Intonation and gesture as bootstrapping devices in speaker uncertainty

Iris Hübscher

Universitat Pompeu Fabra, Barcelona, Spain

Núria Esteve-Gibert

Aix Marseille Université, CNRS, LPL UMR 7309, Aix-en-Provence, France

Alfonso Igualada

Universitat Pompeu Fabra, Barcelona, Spain

Universitat Oberta de Catalunya, Barcelona, Spain

Pilar Prieto

Institució Catalana de Recerca i Estudis Avançats

Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

This study investigates 3- to 5-year-old children's sensitivity to lexical, intonational, and gestural information in the comprehension of speaker uncertainty. Most previous studies on children's understanding of speaker certainty and uncertainty across languages have focused on the comprehension of lexical markers, and little is known about the potential facilitation effects of intonational and gestural features in this process. A total of 102 3- to 5-year-old Catalan-speaking children participated in a comprehension task which involved the detection of uncertainty in materials that combined lexical, intonational, and gestural markers. In a between-subjects design, the children were either administered the lexical condition (where they were exposed to lexical and gestural cues to uncertainty) or the intonation condition (where they were exposed to intonational and gestural cues to uncertainty. Within each condition, three different presentation formats were used (audio-only, visual-only and audio-visual) in a within-subjects design. Our results indicated that all the children performed better overall when they had gestural cues present. Furthermore, in comparison with the older group, the younger group was more sensitive to intonational marking of speaker uncertainty than to lexical marking. This evidence suggests that the intonational and gestural features of communicative interactions may act as bootstrapping mechanisms in early pragmatic development.

Keywords

Intonation, gesture, uncertainty, prosodic development, pragmatic development, language acquisition, belief states, epistemic stance

Introduction

In everyday conversation, speakers are able to rapidly combine multimodal information during utterance comprehension, including verbal content, prosody, and gesture. In particular, in successful social interactions the detection of belief states such as uncertainty (or incredulity, surprise, etc.) is especially important in order to understand the other person's epistemic stance. Epistemic stance refers to the degree of commitment or certainty the speaker has in his or her statements. When inferring uncertainty, listeners can use various cues (depending on the language) such as lexical epistemic markers, morphological marking, gestures such as head nods and facial expressions, or prosodic features such as delays and final rising intonation (e.g., Swerts & Krahmer, 2005 or Borràs-Comes, Roseano, Vanrell, Chen, & Prieto, 2011, for a review). Typical lexical markers in English, for instance, are mental state verbs (such as *think*) and epistemic modal expressions (such as *maybe*). These lexical items convey

information about the epistemic stance of individuals. In many languages intonation plays a key role in shaping the pragmatic meaning of utterances and can encode epistemic and evidential information (Barth-Weingarten, Dehé, & Wichmann, 2009, and Prieto, 2015, for a review of the literature on the prosody-pragmatics interface). Gesture patterns can also play an important part in conveying epistemic information. For example, Swerts and Krahmer (2005) investigated the role of audio-visual prosody for signalling and detecting epistemic information in question answering. The study showed that there are well-defined visual cues that demarcate a speaker's feeling of knowing and that listeners are more capable of estimating another person's knowledge on the basis of visual and auditory information combined than just auditory input alone.

In the study of language development, one of the interesting questions is how and when children develop the ability to recognize an interlocutor's epistemic stance and feeling of knowing. To date, most research has concentrated on children's acquisition of lexical markers of belief states. Moore, Bryant and Furrow's (1989) classical study tested 3- to 8-year-old children in an experimental setting where children had to find an object in one of two boxes as they listened to verbal cues from two different puppets telling them about the place where the object was hidden. Each utterance contained a marker with a different degree of certainty, signposting one or the other box as the location of the hidden object, such as *I know it's in the red box* or *I think it's in the blue box*. The results showed that children aged 4 and above were able

to find the hidden object based on what they heard but 3-year-olds were not. Furthermore, Moore, Pure, and Furrow (1990) also showed that the understanding of modal expressions such as *might* strongly correlates with the understanding of mental verbs such as *think*. Likewise, Noveck, Ho, and Sera (1996) tested 5- to 9-year-old children's understanding of epistemic modals by contrasting *has to* with *might*, etc., and showed that (a) their understanding of modal expressions develops gradually over time; and (b) by 9 years of age children show an adult-like understanding of these modal expressions. There is one exception, though, by Moore, Harris and Patriquin (1993) who compared children's (3- to 6-years-old) comprehension of mental state lexicon to their comprehension of mental state prosody. Children had to listen to contrasting pairs of statements by two puppets and guess the location of a hidden object. Each statement pair either differed with respect to the mental state verbs *know* vs. *think* or *think* vs. *guess* – or with respect to terminal pitch contour – falling or rising. While 3-year-olds were not able to use either lexicon nor prosody to detect where the object was, 4-year-olds started to do so significaly, in the *know* vs. *think* and falling vs. rising contrast condition. Furthermore, the *think* vs. *guess* condition was much harder even for the 5-year-old children, compared to the *know* vs. *think* condition. In a follow-up experiment with 3- to 5-year-old children, the conditions were presented as either matched or mismatched. While in the matched condition lexical items of certainty went together with falling intonation and lexical items of uncertainty were matched with rising

intonation, in the mismatched condition the opposite was applied. While 4-year-olds performed significantly above chance when *know* vs. *think* was matched with the corresponding prosodic cue, they did not show any significant difference between matched vs. mismatched trials. This time 5-year-olds performed a lot better on the *think* vs. *guess* distinction in the matching condition. Also 5-year-old children performed a lot worse in the mismatched condition, showing a certain awareness of prosodic and lexical integration when speaker (un)certainty is expressed. The authors suggested that prosodic and lexical cues to speaker certainty start to be used around the same time by children. Yet, the authors propose that lexical cues to a speaker's belief state initially seem to be more dominant, with prosody playing a secondary role, modulating the effects of the lexical cues. More recent studies have investigated the acquisition of belief states, focussing on other languages such as Korean (Choi, 1995; Papafragou, Li, Choi, & Han, 2007), Cantonese (Lee & Law, 2001; Tardif, Welman, & Cheung, 2004), Turkish and Puerto Rican Spanish (Shatz, Diesendruck, Martinez-Beck, & Akar, 2003), Japanese (Matsui, Yamamoto, & McCagg, 2006), and Japanese and German (Matsui, Rakoczy, Miura, & Tomasello, 2009), yet with a sole focus on lexically encoded mental state information.

There seems to be a general consensus that it is not until age 4 that children are capable of identifying the meaning of modal expressions of uncertainty. Yet, Matsui et al. (2006) investigated children's understanding of knowledge states in Japanese, where

uncertainty can be encoded through both epistemic particles (*yo* = speaker certainty and *kana* = speaker uncertainty) and mental state verbs (such as *shitteru* = know and *omou* = think). They found that 3-year-old Japanese children already comprehended a speaker's knowledge state, but only when they were conveyed by particles. By contrast, at that age their understanding of mental state verbs was still quite poor (see Matsui, 2014, for a detailed overview of children's understanding of epistemicity and evidentiality).

Studies of children's pragmatic development have claimed to take into account gestural cues in the study of communication and language development (e.g., Furman, Kuntay, & Ozyurek, 2014; Guidetti, 2005; Guidetti, & Nicoladis, 2008; Iverson & Goldin-Meadow, 1998; McNeill, 1998; O'Neill, Bard, Linnell, & Fluck, 2005). There is a growing consensus that getsures act as bootstrapping devices in language development (Kelly, 2001; Butcher & Goldin-Meadow, 2000; McNeill, Cassell, & McCullough, 1994). With respect to the acquisition of belief states, some studies seem to suggest an earlier development of uncertainty understanding based on non-linguistic cues. For example, some studies have shown that 3- and 4-year-old children are capable of deciding who to believe based on visual signs of reliability or inference (Koenig, Clements, & Harris, 2004; Koenig & Harris, 2005; Robinson, Mitchell, & Nye, 1995; Robinson & Whitcombe, 2003; Sabbagh & Baldwin, 2001; Whitcombe & Robinson, 2000). There have been two studies that focused on older children (8 to 11 years old) which have investigated the development of their perception and production of facial

gestures as cues to uncertainty (Krahmer & Swerts, 2005; Visser, Krahmer, & Swerts, 2014). More basic forms of epistemic stance comprehension are also found early on in infancy. It has been shown that 12-month-olds are able to distinguish between knowledgeable and ignorant partners (Liszkowski, Carpenter & Tomasello's, 2008). The study explored the ability of 12-month-old infants to point appropriately at an object in order to provide uninformed people with information. To signal ignorance, the experimenter raised his/her hands with the palms upturned. Their results showed that infants pointed more often to an object which the adult had (presumably) not seen fall down and thus needed help to find than an object which the adult had seen fall down and thus could find unassisted.

All these studies suggest that children achieve important communicative milestones initially in the realm of gesture before they do so in speech, and gestures can therefore be seen as helping children to access meaning (e.g., Goldin-Meadow, 2007). While the role of prosody as a syntactic bootstrapper has been highlighted in language acquisition research, that is certain types of prosodic features guide children's initial acquisition of word order and syntactic structure (e.g., those related to constituent or prosodic phrasing; for a conceptualisation see Hirsh-Pasek, Tucker, & Golinkoff, 1996; see also Christophe, Nespor, Guasti, & van Oyeen, 2003). However, so far very little is known about the role of prosody in early pragmatic development and whether it might have a possible bootstrapping effect on the comprehension of pragmatic meaning.

Recent studies on the prosody-pragmatics interface have shown that 12-month-old infants use prosody (together with pointing gestures) to comprehend an adult's basic communicative intentions like expressive, imperative, and informative attention-directing actions (Esteve-Gibert, Prieto, & Liszkowski, 2016), and that 14-month-old infants can use prosody to distinguish between intentional and non-intentional acts (Sakkalou & Gattis, 2012). Also, research has shown that infants as young as two display a basic inventory of target-like intonation contours with an adult-like intentional meaning (e.g., Chen & Fikkert, 2007; Frota, Cruz, Matos, & Vigário, 2016; Prieto, Estrella, Thorson, & Vanrell, 2012). Thus, independent evidence coming from studies investigating the acquisition of pragmatic intonation seems to suggest an initial role for prosody as a bootstrapping mechanism in the early stages of the understanding of pragmatic meaning.

While prosody is a very prominent cue in infancy, studies testing pre-school and school-age children's understanding of prosody have yielded conflicting results. On the one hand research on children's sensitivity to pitch as a cue to emotions has shown that the adult-like ability to judge a speaker's emotional state based on vocal affect is mastered only at 4 years, after children have acquired the lexical semantic meaning of the four basic emotions (happiness, sadness, anger, and fear), which happens around age 3 (Morton & Trehub, 2001; Nelson & Russell, 2011; Quam & Swingley, 2012). Yet, when there are cues in competition regarding the relevant emotion conveyed via either

the lexical meaning of a sentence (Morton & Trehub, 2001; Waxer & Morton, 2011) or the situational context (Aguert, Laval, Le Bigot, & Bernicot, 2010; Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013), the success of pre-schoolers at identifying vocal affect seems compromised. For example, if someone utters "It's Christmas time" with a sad prosody, adults will rely on the prosody and judge the speaker to be sad whereas 6-year-old children will say the speaker is happy. By the same token, Vernice and Guasti (2014) showed that before age 5 children are not able to use prosodic cues in order to decide which referent to mention next. While all these studies hint at a surprisingly late acquisition of certain prosodic cues at the sentence level, a very recent study by Berman, Chambers, and Graham (2016) discovered that when a more implicit methodology such as eye-tracking is used, young children already at age 3 show themselves able to link speech bearing different acoustic cues to emotion. Unfortunately, overall these studies lack a description of the acoustic characterisation of the prosodic differences between the different emotions described, which makes it hard to track which prosodic cues children learn to attend to (with the exception of Quam & Swingley, 2012). And, furthermore, these prosodic cues to emotion might only be subtle cues that do not involve a real change in pragmatic intonation patterns.

On the other hand, however, hardly any research has focused on when and how children understand more complex pragmatic meanings such as epistemicity encoded through prosody and/or gestures. A recent exception is Armstrong (2012) and

Armstrong (2014) which focused on children's comprehension of intonationally-encoded disbelief in polar questions in Puerto Rican Spanish. Particularily relevant for the current study is the study by Armstrong, Esteve-Gibert, and Prieto (2014), which investigated 3- to 5-year-old understanding of disbelief (or incredulity) through three different modalities, visual-only (facial gesture cues), audio-only (intonation), and audio-visual (facial gestures and intonation). The children were exposed to short discourse reactions such as *Una balena?!* ('A whale?!') produced with either incredulous or credulous intonation, and they had to decide between the two meanings. The results showed that 3-year-old children performed the worst on the audio-only task. 4-year-olds performed better, but still showed great variability. Also, a great deal of variability was observed for younger children that received the audio-visual condition, arguably because it was difficult for some of them to integrate the two cues. By contrast, 3- and 4-year-olds performed much better in the visual-only condition compared to the other two conditions. Furthermore, 5-year-old children performed equally well in the audio-only condition. The authors suggest that facial gestures seem to provide children with scaffolding for the detection of speaker disbelief. However, one aspect that this study could not explore was the children's sensitivity to prosodic and gestural features relative to lexical cues, which were not included in the study.

The main purpose of the current study is to assess the relative roles of lexical, intonational, and gestural cues in preschoolers' understanding of uncertainty and to test

the potential bootstrapping role of gestures and intonation in its development. Specifically, we are interested in whether children (1) use gestures as a bootstrapping device in the comprehension of uncertainty and (2) recognise uncertainty more easily through lexical or intonational epistemic markers. To address these questions, we asked 3- to 5-year-olds to select the uncertainty stimuli in a forced-choice task.

A modified version of Armstrong et al.'s (2014) incredulity comprehension task was used here in which children had to decide which speaker was uncertain about something. Importantly, the two experimental conditions (uncertainty/certainty) were tested using stimuli presented in either visual-only, audio-only, or audio-visual modality.

In line with previous studies on the facilitator role of gestures in general (e.g., Furman et al., 2014; Guidetti, 2005; Iverson & Goldin-Meadow, 1998; McNeill, 1998; O'Neill et al., 2005) and in particular in disbelief understanding (Armstrong et al. 2014), we expected that the presence of visual information would bootstrap children's understanding of belief state. Furthermore, contrary to previous studies on the late acquisition of meaning encoded through prosody, it was our position that children younger than 4 years would be sensitive first to intonational cues to uncertainty, then to lexical ones. The results would therefore be important to further our understanding of how pragmatic communication skills develop in children and the role intonation and gesture play in this development.

Methods

Participants

A total of 102 3- to 5-year-old children participated in the experiment. Children were divided into a younger group (N = 51, mean age = 3 years and 9 months, SD = 5.50) and an older group (N = 51, mean age = 5 years and 2 months, SD = 5.27). All the participants were preschoolers at three Catalan public schools located in the Barcelona area. In these schools, the main language of instruction is Catalan. Parents were informed about the experiment's goal and signed a participation consent form. Furthermore, language exposure questionnaires (based on Bosch & Sebastián-Gallés, 2001) were administered to the caregivers in order to ensure that the participating children were predominantly exposed to Catalan (as opposed to Spanish) on a daily basis (mean percentage of overall exposure to Catalan: 87%, SD = 12.0).

Design

The target materials were video recorded by taking into consideration the results of a general knowledge quiz which was constructed to elicit the spontaneous use of utterances conveying different degrees of uncertainty in Catalan, based on Krahmer and Swerts (2005). We analysed the lexical, gestural, and prosodic expressions of certainty in a total of 180 answers (12 questions x 15 participants). Results showed that

participants mainly used two different types of intonation patterns, depending on the certainty condition. In the certain condition they universally used a falling pitch contour L* L% (100% of the cases) and in the uncertain condition they used two variants of a rising pitch contour (L* H% and L+H* H%, which covered 25% of the cases. Furthermore, participants used lexical items (*potser* 'perhaps', *crec que* 'I think', etc.) in 25% of the cases, which went together with a falling intonation (L* L%) when expressing uncertainty (the remaining 50% belonged to less high degrees of uncertainty). Finally, participants produced a head nod when being certain, and a varied group of gestures (e.g., diverted gaze, low/high gaze, raised or furrowed eyebrows, squinted eyes and head tilt) when being very uncertain.

Taking these findings into account, three adult Catalan speakers were videotaped while producing a total of 12 target utterances each (6 trials x 2 epistemic marking conditions; see Appendix), resulting in a grand total of 36 target stimuli (3 speakers × 12 utterances). The epistemic marking conditions consisted of utterances expressing certainty/uncertainty through both lexical and gestural markers (this will henceforth be referred to as the lexical condition), and utterances expressing certainty/uncertainty through only intonation and gestural markers (henceforth the intonation condition). A fourth Catalan speaker was recorded for the familiarization trial.

For the **lexical condition** we used the following lexical epistemic markers[1]: a common adverb signalling uncertainty in Catalan (*potser* 'maybe'), and a very common epistemic construction which signals certainty (*segur que* '[I am] certain that').[2] Thus, the certainty stimuli consisted of a noun phrase preceded by the adverb *segur que* '[I am] certain that' (e.g., *Segur que el tomàquet* '[I am] certain that [it is] the tomato') and accompanied by a head nod gesture suggesting certainty (Figure 1, left-hand panels). The uncertainty stimuli consisted of a noun phrase preceded by the adverb *potser* 'maybe' (e.g., *Potser el tomàquet* 'Maybe [it is] the tomato') and accompanied by gestures suggesting uncertainty (squinted eyes, raised eyebrows, head tilt) (Figure 1, right-hand panels). Crucially, both certainty and uncertainty utterances were produced with the same intonation contour (L* H% associated with the adverb plus final falling intonation, L* L%)

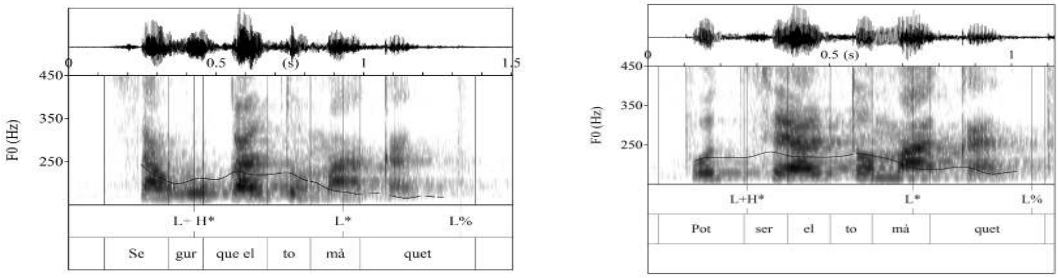

Figure 1: Lexical condition. Upper panels: pitch tracks, spectrograms, and waveforms for the certainty utterance (*Segur que el tomàquet* '[I am] certain that [it's] the tomato') (left-hand panel) and uncertainty utterance (*Potser el tomàquet* 'Maybe [it is] the tomato') (right-hand panel). Lower panels: screenshots of facial expressions corresponding to certainty (left-hand panel) and uncertainty utterances (right-hand panel).

For the **intonation condition**, the certainty stimuli (e.g., *El tomàquet* 'The tomato') were produced with a falling intonation contour (L* L%) and accompanied by a head nod gesture suggesting certainty (Figure 2, left-hand panels). The uncertainty stimuli (e.g., *El tomàquet?* 'The tomato?') were produced with a rising intonation contour (L* H%) and gestures suggestive of uncertainty (squinted eyes, raised eyebrows, head tilt) (Figure 2, right-hand panels). Crucially, in the intonation condition the utterance contained no lexical information such as epistemic adverbs which would help to distinguish certain from uncertain stimuli.

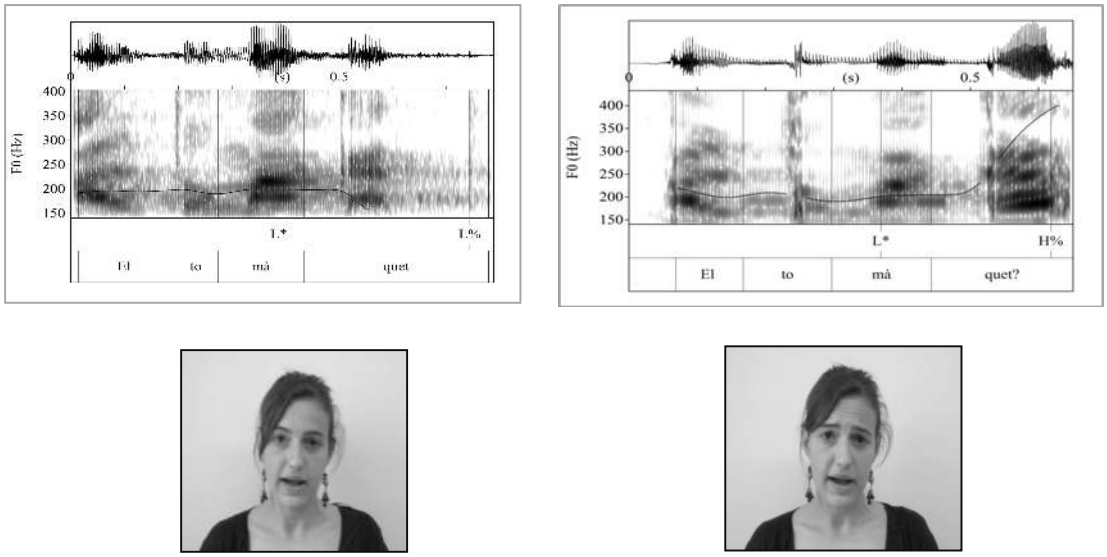

Figure 2: Intonation condition. Upper panels: pitch tracks, spectrograms, and waveforms for the certainty utterance (*El tomàquet* 'The tomato') (left-hand panel) and uncertainty utterance (*El tomàquet?* 'The tomato?') (right-hand panel). Lower panels: screenshots of facial expressions corresponding to certainty (left-hand panel) and uncertainty utterances (right-hand panel).

Each epistemic-marking condition was presented in three different modalities: audio-only, visual-only, and audio-visual. For the audio-only trials, the audio track was played to subjects and the visual information reduced to a minimum by displaying two still photos of the twins speaker with a neutral facial expression. For the visual-only trials, the audio track was removed from the original audio-visual stimuli so that only the visual information was available to subjects. For the audio-visual trials, both the audio track and the accompanying video images were presented.

Table 1 summarizes how the combination of lexical, intonational, and gestural cues differed across the epistemic marking conditions (lexical condition vs. intonation condition) and modalities of presentation (audio-only, visual-only, or audio-visual). This design was intended to allow us to assess the role of the visual cues with respect to the speech cues (be they intonational or lexical cues) in (un)certainty detection.

	Audio-only		*Video-only*		*Audio-visual*	
	Certain	*Uncertain*	*Certain*	*Uncertain*	*Certain*	*Uncertain*
Intonation condition	Falling L* L%	Rising L* H%	Head nod	Squinted eyes, raised eyebrows, head tilt	Falling L* L% Head nod	Rising L* H% Squinted eyes, raised eyebrows, head tilt
Lexical condition	*Segur que* '[I am] certain that' L* L%	*Potser* 'Maybe' L* L%	Head nod	Squinted eyes, raised eyebrows, head tilt	*Segur que* 'I'm certain that' Falling L* L% Head nod	*Potser* 'Maybe' Falling L* L% Squinted eyes, raised eyebrows, head tilt

Table 1: Lexical, intonational, and gestural cues of the stimuli according to epistemic marking condition (intonation vs. lexical condition) and modality of presentation (audio-only, video-only, or audio-visual).

The semantic appropriateness of the stimuli selected was controlled for by running an experiment with the online survey platform SurveyGizmo. Sixty Catalan-speaking adults (30 respondents x 2 epistemic marking conditions) were asked to rate each of the 9 experimental stimuli sets (including both uncertainty and certainty stimuli), yielding a total of 540 tokens. Out of these 540, only two elicited contradictory certainty ratings by

respondents. These two stimuli were subsequently re-recorded, and further testing yielded consistent ratings.

Set-up of the task. This task is an adaptation of the task used in Armstrong et al. (2014). A PowerPoint presentation depicted the story of two twins travelling on a train with their friend Barbara, who plays a game with them to help make the journey pass more quickly. The game consists of her asking the twins if they know about her favourite things. For example, Barbara asks them, 'What is my favourite vegetable?' The answer is then revealed visually as a tomato in a thought bubble (Figure 3, left image), which the experimenter points out to the child. Previous research has shown that 3-year-olds understand thought bubbles as representations of mental contents (Wellman, Hollander, & Schult, 1996).

During the experiment, the child subject was seated in a position to view the screen as a researcher talked and operated the PPT slide show. Once the twins and Barbara had been introduced and the basic guessing game scenario described, the researcher told the child that for each question there was one twin who was sure of the right answer and one who was not, and that the child had to point to the uncertain twin. The child's response was regarded as 'correct' if s/he pointed to the twin who expressed uncertainty (Figure 3, right image).

Figure 3: Sample slides from the PPT presentation used in the comprehension task. Left-hand slide: Barbara is thinking of her favourite vegetable. Right-hand slide: Barbara (top) and the twins (bottom).

Procedure

The children were tested individually in a quiet room at each of the three participating schools. The researcher, a male Catalan-speaking adult (the third author of this paper), was seated beside the child in a room at the child's school, so that both faced the computer screen. The children were administered either the lexical or intonation condition (between-subjects), each containing 3 audio-only, 3 visual-only, and 3 audio-visual trials (within-subjects) in a randomised order. Prior to performing the comprehension task, each participant first went through a familiarisation trial to make sure that they understood what they were supposed to do. They then performed a total of 9 test trials in two counterbalanced orders, either first 3 audio-only, then 3 visual-only, and finally 3 audio-visual, or first 3 visual-only, then 3 audio-only, and finally 3 audio-visual. After each set of three trials, in order to prepare the child for the change in

modality, s/he was shown a filler slide depicting either a photo of an ear (signalling audio-only), an eye (visual-only), or both (audio-visual). In total the procedure lasted at most 10 minutes.

Results

A total of 918 children's responses were obtained from the comprehension task (9 responses x 102 children) and then analysed through a Generalised Linear Mixed Model (GLMM) using IBM SPSS Statistics 21. The dependent variable was 'child's performance', a numerical measure obtained by calculating the mean proportion of correct to incorrect responses. The fixed factors were epistemic marking condition (two levels: intonation condition, lexical condition), modality of presentation (three levels: audio-only, visual-only, audio-visual), age group (two levels: younger group, older group), and all their possible interactions. The random factor was participants.

Figure 4 shows the mean proportion of correct responses broken down by epistemic marking condition (intonation and lexical) and modality condition (audio-only, visual-only, and audio-visual) for the two age groups (younger and older) in the sample. Table 2 reports the mean and standard deviation of the results.

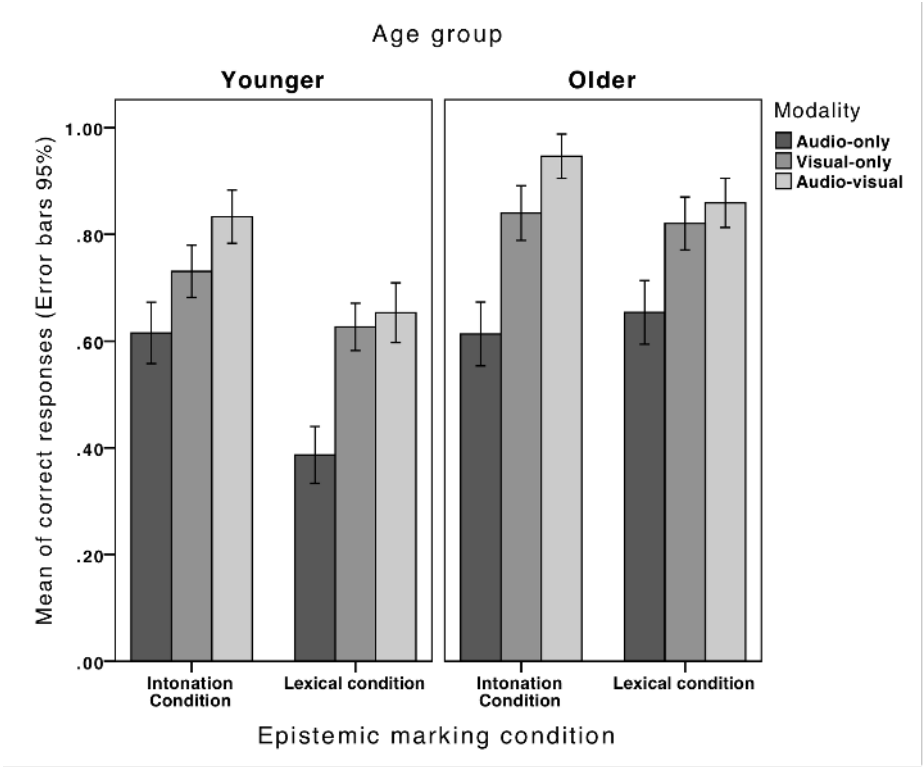

Figure 4: Mean proportion of correct responses to incorrect answers broken down by epistemic marking condition, modality of presentation, and age group.

The GLMM analysis revealed a main effect of age group, $F(1,294) = 21.215$, $p < .001$, with older children performing significantly better than younger children, and a main effect of epistemic marking condition, $F(1,294) = 10.064$, $p < .01$, indicating that when children were presented with the intonation condition they performed significantly better than when they were presented with the lexical condition.

There was also a main effect of presentation modality, $F(2,294) = 20.314$, $p < .001$. Pairwise contrasts showed that when children were presented with visual

modalities (visual-only and audio-visual), they performed significantly better than when presented with the audio-only modality ($p < .001$), with no difference between the two visual modalities ($p = .052$). Having the visual information present clearly helps the children to detect uncertainty better and thus confirms our first hypothesis that gesture has a bootstrapping effect on the child's comprehension of pragmatic meaning.

The model also reported a significant interaction between age group and epistemic marking condition, $F(1,294) = 7.751$, $p < .01$. Pairwise comparisons showed that only the younger children performed significantly better in the intonation condition as compared to the lexical condition ($p < .001$). All the other main effects and possible interactions were not significantly different. This confirms our second hypothesis, namely that younger children are able to detect epistemic meaning first through intonational cues before doing so through lexical cues. Table 2 displays the relevant means and standard deviation for each epistemic marking condition and modality of presentation in both age groups.

Epistemic marking	Modality	Group					
		Younger		Older		Total	
		Mean	SD	Mean	SD	Mean	SD
Intonation condition	Audio-only	1.85	.88	1.84	.90	1.84	.88
	Visual-only	2.19	.75	2.52	.77	2.35	.77
	Audio-visual	2.50	.76	2.84	.62	2.67	.71
	Total	2.18	.83	2.40	.87	2.29	.86
Lexical condition	Audio-only	1.16	.80	1.96	.92	1.57	.94
	Visual-only	1.88	.67	2.46	.76	2.18	.77
	Audio-visual	1.96	.84	2.58	.70	2.27	.83
	Total	1.67	.84	2.33	.83	2.01	.90
Total	Audio-only	1.51	.90	1.90	.90	1.71	.92
	Visual-only	2.04	.72	2.49	.76	2.26	.77
	Audio-visual	2.24	.84	2.71	.67	2.47	.79
	Total	1.93	.87	2.37	.85	2.15	.89

Table 2: Means and standard deviation (SD) of the correct responses.

Discussion and conclusions

The aim of this study was to investigate the role of intonational, lexical, and gestural cues in the early development of epistemic understanding. Overall, the results of the comprehension task with 102 3- to 5-year-old children showed that children make great strides in their comprehension of uncertainty between the ages of 3 and 5, as seen by the fact that children in the older age group performed significantly better than those in the younger age group. These results are in line with previous studies that found that lexical understanding of uncertainty is achieved between the ages of 4 and 5. It is not surprising

that younger children did not perform well in the lexical condition, since it has been documented across languages that children acquire the difference between different degrees of speaker certainty expressed through modal auxiliaries only around age 4 (e.g., Moore et al., 1990).

Yet the main question addressed by this study was whether younger preschool children attain epistemic understanding earlier through gestural and intonational features as compared with lexical features, and thus whether these features give them their first understanding of others' belief states. In the present study, by comparing three modalities of communication (audio-only, visual-only, and audio-visual), it was possible to investigate the relative contributions of gesture, lexical and intonational cues to children's pragmatic comprehension. Our results showed that both younger and older children perform significantly better in both the visual-only and the audio-visual modality than in the audio-only modality. These findings are comparable with those of Armstrong et al. (2014), where facial gestures also seemed to scaffold children's performance in detecting belief state meaning (i.e., incredulity). By the same token they are compatible with the growing consensus that gestures act as bootstrapping devices in language development in general (e.g., Kelly, 2001; Butcher & Goldin-Meadow, 2000; McNeill, Cassell, & McCullough, 1994).

With respect to the contribution of intonation, the novelty of our study lies in the fact that our experimental methodology allowed for a direct comparison between the

children's sensitivity to intonational vs. lexical cues to uncertainty. Crucially, our results showed that 3-year-old children were more sensitive to salient intonational cues to uncertainty (in our case, a rising intonation pattern L* H%) than to lexical cues to uncertainty, regardless of whether visual information was also available or not. This result contradicts previously found results by Moore, Harris, and Patriquin (1993), who regarded prosody as playing a secondary role in children's acquisition of belief state meanings. Furthermore, our results show that 4- and 5-year-olds, by contrast, performed equally well in the audio modality in both epistemic marking conditions, showing that they have acquired an understanding of lexical cues by this age.

These results also seem to point in a different direction than previous studies on children's development of emotional prosody (Quam & Swingley, 2012; Morton & Trehub, 2001; Nelson & Russell, 2011), which have found that children's ability to match the auditory cues to the four basic emotions (happiness, sadness, anger, and fear) seems to appear after children have acquired the lexical-semantic meaning of these emotions. Furthermore, when 5-year-old children are confronted with contrasting lexical cues or additional neutral situational cues juxtaposed on prosodic cues, they rely for their judgments on the lexical or situational cues rather than basing their judgment on the vocal cues encoding emotions. Thus, overall, this research would lead to the interpretation that prosodic cues do not seem to be prominent in the pre-school years in leading children to detect emotional/attitudinal meaning in speech and that children

rather use other cues to guide them (Aguert et al., 2010; 2013; Nelson & Russell, 2011; Waxer & Morton, 2011).

However, these studies deal with emotional prosodic cues (mostly pitch cues of contrasting pitch range) for inferring another person's emotional state, and these are weak prosodic cues not involving distinct pragmatic intonation patterns (Aguert et al., 2013; Quam & Swingley, 2012; Waxer & Morton, 2011). By contrast, our study has shown that 3-year old children are sensitive to intonational contrasts involving final rise (H%) vs. final fall (L%) distinctions for inferring speaker belief. In our data, developmental changes in children's comprehension of belief states become evident first in intonation (and in gesture) and only later in lexical marking. Thus our results seem to suggest that, regardless of the fact that mastering emotional prosody can appear later in development, intonational linguistic contrasts indicating complex pragmatic functions are probably mastered well before children acquire the lexical epistemic markers. Similar to the hypothesis of early *prosodic bootstrapping*, we contend that not only does prosody play a crucial role in the early acquisition of language by helping children to decode syntactic structure but that in later stages of development prosody exerts a different type of bootstrapping effect, namely, it facilitates the acquisition of pragmatic meaning. While our study shows that young children are able to understand epistemic meaning encoded through intonational and gestural cues earlier than through lexical cues, further steps need to be taken to prove whether prosodic abilities are

predictive of later lexical acquisition, that is, whether there exists a direct correlation between early understanding of prosodic cues and the subsequently following lexical comprehension.

To summarise, the results of the current study suggest that not only gesture but also pragmatic prosodic patterns act as an integral part of the language-learning process at the intermediate stages of language development. These prosodic and gestural features can probably be claimed to act as bootstrapping devices in which children ground their early pragmatic development. We thus argue that early sensitivity to and acquisition of intonation and gesture patterns should receive more attention in developmental research in order for us to gain a more complete picture of children's pragmatic development.

Acknowledgments

We would like to thank the staff at the Escola Sant Martí, Escola La Farigola del Clot, and Escola Pública Dr. Estalella Graells for granting us access to and organizing the meetings with preschoolers. We also would like to thank the children and their families for their participation in the experiment. Furthermore, many thanks to Joan Borràs-Comes, Alba Ayneto, Laia Mayol, and Santiago González for participating in the recordings of the experimental materials, and to the participants in the general

knowledge quiz (who were all students and researchers at the UPF). Also thanks to Joan Borràs-Comes again for assistance with the statistical analyses. Finally we are grateful to Marc Swerts, Martine Grice, Judy Reilly and Meghan Armstrong for thoughtful ideas and comments at different international meetings.

Funding

This research has been funded by a grant awarded by the Spanish Ministry of Science and Innovation (FFI2015-66533 BFU2012-31995 "Intonational and gestural meaning in language"), and by a grant awarded by the Generalitat de Catalunya (2014SGR-925) to the Prosodic Studies Group.

Notes

1. Escola Sant Martí in Arenys de Munt, Escola La Farigola del Clot in Barcelona, and Escola Pública Dr. Estalella Graells in Vilafranca del Penedès.

2. Catalan, like other Romance languages, uses a set of epistemic markers and morphosyntactic resources to mark epistemic commitment such as epistemic adverbs (e.g., *potser* 'perhaps'), conditional forms (e.g., *vindria* 'I would come'), verbal tense and subjective mood (e.g., *dubto que vingui* 'I doubt he'd come-subjunctive'), etc.

3. The two forms typically appear in sentence-initial position and thus in an especially prominent position for children to acquire them (for more information about epistemic and evidential marking in Catalan, see González, Borràs-Comes, Roseano, & Prieto, in press).

References

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010). Understanding expressive speech acts: The role of prosody and situational context in French-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research, 53*, 1629–1641.

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Bigot, L. Le. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE*, *8*(12), 1–9.

Armstrong, M.E. (2012). The development of yes-no question in Puerto Rican Spanish. Columbus, Ohio: Ohio State University dissertation.

Armstrong, M.E. (2014). Child comprehension of intonationally-encoded disbelief. BUCLD 38 Proceedings. Somerville, MA: Cascadilla Press.

Armstrong, M.E., Esteve-Gibert, N., & Prieto, P. (2014). The acquisition of multimodal cues to disbelief. *Proceedings of the Speech Prosody 2014*. ISSN: 2333-2042. Dublin, Ireland, May 20-23.

Barth-Weingarten, D., Dehé, N., & Wichmann, A. (2009). *Where prosody meets pragmatics*. Bingley: Emerald.

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, *142*, 391–399.

Borràs-Comes, J., Roseano, P., Vanrell, M. M., Chen, A., & Prieto, P. (2011). Perceiving uncertainty: facial gestures, intonation, and lexical choice. In C. Kirchhof, Z. Malisz & P. Wagner (eds.), *Proceedings of the 2nd Conference on Gesture and Speech in Interaction*. (GESPIN 2011), Bielefeld University, Germany.

Bosch, L., & Sebastián-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*, 29–49.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (ed.), *Language and Gesture* (pp. 235-257). New York: Cambridge University Press.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain, & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 315–320). Dudweiler: Pirrot.

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: John Benjamins Publishing Company.

Christophe, A., Nespor, M., Guasti, M.T., & van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, *6*(2), 211–220.

Esteve-Gibert, N., Prieto, P., & Liszkowski, U. (2016). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy*. Advance online publication. doi: 10.1111/infa.12146.

Frota, S., Cruz, M., Matos, N., Vigário, L. (2016). Early Prosodic Development: Emerging intonation and phrasing in European Portuguese. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Intonational Grammar in Ibero-Romance. Approaches across linguistic subfields* (pp. 299–324). Amsterdam: John Benjamins

Furman, R., Kuntay, A., & Ozyurek, A. (2014). Early language-specificity of children's event encoding in speech and gesture: Evidence from caused motion in Turkish. *Language, Cognition and Neuroscience, 29*, 620–634.

González, M., Borràs-Comes, J., Roseano, P. & P. Prieto (in press), Epistemic and evidential marking in discourse: Effects of register and debatability. *Lingua*.

Goldin-Meadow, S. (2007). The challenge: Some properties of language can be learned without linguistic input. *The Linguistic Review*, *24*, 417–421.

Guidetti, M. (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language*, *32*, 911–924.

Guidetti, M., & Nicoladis, E. (2008). Introduction to special issue: Gestures and communicative development. First Language, 28(2), 107-115.

Hirsh-Pasek, K., Tucker, M. & Golinkoff, R. M. (1996). Dynamic systems theory: reinterpreting ‘‘prosodic bootstrapping’’ and its role in language acquisition. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (449–466). Mahwah, NJ: Erlbaum.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, *28*(2), 325–349.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children’s use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53.

Lee, T. H., & Law, A. (2001) Epistemic Modality and the Acquisition of Cantonese Final Particles. In Mineharu Nakayama (ed.), *Issues in East Asian Language Acquisition* (pp. 67-128). Tokyo: Kurosio Publishers.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve- month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*(3), 732–739.

Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentuality. In D. Matthews (ed.), *Pragmatic Development in First Language Acquisition* (pp. 295–316). Amsterdam: John Benjamins Publishing Company.

Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development, 21*, 158–173.

Matsui, T., Rakoczy, H., Miura, Y., & Tomasello, M. (2009). Understanding of speaker certainty and false-belief reasoning : a comparison of Japanese and German preschoolers, *Developmental Science, 4*, 602–613.

McNeill, D. (1998). Speech and gesture integration. In J. M. Iverson & S. Goldin-Meadow (eds.), *The nature and functions of gesture in children's communication. New directions for child development, No. 79* (pp. 11–27). San Francisco: Jossey-Bass Inc., Publishers.

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction, 27*(3), 223–237.

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

Moore, C., Harris, L., & Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *Journal of Child Language, 20*, 153–167.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*, 722–730.

Morton, J., & Trehub, S. (2001). Children's understanding of emotion in speech. *Child Development, 72*, 834–843.

Nelson, N. L., & Russell, J. A. (2011) Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*, 52–61.

Noveck, I. A., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language, 10*(1), 621–644.

O'Neill, M., Bard, K. a., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, *8*(4), 352–359.

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, *103*, 253–299

Prieto, P. (2015). Multidimensional intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*, 371–381.

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250.

Robinson, E. J., Mitchell, P., & Nye, R. (1995). Young children's treating of utterances as unreliable sources of knowledge. *Journal of Child Language, 22*(3), 663–685.

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

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054 –1070.

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, *27*(1), 1–16.

Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology, 39*, 717–729.

Swerts, M., & Krahmer, E. (2005) Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*(1), 81–94.

Tardif, T., Wellman, H. M., & Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *Journal of Child Language, 31*, 779–800.

Vernice, M., & Guasti, M. T. (2014). Effects of prosodic cues on topic continuity in child language production. *First Language*, *34*, 406–427.

Visser, M., Krahmer, E., & Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Language and Speech, 57*(1), 86–107.

Waxer, M., & Morton, J. B. (2011). Children's judgments of emotion from conflicting cues in speech: Why 6-year-olds are so inflexible. *Child Development, 82*, 1648–1660.

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and of thoughts. Child Development, 67(3), 768–788.

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.

Appendix

Target words and phrases used for the comprehension task with their English translation in the audio-only condition

Intonation Condition

Certain		**Uncertain**	
Catalan	English	Catalan	English
El gos (L* L%)	Dog	El gos (L* H%)	Dog
El futbol (L* L%)	Soccer	El futbol (L* H%)	Soccer
La poma (L* L%)	Apple	La poma (L* H%)	Apple
La guitarra (L* L%)	Guitar	La guitarra (L* H%)	Guitar
El pernil (L* L%)	Ham	El pernil (L* H%)	Ham
La pizza (L* L%)	Pizza	La pizza (L* H%)	Pizza
El tomàquet (L* L%)	Tomato	El tomàquet(L* H%)	Tomato
La platja (L* L%)	Beach	La platja (L* H%)	Beach
El blau (L* L%)	Blue	El blau (L* H%)	Blue

Lexical Condition

Certain		**Uncertain**	
Catalan	English	Catalan	English
Segur que el gos (L* L%)	[I am] certain that [it's] the dog.	Potser el gos (L* L%)	Maybe [it's] the dog.
Segur que el futbol (L* L%)	[I am] certain that [it's] soccer.	Potser el futbol (L* L%)	Maybe [it's] soccer.
Segur que la poma (L* L%)	[I am] certain that [it's] the apple.	Potser la poma (L* L%)	Maybe [it's] the apple.
Segur que la guitarra (L* L%)	[I am] certain that [it's] the guitar.	Potser la guitarra (L* L%)	Maybe [it's] the guitar.
Segur que el pernil (L* L%)	[I am] certain that [it's] ham.	Potser el pernil (L* L%)	Maybe [it's] ham.
Segur que la pizza (L* L%)	[I am] certain that [it's] the pizza	Potser la pizza (L* L%)	Maybe [it's] the pizza.
Segur que el tomàquet (L* L%)	[I am] certain that [it's]	Potser el tomàquet (L* L%)	Maybe [it's] the tomato.

	the tomato		
Segur que la platja (L* L%)	[I am] certain that [it's] the beach	Potser la platja (L* L%)	Maybe [it's] the beach.
Segur que el blau (L* L%)	[I am] certain that [it's] blue	Potser el blau (L* L%)	Maybe [it's] blue.