

Acquisition of uncertainty in Mandarin sentence-final particles

In Mandarin, Sentence-Final Particles (SFPs) - clause-final elements encoding discourse-related meanings - emerge very early in child utterances, many before age 2;0 [1,2]. However, existing studies consist almost entirely of corpus analyses, leaving how children comprehend SFPs unclear. We examine children's comprehension of SFP *ba* (1), which encodes speaker uncertainty in declaratives [3,4], comparing it with the uncertainty-denoting adverb *keneng* 'perhaps' (2) [5]. Results suggest an emerging sensitivity to uncertainty encoded in SFP *ba* by age 4;0. However, children comprehend uncertainty significantly better when encoded in the adverb than the SFP.

- | | |
|--|--|
| 1. ta jintian buhui lai ba . | 2. ta jintian keneng buhui lai. |
| she today will.not come BA | she today perhaps will.not come |
| '(I guess) she will not come today.' | 'Perhaps she will not come today.' |

Background. Regarding production, while *ba* emerges early [1,2], it is unclear if children use it to denote uncertainty in declarative sentences, and when such use emerges (*ba* has different meanings in other clause-types). As for comprehension, *ba* has not been examined, but Mandarin-speaking children recognize uncertainty encoded by *keneng* 'perhaps' by age 5;0 [6].

Corpus. We used the Tong Corpus [7] (age 1;07-3;04), and the Zhou Corpus 3 [8] (age: 1;08-4;00) from CHILDES to investigate whether uncertainty SFP *ba* and adverb *keneng* exist in children's input and production. **Results.** **Table 1** summarizes *ba* and *keneng* frequency in child and adult production, and the age of first use. Both children used *ba* in the uncertainty context before age 3;00. *keneng* appeared in Tong's production immediately after age 3;00 (less frequent than *ba*), but not in Zhou's. Both *ba* and *keneng* occurred in the caregivers' production, confirming presence in input and child production.

Experiment. We had two goals: test whether children comprehend uncertainty encoded in SFPs and determine whether any difficulties stem from SFP-specific issues or uncertainty comprehension in general. We adapted the hidden box task [9-10], and compared children's understanding of uncertainty-marking SFP *ba* with the modal adverb *keneng* 'perhaps'. Selective difficulty with *ba* (but not *keneng*) would indicate SFP-specific comprehension challenges, while poor performance on both items would suggest general difficulty with uncertainty concepts.

Participants saw two puppets making conflicting statements about an object's location and guessed the location of the object. Each pair contrasted a plain declarative (e.g., 'The basketball is in the blue box') with an uncertainty-marked sentence which differs only in location and uncertainty marker (e.g., 'The basketball is *keneng* in the red box' or 'The basketball is in the red box *ba*.') If speakers are sensitive to the uncertainty marker, they would choose the location indicated by the plain declarative; if not, they would choose randomly between the two locations (in this case, the blue box).

Procedure. Participants viewed illustrations (**Fig1**) on screen accompanied by pre-recorded puppet utterances (**Fig2**). The puppet producing the uncertainty-marked utterance was counterbalanced for left-right position. All participants judged 16 items in semi-randomized order (4 fillers (**Fig2-i**), 6 SFPs trials (**Fig2-ii**), 6 adverb trials (**Fig2-iii**)).

Participants. 75 Mandarin speakers were recruited online or from Tianjin kindergartens (CHI3-4: n=19, $M_{age}=4;0$; CHI4-5: n=16, $M_{age}=5;0$; CHI5-6: n=35, $M_{age}=5;11$; Adults (as baseline): n=5, $M_{age}=33;0$). Three child participants were excluded for failing >1 filler or misunderstanding instructions.

Results. **Fig3** summarizes the mean accuracy by condition. Adults achieved 100% accuracy for both items. T-tests against chance (50%) showed that all age groups performed significantly above chance on *keneng* (all $p < .01^{**}$); For *ba*, the youngest group (CHI3-4) showed marginal significance ($p = .052$), whereas older groups performed significantly above chance ($p < .01^{**}$).

Effect of marker type. Wilcoxon signed-rank test showed children performed significantly better on *keneng* (M = 4.77) than *ba* (M = 4.26), $Z = -2.86, p = .004^{**}$. **Effect of age.** Spearman correlations between age and scores were significant for both *keneng* ($r_s = 0.464, p < .001^{***}$) and *ba* ($r_s = 0.415, p < .001^{***}$), showing performance improved with age.

Table 1. First use and frequency of uncertainty expressions.

	First Use		Frequency		
	Tong	Zhou	Tong	Zhou	Adult
<i>ba</i>	2;9	2;10	0.1‰	0.1‰	0.2‰
<i>keneng</i>	3;1	N/A	0.03‰	N/A	0.1‰

Figure 1. Experimental stimuli: trial illustration.

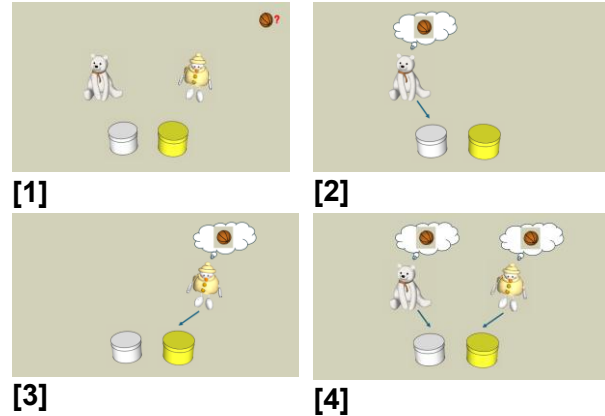
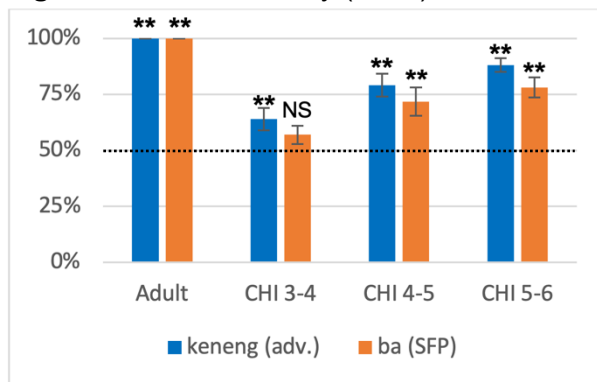


Figure 2. Experimental stimuli: script

(i) Filler trial	(ii) Trial1: SFP <i>ba</i>	(iii) Trial2: adverb <i>keneng</i>
BEAR: The toothbrush is in the red box. { <i>Yashua zai hongse de hezili.</i> } SNOWMAN: The toothbrush is not in the blue box. { <i>Yashua bu zai lansede hezili.</i> } EXP: Where is the toothbrush?	BEAR: (Fig1-[2]) The football is in the red box. { <i>zuqiu zai hongse de hezili.</i> } SNOWMAN: (Fig1-[3]) The football is in the blue box- ba . { <i>zuqiu zai lansede hezili-ba.</i> } EXP: (Fig1-[4]) Where is the football?	BEAR: (Fig1-[2]) The phone is in the red box. { <i>shouji zai hongse de hezili.</i> } SNOWMAN: (Fig1-[3]) The phone is perhaps in the blue box. { <i>shouji keneng zai lansede hezili.</i> } EXP: (Fig1-[4]) Where is the phone?

Figure 3. Mean accuracy (n=72)



Discussion. Our findings with *ba* suggest that Mandarin-speaking children's early uses of SFPs can accompany relevant semantic or pragmatic knowledge, aligning with previous findings with child acquisition of Japanese SFP [10]. However, while [10] proposes that children might extract speaker (un)certainty earlier when it is encoded in SFP than in verbs, our results show such an advantage do not extend to SFP when compared with adverbs; in fact, the results show the contrary: children comprehended adverb *keneng* significantly better than SFP *ba*.

This is intriguing, considering that *ba* is both frequent and occur much earlier than SFPs in child utterances relative to *keneng* that encode similar meanings (Table 1). We suggest that there might be several possibilities: For example, the asymmetry can reflect *ba*'s complexity. *ba* serves multiple functions across clause-types, whereas *keneng* exclusively marks uncertainty. Matching *ba* to its correct pragmatic function in different contexts may prove difficult for children.

References. [1] Zhang et al., 2019. [2] Paul and Yan, 2022. [3] Chu, 1998. [4] Yuan, 2021. [5] Tsai, 2015. [6] Tuan, 2017. [7] Deng and Yip, 2018. [8] Zhang and Zhou, 2009. [9] Hirst and Weil, 1982. [10] Matsui et al., 2006. [11] Hübscher et al., 2017. [12] Ozturk and Papafragou, 2015. [13] Hendriks & Koster, 2010.