General Availability, Local Retreat: Noun-conditioned Age Variation in Mandarin Children's Classifier Choice

Yuxiang Deng [1,#,*] and Litong Qin [2,#]

[1] School of Liberal Arts, Jiangsu University, Zhenjiang, 212013, China
[2] School of Languages, Cultures and Societies, University of Leeds, LeedsLS29JT, UK
Email: 13656177801@163.com (Y.D.); tngp0955@leeds.ac.uk (L.Q.)
[*]Corresponding author;
[#]co-first authors
Manuscript received August 5, 2026; accepted September 15, 2026; published November 23, 2026

***Abstract*—Mandarin-speaking children often continue to use the general classifier *ge* while becoming more selective with noun-linked classifiers. We tested whether this age-related change is global or concentrated in particular lexical contexts. Using frozen Mandarin CHILDES/childes-db 2021.1 records, an adult lexical probe identified nouns with a recurrent specific-classifier competitor. The primary analysis included 7777 high-confidence classifier-choice tokens from 837 children aged 18–75 months. From 48 to 72 months, predicted *ge* probability declined by .213 in stable-competitor contexts but by .027 elsewhere. The difference between these changes was −.185, 95% CrI [−.329, −.026]. The contrast remained negative across extraction, noun-composition and corpus checks. *Ge* therefore remained broadly available, but its relative use declined more sharply in a restricted lexical subspace.**

***Keywords*—Mandarin classifiers, child language, ge, lexical differentiation, CHILDES, Bayesian modelling**

I. Introduction

Mandarin children do not simply learn whether a classifier slot exists. They also learn how a broadly available form, *ge* 个, competes with narrower noun-linked classifiers such as zhī 只, běn 本 and *duǒ* 朵. Classifier–noun associations are not strictly one-to-one [1–4], and adult production can retain *ge* even when a more specific option is available [5, 6]. A developmental account based only on the overall frequency of *ge* therefore risks missing where change actually occurs.

Two observations in previous work are especially relevant. Zhan and Levy [6] note that "there is a general classifier that can be used with most nouns." That broad availability makes the persistence of *ge* unsurprising and shifts the developmental question from disappearance to relative choice. At the child level, Hao *et al.* [7] report "differential accuracy for different classifier-noun combinations." Their result points away from a single global rate: classifier performance can vary item by item. Our design takes these observations together by asking whether age-related change in ge depends on the noun environment.

Earlier studies appear to point in different directions because they count different events. Shape-classifier work reports increasing target performance and frequent *ge* substitutions [7, 8]; broader corpus studies find that *ge* remains highly frequent [9]; analyses restricted to *yī* 'one' distinguish clear overgeneralisation from other uses [10], and longitudinal shape-classifier work documents change within two children [11]. These outcomes are not interchangeable. Noun composition matters as well: animals, books and flowers provide different opportunities for specific alternatives, and classifier distributions vary across registers [12].

Register gives another reason not to read aggregate *ge* frequency too literally. Shi *et al.* [12] found that "everyday speech relied on more generic classifiers," whereas books showed greater classifier diversity. A shift in corpus or activity composition can therefore change the overall denominator even when choice for a comparable noun has not changed. We treat register and corpus as composition factors to be checked, rather than evidence that one setting is linguistically more advanced.

We test two descriptions. A global-retreat account predicts a broadly similar reduction of *ge* across noun contexts. A lexical-redistribution account predicts a larger decline where a stable alternative is available, while *ge* remains productive elsewhere. Stable-competitor status was defined from adult records before fitting the child age model. H1 predicts a negative 48-to-72-month change within stable contexts; H2 predicts that this change is more negative than in other eligible contexts. The adult probe, model specification and focal contrast were frozen before final fitting. Descriptive patterns had already been inspected, so we treat the analysis as theory-guided exploratory rather than confirmatory.

The central result is visible on the probability scale: from 48 to 72 months, *ge* declined by .213 in stable-competitor contexts and by .027 elsewhere. Their difference was −.185, 95% CrI [−.329, −.026]. Thus, the developmental pattern is not well described as uniform replacement of *ge* by specific classifiers.

II. Methods

A. Data and Extraction

Target-child productions came from the Chinese CHILDES collection [13] in the frozen childes-db 2021.1 snapshot [14]. Cantonese corpora were excluded and recorded ages were restricted to 18–75 months. The high-confidence primary dataset contained 7,777 classifier-choice tokens from 837 children, 976 noun types and 1,498 transcripts across 16 corpora.

Candidate extraction required a classifier tag preceded by a numeral, determiner or demonstrative and an overt noun within four following tokens. We retained sortal uses and excluded measure, container, collective, event and verbal uses. Developmental inspection showed why local structure mattered: temporal expressions, pronominal classifier phrases and incomplete compound nouns could otherwise

create false pairings. The revised frame contained 10,476 eligible candidates; the primary analysis used the best-validated local condition, first_local_noun at noun offset 1.

Manual validation assessed whether the algorithm had located a genuine classifier–noun construction, the child's overt classifier, the noun head and their pairing. It did not judge whether the child used an adult-preferred classifier. Corrected complete-frame valid-context PPV was .955 [.932, .975] and end-to-end accuracy .919 [.890, .945]; end-to-end accuracy in the primary domain was .989 [.975, .999]. Two additional coders independently reviewed an 80-record subset. Pre-adjudication agreement was 98.6% for classifier–noun pairing and 95.9% for segmentation/POS, with complete observed agreement on context validity and noun-head identification. Detailed adjudication records are kept outside the conference manuscript.

One post-freeze adjudication identified an invalid candidate in the wider sensitivity samples. Removing that record left the primary dataset unchanged and altered the intermediate/full DID medians by less than .003. The correction is reported as a sensitivity update, not as a redefinition of the validation frame.

B. Outcome and Lexical Context

Each observation was one eligible classifier choice. ge_choice was coded 1 for 个/個 and 0 for another sortal classifier. Stable-competitor status came from a frozen adult lexical probe: a noun needed at least five adult classifier tokens, at least three tokens of its modal non-ge form, a modal non-ge share of at least .50 overall and dominance of at least .75 among non-ge tokens. Nominal/sortal review retained 37 mappings. The label identifies a recurrent distributional competitor, not a uniquely correct classifier or a child-specific input history.

Transcript-level MLU-W was considered because productive-language development can vary among children of similar ages [15–17]. The frozen workflow did not provide a comparable complete-transcript measure across corpora; calculating MLU only from classifier-containing utterances would condition the measure on the construction under study. MLU-W was therefore omitted rather than replaced with a selected proxy.

C. Statistical Analysis

A hierarchical Bayesian logistic model predicted *ge* choice from stable status, four cubic B-spline terms for age and their interactions. Varying intercepts represented child, transcript, noun and corpus, with regularised linear age variation by noun and corpus. The primary fit used PyMC NUTS [18], four chains, 3,000 warm-up iterations and 4,000 retained draws per chain. All varying effects were set to zero for the reported conditional predictions.

H1 used Δstable = P(*ge* | stable, 72) − P(*ge* | stable, 48). H2 used the draw-wise contrast DID = Δstable − Δother. Negative DID values indicate a larger decline in stable-competitor contexts. The same scientific specification was checked across wider extraction sets, common-support nouns, register/corpus exclusions and lexical omissions. A separate supporting model decomposed age within and between corpora.

III. Results and Discussion

A. Main Result

Ge accounted for 6,616 of 7777 primary tokens (85.1%). Stable-competitor contexts were a minority of the sample (687 tokens; 8.8%), which widened uncertainty in that condition but still supported estimation of the focal contrast. From 48 to 72 months, the posterior median change was −.213 in stable contexts and −.027 in other contexts (Table 1). H1 was supported, and the larger stable-context decline supported H2.

Table 1. Conditional probability changes from 48 to 72 months

Contrast	Median	95% CrI	P(contrast < 0)
Δstable (H1)	−.213	[−.367, −.046]	.994
Δother	−.027	[−.062, −.008]	.999
DID (H2)	−.185	[−.329, −.026]	.990

Changes were calculated within posterior draws before summarisation; directional probabilities are not frequentist *p*-values.

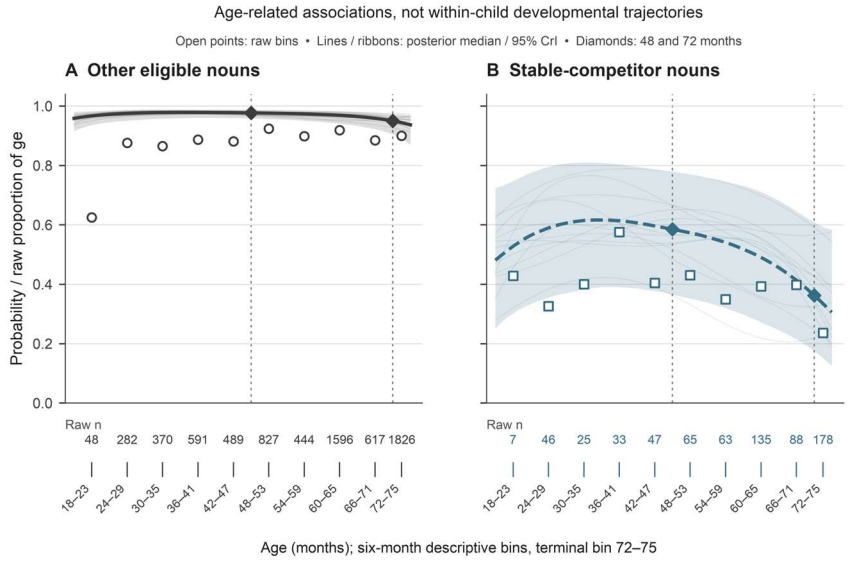


Fig. 1. Raw six-month proportions and conditional age curves. Ribbons show 95% CrIs; vertical guides mark 48 and 72 months. The terminal bin covers 72–75 months.

The fitted trajectories were nonlinear (Fig. 1). Conditional *ge* probability in stable contexts changed from .585 at 48 months to .363 at 72 months; elsewhere it changed from .977 to .950. The resulting DID was −.185, 95% CrI [−.329, −.026], with P(DID < 0) = .990. On the chosen probability scale, the stable-context decline was therefore about 18.5 percentage points larger. Both context types declined; the finding is differential change, not disappearance of *ge*.

B. Robustness and Interpretation

Widening the extraction set did not reverse the main contrast. In the corrected intermediate sample, DID was −.159 [−.295, −.018]; in the corrected full sample it was −.180 [−.308, −.039]. Restricting the analysis to 107 common-support nouns gave −.190 [−.332, −.043]. Leave-one-corpus and lexical omissions also stayed on the negative side, although their intervals differed in width (Fig. 2). We use these fits as stress tests of the primary estimate, not as separate replications.

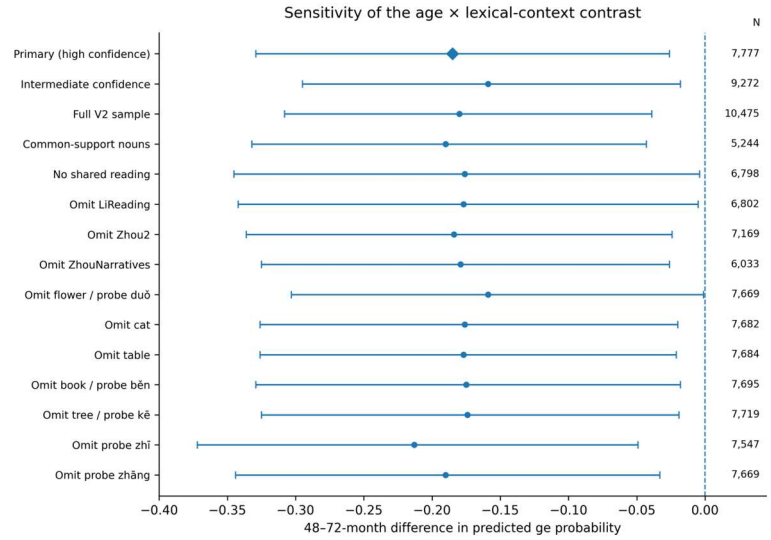


Fig. 2. DID estimates under the sensitivity restrictions. The diamond is the primary estimate; horizontal intervals are 95% CrIs. Intermediate and full rows use the corrected wider-sample fits.

The corpus-age decomposition was less clear-cut. Within a corpus, the 12-month DID was −.099 [−.193, −.006]; across corpus means, it was −.016 [−.125, .095]. A result driven only by corpora with different mean ages would be difficult to square with that within-corpus contrast. Still, this is not longitudinal evidence: transcripts within the same corpus usually come from different children and recording occasions.

Why can performance with specific classifiers improve while corpus counts still show frequent *ge*? Our sample gives one concrete reason. Stable-competitor contexts contributed only 687 of the 7777 primary tokens. Most observations therefore came from noun environments in which the fitted 48-to-72-month decline was small (−.027). The much larger decline (−.213) occurred in the smaller stable-context subset. Pooling the two patterns gives an overall rate that is weighted heavily toward the first group, which can hide the lexical concentration of change.

The theoretically important subset is also the smaller one. With 687 stable-context tokens, its uncertainty is wider than that of the comparison category. The adult probe adds another source of uncertainty: its thresholds create a reproducible contrast, but they are operational cut-offs rather than grammatical boundaries. Nouns below those thresholds may still have conventional specific classifiers.

A different limitation comes from the outcome itself. Every eligible non-*ge* classifier is pooled into one category. We can identify where *ge* loses relative probability, but not whether that probability goes mainly to the probe-linked classifier or is spread across several alternatives. Most observations are also cross-sectional. A three-category outcome—*ge*, the probe-linked form, and other specific classifiers—would be the most informative extension if the cells are large enough.

IV. Conclusion

Across the noun contexts examined here, *ge* did not recede everywhere at the same rate. The larger 48-to-72-month change came from nouns with a stable adult competitor; in other eligible contexts the fitted probability of *ge* was still about .95 at 72 months. An aggregate rate blurs this unevenness.

We use local lexical redistribution as a description of that pattern. The corpus tells us where the relative probability changed. It does not tell us which non-*ge* form received the lost share, which input history produced the pattern, or how an individual child's representation changed over time.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Yuxiang Deng (YD) conceptualized the study, formulated the research questions and hypotheses, and developed the research methodology. YD designed the corpus extraction and coding procedures, coordinated the independent coding and validation process, conducted the adjudication of disputed cases, and supervised the statistical analyses. YD also examined the model diagnostics and sensitivity analyses,

interpreted the findings, and prepared and revised the Methods and Results sections.

Litong Qin (LQ) reviewed the relevant literature and wrote most of the manuscript, including the Abstract, Introduction, Discussion, and Conclusion. LQ also organized the overall structure of the paper and edited the manuscript to ensure consistency and coherence across sections.

Both authors contributed to the interpretation of the findings, critically revised the manuscript, and approved the final version.

REFERENCES

[1] K. Allan, "Classifiers," *Language*, vol. 53, no. 2, pp. 285–311, 1977, doi: 10.2307/413103.

[2] A. Y. Aikhenvald, *Classifiers: A Typology of Noun Categorization Devices*. Oxford, U.K.: Oxford University Press, 2000, doi: 10.1093/oso/9780198238867.001.0001.

[3] C. N. Li and S. A. Thompson, *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA, USA: University of California Press, 1981, doi: 10.1525/9780520352858.

[4] M. S. Erbaugh, "Classifiers are for specification: Complementary functions for sortal and general classifiers in Cantonese and Mandarin," *Cahiers de Linguistique – Asie Orientale*, vol. 31, no. 1, pp. 33–69, 2002, doi: 10.3406/clao.2002.1602.

[5] M. Y. Gao and B. C. Malt, "Mental representation and cognitive consequences of Chinese individual classifiers," *Language and Cognitive Processes*, vol. 24, no. 7–8, pp. 1124–1179, 2009, doi: 10.1080/01690960802018323.

[6] M. Zhan and R. P. Levy, "Availability-based production predicts speakers' real-time choices of Mandarin classifiers," in *Proc. 41st Annu. Meeting Cogn. Sci. Soc.*, 2019, pp. 1268–1274. [Online]. Available: https://escholarship.org/uc/item/8h77s696

[7] Y. Hao, L. Bedore, L. Sheng, P. Zhou, and L. Zheng, "Exploring influential factors of shape classifier comprehension and production in Mandarin-speaking children," *First Language*, vol. 41, no. 5, pp. 573–604, 2021, doi: 10.1177/01427237211026435.

[8] P. Li, B. Huang, and Y. Hsiao, "Learning that classifiers count: Mandarin-speaking children's acquisition of sortal and mensural classifiers," *Journal of East Asian Linguistics*, vol. 19, no. 3, pp. 207–230, 2010, doi: 10.1007/s10831-010-9060-1.

[9] A. Huang and Y. He, "Integrating the syntactic and semantic acquisition of Chinese classifiers in Mandarin-speaking children: A corpus study," *First Language*, vol. 46, no. 2, pp. 256–287, 2026, doi: 10.1177/01427237251371761.

[10] S. Wang and Y. Zhang, "A study on the acquisition of shape classifiers by Mandarin-speaking children," *Lingua*, vol. 332, 104106, 2026, doi: 10.1016/j.lingua.2025.104106.

[11] N. Lin and M. Zhou, "L1 Mandarin children's acquisition of the 'One + Classifier' structure: Empirical evidence and insights from special language domain theory," *Frontiers in Psychology*, vol. 17, 1763771, 2026, doi: 10.3389/fpsyg.2026.1763771.

[12] J. Shi, Y. Hsiao, Y. Yang, E. Wonnacott, and K. Nation, "Learning from the input: A corpus-based investigation of Chinese classifiers in children's books and child-directed speech," *Journal of Child Language*, early access, 2026, doi: 10.1017/S030500092610049X.

[13] B. MacWhinney, *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2000.

[14] A. Sanchez, S. C. Meylan, M. Braginsky, K. E. MacDonald, D. Yurovsky, and M. C. Frank, "childes-db: A flexible and reproducible interface to the Child Language Data Exchange System," *Behavior Research Methods*, vol. 51, no. 4, pp. 1928–1941, 2019, doi: 10.3758/s13428-018-1176-7.

[15] S.-Y. Wu, "Mean length of utterance among Mandarin-speaking children with and without DLD," *Child Language Teaching and Therapy*, vol. 36, no. 3, pp. 165–179, 2020, doi: 10.1177/0265659020945366.

[16] M. L. Rice, F. Smolik, D. Perpich, T. Thompson, N. Rytting, and M. Blossom, "Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 2, pp. 333–349, 2010, doi: 10.1044/1092-4388(2009/08-0183).

[17] C.-Y. Chu, P.-H. Chen, Y.-S. Tsai, C.-A. Chen, Y.-C. Chan, and Y.-J. Ciou, "Effect of sample length on MLU in Mandarin-speaking hard-of-hearing children," *Journal of Deaf Studies and Deaf Education*, vol. 29, no. 3, pp. 388–395, 2024, doi: 10.1093/deafed/enae007.

[18] O. Abril-Pla *et al.*, "PyMC: A modern, and comprehensive probabilistic programming framework in Python," *PeerJ Computer Science*, vol. 9, e1516, 2023, doi: 10.7717/peerj-cs.1516.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).