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

Yuanfan Ying[1,3], Xiaolu Yang[1], Rushen Shi[2]
[1]Tsinghua University [2]Université du Québec à Montréal [3]University of Maryland

THE POINT

- 19-month-olds categorize words based on a following *function morpheme* (i.e., X←**f**)
- They infer syntactic categories of adjacent words from syntactic relations, not just bigram probabilities.

INTRODUCTION

Infants are sensitive to function morphemes from a young age [1, 2]
Function morphemes = function words + affixes
- Segmentation: 11mos (determiners in German, English, French) [3, 4, 5]
- Tracking distribution: 12mos (in natural and artificial languages) [6, 7, 8]

They use function morphemes to infer syntactic categories [8, 9, 10, 11, 12]

Method	Language	Age/mo	Context	Category	Result
Mixed contexts (N & V)	English	12	*the _ is* *to _ it*	'the X_N is' 'to X_V it'	N: × V: √
		14-18	*my _* *can _*	'my X_N' 'can X_V'	N: √ V: √
	Mandarin	12	*zhege _* *ni bie _*	'this-Cl X_N' 'You don't X_V'	N: √ V: √
Separate contexts (N / V)	German	14-16	*ein _* *sie _*	'a X_N' 'she X_V'	N: √ V: ×
	French	14	*le _* *je _*	'the X_N' 'I X_V'	N: √ V: ×

- General success with nouns but not with singled-out *verbs*
- Why failure in *using pronouns for inferring following verbs* in two-word sequences?
  - former account: *bigram frequency*
  - **Q:** *Is syntactic relation at play?*

HYPOTHESIS

Syntactic relations outrank sequence patterns in infants' category inferencing, and bigram order (fX or Xf) should not matter.

word-internal suffixes
phrase-internal function words
-ing$_{Asp}$, *the*$_{Det}$, *can/will*$_{Aux}$

subject-verb template
[$_{TP}$ She$_{Pron}$ [$_{VP}$ ran$_V$]]

- *Test* – Mandarin

Prediction of categorization
syntactic relations vs. bigram probabilities

function morphemes	Frequency /million	Syntactic environment	Linear co-occurrence (bigram)	Linear co-occurrence (prosody-filtered bigram)
ye (focus particle)	4209	[$_{FocP}$ X_N *ye*$_{Foc}$ […]]	0.957	1
***le* (aspect marker)**	23422	[$_{AspP}$ X_V-*le*$_{Asp}$ […]]	0.594	0.75
bu (negation marker)	22873	[$_{TP}$ X_N [$_{NegP}$ *bu*$_{Neg}$ …]]	0.429	0.75

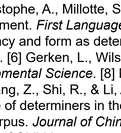
Where noise comes in and predictions diverge!

Note: The probability of 'le' following a noun is 0.18.
Tong's corpus, CHILDES (age 1;7-1;8, 5465 total word tokens) [13]

EXPERIMENTS

Participants. Mandarin-learning toddlers, n = 36 (19 male), mean age = 1;7;17 (range: 1;6;5-1;8;26), between-subject
Method. Visual fixation procedure. (infant-controlled)
Design. Three conditions: *ye*, *le*, and *bu*.

① A lip-sync puppet uttered sentences where the target words were focused in the sequence to license only the intended parsing.
② Nonce words were used. ③ Sentence prosodies were controlled.

Visual	Audio
	**target word** [*shai1* 'sieve; to sift', *man2* 'eel; to conceal'] **XfY** **filler word** [*tong3ji4* 'calculation; to calculate', *jian3yan4* 'examination; to examine'] **function morpheme** [aspect *-le*, focus *ye*, negator *bu*] (neutral tone syllables)

Timecoure

Familiarization *30s* | **Test** *10 trials × 17.6s/trial* | **Post-test** *17.6s*

***ye* condition** (n=12)	$shai_N$ *ye* tongji man_N *ye* tongji man_N *ye* jianyan $shai_N$ *ye* jianyan
***le* condition** (n=12)	$shai_V$ *le* tongji man_V *le* tongji man_V *le* jianyan $shai_V$ *le* jianyan
***bu* condition** (n=12)	$shai_N$ *bu* tongji man_N *bu* tongji man_N *bu* jianyan $shai_N$ *bu* jianyan

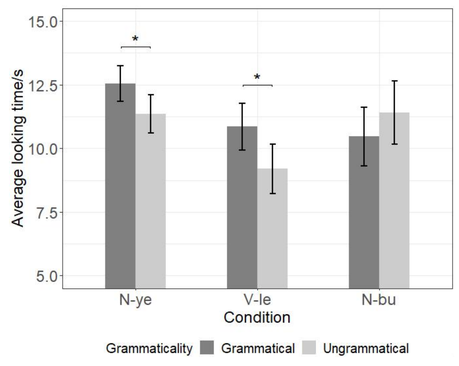
Trial A
shi zhege $shai_N$
shi zhege man_N
Grammatical - ***ye, bu,***
Ungrammatical - ***le***

Trial B
dou keyi $shai_V$
dou keyi man_V
Ungrammatical - ***ye, bu,***
Grammatical - ***le***

(alternating & counterbalanced)

Translation:
Familiarization - *X ye*$_{Foc}$ *Y* 'even **X**$_N$ Y' || *X le*$_{Asp}$ *Y* 'have **X**$_V$-ed Y' || *X bu*$_{Neg}$ *Y* '**X**$_N$ doesn't Y'.
Test - *shi*$_{BE}$ *zhege*$_{Det}$ *X* 'It's this-Cl **X**$_N$' || *dou*$_{Adv}$ *keyi*$_{Aux}$ *X* 'all may **X**$_V$'.

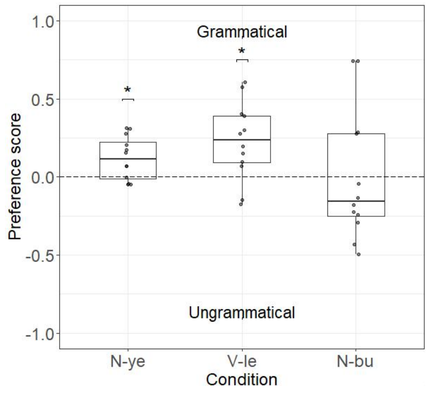
RESULT



19-month-olds succeeded with *le* and *ye* but not with *bu*.

- Supported by syntactic relations, they categorized nonce words backwardly regardless of them being nouns or verbs.
- The results were predicted by syntactic relations, not bigram probabilities.

[**Xf** […]] > [**X** [**f** …]]
within-phrase cross-phrase



Note: Our samples are all normally distributed according to the Shapiro-Wilk normality test.

TAKE-HOMES

19-month-olds use function morphemes to categorize preceding nonce words despite constraints of working memory, [14, 15] but they only do this when the nonce words are structurally related to their following function morphemes.

SELECTED REFERENCES

[1] Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*. [2] Dye, C., Kedar, Y., & Lust, B. (2019). From lexical to functional categories: New foundations for the study of language development. *First Language*. [3] Höhle, B., & Weissenborn, J. (2000). The origins of syntactic knowledge: Recognition of determiners in one year old German children. In *BUCLD 24 Proceedings*. [4] Shi, R., Cutler, A., Werker, J. F., & Cruickshank, M. (2006). Frequency and form as determiants of functor sensitivity in English-acquiring infants. *Journal of Acoustical Society of America*. [5] Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*. [6] Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cue to form syntactic categories. *Journal of Child Language*. [7] Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*. [8] Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*. [9] Hicks, J. P. (2006). *The impact of function words on the processing and acquisition of syntax*. Doctoral dissertation. [10] Zhang, Z., Shi, R., & Li, A. (2014). Grammatical categorization in Mandarin-Chinese-learning infants. *Language Acquisition*. [11] Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional element in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*. [12] Shi, R., & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*. [13] Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics*. [14] Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*. [15] Sundara, M. (2018). Why do children pay more attention to grammatical morphemes at the ends of sentences? *Journal of Child Language*.

Acknowledgements: Thanks to all members of the Language Acquisition Lab at Tsinghua, the Chinese National Funding of Social Sciences #11BYY080, to all parents and children who participated in the study, and to all those who have contributed to the study.

BUCLD 45

Contact: jackying@umd.edu