Structures of Function Morphemes Guide Mandarin-Learning 19-month-olds In Backward Syntactic Categorization

Syntactic categorization is one of the challenges facing infants in language acquisition. To acquire syntax, they need to classify words in continuous speech stream into abstract syntactic categories such as nouns and verbs. Previous studies show that infants store function morphemes based on their acoustic and distributional properties (e.g., Hallé, Durand, & de Boysson-Bardies, 2008; Shi, Werker, & Cutler, 2006) and use them to identify phrases and categorize adjacent words based on their co-occurrence patterns (e.g., Höhle et al., 2004; Shi & Melançon, 2010). We inquire whether infants infer syntactic categories from syntactic relations or just bigram probabilities (*P*). Using a visual fixation procedure, we examined whether Mandarin-learning 19-month-olds categorize target word X by using a following function morpheme f in a 3-word sequence X-f-Y.

In our study, 19-month-olds (n = 36, mean age: 1; 7; 17) were randomly familiarized with one out of three types of prosody-neutral contexts, X-*ye*$_{\text{Foc}}$-Y 'even X_N Y_V', X-*le*$_{\text{Asp}}$-Y 'have X_V-ed Y_N', or X-*bu*$_{\text{Neg}}$-Y 'X_N doesn't Y_V', where unfamiliar ambicategorial words (*shai1* and *man2* for the target X; *tong3ji4* and *jian3yan4* for the filler Y) were used to make distribution the only cue to category inferencing. They were then tested on grammatically contrasting contexts presenting X as a noun (*shi*$_{\text{V}}$ *zhege*$_{\text{Det}}$ *X* 'be this-Cl X_N') or as a verb (*dou*$_{\text{Adv}}$ *keyi*$_{\text{Aux}}$ *X* 'all may X_V') and were expected to discriminate between the two types of contexts upon categorization of X. Bigram probability of X-f only predicts categorization with *ye*, given that $P(X_N \mid ye) = 0.957$, $P(X_V \mid le) = 0.516$, $P(X_N \mid bu) = 0.429$.

We found that 19-month-olds succeeded in the *ye* and *le* condition, contrary to what bigram probabilities predict. As shown in Figure 1, toddlers familiarized with *ye* sequences looked significantly longer to grammatical trials than to ungrammatical trials, $t(11) = 2.913$, $p = .014$; the same looking preference replicated with toddlers familiarized with *le* sequences, $t(11) = 2.503$, $p = .029$, but not with those familiarized with *bu* sequences, $t(11) = -.710$, $p = .493$. This was confirmed by the results of preference scores (Figure 2).

The results that sequences with focus particle *ye* ([$_{\text{FocP}}$ X_N ***ye***$_{\text{Foc}}$ […]]) and aspect marker *le* ([$_{\text{AspP}}$ X_V ***le***$_{\text{Asp}}$ […]]) helped the categorization of target X, whereas those with negation marker *bu* ([$_{\text{TP}}$ X_N [$_{\text{NegP}}$ ***bu***$_{\text{Neg}}$ …]]) did not, demonstrate that 19-month-olds can use both free and bound function morphemes to categorize preceding words (nouns or verbs) despite constraints of working memory and unfavorable bigram probabilities. Crucially, they are inclined to categorize preceding targets that are structurally adjacent, not just linearly adjacent by probability, to their following function morphemes, which implies that toddlers represent structures of function morphemes beyond bigram co-occurrences. Furthermore, our findings lend cross-linguistic support to the view that toddlers harness functional elements to trace the structural skeletons of human language and bootstrap the syntactic categories of unfamiliar words.

(477 words)

Table 1. Familiarization conditions (*ye*, *le*, and *bu*) and test items

Familiarization (T_{max}=30 s, ISI = 1200ms)		
***ye* condition**	***le* condition**	***bu* condition**
shai *ye* tongji	shai *le* tongji	shai *bu* tongji
man *ye* tongji	man *le* tongji	man *bu* tongji
man *ye* jianyan	man *le* jianyan	man *bu* jianyan
shai *ye* jianyan	shai *le* jianyan	shai *bu* jianyan
Test (10 trials; T_{max}=17.6 s, ISI = 1000 ms)		
Trial 1 *shi zhege* shai *shi zhege* man (grammatical – *ye* & *bu*; ungrammatical – *le*)		
Trial 2 *dou keyi* shai *dou keyi* man (ungrammatical – *ye* & *bu*; grammatical – *le*)		

Note: T_{max} represents 'maximal trial length'; ISI represents 'inter-stimulus interval'.

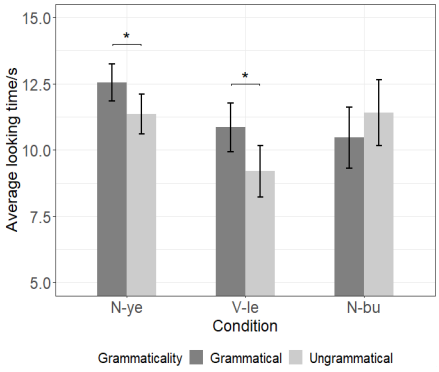

Figure 1. Mean looking times (and standard errors) during the two test trial types (grammatical vs. ungrammatical) for the three conditions (*ye*, *le*, and *bu*)

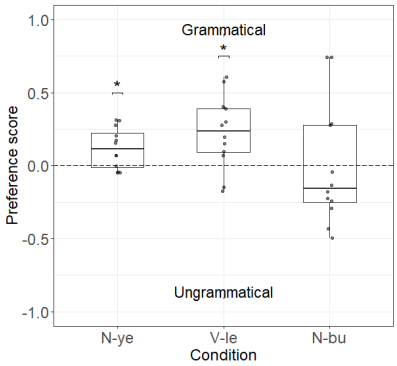

Figure 2. Preference scores for the three conditions (*ye*, *le*, and *bu*)

Selected references

Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech, 51*, 23-44.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, *5*, 341-353.

Shi, R., & Melançon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy, 15*(5), 517-533.

Shi, R., Werker, J. L., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10*(2), 187-198.