g. perceptual) utterances were coded as attention-getters as well. Positive consequence and negative consequence categories contain utterances that mention a positive or negative consequence of the desired/unwanted behavior respectively (e.g. çorbanı yersen abla sana çok güzel bir şey ver-ecek-miş ‘if you eat your soup, she is (reportedly) going to give you something very nice’ and abla kız-iyor-miş ‘she is getting angry’).

Finally, utterances sometimes conveyed surprise (e.g. evime ne yap-miş-sınız ‘what did you do to my house’), irony (e.g. halıyı çok güzel ıslat-miş-sın ‘you wet the carpet very well’), or compliment (e.g. iyi yap-miş-sın ‘you did it well’).

6. Results and Discussion
6.1. Source of Information

Our first goal was to examine the distribution of types of source of information (SoI) conveyed by the evidential –miş within CHI and CDS, and the extent to which CHI and CDS showed parallel distributions.

Table 3 shows the distribution of SoI within CHI and CDS across all children. A Pearson product-moment correlation was conducted on the arcsine transformations of the observed percentages. Results showed a significant relationship between the distributions of source of information in CDS and CHI, $r = 0.82$, $n = 24$, $p < .001$, indicating similar overall distributions of different evidential functions within CDS and CHI in terms of SoI.

Table 3. Overall distributions of SoI in CDS and CHI.*

<table>
<thead>
<tr>
<th></th>
<th>Perceptual</th>
<th>Inference</th>
<th>Hearsay</th>
<th>Nonfactual Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall CDS</td>
<td>24.2%</td>
<td>15.8%</td>
<td>9.9%</td>
<td>49.6%</td>
</tr>
<tr>
<td>Overall CHI</td>
<td>23.2%</td>
<td>13.0%</td>
<td>11.6%</td>
<td>47.4%</td>
</tr>
</tbody>
</table>

*Rows do not add up to 100% due to errors in usage (e.g. talking about a directly experienced event in –miş form).

Except for one child (C1), the nonfactual usage was the most frequent and the hearsay was the least frequent category in CDS. Although each child committed a certain number of errors when using –miş, one child (C1) had the
highest percentage of erroneous uses. This child was the one who received the least amount of child-directed input among the 6 children (see Table 2).

Table 4. Distribution of different types of source of information in CDS and CHI.

<table>
<thead>
<tr>
<th></th>
<th>Perceptual Function</th>
<th>Inference Function</th>
<th>Hearsay Function</th>
<th>Nonfactual Uses</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CDS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>32.5%</td>
<td>16.0%</td>
<td>19.5%</td>
<td>32.0%</td>
<td>-</td>
</tr>
<tr>
<td>C2</td>
<td>23.3%</td>
<td>19.4%</td>
<td>8.4%</td>
<td>48.9%</td>
<td>-</td>
</tr>
<tr>
<td>C3</td>
<td>28.2%</td>
<td>17.0%</td>
<td>10.8%</td>
<td>41.1%</td>
<td>3.0%</td>
</tr>
<tr>
<td>C4</td>
<td>18.5%</td>
<td>17.7%</td>
<td>9.9%</td>
<td>53.7%</td>
<td>0.1%</td>
</tr>
<tr>
<td>C5</td>
<td>22.6%</td>
<td>13.4%</td>
<td>6.6%</td>
<td>57.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>C6</td>
<td>20.2%</td>
<td>11.3%</td>
<td>4.2%</td>
<td>64.3%</td>
<td>-</td>
</tr>
<tr>
<td><strong>CHI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>25.3%</td>
<td>12.5%</td>
<td>24.8%</td>
<td>17.7%</td>
<td>19.6%</td>
</tr>
<tr>
<td>C2</td>
<td>26.6%</td>
<td>14.6%</td>
<td>11.2%</td>
<td>46.1%</td>
<td>1.5%</td>
</tr>
<tr>
<td>C3</td>
<td>9.1%</td>
<td>3.8%</td>
<td>14.4%</td>
<td>69.7%</td>
<td>3.0%</td>
</tr>
<tr>
<td>C4</td>
<td>23.3%</td>
<td>18.0%</td>
<td>4.6%</td>
<td>52.3%</td>
<td>1.8%</td>
</tr>
<tr>
<td>C5</td>
<td>28.4%</td>
<td>8.9%</td>
<td>6.2%</td>
<td>54.1%</td>
<td>2.5%</td>
</tr>
<tr>
<td>C6</td>
<td>26.6%</td>
<td>19.9%</td>
<td>8.1%</td>
<td>44.8%</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

Figure 2 shows how the distribution of SoI in CDS changes over time collapsed across children. It is observed that perceptual uses which function to invite the child to attend to new information decrease with age whereas nonfactual uses that include play and narrative activities and behavior regulation increase with age. The proportion of inference and hearsay utterances appears to be relatively more stable over time.

Figure 2. The distribution of different evidential functions in CDS collapsed across children over time.
6.2. Age of Emergence

Our second goal was to examine when children started to use the evidential –mIş productively. Furthermore, we were interested in the order of emergence of different types of source of information, and whether this order was related to the frequency in the input. For each child, we examined the age of emergence of each category of SoI by disregarding imitative uses. An imitative use was defined as an utterance in which the child uses the same verbal root found in an adult’s utterance with –mIş in the preceding 15 utterances before the child’s use (e.g. adult-to-child: çorba al-miş mı?, ‘did (s/he) buy soup?’, child-to-adult: al-ma-miş, ‘(s/he) did not buy (soup)’).

Corroborating previous semi-naturalistic findings (Aksu-Koç, 1988), all children were able to use evidentials productively before age three. Perceptual and nonfactual functions were the first to emerge.

For each SoI category and child, Table 5 shows the first month of productive usage. In terms of production there is not a definitive order of emergence such as perceptual-inference-hearsay or perceptual-hearsay-inference.

Table 5. First month of productive usage for each child for each SoI function.

<table>
<thead>
<tr>
<th>Child</th>
<th>Perceptual</th>
<th>Inference</th>
<th>Hearsay</th>
<th>Nonfactual Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>C2</td>
<td>25</td>
<td>27</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>C3</td>
<td>29</td>
<td>36</td>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>C4</td>
<td>26</td>
<td>26</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>C5</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>C6</td>
<td>23</td>
<td>28</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

As mentioned before, Ozturk and Papafragou (2015) observed that children found it more difficult to use the inferential function in comparison to the hearsay function when tested in an experimental setting. This was attributed to the inferential function being less frequently encountered than the hearsay function in CDS. As shown in Table 4, our findings showed that the inferential function is actually more frequent than the hearsay function in CDS for five out of six children. In fact, across these five children, the frequency in the input follows the same order, where –mIş is most frequently used in a nonfactual sense followed by perceptual, inferential and hearsay usages. However, each child showed a different pattern of emergence of the evidential functions. They all started with the more frequent functions in CDS, either the nonfactual or the perceptual. The third function to emerge was the inferential in the speech of four
of the children and the hearsay in the speech of two. These results indicate that the distribution in the input was not the only determining factor of age of emergence.

6.3. Nonfactual Uses

Since the frequency of utterances expressing surprise, irony and compliments was very low in both CDS and CHI, analyses were not carried out on these pragmatic categories. However, we took a more detailed look into the nonfactual usage of the evidential –mış, since it was the most frequently used category for five out of six children and their caregivers.

In the analyses we report below, we merged ‘pretend play’ and ‘narrative’ under a new category called ‘activities’ to denote that the child is engaged in some activity like listening to or telling a story or spending time in play. The three subcategories ‘attention-getter’, ‘positive consequence’, and ‘negative consequence’ were treated together under ‘behavior regulation’.

Since we noticed that families with lower and higher SES tended to differ from each other with respect to their purpose of the nonfactual usage, we conducted analyses to compare low and high SES groups. Overall, the nonfactual usage was less common in families with lower SES, \( t(4) = 3.03, p = .039, d = 2.47 \). Furthermore, families with higher SES preferred to use –mış in nonfactual utterances more frequently during activities, whereas families with lower SES preferred the behavior regulation function more often, \( t(4) = 3.88, p = .018, d = 3.16^3 \) (see Figure 3).

![Figure 3. Distribution of the purpose of the nonfactual usage in CDS according to SES.](image)

Note: Some columns do not add up to 100% since some of the nonfactual uses were idioms and not classifiable under any category.

---

3 T-tests are based on the arcsine transformations of percentage data.
Figure 4 shows a more detailed distribution of the pragmatic functions of nonfactual utterances for each child and caregiver. While the nonfactual uses of the caregivers are differentiated in terms of purpose, children use mainly two categories, namely narrative and pretend play.

![Figure 4](image)

**Figure 4. Distribution of nonfactual uses according to pragmatic function in CDS and CHI.**

Note: Some columns do not add up to 100% since some of the nonfactual uses were idioms and not classifiable under any category.

7. Conclusion

In this corpus study, we used a relatively dense longitudinally obtained data to investigate the acquisition of the evidential marker –mIş by six Turkish-speaking children between 8 and 36 months of age. First, we examined how different sources of information (perceptual, inferential, reportative/hearsay, and nonfactual uses) were distributed in child-directed and child’s speech. For 4 out of 6 children, we observed highly similar distributions between CDS and CHI. For five out of six children the frequencies of different sources of information in the input followed the same order: Nonfactual uses formed the most commonly used category followed by perceptual, inferential, and hearsay uses. Despite these similarities, children started encoding different sources of information at different months of age. All children could productively use the evidential marker before age three. Except for one child, erroneous uses were not common, and usually occurred due to children using –mIş instead of –DI for directly experienced events.

Finally, we were interested in nonfactual uses, since these were very frequent in both CDS and CHI. We looked into the purpose of using –mIş in a nonfactual sense, and defined a new category that we refer to as ‘behavior regulation’, where the goal is to manipulate the behavior of the addressee by using the evidential –mIş. In such utterances –mIş allows the speaker to make promises or threats without real commitment to its actualization. Caregivers in
low and high SES families differed from each other in terms of their practices of employing the nonfactual usage, where caregivers with higher SES preferred to use the nonfactual –mIş mostly during pretend play and story-telling, and caregivers with lower SES preferred to use the nonfactual –mIş mostly for behavior regulation to divert the child’s attention and mention outcomes of the child’s behavior.

References

Küntay, Aylin C.; Koçbaş, Dilara; & Taşçi, Süleyman S. (2015). Koç University Longitudinal Language Development Database on language acquisition of 8 children from 8- to 36-months of age (approx. 1,000,000 words transcribed out of which 250,000 are glossed).
